INFORMS Journal on Computing
Vol. 18, No. 3, Summer 2006, pp. 366–376
ISSN 1091-9856 | EISSN 1526-5528 | 06 | 1803 | 0366

DOI 10.1287/ijoc.1040.0129

Base Station Location and Service Assignments in W–CDMA Networks

Joakim Kalvenes
Edwin L. Cox School of Business, Southern Methodist University, Dallas, Texas 75275-0333, USA,
kalvenes@smu.edu

Jeffery Kennington, Eli Olinick
Department of Engineering Management, Information, and Systems, School of Engineering,
Southern Methodist University, Dallas, Texas 75275-0123, USA {jlk@engr.smu.edu, olinick@engr.smu.edu}

Designing a wideband code division multiple access (W–CDMA) network is a complicated task requiring the selection of sites for radio towers, analysis of customer demand, and assurance of service quality in terms of signal-to-interference ratio requirements. This investigation presents a net-revenue maximization model that can help a network planner with the selection of tower sites and the calculation of service capacity. The integer programming model takes as input a set of candidate tower locations with corresponding costs, a number of customer locations with corresponding demand for traffic, and the revenue potential for each unit of capacity allocated to each demand point. Based on these data, the model can be used to determine the selection of radio towers and the service capacity of the resulting radio network. The basic model is a large integer program and requires a special algorithm for practical solution. Our algorithm uses a priority branching scheme, an optimization-gap tolerance between 1% and 10%, and two sets of global valid inequalities that tighten the upper bounds obtained from the linear programming relaxation. The algorithm has been implemented in software for the AMPL/CPLEX system and an empirical investigation has been conducted. Using over 300 problem instances with up to 40 towers and 250 service locations, various combinations of algorithm settings have been evaluated. Using the recommended setting results in a design tool that generally runs in under 20 minutes on a 667 MHz AlphaStation.

Key words: communications; programming, integer; applications
History: Accepted by Prakash Mirchandani, Area Editor for Telecommunications and Electronic Commerce; received October 2002; revised September 2003, November 2004; accepted December 2004.

1. Introduction

During more than 20 years of operation of radio-based mobile telephone communication systems, the mobile telephony industry has experienced tremendous growth and technological development. Due to the limited availability of radio frequency spectrum and the limitations in radio transmission technologies, the communication bandwidth for mobile radio service is severely limited. This was the background for the development of cellular radio systems in which a service area is divided into smaller areas (called cells), each of which is serviced by a radio tower using a fraction of the available bandwidth. By limiting transmission power and, thus, the reach of the radio signal transmitted by mobile telephone users and radio towers, the same portion of the bandwidth can be reused by mobile telephone users and towers located sufficiently far away from one another. Based on this concept, it is possible to increase system capacity by replicating the network infrastructure.

In first-generation cellular systems, the bandwidth was shared between users by dividing the bandwidth into distinct frequency channels. This technology is referred to as *frequency division multiple access* (FDMA). To maximize system capacity, assigning channels to cells was an important resource-management problem of great computational difficulty. The frequency-assignment problem is closely related to the graph-coloring problem and has been studied extensively (see Murphey et al. 1999 for a survey). Typical first-generation cellular systems allow for the use of the same frequency approximately in every seventh cell. Assignment of frequency channels to radio towers determines which customers are serviced from which tower. Since total system capacity is a function of the infrastructure investment, tower location is another problem of interest in cellular network design (Mathar and Niessen 2000). Traditionally, these two problems have been solved independently with frequency assignment following tower-location decisions.

In second-generation cellular systems, radio signaling was changed from analog to digital encoding. As a consequence, the effects of interference were reduced so that a frequency channel could be reused on average in every four cells. Two dominant technologies emerged. In *time division multiple access* (TDMA), a frequency channel is divided into several time slots, each of which is assigned to a different user. Users share the frequency channel and take turns using it. Two implementations exist in the United States. The U.S. *digital cellular standard* (IS-54) splits a regular frequency channel into three time slots. In *global system for mobile* (GSM), a wider frequency channel accommodates eight simultaneous users. The twin problems of tower location and frequency assignment remained largely the same.

The other technology used in second-generation cellular systems is *code division multiple access* (CDMA). In CDMA, the entire bandwidth available to a service provider is shared by all users of the system using a technique called direct sequence pseudorandom noise spreading. CDMA network design differs considerably from FDMA and TDMA network design in that channel allocation is not an explicit issue. In each cell, all of the bandwidth available to the service provider can be used. The features in CDMA making this possible are stringent power control of all system devices (including user handsets) and the use of orthogonal codes to ensure minimal interference between simultaneous sessions. CDMA systems operate so as to keep the signal-to-interference ratio of each user at an acceptable level. As each new call causes incremental noise, new calls are admitted into the system only if the signal-to-interference ratio will remain within a reasonable threshold for all calls in the system. The power transmitted by a user's handset depends to a large extent on the distance between the handset and the radio tower providing the service. Thus, as indicated by Amaldi et al. (2001a), the issues of tower location and customer service assignments (i.e., which customer locations will be serviced by which towers) must be solved simultaneously.

Third-generation cellular systems currently under development and testing are all based on CDMA technology. The two major standards that have emerged are wideband CDMA (W–CDMA), which uses 5 MHz radio carriers, and CDMA2000, which uses multiple 1.25 MHz radio carriers. Compared to second-generation systems, third-generation systems will be able to provide more communication bandwidth per user, so as to provide mobile users with high-speed data services at rates up to 100 times those of second-generation voice channels. Compared to second-generation services, third-generation wireless services require orders of magnitude more bandwidth per communication session. Given that only limited additional bandwidth is available for third-generation services, it is clear that operators of third-generation networks must increase system capacity by increasing bandwidth reuse. This, in turn, implies that third-generation network operators must make considerable investments in infrastructure to reduce the reach of each radio tower and increase tower density in the service area.

Little research has been reported on the simultaneous selection of radio tower (or base station) locations and customer service assignments to base stations. Kalvenes et al. (2005) present a profit-maximization model for base station location, frequency assignment, and customer service allocation in hierarchical cellular networks based on the FDMA or TDMA technology. Galota et al. (2001) propose a profit-maximization model for base station location and customer service allocation in CDMA networks. They further analyze the computational complexity of the problem and develop a polynomial-time approximation scheme to solve it. Mathar and Schmeink (2001) maximize system capacity subject to a budget constraint. However, their interference model only accounts for the towers being utilized and does not include the number of customers serviced by the respective towers. In a similar vein, Amaldi et al. (2001a) develop a cost-minimization model that explicitly considers the signal-to-interference conditions generated by the base station location and customer service allocation choices. The signal-to-interference levels are incorporated into the objective function rather than as a system capacity constraint. They subsequently produce feasible solutions with a heuristic randomized add-drop algorithm. In a companion paper, Amaldi et al. (2001b) develop a tabu-search-based procedure to find improved feasible solutions.

Building on the work by Amaldi et al. (2001a, b), we propose an enhanced model that maximizes service provider net revenues for *the base station location and service assignment problem*. While the previous work by Galota et al. (2001) and Amaldi et al. (2001a, b) use signal-to-interference as a penalty term in their respective objective functions, we use the signal-to-interference ratio as a hard constraint. We also incorporate minimum service requirements based on the licensing rules developed by the Federal Communications Commission (FCC) for service providers in the United States.

The contributions of this work are threefold. First, we provide a new and improved model for the base station location and service assignment problem. When selecting base station locations, our model explicitly considers the trade-off between the revenue potential of each tower with its cost of installation and operation. This trade-off is subject to quality-of-service constraints in terms of sufficient signal-to-interference ratio at all towers in the service area. Second, we develop an efficient algorithm using AMPL and CPLEX to produce very high quality solutions with known error bounds. The computational viability of our algorithm is demonstrated on both test cases from the literature and data from a real-world market in the United States. Finally, we demonstrate the use of our design tool on two problems arising from the use of existing second-generation network infrastructure in the development of a third-generation service network. Hence, our design tool can be used not only for the initial design for a third-generation system, but can also assist with the migration problem from second to third generation.

2. Tower-Selection Model

In this section, we present our base model. It differs from previous work on CDMA network design (with the exception of Galota et al. 2001, which uses a limited-interference model) in that it maximizes net revenue, rather than minimizes cost. The basic constraints include minimum service requirements as mandated by the FCC in the United States, and minimum quality-of-service requirements as dictated by technology. To improve the upper bound provided by the continuous relaxation, a pair of global cuts are developed. An empirical analysis using several hundred test problems is presented.

2.1. Minimum Service Requirements

In the United States, the FCC regulates the telecommunications industry. Rather than stipulating the proportion of the population in each market (the larger geographical area for which a service provider has a license) that has to have access to service, the FCC has regulated the rate at which the service provider develops the market. The stipulations are as follows.

1. From the date the license is granted, each service provider has five years to develop a network infrastructure to service the allocated market. For a service provider with a license for 30 MHz of bandwidth, the network has to be able to reach geographical areas that combined have at least 1/3 of the population in the market. For a service provider with a license for 15 MHz of bandwidth, the corresponding requirement is at least 1/4 of the population in the market.
2. At the end of this five-year period, any geographical area within a market that is not covered with sufficient signal strength to offer service in this area will revert back to the FCC for allocation to other service providers.
3. Within 10 years from the date the license is granted, a 30 MHz licensee must cover geographical areas that combined have at least 2/3 of the population in the market. For other service providers, there are no additional service requirements.

In other words, a licensee can choose not to offer service in a limited geographical area if it is not profitable to do so. The service requirements do not stipulate how many customers should actually be able to receive service. It only specifies the land coverage of the market and not the system capacity in the covered areas.

Although there are two minimum-service requirements separated in time by five years, we model only a single period decision for a given minimum-service requirement. We assume that the service provider will make several investments to upgrade the coverage during that five-year span and, when appropriate, will switch from the initial service requirement to the second one.

2.2. Parameters Used in the Model

Let L denote the set of candidate locations for tower construction and M denote the set of subscriber locations. The demand for service in customer area $m \in M$ is denoted by d_m. This is the number of channels required to service the population in the area at an acceptable service level (call-blocking rate). Let r denote the annual revenue generated by each channel equivalent utilized in a customer area. The mandated minimum-service requirement, given as a proportion of the population in the market the service provider operates, is denoted by ρ. We assume that the population and customer demand are essentially distributed the same geographically so that the minimum-service requirement can be written in terms of customer demand points covered. The cost (amortized annually) of building and operating a tower at location $\ell \in L$ and connecting it to the backbone network is given by the parameter a_ℓ. Operating cost includes the cost of transmission power, marketing, accounting, customer aquisition and retention, and any other cost that is contingent upon operating a tower. When a subscriber in location m is serviced by tower ℓ, the subscriber's handset must transmit with sufficient power so that the tower receives it at the target power level P_{target}. Due to attenuation, the signal transmitted weakens over the path from the handset to the tower based on the relative location of the origin and destination (depending on distance, topography, local conditions, etc.). The attenuation factor from subscriber location m to tower location ℓ is given by the parameter $g_{m\ell}$. To ensure proper received power, P_{target}, at the tower location, the handset will transmit with power level $P_{\text{target}}/g_{m\ell}$. At each tower location, signals are received from many subscriber handsets in the surrounding neighborhood. For the voice packets to be processed with a reasonable error rate, the signal-to-interference ratio (SIR) for any active session must be more than the threshold value $\text{SIR}_{\min}$ (a derivation of $\text{SIR}_{\min}$ can be found in, e.g., Lee and Miller 1998, p. 1040).

The set $C_m \subset L$ denotes the set of candidate towers that can service customers in location $m \in M$. The set C_m includes all locations $\ell \in L$ such that

$g_{m\ell} > g_{\min}$, where $g_{\min}$ is derived from the maximum transmission power of the handset and the target received power level at the tower. Similarly, for every $\ell \in L$, $P_\ell \subset M$ denotes the set of customer locations that can be serviced by tower ℓ.

2.3. Decision Variables Used in the Model

The decision variables in this model include general integer and binary variables. The decision to build a tower at a candidate location is represented by variable y_ℓ, which is one if a tower is built at location $l \in L$, and zero otherwise. The integer variable $x_{m\ell}$ denotes the maximum number of customers at $m \in M$ that can be serviced by a tower at $l \in L$. That is, $x_{m\ell}$ represents a capacity assignment at tower ℓ for service of customers at location m. The variables are related so that $x_{m\ell} \geq 1$ only if $y_\ell = 1$, that is, customers in location m can be assigned to tower ℓ for service only if tower ℓ is constructed. The variable $x_{m\ell}$ is also limited by demand for service in location m so that $\sum_{\ell \in C_m} x_{m\ell} \leq d_m$. For computational-efficiency reasons, we have chosen a general integer variable for service assignment for each demand location instead of binary variables for each unit of customer demand so as to reduce the number of variables and the number of constraints in the formulation. Finally, q_m is an indicator variable that is one if customer location m can be serviced by at least one tower, and zero otherwise.

2.4. Quality-of-Service Constraint

Suppose that $x_{m\ell}$ customers in location m are assigned to tower ℓ for service. Each user's handset will transmit with power $P_{\text{target}}/g_{m\ell}$ so that the received power at tower ℓ is $g_{m\ell}P_{\text{target}}/g_{m\ell} = P_{\text{target}}$ from each of the customers in location m. Similarly, suppose that x_{nj} customers in location n are assigned to tower j for service. Each user's handset will transmit with power P_{target}/g_{nj} so that the received power at tower ℓ is $g_{n\ell}P_{\text{target}}/g_{nj}$ from each of the customers in location n. The total received power at tower location ℓ, P_ℓ^{TOT}, from all customers in the market receiving service is given by

$$P_\ell^{\text{TOT}} = P_{\text{target}} \sum_{m \in M} \sum_{j \in C_m} \frac{g_{m\ell}}{g_{mj}} x_{mj}, \tag{1}$$

where the right-hand side sums the received power at tower ℓ from all customer locations that are serviced by some tower.

For a given session assigned to tower ℓ, P_{target} represents the signal for the session, while all other signals received at tower ℓ, $P_\ell^{\text{TOT}} - P_{\text{target}}$, represents interference for that session (Amaldi et al. 2001b). Thus, a quality-of-service constraint based on the threshold signal-to-interference ratio for each session assigned to tower ℓ is given by

$$\frac{P_{\text{target}}}{P_\ell^{\text{TOT}} - P_{\text{target}}} \geq \text{SIR}_{\min}, \tag{2}$$

provided that tower ℓ is constructed. Since the tower is built only if $y_\ell = 1$, this constraint can be written as

$$\sum_{m \in M} \sum_{j \in C_m} \frac{g_{m\ell}}{g_{mj}} x_{mj} \leq 1 + \frac{1}{\text{SIR}_{\min}} + (1 - y_\ell)\beta_\ell \quad \forall\, \ell \in L, \tag{3}$$

where $\beta_\ell = \sum_{m \in M} d_m \{\max_{m \in C_m \setminus \{\ell\}} (g_{m\ell}/g_{mj})\}$ and $\max_{j \in C_m \setminus \{\ell\}} (g_{m\ell}/g_{mj}) = 0$ if $C_m \setminus \{\ell\} = \varnothing$. The second term on the right-hand side is zero when a tower is built ($y_\ell = 1$), so that the signal-to-interference requirement must be met at tower ℓ. When $y_\ell = 0$, the right-hand side is so large that the constraint is automatically satisfied.

2.5. Integer Program

The objective of the model is to maximize the total annual revenue generated by the cellular network less the cost of building, maintaining, and operating it. Mathematically we have

$$\text{maximize } \underbrace{r \sum_{m \in M} \sum_{\ell \in C_m} x_{m\ell}}_{\text{revenue}} - \underbrace{\sum_{\ell \in L} a_\ell y_\ell}_{\text{cost}}. \tag{4}$$

There are nine sets of constraints that define the model. The first set ensures that customers can be serviced only if there are towers that cover the demand area:

$$x_{m\ell} \leq d_m y_\ell \quad \forall\, m \in M,\ \ell \in C_m. \tag{5}$$

The next set of constraints ensures that one cannot serve more customers in a location than there is demand for service:

$$\sum_{\ell \in C_m} x_{m\ell} \leq d_m \quad \forall\, m \in M. \tag{6}$$

The minimum-service restrictions are handled by three sets of constraints. The first set states that customers cannot be serviced in location m if no towers are built that can reach demand area m. Second, if there is at least one tower that can reach area m, then customers in this location can be serviced. Finally, the third constraint ensures that service is available in demand areas that have at least a proportion ρ of all customers in the operator's total service area. Note, however, that there does not have to be a sufficient capacity to service all of the customers that can be reached by the network:

$$q_m \leq \sum_{\ell \in C_m} y_\ell \quad \forall\, m \in M, \tag{7}$$

$$q_m \geq y_\ell \quad \forall\, m \in M,\ \ell \in C_m, \tag{8}$$

$$\sum_{m \in M} d_m q_m \geq \rho \sum_{m \in M} d_m. \tag{9}$$

The next set of constraints enforces the quality-of-service restrictions on received signal quality at the towers:

$$\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x_{mj}\le s+(1-y_\ell)\beta_\ell\quad \forall\,\ell\in L, \tag{10}$$

where $s=1+1/\mathrm{SIR}_{\min}$.

The last three sets of constraints provide the domains for the variables:

$$q_m\in\{0,1\}\quad \forall\,m\in M, \tag{11}$$

$$y_\ell\in\{0,1\}\quad \forall\,\ell\in L, \tag{12}$$

$$x_{m\ell}\in\mathbb{N}\quad \forall\,m\in M,\ \ell\in C_m. \tag{13}$$

2.6. Valid Inequalities

To improve computational performance, we add valid inequalities to strengthen the LP bound. The first set of valid inequalities is based on the property that an assignment of a customer to the tower that requires the least handset transmission power is also the assignment that causes the least interference at any tower location and, hence, minimizes system resource consumption.

Proposition 1. *If $y_u=y_v=1$, $\{u,v\}\in C_n$ and $g_{nu}<g_{nv}$, then there exists an optimal solution with $x_{nu}=0$.*

Proof. Consider a feasible solution $\mathbf{x}$ to an instance of (4)–(13) so that $x_{nu}=\bar{x}>0$. Without loss of generality, let $v=\arg\max_{\ell\in C_n:y_\ell=1}\{g_{n\ell}\}$. Define $\mathbf{x}'$ so that $x'_{m\ell}=x_{m\ell}\ \forall\,m\in M,\ \ell\in L$, except $x'_{nu}=0$ and $x'_{nv}=x_{nv}+\bar{x}$. The objective-function value is the same for $\mathbf{x}$ and $\mathbf{x}'$. Since constraints (5)–(9) are satisfied for $\mathbf{x}$, they are also satisfied for $\mathbf{x}'$. Consider constraint (10):

$$\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x'_{mj}=\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x_{mj}+\frac{g_{n\ell}}{g_{nv}}\bar{x}-\frac{g_{n\ell}}{g_{nu}}\bar{x}\quad \forall\,\ell\in L.$$

Since $\bar{x}>0$, $g_{nu}<g_{nv}$ and $g_{m\ell}\ge 0\ \forall\,m\in M,\ \ell\in L$, it follows that $g_{n\ell}\bar{x}/g_{nv}-g_{n\ell}\bar{x}/g_{nu}\le 0$ and

$$\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x'_{mj}\le\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x_{mj}\quad \forall\,\ell\in L.$$

Since (10) is satisfied for $\mathbf{x}$, it is also satisfied for $\mathbf{x}'$ and the proposition follows. □

Proposition 1 indicates that if customers at site n are served, then profit can be maximized by assigning them to the available tower that has largest attenuation factor. That is, the objective function value cannot be increased by assigning service of customers at site n to a tower with a smaller attenuation factor. Hence, the following valid inequalities can be added to the formulation:

$$x_{m\ell}\le d_m(1-y_j)\quad \forall\,m\in M,\ell,j\in C_m\ \text{such that}\ g_{m\ell}<g_{mj}. \tag{14}$$

The second set of valid inequalities is based on the quality-of-service constraint (3). For each selected tower location $\ell\in L$, the customer assignments that contribute the most to the interference level at this location are the ones that receive service from tower ℓ. Thus, instead of calculating the total interference at tower ℓ from all serviced customers, we calculate only the interference generated by those customer locations m that have been assigned to tower ℓ for service, i.e., $x_{m\ell}\ge 1$. The left-hand side of (3) can be re-written as

$$\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x_{mj}$$

$$=\sum_{m\in M|\ell\in C_m}\frac{g_{m\ell}}{g_{m\ell}}x_{m\ell}+\sum_{m\in M}\sum_{j\in C_m\setminus\{\ell\}}\frac{g_{m\ell}}{g_{mj}}x_{mj}\quad \forall\,\ell\in L.$$

In other words, we have separated the interference caused by the customers that are serviced by tower ℓ from the interference caused by other customers. Observing that $g_{mj}\ge 0\ \forall\,m\in M,\ j\in L$, and $x_{m\ell}\in\mathbb{N}$, we obtain

$$\sum_{m\in P_\ell}x_{m\ell}\le\sum_{m\in M}\sum_{j\in C_m}\frac{g_{m\ell}}{g_{mj}}x_{mj}\quad \forall\,\ell\in L.$$

Thus, the following set of valid inequalities can be added to the formulation:

$$\sum_{m\in P_\ell}x_{m\ell}\le 1+\frac{1}{\mathrm{SIR}_{\min}}\quad \forall\,\ell\in L. \tag{15}$$

Although this set of valid inequalities is weaker than (3), it helps speed up the pruning of the branch-and-bound tree.

2.7. NP-Hardness

Proposition 2. *The W–CDMA base station location and service assignment problem represented by* (4)–(13) *is NP-hard.*

Proof. Restrict the problem instances represented by (4)–(13) so that $r=0$, $M=L$, and $C_m=\{m\}$. The problem reduces to

$$\max -\sum_{\ell\in L}a_\ell y_\ell \tag{16}$$

subject to

$$q_m\le\sum_{\ell\in C_m}y_\ell\quad \forall\,m\in M, \tag{17}$$

$$q_m\ge y_\ell\quad \forall\,m\in M,\ \ell\in C_m, \tag{18}$$

$$\sum_{m\in M}d_m q_m\ge\rho\sum_{m\in M}d_m, \tag{19}$$

$$q_m\in\{0,1\}\quad \forall\,m\in M, \tag{20}$$

$$y_\ell\in\{0,1\}\quad \forall\,\ell\in L. \tag{21}$$

Observing that since $C_m=\{m\}\ \forall\,m\in M$, constraints (17) and (18) imply $q_m=y_m$ and the problem is simplified to

$$\min \sum_{m\in M} a_m q_m \tag{22}$$

subject to

$$\sum_{m\in M} d_m q_m \geq \rho \sum_{m\in M} d_m, \tag{23}$$

$$q_m \in \{0, 1\} \quad \forall\, m \in M. \tag{24}$$

This is the knapsack problem, which is known to be NP-hard. □

2.8. Solver Parameter Settings

Instances of the base station location and service assignment problem with the added valid inequalities were solved with the branch-and-bound based CPLEX solver (http://www.cplex.com). CPLEX offers the choice of several solver parameter settings, including the branching order in the branch-and-bound tree, as well as a user-determined acceptable gap between the best known feasible solution and the current upper bound on the solution (CPLEX will terminate the solution procedure when this gap requirement is met).

2.8.1. Branching Order. There are two important sets of decision integer variables in our formulation of the base station location and service assignment problem. These are the tower-selection variables (y_ℓ) and the customer-assignment variables ($x_{m\ell}$). If the tower locations have not been selected, then the number of possible customer assignments is very large. However, based on Proposition 1, if the tower locations have been selected (i.e., for a given set of values for the y_ℓ variables), the problem of customer assignment reduces to a choice of which customers will be serviced (i.e., a multiconstrained integer knapsack problem). Since the number of candidate tower locations is small compared to the number of customer locations, a reasonable conjecture would be that branching on y_ℓ before $x_{m\ell}$ would reduce computational times. The computational study tests the effect of branching order on solution times.

2.8.2. Optimality Gap. The CPLEX solver permits the user to set a gap for termination between the best feasible solution obtained and the current upper bound. The gap is defined as (upper bound − best feasible solution)/upper bound. The larger the gap, the more rapidly the branch-and-bound tree is expected to be pruned. The user must determine an acceptable optimality gap for the design problem under investigation. Typical settings are 1%, 5%, or 10%. The computational study tests the effect of optimality gap on solution times.

2.9. Postprocessing Procedure

Since the values of the attenuation factors have a large range, the coefficients in (10) may differ in magnitude by as much as 10^9. This may result in numerical instability and we observed this when CPLEX was applied to problem instances. To ensure that feasibility is achieved within a reasonable tolerance, we created a postprocessing procedure that drops a few customers in exchange for satisfying the signal-to-interference restrictions. CPLEX scales the coefficient matrix so that all coefficients of the constraint set have an absolute value between zero and one. The scaled problem is solved and then the reverse procedure is applied to obtain the solution. Solving the scaled problem is Phase I of our procedure. Even though the solution to the scaled problem satisfies feasibility within the default tolerances, the solution may not be feasible for the original (unscaled) problem. To address these infeasibilities automatically, we have added the following postprocessing procedure that eliminates infeasibilites in a solution, if present.

Procedure: Phase II: Eliminate Infeasibilities

Inputs: $\bar{x}$ and $\bar{y}$ are an optimum for the scaled problem

Output: x^* is the best feasible solution obtained

Step 1: $LV \leftarrow \{\ell\colon$ (10) is not satisfied$\}$

Step 2: $A \leftarrow \{(m, \ell)\colon \bar{x}_{m\ell} > 0\}$

Step 3: Let $\bar{\delta}_{m\ell}$ for all $(m, \ell) \in A$ denote an optimum for

$$\min \sum_{(m,\ell)\in A} \delta_{m\ell} \tag{25}$$

subject to

$$\sum_{(m,j)\in A} \frac{g_{m\ell}}{g_{mj}} (\bar{x}_{mj} - \delta_{mj}) \leq s + (1 - \bar{y}_\ell)\beta_\ell \quad \forall\, \ell \in LV, \tag{26}$$

$$\delta_{m\ell} \in \mathbb{N} \quad \forall\, (m, \ell) \in A. \tag{27}$$

Step 4: $x^*_{m\ell} \leftarrow \bar{x}_{m\ell} - \bar{\delta}_{m\ell}$

2.10. Computational Results

On a Compaq AlphaServer DS20E with dual EV6.7 (21264A) 667 MHz processors and 4,096 MB of RAM, we used CPLEX 6.6.0 to solve 20 instances of the test problem proposed by Amaldi et al. (2001a, b). The test data are summarized in Table 1. Each of the 20 problem instances was generated by drawing different random sets of tower and customer locations. Based on the propagation model for urban areas developed by Hata (1980), we calculated the path attenuation coefficients $g_{m\ell}$. That is,

$$\begin{aligned} A^u_{m\ell} = {} & 69.55 + 26.16\log(f) - 13.82\log(H_\ell) + \\ & - [(1.1\log(f) - 0.7)H_m - (1.56\log(f) - 0.8)] \\ & + [44.9 - 6.55\log(H_\ell)]\log s_{m\ell}, \end{aligned} \tag{28}$$

Table 1 **Data for Test Problems**

Data item	Value or range	Description
Grid size	400 m $\times$ 400 m	Coverage area for test cases.
$\|L\|$	22	Number of potential base stations randomly placed in the coverage area.
$\|M\|$	95	Number of potential subscribers randomly placed in the coverage area.
C_m	L	The potential base stations to which subscriber m can be assigned.
d_m	1	The number of potential subscribers in location m.
r	$42,820	Annual revenue for each subscriber serviced.
ρ	0.25	Mandated minimum service requirement.
a_ℓ	$145,945	Annualized cost for installing a base station in location ℓ.
$\mathrm{SIR}_{\min}$	0.009789	Minimum signal-to-interference ratio required.
f	2,000 MHz	Operating frequency.
H_b	10 m	Height of base station antenna.
H_m	1 m	Height of mobile device antenna.
β	95	The large number used in constraints of type (10).

where f is the center frequency used for transmission, H_ℓ is the height of tower ℓ, H_m is the transmitter height of mobile customers in location m, and $s_{m\ell}$ is the distance between customer location m and tower location ℓ. $A^u_{m\ell}$ is given in dB and the conversion to the attenuation value is given by $g_{m\ell} = 10^{-0.1A^u_{m\ell}}$.

The computational results are summarized in Table 2. We ran 12 different experiments with the same set of 20 test cases, varying the optimality gap and the inclusion of the two valid inequalities. The first column in the table identifies the specific test run, while columns two, three, and four indicate the experimental settings. Columns five, six, and seven provide the average results for the 20 test problems with respect to the percentage of customer demand satisfied, objective function value, and CPU time consumed. Column eight indicates the number of test problems that required the second phase of the procedure to eliminate infeasibilities caused by CPLEX scaling of the input data. The last two columns provide the average and maximum gap, respectively, compared to our best upper bound on the solution. We observe that the choice of optimality gap has little impact on the computational times, with the exception of the test cases that were not using inequalities (15). We also note that the addition of inequalities (14) alone improves the solution quality, but also increases computational time significantly. In contrast, the use of inequalities (15) improves the solution quality substantially without disproportionally increasing the computational time. However, using both sets of inequalities provides the best solutions at reasonable computational times (22 minutes on average). Our conclusion is that the best choice of solution parameters is to use an optimality gap of 1% and both valid inequalities.

In a second series of computational experiments, we used a set of 40 tower locations first presented in Farmehr (2000). These tower locations are from a market in the northern plains of the United States. For this set of candidate tower locations, we drew 100 sets of 250 customer locations. For each customer location, we generated different sets of demands drawn from a uniform distribution with varying parameters. Uniform on the range $[1, 8]$ is denoted by $U[1, 8]$. While in the previous experiments, each customer location could be serviced by any candidate tower (i.e., $|C_m| = |L|$), the distances in the second series of experiments were such that each customer location could be serviced on average by between 1.7 and 2.4 tower locations (i.e., $1.7 \le |C_m| \le 2.4$). The computational results are summarized in Table 3.

We observe that as the subscriber density increases the problems become more difficult. At low customer densities, branching rules and the addition of valid inequalities are of minor importance and all problems are solved within a couple of CPU seconds. As customer density increases, so do both computational times and the number of cases that require post-processing in Phase II of the algorithm to eliminate

Table 2 **Computational Results for a Dense Set of Customer Locations**

Test run	Opt. gap (%)	Ineq. (14)	Ineq. (15)	Average Demand (%)	Average Profit ($)	Average CPU	Phase II	Avg. gap (%)	Max. gap (%)
1	1	no	no	61.21	2,154,309	17:47	15	39.25	65.09
2	1	yes	no	63.05	2,221,947	27:55	14	37.34	65.09
3	1	no	yes	93.79	3,238,779	4:30	9	8.76	33.85
4	1	yes	yes	97.47	3,381,352	22:05	9	4.76	15.82
5	3	no	no	66.05	2,322,092	2:41	12	34.48	65.09
6	3	yes	no	69.63	2,453,086	11:43	10	30.79	65.09
7	3	no	yes	95.00	3,280,725	3:50	8	7.57	33.85
8	3	yes	yes	97.32	3,374,929	20:17	8	4.94	15.82
9	5	no	no	66.16	2,326,374	3:33	12	34.36	65.09
10	5	yes	no	70.05	2,455,619	9:32	10	30.73	65.09
11	5	no	yes	95.11	3,285,007	3:19	8	7.48	33.85
12	5	yes	yes	96.16	3,335,124	24:40	8	6.07	28.49

Table 3 **Computational Results for a Sparse Set of Customer Locations**

Test case	d_m	Opt. gap (%)	Branch rule	Ineq. (14)	Ineq. (15)	Average Demand satisfied (%)	Average CPU time	Average Infeas. (%)	Ph. II	Avg. gap (%)	Max. gap (%)
1	U[1, 8]	1	no	no	no	55.28	0:02	3.17	2	1.01	1.62
2	U[1, 8]	1	yes	no	no	55.83	0:01	0.34	3	1.01	1.62
3	U[1, 8]	1	no	yes	no	55.83	0:02	0.03	1	1.01	1.62
4	U[1, 8]	1	no	no	yes	55.83	0:02	1.61	3	1.00	1.00
5	U[1, 8]	1	no	yes	yes	55.83	0:02	0.03	1	1.00	1.00
6	U[1, 8]	1	yes	no	yes	55.83	0:01	0.35	3	1.00	1.15
7	U[1, 8]	1	yes	yes	no	55.84	0:01	0.03	1	1.01	1.62
8	U[1, 8]	1	yes	yes	yes	55.84	0:01	0.03	1	1.00	1.00
9	U[1, 16]	1	yes	no	yes	54.49	0:09	0.70	13	1.07	2.67
10	U[1, 16]	1	yes	yes	no	54.45	0:09	0.79	17	1.08	3.01
11	U[1, 16]	1	yes	yes	yes	54.48	0:17	0.52	10	1.05	2.68
12*	U[1, 32]	0	no	no	no	37.76	7:13:50	3.55	3	15.41	49.69
13**	U[1, 32]	1	yes	no	yes	44.38	51:28	5.07	15	1.12	3.98
14	U[1, 32]	1	yes	yes	no	44.13	36:15	5.60	29	1.33	8.75
15	U[1, 32]	1	yes	yes	yes	44.19	13:00	2.86	31	1.18	4.27
16	U[1, 32]	5	yes	no	yes	43.43	47:55	3.81	26	5.14	7.86
17	U[1, 32]	5	yes	yes	no	43.45	27:15	6.45	25	5.37	12.43
18	U[1, 32]	5	yes	yes	yes	43.48	14:05	3.23	34	5.20	8.27
19	U[1, 32]	10	yes	no	yes	42.83	32:24	3.85	27	10.13	12.71
20	U[1, 32]	10	yes	yes	no	42.55	18:53	6.23	25	10.34	16.54
21	U[1, 32]	10	yes	yes	yes	42.71	8:25	2.60	31	10.15	13.10

*Only 10 problems were attempted. The computational resources were eight hours of CPU time and 1.8 GB of RAM.
**Only 69 of 100 problems were solved within eight hours of CPU time and 1.8 GB of RAM.

CPLEX scaling inaccuracies. At the highest customer density levels, with demand uniformly distributed between 1 and 32 customers in each of the 250 locations (yielding an average of 4,125 simultaneous customers in the system), both the branching rule and the addition of valid inequalities become significant for the performance of the solution procedure. We observe that if we let CPLEX run without specifying a branching rule and without adding the valid inequalities, we simply cannot solve the problem (case 12). Due to the very long computational times required, we attempted only 10 problem instances without using the branching rule or the valid inequalities (14). The average CPU time for these 10 problem instances exceeded seven hours. Using both the branching rule and the valid inequalities (14) and (15), we can solve all problem instances (case 15) with an average gap of 1.18% and a maximum gap of 4.27% between the best feasible solution and the upper bound. The average CPU time to do so was 13 minutes.

Based on the two series of experiments, we conclude that using a branching rule in CPLEX that considers tower location variables before customer assignment variables improves solution times significantly. We also conclude that the valid inequalities (14) and (15) are helpful in both series of experiments. However, the valid inequalities (14) add significant computational time to the first series of test problems while it reduces computational times in the second series. We believe that the reason is associated with the number of towers to which each customer location can be assigned. When this number is high, the valid inequalities (15) that consider the interference level at each tower generated by its own customer assignments are efficient at eliminating customer assignment choices. Adding the valid inequalities (14) does not reduce the choice set considerably, but adds to the computational time. However, when the number of possible tower assignments is limited (as is the case in our second test series), assigning customers to the closest tower reduces the choice set of customers who may receive service.

3. Service Capacity of Existing Infrastructure

Motivated by a problem brought to our attention by a U.S. service provider, we next consider the following special case of the model introduced in the previous section: *Given an existing infrastructure for second-generation wireless service, how many customers in the service area can be accommodated with W–CDMA service and how should resources be allocated*?

3.1. Test Data

The tower locations are identical to those used in Table 2. There are 250 customer location test points randomly distributed over a 14 km by 10 km geographical area. With each test point is an associated set of path-attenuation coefficients corresponding to each of the 40 tower locations. Each of the customer locations has an artificial demand for simultaneous service of 100 customers, for a total of 25,000 demand units at any point in time. This number far exceeds what the infrastructure can service and, thus, only a subset of the demand will be assigned to towers. The purpose of the high demand numbers is to find the maximal number of customers the infrastructure can support.

We assume that the cost of towers and equipment at the tower locations is zero. That is, the network designer's decision problem is to service as many customers as possible in the market from an existing infrastructure of towers and equipment. The objective function is to maximize the revenue generated by the customers serviced. We use a signal-to-interference ratio requirement of 0.009789.

3.2. Results

Figure 1 illustrates the solution to one of the test cases in graphical form. Each tower location is marked with a T and an index number. Each x represents a demand point that receives service, while each · represents a demand point that is not serviced by any tower. From the figure, it is apparent that customers close to the tower locations have been given priority over those located far away from the towers. The reason for this is that customers located far away from a tower must transmit at a higher power than those that are close to the tower. Compared to nearby customers, the higher transmission power of far-away customers will contribute more to the total interference level in the service area, thus reducing the overall capacity of the infrastructure. Overall, 36 of the 40 towers in the market will be used to service customers, while 51 of the 250 demand points will receive service. The system can service a total of 3,245 simultaneous customers.

4. Expansion of Existing Infrastructure

In this section, we consider a special case of the basic model. For an existing infrastructure, the service provider wants to add resources so as to service a changing demand population subject to a budget constraint on the capacity expansion. For the existing infrastructure, the corresponding annualized cost $\bar{a}_\ell$ is

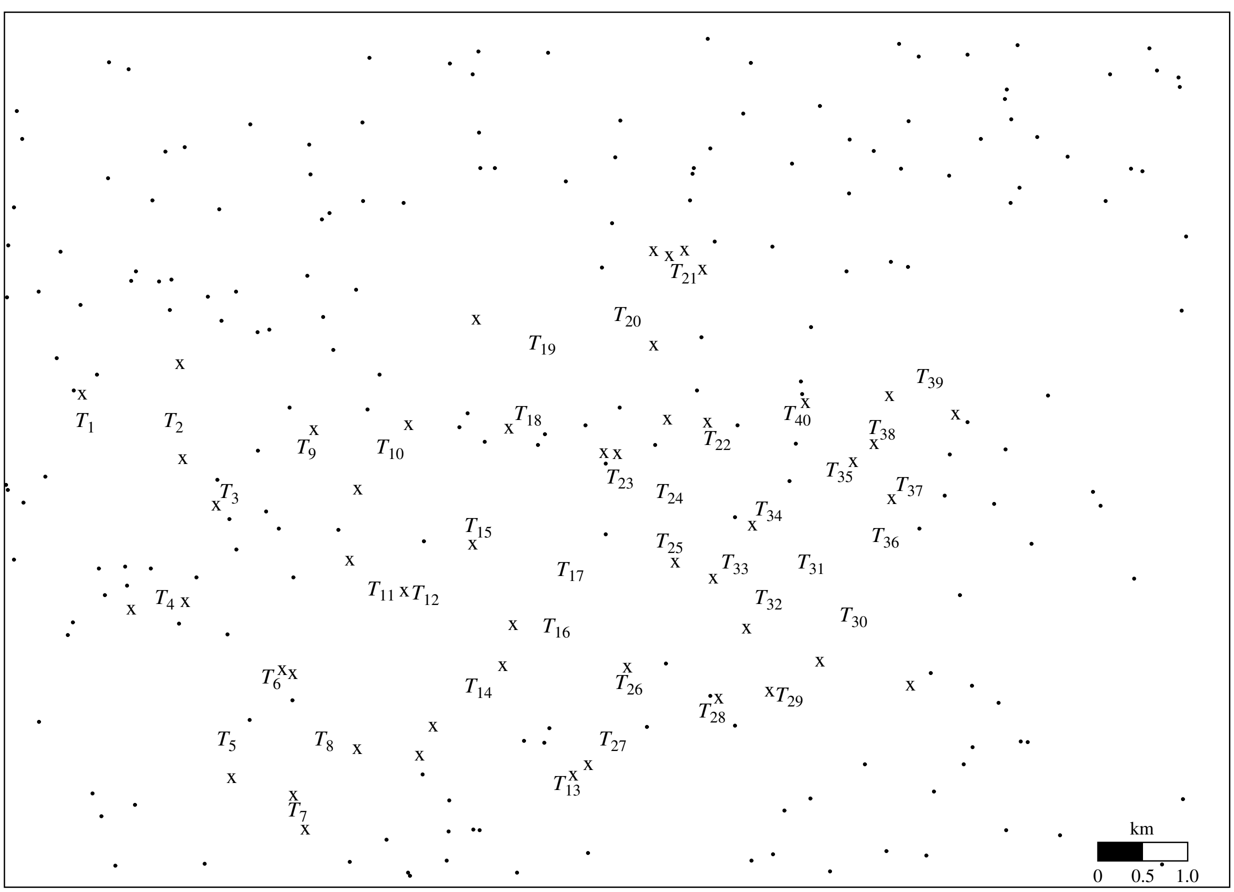

Figure 1 Sample Solution for Tower Location and Service Assignment in a W–CDMA Network

zero so that for existing towers, $y_\ell = 1$. We also add the constraint

$$\sum_{\ell \in L} \bar{a}_\ell y_\ell \le B, \tag{29}$$

where B is the annualized capital investment budget.

4.1. Test Data

The test data for this sample problem are essentially the same as for the existing infrastructure example examined in the previous section. However, 30 new candidate tower locations were added, as indicated in Figure 2. The cost per new candidate tower location was so that $\bar{a}_\ell = a$, while the budget was ten times this number, i.e., $B = 10a$.

4.2. Results

The solution is illustrated in Figure 2 for one test case. Existing tower location i is indicated with the symbol T_i, selected candidate tower location j is indicated with the symbol S_j, while unselected tower location k is indicated with the symbol C_k. Demand points receiving service are marked with the symbol x. Unserviced demand points are marked with $\cdot$. The 10 selected new towers are spread out geographically, demonstrating the propensity of the model to maximize the number of customers serviced by minimizing the total system interference level. The computational time for this seventy-tower example was approximately two hours.

5. Summary and Conclusions

This paper presents a new model and computational procedure for solving the W–CDMA base station location and service assignment problem. Our model takes as input candidate tower locations and customer demand points and determines the revenue-maximizing tower configuration and customer assignment. The solution procedure was implemented and tested on a Compaq AlphaServer, using as a base the commercially available integer linear programming software package CPLEX. While CPLEX can solve small instances of the problem, it cannot solve larger problem instances. To address this issue, we developed two sets of valid inequalities that were added to the model. In addition, we developed a branching rule that results in faster pruning of the branch-and-bound tree. With these enhancements, we could use CPLEX to solve realistically sized problem instances within reasonable computational times.

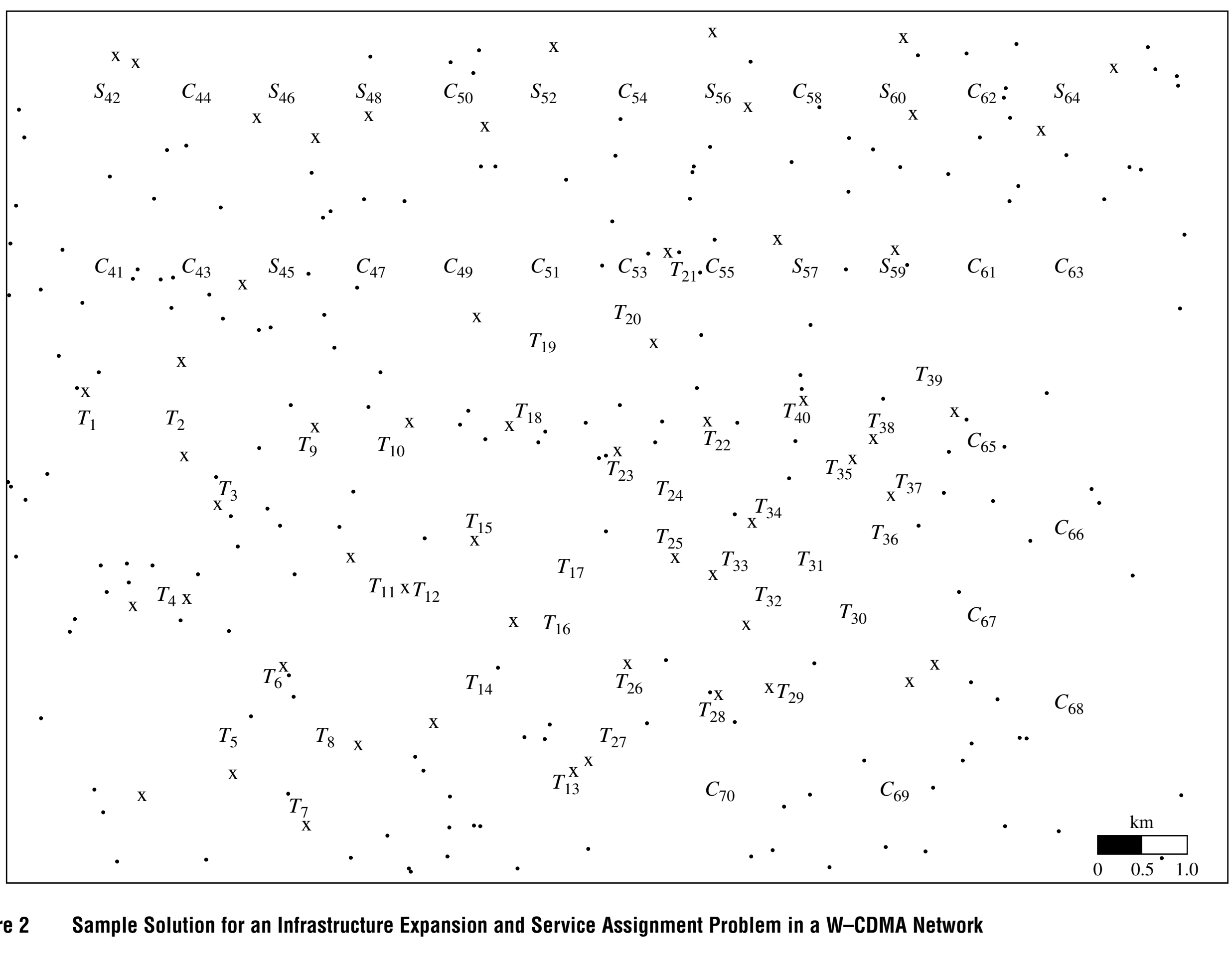

Figure 2 Sample Solution for an Infrastructure Expansion and Service Assignment Problem in a W–CDMA Network

Three variations of the problem were considered. First, we addressed the case of entirely new infrastructure development. We solved 20 test problems with dense candidate tower locations and customer demand points using different selections of branching rules and valid inequalities to learn which ones were the most effective. We concluded that branching on tower locations first was beneficial. We further concluded that both the distance-based and the interference-based valid inequalities also worked well for this type of problem. However, the distance-based valid inequalities added considerable computational time for this set of test problems. In contrast to previous work, which is based on randomized local search procedures, we find feasible solutions that are provably within 5% of the optimal solution on average and within 16% of the optimal solution in the worst case.

We also solved a set of 300 test problems with sparser candidate tower and customer location structure. Starting with 40 candidate tower locations from a wireless market in the northern plains of the United States, we generated 100 sets of 250 randomly located customer demand points. Each of the 250 customer demand points were subsequently loaded with service demand drawn from a uniform distribution. For each of the 100 problems, three different demand levels were considered, resulting in a total of 300 test problems. For this set of test problems, we found that both the branching rule and the two sets of valid inequalities were essential to solving the test problems. The obtained solution quality for this set of problems was excellent. The gap between the best feasible solution and the upper bound was 1.2% on average and 4.3% in the worst case. The average CPU time was only 13 minutes.

Motivated by a real-world problem encountered by a W–CDMA provider, we considered two additional applications of the design tool. In the first application, we found the maximum capacity of an existing second-generation wireless tower infrastructure that is refurbished with new transmission equipment to provide W–CDMA service. In the second application, we considered expansion of an existing tower infrastructure subject to a budget constraint. Both applications were solved within acceptable computation times.

Acknowledgments
The authors gratefully acknowledge the many helpful comments provided by the anonymous referees, one of whom provided the Proof of Proposition 1. This research was supported in part by Office of Naval Research Award N00014-96-1-0315.

References
Amaldi, E., A. Capone, F. Malucelli. 2001a. Discrete models and algorithms for the capacitated location problems arising in UMTS network planning. P. Crescenzi, B. Yener, eds. *Proc. 5th Internat. Workshop Discrete Algorithms Methods Mobile Comput. Comm.*, ACM Press, New York, 1–8.

Amaldi, E., A. Capone, F. Malucelli. 2001b. Improved models and algorithms for UMTS radio planning. *IEEE 54th Vehicular Technology Conf. Proc.*, IEEE, Piscataway, NJ, 920–924.

Farmehr, A. 2000. Design strategies for both survivable and non-survivable personal communication networks. Praxis Report, School of Engineering and Applied Science, Southern Methodist University, Dallas, TX.

Galota, M., C. Glasser, S. Reith, H. Vollmer. 2001. A ploynomial-time approximation scheme for base station positioning in UMTS networks. P. Crescenzi, B. Yener, eds. *Proc. 5th Internat. Workshop Discrete Algorithms Methods Mobile Comput. Comm.*, ACM Press, New York, 52–59.

Hata, M. 1980. Empirical formula for propagation loss in land mobile radio service. *IEEE Trans. Vehicular Technology* **29** 317–325.

Kalvenes, J., J. Kennington, E. Olinick. 2005. Hierarchical cellular network design with channel allocation. *Eur. J. Oper. Res.* **160** 3–18.

Lee, J. S., L. E. Miller. 1998. *CDMA Systems Engineering Handbook*. Artech House Publishers, Boston, MA.

Mathar, R., T. Niessen. 2000. Optimum positioning of base stations for cellular radio networks. *Wireless Networks* **6** 421–428.

Mathar, R., M. Schmeink. 2001. Optimal base station positioning and channel assignment for 3G mobile networks by integer programming. *Ann. Oper. Res.* **107** 225–236.

Murphey, R. A., P. M. Pardalos, G. C. Resende. 1999. Frequency assignment problems. D.-Z. Zu, P. M. Pardalos, eds. *Handbook of Combinatorial Optimization*. Kluwer Academic Publishers, Boston, MA, 1–55.