Budget-constrained, Capacitated Hub Location to Maximize Expected Demand Coverage in Fixed-wireless (Broadband-access) Telecommunication Networks

Ramesh Bollapragada • Yanjun Li • Uday S. Rao

College of Business, San Francisco State University, CA and Bell Laboratories, Lucent Technologies, Holmdel, NJ, USA

Department of Quantitative Analysis and Operations Management, College of Business Administration, University of Cincinnati, Cincinnati, Ohio 45221-0130, USA

Krannert School of Management, Purdue University, IN, USA

rameshb@sfsu.edu • li14@mgmt.purdue.edu • uday.rao@uc.edu

This paper studies a model for telecommunication network installation by companies in the Broadband-access business, specialized to the fixed-wireless case. Under stochastic demand (modeled using scenarios), we maximize expected demand coverage subject to a budget constraint on hub installation, and technological (crane-rain radius, line-of-sight, and limited capacity) constraints at each demand location where a hub is installed. There are multiple hub-types, differing in costs and capacities. A practical greedy solution heuristic based on the budgeted maximum coverage problem is presented and its worst-case performance is analyzed. For special cases with a single hub-type or a single demand scenario, we show that a guarantee of $1 - \frac{1}{e}$ or 63.2% applies to our greedy heuristic. For the general case we develop a data-dependent bound on the greedy's worst-case performance (which is, in practice, within a near-factor of $1 - \frac{1}{e}$ of the optimal). Further, our computational results show that the greedy's empirical performance is on average within two percent of optimal.

(*Telecommunications; Broadband-access, fixed-wireless; location-allocation; network planning; stochastic programming; greedy approximation algorithms.*)

---

1. Introduction

Worldwide service revenues for broadband wireless will be worth nearly 42 billion dollars in the year of 2005, according to a recent research report by ARC Group (www.arcgroup.com). The US market is forecast to become highly competitive due to intense competition from alternative technologies and operators. Further, many of today's broadband deployments are inherently inflexible and will rapidly become tomorrows last-mile bottlenecks. The ARC Group reports that "Fixed wireless systems will increasingly become a key feature of the access landscape, with lower frequency systems catering for the mass market. Broadband wireless offers a unique combination of flexibility and high performance that is unmatched by other high bandwidth access technologies. High speed, high capacity coupled with true flexibility and scalability are the key features that are proving to be highly attractive to both service providers and end-users. The ability to rapidly deploy networks, whilst avoiding the complexity and restrictions of operating in the local loop, is also a huge plus."

This paper investigates a model for telecommunication network installation for companies in the Broadband-access business. More specifically, we conduct our analysis in the context of a fixed-wireless problem. However, the approaches we present may be modified to solve the problem of locating different hubs (aggregation nodes) in network planning of other Broadband-access networks such as DSL, Cable, and Fiber-to-the-Home. The terminology and notation vary from one access technology to another and hence we focus on one specific access technology, fixed-wireless, in this paper. The research presented here includes the design and analysis of algorithms that aid in strategic scenario-based network planning.

Faced with rising maintenance costs associated with aging outside of regular copper wire technology in urban areas, planners in many metropolitan areas are replacing the drop wire with radio technology combined with a high-capacity distribution system. Since this wireless technology is cost-effective, companies have been investing extensively in expanding their wireless telecommunication networks. Given the growing demand for wireless services and the increasing competition, an analytical framework for network planning (such as the hub location considered in this paper) will aid service providers in achieving their objectives.

Currently, telephone companies that provide local toll services through wireless technology use building specific census data to determine the optimal location of hubs in the network. These companies are in need of quantitative models that help determine the optimal placement of local hubs at building roof tops for fixed-wireless service. The hub placements determine the coverage area for the local loop wireless service. The service is provided to surrounding buildings through receiving antennas in a radius determined by the "crane-rain" radius, which is based on environmental and weather conditions and varies from region to region, e.g., this radius is larger in California where there is more dry air than New York. The greater the area of coverage by a hub, the greater the likelihood of extending its wireless network. Some other important considerations in the placement of hubs include the capacity of the hub and a "line-of-sight" constraint associated with the hub and the buildings within its crane-rain radius.

Designing a network of hubs for fixed-wireless access is inherently different from mobile wireless or wire-line access because of different technological constraints. In addition, the coverage area and number of subscribers are not known before the hubs are installed. Since the network is built before subscriber commitment, there is substantial market size uncertainty. Thus the problem to be addressed is stochastic in nature. Moreover, it is clear that demand and revenue growth issues need to be modeled upfront to account for the uncertain nature of the problem (so as to develop robust solutions). This problem comes in two flavors: (1) A "single-stage" problem where the hub placement is done only once and no expansions are considered over the planning horizon; in different time periods, realized building demands may be re-assigned to different installed hubs. (2) A "multi-stage" problem that involves multiple stages of hub placements and expansions, over a longer planning horizon. In this paper, we will address the single-stage, multi-period, multi-scenario problem. Based on insights developed from the single-stage model, our future research will explore the general multi-stage problem.

In this paper, we illustrate the development and analysis of practical fixed-wireless network planning algorithms that help in determining near optimal placement of different types of hubs. The main contributions of this work include:

1. A single-stage, multi-period model for capacitated hub location under multiple demand scenarios, technological line-of-sight and crane-rain radius constraints, and a limited budget for hub installation. The model can be used for developing and analyzing fixed-wireless networks.

2. A simple greedy solution heuristic for determining: (i) The optimal location of hubs in the fixed-wireless network. (ii) The number of hubs of each hub-type at each location. (iii) The assignment of realized demands to hubs in each period.

3. An analysis of the worst-case performance of the proposed greedy heuristic. This analysis applies and extends results on the budgeted maximum coverage problem to hub location.

4. A computational study of the greedy algorithm's average-case performance as compared to the optimal and linear programming based bounds on the optimal coverage.

In Section 2, we review relevant literature; in Section 3, we present the model. We develop problem properties and a greedy approach in Section 4, along with worst-case performance guarantees. We computationally test the greedy in Section 5 and conclude in Section 6.

2. Literature Review

A large body of research on various location models has appeared in the literature over the past three decades (see for instance, Cornuéjols *et al* 1977 or Labbe *et al* 1995). Recent books and survey papers include Brandeau and Chiu (1989), Mirchandani and Francis (1990), Francis *et al* (1992), Louveaux (1993), Daskin (1995), Drezner (1995), ReVelle and Laporte (1996), Hamacher and Nickel (1998), Drezner and Hamacher (2002). The problem we study in this paper is most closely related to a budgeted version of the capacitated maximum covering location problem, somewhat related to facility location problems such as the p-median problem, and only remotely related to the traditional (transportation) hub location problems. Hence, in this section, we restrict attention to covering problems.

A review of results on covering problems can be found in Schilling *et al* (1993). Two covering location problems have appeared in the literature: set covering location problem (SCLP) and maximum covering location problem (MCLP). In these problems, a demand node is covered if it is within a given distance from a facility. The SCLP minimizes total cost of facilities located such that all demand nodes are covered. The MCLP maximizes total demand covered given that only a fixed number of facilities can be located. Capacitated versions have been studied by Current and Storbeck (1988) and Pirkul and Schilling (1991). Dynamic (multi-period) versions and stochastic versions of the uncapacitated problems are considered, respectively, by Gunawardane (1982) and Batta *et al* (1989). Schilling (1982) extends the uncapacitated MCLP problem to a stochastic case where demand at each node is stochastic and specified by a set of possible scenarios. Two stages are involved: in the first stage, locate a common set of facilities for all scenarios, and in stage two, locate scenario-dependent facilities. The objective is to maximize total weighted (expected) demand coverage over all scenarios in the MCLP given that a certain number of common facilities must be located in the first stage. Daskin (1983) and Batta *et al* (1989) consider another stochastic covering model where a facility is busy with some probability and thus demand at a node is covered by a facility with some probability.

Ideas for heuristics to solve the capacitated MCLP are briefly described in Current and

Storbeck (1988), Chung (1986) and Daskin (1995). Two main approaches are employed: (i) Greedy algorithms and (ii) Mathematical Programming methods, typically using Lagrangean Relaxation. Papers exploiting mathematical programming approaches for solving telecommunications and location problems include: Balakrishnan, Magnanti, and Wong (1995), Balakrishnan *et al* (2002), and Barahona and Chudak (1999). Our focus will be on greedy heuristics but not on mathematical programming approaches. We describe a simple greedy approach tailored to our hub location problem. We present an analytical worst-case performance guarantee for the greedy heuristic and demonstrate that the empirical performance (based on computational experiments) is on average within 2% of the optimum. As we show later, our hub location problem may be viewed as a maximum coverage problem (Hochbaum 1997). Consequently some of our analysis will draw upon related past research on the budgeted maximum coverage problem (Khuller, Moss and Naor 1999).

3. Problem Modeling

The modeling methodology will determine the hub placement and hub assignment decisions in "single-stage" systems (in which hub placement is done exactly once at the beginning of the planning horizon and no subsequent hub expansions are permitted, but hubs may be assigned to different building demands over time). The objective is to maximize the total expected demand covered by the hubs over the planning horizon subject to technological (crane-rain and line-of-sight) constraints and the budget constraint on hubs installed. This formulation was considered more appropriate than the alternative of minimizing cost of hubs installed subject to covering a fixed percentage of total demand.

There are N buildings $\mathcal{N} = \{1, 2, ..., N\}$ in the region being considered. There are H different types of hubs, $\mathcal{H} = \{1, 2, ..., H\}$; type-$h$ hub has unit cost C_h and capacity D_h. Let R be the budget available to install hubs. All hubs have the same crane-rain radius (a building within this radius can communicate with this hub if the line-of-sight constraint is satisfied). Our model incorporates hub-to-antenna line-of-sight and crane-rain radius constraints using a matrix M with $(i,\ j)^{th}$ entry M_{ij} equal to 1 iff building i is within the line-of-sight and crane-rain radius of building j, and 0 otherwise. Alternatively, let $A_j \subseteq \mathcal{N}$ denote the set of buildings that can be covered by a hub located on the top of building j. Let $B_i = \{j | i \in A_j\}$ be the set of all buildings from which a hub can cover building i. That is, $A_j = \{i | M_{ij} = 1\}$, and $B_i = \{j | M_{ij} = 1\}$.

Demand and hub capacities are measured in DS0 equivalents – A DS0 is equivalent to a band-width of 64 kbps (kilobits per second), refer to Newton and Horak (2002). We assume that each building can host multiple hubs. Further, the demand at a building may be partially or fully covered by one or more hubs. Building demand is uncertain and could vary with time. Let Ω be the sample space of random variables that affect demand; a specific scenario of uncertainty will be denoted by ω, for $\omega \in \Omega$. Let $d_{jt}(\omega)$ be the realized demand at building j in time period t under scenario ω. Note that each scenario ω yields a realization of demand over the entire horizon. (This model may be easily generalized to the case where realization ω_t is different for each t. In this case, the sample space Ω must be enlarged to be T-dimensional.) Assignment of demand to installed hubs is a dynamic recourse action (which assumes that the actual realized demand is unknown before period t and becomes known at t when the assignment or re-assignment decision is made).

The problem is to determine: (i) The hub placement decision, i.e., how many hubs of each type should be placed on the top of each building, at the beginning of the planning horizon, without violating the budget constraint. We use x_{jh} to denote the number of type-h hubs placed on building j. (ii) The demand assignment decision, i.e., how much realized demand from which buildings should be assigned to which hubs, in each time period, to maximize the total expected demand covered over the planning horizon subject to the capacity, crane-rain radius, and line-of-sight constraints. We use $y_{ijht}(\omega)$ to denote the amount of demand $d_{it}(\omega)$ at building i in scenario w that is covered by type-h hubs placed on building j. With this notation, the problem may be formulated as the following two-level stochastic program:

Level 1: Hub placement decision (made before customers actually subscribe to the service, i.e., before demands are known),

$$[\mathbf{L1}]: \quad \mathrm{Max}_{x \geq 0} \;\; E_{\omega \in \Omega}[\sum_{t=1}^{T} Q_t(x, \omega)] \tag{1}$$

$$\text{s.t.} \quad \sum_{j \in N} \sum_{h \in \mathcal{H}} C_h x_{jh} \leq R, \tag{2}$$

where $Q_t(x, \omega)$ is the total demand covered in period t after the demand assignment decision (see the Level 2 problem described below). The objective, (1), maximizes expected total demand covered over the entire planning horizon; (2) enforces the budget constraint on hub placement, with R denoting the budget available.

Level 2: Hub assignment decisions for time periods $t = 1, 2, \cdots, T$, are recourse actions made after hubs are placed and demands materialize.

$$
\begin{aligned}
[\mathbf{L2t}\omega]: \quad Q_t(x,\omega) &= \mathrm{Max}_{y\geq 0} \sum_{i\in N}\sum_{j\in B_i}\sum_{h\in\mathcal{H}} y_{ijht}(\omega), && (3)\\
\text{s.t.} \quad \sum_{j\in B_i}\sum_{h\in\mathcal{H}} y_{ijht}(\omega) &\leq d_{it}(\omega), \ \text{for } i\in N, && (4)\\
\sum_{i\in A_j} y_{ijht}(\omega) &\leq D_h x_{jh}, \ \text{for } j\in N, \ h\in\mathcal{H}. && (5)
\end{aligned}
$$

Problem $\mathbf{L2t}\omega$ maximizes the total demand coverage in period t, given that the hub placement decision x is already fixed, and the demand realization follows sample ω. Constraints (4) enforce the fact that each building i has limited demand and (5) represents the coverage capacity limit on type-h hubs placed at building j. Given the demand scenarios, the $T|\Omega|$ problems in $\mathbf{L2t}\omega$, for $t = 1,\cdots,T$, and $\omega\in\Omega$ may be solved independently.

4. Solution Approach and Analysis

We first characterize a relation between hub cost and capacity data that allows us to restrict the set of problem instances we need to consider. We also present an equivalence between the T-period problem and a single-period problem, assuming all scenarios are known at time 0. Next, we demonstrate that our hub location problem **L1** is NP-hard. Finally, we present a practical greedy solution algorithm and analyze its worst-case performance.

4.1 Model Properties

Wlog, we assume that the hub types are ordered in increasing unit cost so that $C_1 < C_2 < \cdots < C_H$. This implies that hub capacities are also ordered $D_1 \leq D_2 \leq \cdots \leq D_H$, since a costlier hub with lower capacity should not be selected in an optimal solution.

Property 1 *Given a problem instance with two different hub types, h and h', $1\leq h < h' \leq H$, satisfying*

$$C_{h'} \geq C_h \left\lceil \frac{D_{h'}}{D_h} \right\rceil,$$

there exists an optimal solution which does not use hub h'. Hence, wlog, type-h' hubs need not be considered.

Proof. By construction, see Appendix A for details. ■

Henceforth, wlog, we restrict attention to problem instances in which any pair of hubs h and h', $1\leq h < h' \leq H$, satisfy $C_{h'} < C_h \left\lceil \frac{D_{h'}}{D_h} \right\rceil$. The following two properties present equivalences between **L1** and a single period problem or a single scenario problem. In particular, Property 2 allows us to restrict attention to single period problem instances.

Property 2 *The T-period $|\Omega|$-scenario hub location problem* **L1** *is equivalent to the following single period $T|\Omega|$ hub location problem:*

$$Max_{x,y\geq 0} \frac{1}{|\overline{\Omega}|} \sum_{i\in N} \sum_{j\in B_i} \sum_{h\in \mathcal{H}} \sum_{\overline{\omega}\in\overline{\Omega}} y_{ij h\overline{\omega}} \tag{6}$$

$$s.t. \quad \textstyle\sum_{j\in B_i} \sum_{h\in\mathcal{H}} y_{ijh\overline{\omega}} \leq d_{i\overline{\omega}}, \ for\ i \in N,\ \overline{\omega}\in\overline{\Omega}, \tag{7}$$

$$\textstyle\sum_{i\in A_j} y_{ijh\overline{\omega}} \leq D_h x_{jh}, \ for\ j\in N,\ h\in\mathcal{H},\ \overline{\omega}\in\overline{\Omega}, \tag{8}$$

$$\textstyle\sum_{j\in N}\sum_{h\in\mathcal{H}} C_h x_{jh} \leq R, \tag{9}$$

where $\overline{\Omega} \equiv \{\overline{\omega} = (\omega, t)|\omega\in\Omega,\ t\in T\}$ denotes the set of new scenarios and $|\overline{\Omega}|$ is its cardinality.

Proof. The only difference between the formulation for **L1** in (1-5) and the one in (6-9) above, is that we have replaced each ω, t pair with $\overline{\omega}$ and expanded the scenario set accordingly to $\overline{\Omega} \equiv \{\overline{\omega} = (\omega, t)|\omega\in\Omega,\ t\in T\}$ with $d_{i\overline{\omega}} = d_{it}(\omega)$. No constraints have been added or deleted, and the objective in (6) is clearly within a constant factor of the objective in (1). ■

Property 3 *The T-period $|\Omega|$-scenario hub location problem* **L1** *is equivalent to a single scenario (deterministic) hub location problem with $T|\Omega|$ time periods.*

Property 4 *The hub location problem* **L1** *is NP-hard.*

Proof. See Appendix A for a reduction from the minimum dominating set problem. ■

In the remainder of this section, we present a greedy solution algorithm for **L1** and investigate its worst-case performance.

4.2 A Greedy Algorithm for Hub Location

There are many alternative approaches available to solve hub location problems. In this paper, we do not explore different solution methods to evaluate how they perform relative to each other. Instead, our goal is to describe a simple greedy algorithm (akin to the ones commonly used in practice) and to evaluate its worst-case and empirical average-case performance relative to the optimal. Note that the greedy algorithm presented below is an enhanced version of the greedy approaches we have seen in practice. For instance, our greedy solves the level 2 assignment problem optimally using linear programming, whereas, in a practical greedy, the assignment problem may also be solved approximately. Further, we combine two approaches (Plan 1 and Plan 2) into our greedy approach.

Our greedy algorithm selects a solution with the higher expected demand coverage from among the following two hub location plans:

Plan 1: Let Q^R_{hi} be the maximum *extra* expected demand covered by a type-h hub located at building i. Q^R_{hi}, which does not include demands already covered by previously located hubs, may be computed by solving a linear program akin to (6)-(9).

Select (h^*, i^*) that maximizes Q^R_{hi}/C_h and satisfies $C_h \leq R$. Ties in h are broken by choosing the hub of lower cost; ties in choice of building i are broken arbitrarily.

If $Q^R_{h^*i^*} = 0$ or $C_h > R$ for every $1 \leq h \leq H$, then Plan 1 terminates; else locate a type-h^* hub at the building i^*, set $R = R - C_{h^*}$, and repeat the above procedure.

Plan 2: Locate hubs as in Plan 1 using the criterion of maximizing Q^R_{hi} instead of Q^R_{hi}/C_h.

We use Plan 2 in the above algorithm because there are extreme problem instances in which it is optimal to pick the smallest hub that can cover the largest extra demand. However, Plan 1 may not pick this hub when its coverage to cost ratio (Q^R_{hi}/C_h) prevents this optimal hub from being chosen. In such cases, Plan 2 becomes necessary to ensure that a performance guarantee (of 39.4%) will apply. Our analysis of the above greedy algorithm will require the following property of Plan 1.

Lemma 5 *For the single scenario, single period, hub location problem* **L1** *with integer hub cost and budget data, Plan 1 selects hubs from more expensive hub types to less expensive hub types (in the order of decreasing costs and capacities).*

Proof. See Appendix A for a proof by contradiction. ■

As shown later, the above property and existing results immediately yield the well-known worst-case performance guarantee of $1 - \frac{1}{e}$ for the single scenario, single period special case. The performance of the simple greedy for the multi scenario general case is somewhat more complicated since, as illustrated below, results such as Lemma 5 do not apply:

Example 6 *Consider a single period instance with two types of hubs with unit costs and capacities as follows:* $C_1 = 1\$/unit$, $D_1 = 3$; $C_2 = 1.2\$/unit$, $D_2 = 4$. *There are two buildings with two demand scenarios. For building 1, the two scenarios are 5 and 1, for building 2, the scenarios are 3 and 3. The total budget is at least 2.2$ and both buildings are within the crane-rain radius and line-of-sight limits. With this data, it is easy to verify that the greedy will first pick a smaller type-1 hub to cover the demand scenarios of building 2* ($Q^R_{12} = 3$) *and then a larger type-2 hub to cover those of building 1* ($Q^R_{21} = 2.5$)*. However, even though the property in Lemma 5 is violated, the hubs installed by the greedy in this example are optimal.*

In addition to the above, as we explain later, even for the single scenario case, existing semi-enumerative greedy algorithms in the literature for the budgeted maximum coverage problem (Khuller *et al* 1999) require impractical enumeration to assure good performance. Developing effective guarantees for our simple greedy approach applied to the general case of **L1** is the topic of the next section.

4.3 Analysis of the Greedy Algorithm

We first review relevant results from the literature on performance of the greedy approach for our telecommunication hub location problem. We then present improved, data-dependent worst-case performance guarantees for our simple greedy algorithm.

4.3.1 Preliminary Results

Note that the telecommunication hub location problem is related to the following set covering problem: Given a set family $\mathcal{S}$, with each set consisting of weighted elements, and an integer $k > 0$, the maximum coverage problem (MCP) is to find k sets from $\mathcal{S}$ that maximize the total weight of elements covered by the k sets. This problem is clearly NP-hard (see Hochbaum 1997). However, it is well-known that there is a greedy algorithm for MCP which selects k sets by iteratively picking the set that covers the maximum weight group of currently uncovered elements. Let $wt(\text{GREEDY})$ and $wt(\text{OPT})$ denote, respectively, the total weight covered by the set family GREEDY $\subseteq \mathcal{S}$ selected by the greedy algorithm and the total weight of the optimal coverage. Then, we have,

$$wt(\text{GREEDY}) \geq (1 - \frac{1}{e})wt(\text{OPT}), \tag{10}$$

where $1 - \frac{1}{e} \approx 62.3\%$.

Our telecommunication hub location problem is related to the above maximum coverage problem in that we have demands that need to be covered and buildings and hub-types that need to be selected for hub installation. However, our problem has the following additional features relative to the traditional maximum coverage problem: (i) The cardinality of the set family $\mathcal{S}$ is very large since there are uncountably many ways of forming sets to cover demands. (ii) Different selected sets incur different costs (which are limited to a set of H hub-installation costs) and there is a finite total budget, (iii) There are technological (e.g., capacity) constraints on coverage, and (iv) We have multiple demand scenarios and the weight covered by a selected set depends on these scenarios.

Feature (ii) above was incorporated in Khuller, Moss and Naor (1999). They analyzed the following budgeted maximum coverage problem: Given a collection $\mathcal{S}$ of sets with associated costs defined over a domain of weighted elements, and a budget L, find a subset of $\mathcal{S}' \subseteq \mathcal{S}$ such that the total cost of sets in $\mathcal{S}'$ does not exceed L, and the total weight of elements covered by $\mathcal{S}'$ is maximized. Their work proposed a simple greedy algorithm and proved that its worst-case guarantee is $1 - \frac{1}{\sqrt{e}} \approx 39.4$ %; this guarantee is nearly tight (they identify an instance for which greedy performance was as low as 44%). The performance guarantee can be improved to $1 - \frac{1}{e}$ by enumerating all subsets of $\mathcal{S}$ of cardinality k (for $k \geq 3$) which have cost at most L. As shown below, the collection of sets $\mathcal{S}$ for hub location is large. Consequently the enumerative algorithm in Khuller *et al* turns out to be impractical for our hub location problem since there can be uncountably many subsets of size k.

In order to better understand how existing results on the maximum coverage problem apply to our hub location problem, it is insightful to map an instance of hub location to an instance of budgeted maximum coverage. We explain this mapping in the remainder of this section. For simplicity, our analysis will restrict attention to integer hub cost data. By Property 2, wlog, it suffices to consider the single period problem (so we drop the t subscript). For notational convenience, henceforth, we will use ω and Ω synonymously with $\overline{\omega}$ and $\overline{\Omega}$ (defined in Property 2). For $i = 1, \ldots, B$ and $\omega \in \Omega$, we represent building i demand in scenario ω as a planar region with area equal to the demand magnitude $d_{i\omega}$. Demand from different scenarios correspond to separate parallel planes. For each scenario ω, demands from different buildings correspond to different nonintersecting regions in the ω-th plane. Let Γ be the set of all points in the $|B||\Omega|$ regions of the $|\Omega|$ planes described above.

Let $\mathrm{mcg}(h, i, \omega)$ be an arbitrary point set in Γ that corresponds to a maximum demand coverage in scenario ω by a type-h hub located at building i. This coverage implicitly incorporates hub capacity, crane-rain and line-of-sight constraints. Let $\mathrm{mcg}(h,i) \equiv \bigcup_{\omega\in\Omega} \mathrm{mcg}(h,i,\omega)$. In general, there are uncountably many $\mathrm{mcg}(h,i)$ for each h and i because of the uncountably many ways of maximally covering demand scenario ω using a type-h at building i. Let $\alpha_{hi} \equiv |\mathrm{mcg}(h,i)|$ denote the coverage amount of the set $\mathrm{mcg}(h,i)$. Then, for the maximum coverage problem, the weight of element $\mathrm{mcg}(h,i)$ is set to α_{hi} and its cost is C_h. For any pair (h,i), define a set family $S_{hi} \equiv \{T | T \in \mathrm{mcg}(h,i) \text{ for some } \mathrm{mcg}(h,i)\}$. Let $\mathcal{S} \equiv \bigcup_{\substack{1\le i\le B}}^{1\le h\le H} S_{hi}$. Finally setting $L \equiv R$ yields an instance of the budgeted max coverage problem.

In the next section we show how the above translation provides several performance guarantees for our greedy hub location algorithm.

4.3.2 Performance Guarantees for Plan 1

Given the translation of hub location to a specialized budgeted maximum coverage problem, we see that our greedy algorithm may be viewed as a simple translation of Khuller *et al*'s basic greedy algorithm to fit our environment. Consequently, we can obtain the following important result (which follows from their proofs of performance guarantees for the budgeted maximum coverage problem):

Theorem 7 *If the (h^*,i^*) that yields the maximum Q^R_{hi}/C_h always satisfies $C_{h^*} \le R$ during Plan 1 until the budget R entirely runs out, then the expected demand coverage by Plan 1 is at least $1-\frac{1}{e}$ of the optimal demand coverage.*

Special case: Consider a deterministic single period hub location problem (with only one demand scenario). Assume that $C_h \equiv C_0 z^{f(h)}$, where $C_0 \in R_+$, $z \in Z_+$, and $f(h) : Z_+ \to Z_+$ is a strictly increasing function of h; let $R \equiv pC_H$, where $p \in Z_+$.

Let $cg(\mathrm{PLAN1})$ be the total expected demand coverage for the hub location solution obtained by Plan 1 and let $cg(\mathrm{OPT})$ be the maximum expected demand coverage. Then, the following result states that the greedy solution covers at least 63.2% of the optimal demand coverage.

Theorem 8 *$cg(PLAN1) \ge (1-\frac{1}{e})cg(OPT)$ in the single-scenario special case defined above.*

Proof. The result follows from Lemma 5 and Theorem 7: The special cost-and-budget structure and Lemma 5 ensure that the (h^*,i^*) that yields the maximum Q^R_{hi}/C_h always

satisfies $C_{h^*} \leq R$ during Plan 1 until R runs out. Now Theorem 7 yields the required result.
■

Thus, Theorem 8 shows that for the single period deterministic version of our hub location problem, the greedy in Section 4.2 has a performance guarantee of at least 63.2%. Since Lemma 5 does not always hold true in the multi-scenario case, the above bound does not follow. In the remainder of this section, we develop a valid bound for the general case of problem **L1**.

Wlog, we assume that C_i and R are all positive integers. Define $\gamma_{hh'} \equiv \min(\frac{D_h/C_h}{D_{h'}/C_{h'}}, 1)$, for any $1 \leq h < h' \leq H$, and let $\gamma_{\min} \equiv \min\{\gamma_{hh'} : 1 \leq h \leq h' \leq H\}$. Given a hub location problem, let (h^*, i^*) be the first pair of hub type and building that Plan 1 selected according to the criterion Q^R_{hi}/C_h but could not locate because C_{h^*} exceeded the residual budget. Let β denote the amount of this residual budget, and let p^* be the number of iterations that Plan 1 takes to reduce the budget to β.

Let $\mathcal{U}$ be the amount of expected demand coverage in the first p^* iterations of Plan 1. Let β_h, $1 \leq h \leq h^* - 1$, be the total cost of the type-h hubs located by Plan 1 after the p^*-th iteration. Let $\psi \equiv \frac{\mathcal{U}}{\mathcal{U}_{i^*}}$, where $\mathcal{U}_{i^*} \equiv \frac{Q}{C_{h^*}}$ and Q denotes the amount of extra expected demand that would be covered in the (p^*+1)-th iteration of Plan 1 if a type-h^* hub were located at building i^*. Then we have the following data-dependent worst-case performance guarantee:

Theorem 9 *Let PLAN1 be the solution generated by Plan 1 for hub location and let OPT be the optimal solution. Then*

$$cg(PLAN1) \geq \left(\frac{\psi + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h}{\psi + \beta}\right)\left(1 - \frac{1}{e}\right) cg(OPT).$$

Theorem 9, which follows from the lemmas stated below, allows assessment of the error when the greedy algorithm is run. (The proofs of results stated below are included in Appendix A.)

Lemma 10 $\mathcal{U} + \beta\mathcal{U}_{i^*} \geq (1 - \frac{1}{e})cg(OPT)$.

By the definition of β_h, there are β_h/C_h type-h hubs located by Plan 1 since the (p^*+1)-th iteration. Now consider an alternative hub location plan. The hubs that are selected and located in the first p^* iterations by Plan 1 remain unchanged, whereas the hubs that are selected after the p^*-th iteration by Plan 1 are all located at building i^*. Call this alternative plan Plan 3, and let cg(PLAN3) be the total expected demand coverage by this new hub location plan.

Lemma 11 *$cg(PLAN1) \geq cg(PLAN3)$.*

Lemma 12

$$cg(PLAN3) \geq \left(\frac{\psi + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h}{\psi + \beta}\right)(\mathcal{U} + \beta \mathcal{U}_{i^*}).$$

Lemmas 10-12 together prove Theorem 9. Although the result is data-and-process dependent, it shows that the performances of Plan 1 and the greedy algorithm can be enhanced in many hub location situations over the $1 - \frac{1}{\sqrt{e}}$ obtained from the general budgeted maximum coverage context (Khuller *et al*). In the remainder of this section, we illustrate some useful implications of Theorem 9. In the text below, we assume that $C_1 = 1$ in Corollary 13 below. This is true wlog in most practical cases and it ensures that the total budget will be exhausted whenever the budget is insufficient to cover all the demands. Under this assumption we have $\beta = \sum_{h=1}^{h^*-1} \beta_h$. Then

Corollary 13 $$cg(PLAN1) \geq \delta \left(1 - \frac{1}{e}\right) cg(OPT),$$

for $(i)\ \delta = \left(1 - \frac{(C_{h^*} - 1)(1 - \gamma_{\min})}{R}\right)$ *or* $(ii)\ \delta = \max\{1 - \frac{C_{h^*}}{R}, \gamma_{\min}\}$.

Proof: We prove (i), from which (ii) follows directly. We know $\frac{\psi + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h}{\psi + \beta} \geq \frac{(R-\beta)+\gamma_{\min}\beta}{(R-\beta)+\beta} = 1 - \frac{\beta(1-\gamma_{\min})}{R}$, since $\frac{\psi + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h}{\psi + \beta}$ is a nondecreasing function of ψ and $\sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h$. Because $1 - \frac{\beta(1-\gamma_{\min})}{R}$ is a nonincreasing function of β, we have $1 - \frac{\beta(1-\gamma_{\min})}{R} \geq 1 - \frac{(C_{h^*}-1)(1-\gamma_{\min})}{R}$. ■

A slightly weaker *a priori* bound than Corollary 13 may be obtained by replacing C_{h^*} with C_H. If all hub types are close to being identical, then $\gamma_{\min}$ is approximately 1 and we get a performance bound near $1 - \frac{1}{e}$. As the budget R increases relative to C_H (or C_{h^*}), the performance bounds also improve up to $1 - \frac{1}{e}$. Finally, note that, since the greedy algorithm in Section 4.2 chooses the better one of Plan 1 and Plan 2, its performance is never worse than that of Plan 1. Therefore, all results from Theorem 9 to Corollary 13 hold for our greedy algorithm as well. In the next section, we investigate the empirical, average-case performance of the greedy algorithm for hub location.

5. Computational Results

First we describe the test problem instances, next we summarize performance of the greedy heuristic from Section 4.2. Note that, we do not describe other heuristics we tested such as an LP-based "price-and-round" approach which rounds the solution obtained from the LP relaxation of **L1**. In preliminary testing, the simple greedy outperformed all our simple LP rounding methods.

5.1 Data Generation

Two sets of problems were generated: Small and Large. Small problems were mostly solved to optimality using CPLEX. In this case, the greedy heuristic's solution can be compared to the optimal and an upper bound (from a LP relaxation of **L1**). Large problems are more representative of industrial sized problems, however, optimal solutions are clearly difficult to find (in preliminary testing, CPLEX could not solve these problems even after five days). In the latter case, solution quality of the greedy may only be assessed by comparison with the LP upper bound.

In all problem instances, data was generated to reflect industrial problems. Number of hub types was 4; their unit costs and capacities followed Property 1 (so that capacity/cost was increasing, corresponding to economies of scale; for instance, the small problem instances had $C_h = 1, 2, 4, 6$ with $D_h = 200, 440, 1000$, and 1800). Values of other parameters such as number of buildings, scenarios, periods are shown in Table 1; with data typically generated using a uniform distribution. The choice of number of scenarios and periods was made to ensure that the corresponding linear programming relaxation of **L1** has a reasonable number of variables and constraints. Further, changing these parameters did not appear to affect quality of solutions obtained by the greedy heuristic. The locations of the buildings were randomly generated within a square region using longitude, latitude and distance data from the industrial application. Both urban and suburban problem instances were generated by adjusting the size of the square and the density of buildings. A typical crane-rain radius was approximately one mile (for a square region with total area of nine to twenty-five square miles). Given the building locations and the crane-rain radius, the building-to-building incidence relations are obtained. Line-of-sight restrictions were imposed using a separate randomly generated incidence matrix.

The demand at any building is partitioned into voice demand and data demand. The

Table 1: Generated Problem Data for Computational Experiments

	Small Problems	**Large Problems**
Total Number of Instances	100	100
Number of buildings, B	$10 \leq B \leq 20$	$200 \leq B \leq 400.$
Number of Scenarios, $\lvert\Omega\rvert$	2 to 8	2
Number of Periods, T	2 to 6	2
Budget, R	Low, Medium, High	High

demands in the first period are randomly generated (uniform between 100 and 450) with voice representing about 80% of the total demand and data contributing the remaining 20%. In the following time periods, voice demand increases by 4% to 8% (uniform between 4% to 8% in each period), while data demand grows by 20% to 25% in each period. For each problem instance, we solve a linear program to approximate the minimum budget level, $R_{\min}$, required to satisfy all demand. Small problem instances were solved for three budget levels with $R = kR_{\min}$, and $k = 80\%$ (high), 50% (medium), and 20% (low). Large problems were solved with only the high budget (which was considered most representative of real problems).

5.2 Solution Quality

Let $cg(\text{OPT})$ be the optimal expected coverage for the hub location problem, and let $cg(G)$ be the corresponding coverage by the greedy solution. Let $cg(\text{LP})$ denote the optimal value of the LP relaxation of **L1**. Clearly, $cg(G) \leq cg(\text{OPT}) \leq cg(\text{LP})$. We report solution quality of our greedy heuristic using the following ratios:

$$\epsilon_{\text{LP}}^{G} \equiv \frac{cg(G)}{cg(\text{LP})}, \quad \epsilon_{\text{OPT}}^{G} \equiv \frac{cg(G)}{cg(\text{OPT})}, \quad \epsilon_{\text{LP}}^{\text{OPT}} \equiv \frac{cg(\text{OPT})}{cg(\text{LP})}. \tag{11}$$

Clearly, $\epsilon_{\text{LP}}^{G} \leq \epsilon_{\text{OPT}}^{G} \leq 1$ and $\epsilon_{\text{LP}}^{G} \leq \epsilon_{\text{LP}}^{\text{OPT}} \leq 1$. Hence, ϵ_{LP}^{G} is a lower bound on both $\epsilon_{\text{OPT}}^{G}$ and $\epsilon_{\text{LP}}^{\text{OPT}}$. Further, the larger the ratio $\epsilon_{\text{OPT}}^{G}$ is (100% being the maximum permissible value), the better the quality of the greedy solution. The ratio $\epsilon_{\text{OPT}}^{G}$ is only reported for small problems for which the optimal solution could be determined; for large problems we report the ratio ϵ_{LP}^{G} which is a lower bound on the performance of the greedy (when compared to the optimal).

Small problems: Of the generated 100 small test problem instances, the optimal solutions of eight could not be obtained using CPLEX within a run time of twelve hours on a Pentium III computer (800 MHz Precision 220). In these cases, we simply take the best identified solution as the optimal solution and 43,200 seconds as the corresponding running time. The results are presented in Table 2, where we see that, on average, the greedy solution was within 0.4% of optimal and within 3.4% of the LP upper bound. Thus it appears that the greedy solution is very close to optimal, which itself is somewhat close to the upper bound. The heuristic is two orders of magnitude faster than the optimal.

Table 2: Greedy's Performance and Run-Time vs. Optimal (small problems)

	$\epsilon_{\mathrm{LP}}^{\mathrm{OPT}}$	$\epsilon_{\mathrm{OPT}}^{G}$	$\epsilon_{\mathrm{LP}}^{G}$	Greedy Time (sec.)	CPLEX Time (sec.)
Average	97.09%	99.67%	96.77%	15.91	4122.53
Std. Dev.	2.51%	0.86%	2.45%	19.72	11994.12
Min.	90.37%	95.99%	90.37%	1.20	0.4
Max.	100%	100%	100%	106.50	43200.00

Large problems: For the generated 100 large problem instances, in Table 3, we summarize the $\epsilon_{\mathrm{LP}}^{G}$ values and the running time of the greedy algorithm. For these large problem instances, the main portion of the running time is due to solving several LPs per iteration of the greedy step. The number of iterations, which depends on the budget R, was relatively small (our experience with this problem environment indicated that typical demand regions should be covered by a total of less than 30 hubs). Note that, the LPs for large instances had thousands of constraints and hundreds of thousands of variables; the greedy algorithm solves many large-scale linear programs to determine the hub type and location that maximizes the extra demand coverage. Hence, the greedy algorithm takes nearly 6000 seconds on average to solve each problem. Given that the hub location problem is a strategic problem, solved once over the planning horizon (which exceeds six months), a running time of under six hours was considered acceptable. Further, in preliminary testing, the corresponding optimal CPLEX procedure failed to terminate with an optimal solution after five days; such a running time (of the order of one week) was considered unacceptably high since it did not permit the necessary what-if analyses on problem parameters.

Table 3: Greedy's Performance and Run-Time (large problems)

	$\epsilon_{\mathrm{LP}}^{G}$	Greedy Time (sec.)
Average	98.05%	5986.79
Std. Dev.	1.81%	6472.42
Min.	91.8%	534.40
Max.	100%	44123.20

Based on results in Table 3, we see that the greedy solution was near optimal (within 2% of the upper bound and hence within 2% of optimal). Consequently, we conclude that even for industrial-sized problem instances the greedy algorithm will likely exhibit good performance (whose quality can be assessed by comparison with the LP upper bound or by Theorem 9), effectively balancing the trade-off between solution quality and run time. Finally, we note that the performance of the greedy heuristic was not affected by the tested problem parameters such as urban vs. suburban building densities, demand growth rates, and budgets. We conclude that the proposed greedy heuristic is a robust, effective, and efficient method for the fixed-wireless hub location problem.

6. Summary

In this paper, we presented a model for telecommunication network installation by companies in the Broadband-access business, specialized to the fixed-wireless case. Our multi-period model for fixed-wireless capacitated hub location under multiple demand scenarios incorporated technological line-of-sight and crane-rain radius constraints, and a limited budget R for hub installation. We presented a greedy solution heuristic and showed that its worst-case performance was better than $(1-\frac{C_H}{R})(1-\frac{1}{e})$, where C_H is the cost of the most expensive hub. This analysis applied and extended results from the budgeted maximum coverage problem to our hub location problem. We demonstrated computationally that, over a wide variety of problem instances, the greedy algorithm produced solutions that were on average within 2% of the optimal; performance of the greedy heuristic did not deteriorate significantly with problem size or other problem parameters.

Possible extensions of this paper include: (i) Investigation of mathematical programming based heuristics. (ii) Studying the multi-stage problem in which hub installations are permitted at several time points within the planning horizon. (ii) Developing managerial insights such as the effect on profitability of different demand scenarios (either different number of scenarios or different voice and data demand growth models) and varying budget values. (iii) Linking the hub location problem to other network planning problems such as hub configuration (engineering) and long-haul backbone network design. (The hub location Broadband-access problem we address in this paper is only one part of network planning. Given the location of hub types, the hub configuration problem engineers the different hubs to be installed. The backbone network problem (Klincewicz 1998) identifies the service nodes to be installed that will interact with the hubs.) (iv) Providing methods to help wireless service providers to prioritize and size the different regions to expand into. Some of these problems form the basis of our current research.

A. Appendix: Proof of Results

Proof of Property 1: Consider an optimal solution using u_h hubs of type-h and $u_{h'}$ hubs of type-h', with $C_{h'} \geq C_h \lceil \frac{D_{h'}}{D_h} \rceil$. Total relevant cost of using type-h and h' hubs is $C_h u_h + C_{h'} u_{h'}$. Construct a new solution by replacing one type-h' hub with $\lceil \frac{D_{h'}}{D_h} \rceil$ type-h hubs, while keeping everything else unchanged. The incremental total cost is $C_{h'}(u_{h'} - 1) + C_h(u_h + \lceil \frac{D_{h'}}{D_h} \rceil) - C_{h'} u_{h'} - C_h u_h = -C_{h'} + C_h \lceil \frac{D_{h'}}{D_h} \rceil \leq 0$. Hence, the new solution does not exceed the given budget. Furthermore, the additional $\lceil \frac{D_{h'}}{D_h} \rceil$ type-h hubs can cover at least the demand that the deleted type-h' hub covered, because $\lceil \frac{D_{h'}}{D_h} \rceil D_h \geq D_{h'}$. Hence we see that all $u_{h'}$ type-h' hubs could be replaced by $u_{h'} \lceil \frac{D_{h'}}{D_h} \rceil$ type-h hubs while maintaining the total demand coverage and decreasing the total cost. ■

Proof of Property 4: Our proof is based on a reduction from the minimum dominating set problem (MDSP) on graph $G(V, E)$. Note that, a dominating set for G is a subset $V' \subseteq V$ such that for all $u \in V \backslash V'$ there is a $v \in V'$ for which $(u, v) \in E$. The minimum dominating set problem is to find a dominating set for G of minimum cardinality, i.e., $|V'|$ is minimum.

Given an instance of MDSP, we construct a single-period, single-scenario instance of **L1** as follows: $H = 1$, there is only one type of hub of capacity $D_h = N = |V|$, cost $C_h = 1$, and total budget R (an arbitrary integer). Each vertex in G corresponds to a building in **L1**; demand at each building is 1. Finally, $j \in B_i$ iff $(i, j) \in E$.

Clearly, the minimum dominating set for G has size less than or equal to R iff the maximum demand coverage in **L1** is N. Since MDSP is NP-hard (Johnson 1974), deciding whether there exists a dominating set with size less than or equal to R is also NP-hard. Hence the constructed special version of **L1** is NP-hard. ■

Proof of Lemma 5: Suppose not. The greedy selects a type-h' hub after selecting a type-h hub, where $h < h'$, with the type-h' hub located at building i. Since $C_h < C_{h'}$, selecting the type-h' hub rather than the type-h hub is not due to the budget constraint. Hence, we can deduce that $Q^{\overline{R}}_{h'i} > Q^{\overline{R}}_{hi}$, where $\overline{R}$ is the remaining budget before selecting the type-h' hub. Further, when there is a single scenario, $Q^{\overline{R}}_{hi} = D_h$, since otherwise $Q^{\overline{R}}_{h'i} = Q^{\overline{R}}_{hi}$ and the greedy would not select the more expensive type-h' hub. Hence,

$$\frac{Q^{\overline{R}}_{h'i}}{C_{h'}} > \frac{Q^{\overline{R}}_{hi}}{C_h} = \frac{D_h}{C_h}.$$

Now consider the type-h hub that the greedy located before, say, at building j. Since $Q^{\overline{R}}_{hi} = D_h$ and $\widetilde{R} \geq \overline{R}$, we must have had $Q^{\widetilde{R}}_{hj} = D_h$, where $\widetilde{R}$ is the remaining budget before locating this type-h hub. Consequently, we have

$$\frac{D_h}{C_h} = \frac{Q_{hj}^{\widetilde{R}}}{C_h} \geq \frac{Q_{h'i}^{\widetilde{R}}}{C_{h'}} \geq \frac{Q_{h'i}^{\overline{R}}}{C_{h'}} > \frac{D_h}{C_h}, \tag{12}$$

where the first inequality follows from the fact that the greedy picks the hub type h and building j with the largest $Q_{hj}^{\widetilde{R}}/C_h$ ratio. The second inequality in (12) follows from $\widetilde{R} \geq \overline{R}$. Finally, we note that inequality (12) yields a contradiction, which proves the lemma. ■

In order to prove Lemma 10, we first need to consider the following modified version of the standard greedy algorithm for MCP. Note that the only difference between the "greedy with deletion" described below and the standard greedy from MCP is that in the latter algorithm no sets are deleted ($\tilde{S}_l = \emptyset$ for all iterations l). Let S^0 denote the initial set family.

Greedy with deletion: $\overline{\text{GREEDY}} \leftarrow \emptyset$, $S^1 \leftarrow S^0$. For $l = 1, \cdots, k$ do

{ Select $G_l \in S^l$ that maximizes $wt(\overline{\text{GREEDY}} \cup G_l)$. Set $\overline{\text{GREEDY}} \leftarrow \overline{\text{GREEDY}} \cup G_l$, $S^{l+1} \leftarrow S^l \backslash \tilde{S}_l$, where $\tilde{S}_l$ satisfies $S^l \backslash \tilde{S}_l \neq \emptyset$, for $l \leq k-1$, and $G_i \notin \tilde{S}_l$ for $1 \leq i \leq l$. }

The output of the above algorithm is the set family $\overline{\text{GREEDY}}$. Let $\tilde{S} \equiv \cup_{l=1}^{k} \tilde{S}_l$ and $S \equiv S^0 \backslash \tilde{S}$. Inequality (10) yields the following:

Lemma 14 *If $\overline{GREEDY}$ is the output of "Greedy with deletion" and OPT is an optimal solution to the maximum coverage problem with set family S and integer k, then* $wt(\overline{GREEDY}) \geq (1 - \frac{1}{e}) wt(OPT)$.

Proof of Lemma 14: Since $G_l \in S^0 \backslash \tilde{S} \equiv S$, $\overline{\text{GREEDY}} \subseteq S$, so $\overline{\text{GREEDY}}$ is feasible for the maximum coverage problem with set family S. Further, in the l-th iteration, from all the possible sets, greedy with deletion picks $G_l \in S_l$ that maximizes the extra coverage. Since, for all l, $S_l \supseteq S$ and $G_l \in S$, G_l is the set in S that maximizes extra coverage. Hence, we conclude that $wt(\overline{\text{GREEDY}}) \geq wt(\text{GREEDY})$. Now the required result follows from inequality (10). ■

Proof of Lemma 10: We will prove this lemma by applying the greedy with deletion to some specially designated set family S^0 and number k. Let the sets $\text{mcg}(h,i,\omega)$, $\text{mcg}(h,i)$, and Γ be as defined in the hub location to budgeted maximum coverage translation at the end of Section 4.3.1. As before, let $\alpha_{hi} \equiv |\text{mcg}(h,i)|$ and $\alpha_{hi\omega} \equiv |\text{mcg}(h,i,\omega)|$ respectively denote the coverage amounts of the sets $\text{mcg}(h,i)$ and $\text{mcg}(h,i,\omega)$. For any set $\text{mcg}(h,i,\omega)$, let $S_{\text{mcg}(h,i,\omega)}$ be any family of C_h nonintersecting subsets of $\text{mcg}(h,i,\omega)$ which partition it into equal pieces of area $\frac{\alpha_{hi\omega}}{C_h}$. Similarly define $S_{\text{mcg}(h,i)}$ as a partitioning of $\text{mcg}(h,i)$ into equal subsets of size $\frac{\alpha_{hi}}{C_h}$. Each subset in $S_{\text{mcg}(h,i)}$ is a union of $|\Omega|$ sets respectively from $S_{\text{mcg}(h,i,\omega)}$, for $\omega \in \Omega$. There are uncountably many $S_{\text{mcg}(h,i)}$ because there are uncountably many $S_{\text{mcg}(h,i,\omega)}$ for any set $\text{mcg}(h,i,\omega)$. For any pair (h,i), define a set family $S_{hi} \equiv \{T | T \in S_{\text{mcg}(h,i)}$ for some $S_{\text{mcg}(h,i)}$ from some $\text{mcg}(h,i)\}$. Let $S^0 \equiv \bigcup_{\substack{1 \leq h \leq H \\ 1 \leq i \leq B}} S_{hi}$. Let $k \equiv R \in Z_+$.

Now consider the greedy with deletion starting with S^0 and k. Let $T_1 \in S_{\mathrm{mcg}(h,i)}$ be a set with the largest demand coverage in S^0, i.e., $\frac{Q^R_{hi}}{C_h}$ is the largest. Hence, the greedy will select one such set T_1 as G_1. By the definition of $S_{\mathrm{mcg}(h,i)}$, there are totally C_h sets in $S_{\mathrm{mcg}(h,i)}$ of the same size as T_1 and these sets are nonintersecting. Hence, wlog, in the first C_h steps of the greedy, it selects these C_h sets sequentially until all these subsets in $S_{\mathrm{mcg}(h,i)}$ have been selected. Note that, $\tilde{S}_j = \emptyset$ for $1 \leq j \leq C_h - 1$. That is, no sets are deleted during the first $C_h - 1$ steps of the greedy with deletion.

At the end of the C_h-th step, let S_1 be the set of the points that are covered by those C_h selected sets in $S_{\mathrm{mcg}(h,i)}$. Let $\tilde{S}_{C_h}$ be a set family that consists of those families $S_{\mathrm{mcg}(\tilde{h},\tilde{i})}$ for which $|(P_1 \cap P(\omega))\backslash S_1| \neq |(P_2 \cap P(\omega))\backslash S_1|$ holds for some two sets P_1 and P_2 in $S_{\mathrm{mcg}(\tilde{h},\tilde{i})}$ and some plane $P(\omega)$. Thus, $\tilde{S}_{C_h}$ consists of those partitions of some demand coverage set $\mathrm{mcg}(\tilde{h},\tilde{i})$ whose intersections with some plane $P(\omega)$ result in unequal-sized subsets if S_1 is removed. Such partitions need not be considered by the greedy with deletion in later steps and are deleted. That is, let $S^{C_h+1} \equiv S^{C_h}\backslash\tilde{S}_{C_h} = S^0\backslash\tilde{S}_{C_h}$. Finally, note that, in the first C_h steps, the greedy with deletion performs exactly as the Plan 1 does in its first step (the latter chooses a type-h hub and then reduces R to $R - C_h$).

Now the greedy with deletion goes to the step $C_h + 1$. Starting at this step, it will do exactly as it did in the previous C_h steps, except that it might have another C_q steps with $\tilde{S}_j = \emptyset$ for $C_h + 1 \leq j \leq C_h + C_q - 1$, where C_q nonintersecting sets are sequentially selected from some $S_{\mathrm{mcg}(q,j)}$ and S_2 is the union of S_1 with the set of points covered by those C_q selected sets. At the end of the (C_h+C_q)-th step, we delete the corresponding set family $\tilde{S}_{C_h+C_q}$ from S^{C_h+1} in a manner analogous to the first C_h steps, where we use S_2 to define $\tilde{S}_{C_h+C_q}$. In other words, those families $S_{\mathrm{mcg}(h',i')}$ will be put into $\tilde{S}_{C_h+C_q}$ if there exist two sets $P_1, P_2 \in S_{\mathrm{mcg}(h',i')}$ and a plane $P(\omega)$ such that $|(P_1 \cap P(\omega))\backslash S_2| \neq |(P_2 \cap P(\omega))\backslash S_2|$. These C_q steps of greedy with deletion correspond to the second step of Plan 1, at the end of which R get reduced by C_q.

Having gone through all those p^* steps of Plan 1, similarly we see that they map to the $R - \beta$ steps of the greedy with deletion. When the greedy with deletion ends with another two steps with no sets deleted after each, then $wt(\overline{\mathrm{GREEDY}}) = \mathcal{U} + \beta\mathcal{U}_{i^*}$.

At the end of the greedy with deletion, let $\tilde{S} \equiv \cup_{l=1}^{k}\tilde{S}_l$ and $S \equiv S^0\backslash\tilde{S}$. We see that the deletion procedure above guarantees that there exists some $S_{\mathrm{mcg}(h',i')} \in S$ for any set $\mathrm{mcg}(h',i')$. Therefore, any optimal solution to the hub location problem corresponds to a feasible set cover under S and k. Therefore, $wt(\mathrm{OPT}) \geq cg(\mathrm{OPT})$. Finally, by Lemma 14, $wt(\overline{\mathrm{GREEDY}}) \geq (1 - \frac{1}{e})wt(\mathrm{OPT})$ and the required result follows. ■

Proof of Lemma 11: Wlog, suppose that k is the number of hubs located by Plan 1 after the p^*-th iteration. Let $cg(\mathrm{PLAN1}(p^* + j))$ be the expected demand coverage by Plan 1 as of the $(p^* + j)$-th iteration, and, correspondingly, let $cg(\text{PLAN3}(p^* + j))$ be the expected demand coverage by Plan 3 as of the $(p^* + j)$-th iteration, where $1 \leq j \leq k$. Since $cg(\text{PLAN1}) = cg(\text{PLAN1}(p^* + k))$ and $cg(\text{PLAN3}) = cg(\text{PLAN3}(p^* + k))$, we will prove the lemma if we show that $cg(\text{PLAN1}(p^* + j)) \geq cg(\text{PLAN3}(p^* + j))$ for every $1 \leq j \leq k$. We do so using induction on j.

When $j = 1$, the hub chosen and located by Plan 1 in the $(p^* + 1)$-th iteration covers the maximum extra expected demand, which is clearly not less than the extra expected demand coverage by the same type of hub located at building i^*. Let $ecg(\text{PLAN1}(p^* + 1))$ and $ecg(\text{PLAN3}(p^* + 1))$ be the corresponding extra expected demand coverages, then the above shows $ecg(\text{PLAN1}(p^*+1)) \geq ecg(\text{PLAN3}(p^*+1))$. Since $cg(\text{PLAN1}(p^*+1)) = \mathcal{U} + ecg(\text{PLAN3}(p^*+1))$ and $cg(\text{PLAN3}(p^*+1)) = \mathcal{U} + ecg(\text{PLAN1}(p^*+1))$, we see that $cg(\text{PLAN1}(p^*+1)) \geq cg(\text{PLAN3}(p^*+1))$.

Now we assume $cg(\text{PLAN1}(p^* + j)) \geq cg(\text{PLAN3}(p^* + j))$, and we will show $cg(\text{PLAN1}(p^* + j + 1)) \geq cg(\text{PLAN3}(p^* + j + 1))$, where $1 \leq j \leq k - 1$. Let $\text{PLAN1}^\omega(p^* + j)$ be the amount of the demand in scenario $\omega \in \Omega$ that can be covered by any hub located at building i^* but has not be covered yet as of the $(p^* + j)$-th iteration of Plan 1, where $0 \leq j \leq k$. Define $\text{PLAN3}^\omega(p^* + j)$ analogouly using Plan 3.

Now consider the $p^* + i + 1$-th iteration of Plan 1 and Plan 3. From the $(p^* + 1)$-th iteration until the $p^* + i$-th iteration, Plan 1 and Plan 3 locate the same set of hubs. The only difference between two plans is that Plan 1 locates the hubs at different buildings while Plan 3 just locates the hubs at building i^*. Hence we see that $\text{PLAN1}^\omega(p^* + j) \geq \text{PLAN3}^\omega(p^* + j)$ for every $\omega \in \Omega$. Even in the $p^* + j + 1$-th iteration, locating the hub selected by Plan 1 at building i^* would result in an extra demand coverage (relative to $\text{PLAN1}^\omega(p^* + j)$) that is at least as large as the extra demand coverage (relative to $\text{PLAN3}^\omega(p^* + j)$). Consequently, the hub selected and located by Plan 1 in the $(p^* + j + 1)$-th iteration has at least as much extra demand coverage from demand scenario ω as it has when located at building i^*. Since this happens for all $\omega \in \Omega$, $cg(\text{PLAN1}(p^* + j + 1)) - cg(\text{PLAN1}(p^* + j)) \geq cg(\text{PLAN3}(p^* + j + 1)) - cg(\text{PLAN3}(p^* + j))$. By the induction assumption, we conclude that $cg(\text{PLAN1}(p^* + j + 1)) \geq cg(\text{PLAN3}(p^* + j + 1))$. ■

Proof of Lemma 12: Let Q^ω be the extra amount of demand coverage from scenario $\omega \in \Omega$ in the $(p^* + 1)$-th iteration of Plan 1 if a type-h^* hub were located at building i^*. Let $\mathcal{U}_{i^*}^\omega \equiv \frac{Q^\omega}{C_{h^*}}$. Then we have $Q = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} Q^\omega$ and $\mathcal{U}_{i^*} = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \mathcal{U}_{i^*}^\omega$.

Now let us consider a feasible demand coverage of those $\frac{\beta_h}{C_h}$ type-h hubs located at the building i^*. Let each type-h hub cover extra demand $Q_h^\omega \equiv \min\{C_h \mathcal{U}_{i^*}^\omega, D_h\}$ from scenario $\omega \in \Omega$. Then the extra expected demand covered by each type-h hub is $Q_h \equiv \frac{1}{|\Omega|} \sum_{\omega \in \Omega} Q_h^\omega$. If $C_h \mathcal{U}_{i^*}^\omega \leq D_h$, then $Q_h^\omega = C_h \mathcal{U}_{i^*}^\omega$, *i.e.*, $\frac{Q_h^\omega}{C_h} = \mathcal{U}_{i^*}^\omega$. If $C_h \mathcal{U}_{i^*}^\omega \geq D_h$, then $Q_h^\omega = D_h$, *i.e.*, $\frac{Q_h^\omega}{C_h} = \frac{D_h}{C_h} = (\frac{D_h/C_h}{D_{h^*}/C_{h^*}}) \frac{D_{h^*}}{C_{h^*}} \geq (\frac{D_h/C_h}{D_{h^*}/C_{h^*}}) \mathcal{U}_{i^*}^\omega$. By the definition of γ_{hh^*}, we see $\frac{Q_h^\omega}{C_h} \geq \gamma_{hh^*} \mathcal{U}_{i^*}^\omega$. Hence $Q_h \geq \frac{1}{|\Omega|} \sum_{\omega \in \Omega} C_h \gamma_{hh^*} \mathcal{U}_{i^*}^\omega = \gamma_{hh^*} C_h \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \mathcal{U}_{i^*}^\omega = \gamma_{hh^*} C_h \mathcal{U}_{i^*}$. Since we totally have $\frac{\beta_h}{C_h}$ type-h hubs located at the building i^*, by this feasible demand coverage we know that the amount of extra expected demand covered by these hubs is $Q_h \frac{\beta_h}{C_h} \geq \gamma_{hh^*}\beta_h \mathcal{U}_{i^*}$.

Because this is just a feasible demand coverage when $\frac{\beta_h}{C_h}$ type-h hubs are located at the building i^*, where $1 \leq h \leq h^* - 1$,

$$\begin{aligned} cg(\text{PLAN3}) &\geq \mathcal{U}^g + \textstyle\sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h \mathcal{U}_{i^*} = \psi \mathcal{U}_{i^*} + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h \mathcal{U}_{i^*} \\ &= (\psi + \textstyle\sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h)\mathcal{U}_{i^*} \\ &= \left(\frac{\psi + \sum_{h=1}^{h^*-1} \gamma_{hh^*}\beta_h}{\psi+\beta}\right)(\mathcal{U} + \beta \mathcal{U}_{i^*}). \end{aligned}$$

■

Acknowledgement: The first author thanks Zhi-Long Chen and Karen Donohue for providing insights, ideas, and encouragement on this problem. The second author thanks Gérard Cornuéjols for his advice and support.

References

[1] A. Balakrishnan, T.L. Magnanti, and R.T. Wong, *A decomposition algorithm for local access telecommunications network expansion planning*, Operations Research, **43**, 1995, 58-76.

[2] A. Balakrishnan, T.L. Magnanti, J.S. Sokol, and W. Yi, *Spare-capacity assignment for line restoration using a single-facility type*, Operations Research, **50**, 2002, 617-35.

[3] F. Barahona and F. Chudak, *Near-optimal solutions to large scale facility location problems*, IBM Research Report, RC 21606, 1999.

[4] R. Batta, J.M. Dolan and N.N. Krishnamurthy, *The maximal expected covering location problem: Revisited*, Transportation Science, **23**, 1989, 277-287.

[5] M.L. Brandeau and S.S. Chiu, *An overview of representative problems in location research*, Management Science, **35**, 1989, 645-674.

[6] C.-H. Chung, *Recent applications of the maximal covering location planning (MCLP) model*, Journal of Operational Research Society, **37**, 1986, 735-746.

[7] G.P. Cornuéjols, M.L. Fisher and G.L. Nemhauser, *Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms*, Management Science, **23(8)**, 1977, 789-810.

[8] J.R. Current and J.E. Storbeck, *Capacitated covering models*, Environment and Planning B: Planning and Design, **15**, 1988, 153-163.

[9] M.S. Daskin, *A maximum expected covering location model: Formulation, properties and heuristic solution*, Transportation Science, **17**, 1983, 48-70.

[10] M.S. Daskin, *Network and Discrete Location: Models, Algorithms, and Applications*, Wiley, New York, 1995.

[11] Z. Drezner, *Facility Location: A Survey of Applications and Methods*, Springer-Verlag, Berlin, 1995.

[12] Z. Drezner and H.W. Hamacher, *Facility Location: Applications and Theory*, Springer, 2002.

[13] R.L. Francis, L.F. McGinnins and J.A. White (Eds.), *Facility Layout and Location: An Analytical Approach*, Prentice Hall, Englewood Cliffs, NJ, 1992.

[14] G. Gunawardane, *Dynamic versions of set covering type public facility location problems*, European Journal of Operational Research, **10**, 1982, 190-195.

[15] H.W. Hamacher and S. Nickel, *Classification of location models*, Location Science, **6**, 1998, 229-242.

[16] D.S. Hochbaum, *Approximating covering and packing problems: set cover, vertex cover, independent set, and related problems*, in Approximation Algorithms for NP-hard Problems, PWS Publishing Company, Boston, 1997, 94-143.

[17] D.S. Johnson, *Approximation algorithms for combinatorial problems*, Journal of Computer and System Sciences, **9**, 1974, 256-278.

[18] S. Khuller, A. Moss and J. Naor, *The budgeted maximum coverage problem*, Information Processing Letters, **70**, 1999, 39-45.

[19] J.G. Klincewicz, *Hub location in backbone/tributary network design: A review*, Location Science, **6**, 1998, 307-335.

[20] M. Labbe, D. Peeters and J.-F. Thisse, *Location on networks*, In: M.O. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser (eds.), Network Models, Handbooks in Operations Research and Management Science, North-Holland, Amsterdam, 1995.

[21] F.V. Louveaux, *Stochastic location analysis*, Location Science, **1**, 1993, 127-154.

[22] P.B. Mirchandani and R.L. Francis (Eds.), *Discrete Location Theory*, Wiley, New York, 1990.

[23] H. Newton and R. Horak, *Newton's telecom dictionary : The Authoritative Resource for Telecommunications, Networking, the Internet and Information Technology*, 18th Edition, Telecom, New York, 2002.

[24] H. Pirkul and D.A. Schilling, *The maximal covering location problem with capacities on total workload*, Management Science, **37**, 1991, 233-248.

[25] C.S. ReVelle and G. Laporte, *The plant location problem: New models and research prospects*, Operations Research, **44**, 1996, 864-874.

[26] D.A. Schilling, *Strategic facility planning: The analysis of options*, Decision Sciences, **13**, 1982, 1-14.

[27] D.A. Schilling, V. Jayaraman and R. Barkhi, *A review of covering problems in facility location*, Location Science, **1**, 1993, 25-55.