DEPARTMENT OF OPERATIONS RESEARCH
UNIVERSITY OF AARHUS

Working Paper no. 2001/2

Network planning in telecommunications: A stochastic programming approach

Morten Riis
Anders J.V. Skriver
Jørn Lodahl

ISSN 1600-8987

Department of Mathematical Science
Telephone: +45 8942 3188
E-mail: institut@imf.au.dk

Building 530, Ny Munkegade
DK-8000 Aarhus C, Denmark
URL: www.imf.au.dk

Network planning in telecommunications: A stochastic programming approach

Morten Riis*and Anders J.V. Skriver
Department of Operations Research
University of Aarhus, Building 530
Ny Munkegade
DK - 8000 Århus C
Denmark

Jørn Lodahl
Sonofon
Skelagervej 1
DK - 9100 Ålborg
Denmark

Abstract

We consider a network design problem arising in mobile communications. The problem consists in deploying a number of new MSCs and allocating existing BSCs to MSCs, so as to minimize the incurred costs while meeting customer demand and observing the capacity restrictions. We formulate this problem as a two-stage stochastic program with mixed-integer recourse. To solve the problem we apply a dual decomposition procedure, solving scenario subproblems by means of branch and cut. The solution procedure has been tested on a real life problem instance provided by Sonofon, a Danish mobile communication network provider, and we report some results of our computational experiments.

Keywords: Network planning; Telecommunication; Stochastic Programming; Dual Decomposition; Branch and Cut.

1 Introduction

Mobile telecommunication network providers have been facing a rapid growth in demand for several years and this trend seems likely to continue. This forces the network provider to

*Corresponding author. Email: riis@imf.au.dk

constantly expand the capacity of the network in order to provide an acceptable grade of service to customers. There is a vast amount of literature concerning the optimal expansion of link capacities in a telecommunications network. We refer to papers by e.g. Balakrishnan, Magnanti and Wong [1], Bienstock and Günlük [2], Chang and Gavish [4] and Dahl and Stoer [5] for different approaches to such types of problems. The link capacities do not constitute the only potential bottleneck in a telecommunications network, however, since capacity restrictions may be imposed not only on traffic but also on the number of customers served by the network. In this paper we study a network design problem in which some capacity constraints are imposed to restrict traffic on links in the network while others are imposed to restrict the number of customers served by nodes in the network.

We study a mobile communications network. The base transceiver stations (BTSs) are each connected to one base station controller (BSC). Each BSC serves a number of BTSs and is connected to one mobile switching center (MSC). Finally each MSC serves a number of BSCs and the MSCs are connected internally. The network is illustrated in Figure 1.

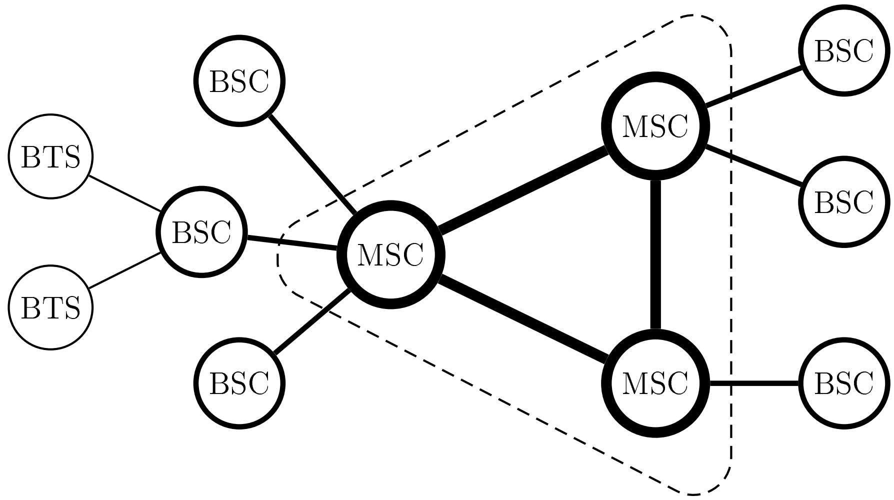

Figure 1: Illustration of a mobile telecommunications network.

The visitor location register (VLR) of an MSC, a database handling all information about clients, has a limited capacity, thus restricting the number of customers that can be served (through BTSs and BSCs) by an MSC. Thus the network provider not only has to expand the link capacities but should consider when and where to deploy new MSCs in order to be able to serve the increasing number of customers.

We will consider the problem of deploying a number of new MSCs and allocating the BSCs to new and existing MSCs, thus treating the number and locations of BTSs and BSCs as exogenous. The deployment of MSCs must be carried out so as to minimize the incurred costs while meeting customer demand and observing the capacity restrictions. The cost function will include four terms:

1. The cost of new MSCs.

2. The cost of connecting BSCs to MSCs.

3. The cost of expanding the capacity of links connecting the MSCs.

4. A penalty cost for handovers that occur among BSCs that are connected to different MSCs.

Tzifa et al. [17] study a similar problem in which only the access network is considered, thus ignoring the third cost term mentioned above. Also, the problem of optimally assigning BSCs to MSCs has been addressed by several authors such as Saha, Mukherjee and Bhattacharya [15] and Merchant and Sengupta [8]. Apart from minimizing the incurred costs of connecting BSCs to MSCs and the handover cost, it is customary to enforce some degree of load balancing among the MSCs. Tzifa et al. and Saha, Mukherjee and Bhattacharya explicitly include a penalty cost on uneven loads in the objective function, whereas Merchant and Sengupta propose to handle the load balancing problem parametrically. We do not explicitly consider load balancing but the parametric approach of Merchant and Sengupta may easily be adopted in our setting.

All of the above-mentioned authors follow a deterministic approach in the sense that the cost parameters, the number of customers and the demand for bandwidth are all assumed to be known at the point of decision. It is a fact, however, that the time that passes from the moment at which deployment of MSCs is resolved on, until the equipment is actually in place and available for use, is rather long (about a year). This means that at the time the decision has to be made, the network provider does not have full knowledge of several important parameters of the model. For this reason the network provider should put off the definitive decision on allocation of BSCs to MSCs for as long as possible, allowing uncertainty to be at least partially revealed. This is the incentive for us to model the problem as a two-stage stochastic program. In this formulation uncertain parameters are replaced by random variables and decisions are organized in two stages. The first stage consists of deployment of MSCs which must be resolved on before uncertainty has been revealed and hence must be based on the distribution of random parameters only. In the second stage outcomes of all random parameters have been observed and an optimal allocation of BSCs to MSCs and a corresponding routing of traffic in the resulting network is determined.

The importance of including uncertainty in the problem formulation when modeling capacity expansion problems is well recognized. Stochastic programming has been used as a modeling tool for such problems in telecommunications by several authors. Sen, Doverspike and Cosares [16] study a capacity expansion problem in which the expected number of unserved requests is minimized subject to limitations on the total capacity expansion. Riis and Andersen [11, 12] use stochastic programming to solve two different capacity expansion problems in which additional capacity, required to meet customer demand, should be installed on edges of the network in modularities of fixed batch sizes. Finally, Dempster, Medova and Thompson [6] use chance-constrained programming to solve a capacity expansion problem subject to certain grade of service constraints assuming that the arrival process of calls is known. The main emphasis in previous studies has been on the capacity expansion of links, while less has been said about the network design problem considered in this paper.

This paper is organized as follows. We start out by formalizing the problem formulation and describing the parameters involved in Section 2. Extensions of the basic model to hedge against potential node and edge failures by imposing survivability constraints are discussed in Section 3. Next, in Section 4 we briefly outline the concept of dual decomposition (or scenario decomposition). Dual decomposition techniques have been applied in the context of stochastic programming by numerous authors including Carøe and Schultz [3], Mulvey and Ruszczynski [9] and Rockafellar and Wets [14]. The seminal idea is to use variable splitting to make the problem separable into independent subproblems which are easily solved. In our case, the subproblems are solved by means of branch and cut, using valid inequalities derived in Section 5 as cutting planes. In Section 6 our application is described along with some of the practical difficulties concerning implementation of the algorithm. Finally, we give some concluding remarks in Section 7.

2 Problem Formulation

To give a formal formulation of the capacity expansion problem introduced in the previous section, we will consider a finite number of potential locations for new MSCs and hence the basic setup will be described by three finite sets of nodes representing the locations of MSCs and BSCs:

- V_1 The set of locations of existing MSCs.
- V_2 The set of potential locations for new MSCs.
- W The set of locations of BSCs.

Note that a given location may very well be represented as a node in more than one of the sets (even in all of them). In fact, the model allows for a single location to be represented as several nodes in one set, for example if we wish to deploy more than one MSC at a location.

The network interconnecting the MSCs is modeled as an undirected graph $G = (V, E)$. The nodeset $V = V_1 \cup V_2$ represents the existing and potential locations of MSCs, and the edge set E represents the existing and potential links $\{i, j\}$ between nodes $i, j \in V$. We will consider demand at BSC level. Even though we assume that traffic is bidirectional, we will find it convenient to use directed flow for modeling purposes. Hence we shall assign an arbitrary direction to each point-to-point demand and refer to its origin and destination. Also, each undirected edge $\{i, j\} \in E$ will correspond to two (conceptual) directed edges (i, j) and (j, i), each of which can carry flow. Still, to allow for the appropriate bidirectional traffic, edge capacities are dimensioned with respect to the total traffic on the given edge, disregarding the arbitrarily assigned directions of flow.

Demand for bandwidth on the connections will be described by a set K of commodities. Two main approaches for defining such commodities have been used in the literature. One possibility is to define a commodity for each point-to-point demand resulting in a total of $O(|W|^2)$ commodities. In general we find it more convenient, though, to reduce the number of variables by working with an aggregated formulation containing a total of only $O(|W|)$ commodities. This is achieved by letting each commodity $k \in K$ correspond to demand originating at a given BSC with respect to the arbitrary directions assigned to traffic. If one wishes to impose survivability constraints, however, it turns out that the disaggregated formulation may be more convenient. We will return to this issue in Section 3.

As previously discussed, several parameters of the model are not known with certainty at the time the decision on deployment of MSCs has to be made. In particular, the only information about future demand available at the point of decision, comes from past observations and some form of forecast model. This inherent uncertainty will be incorporated in the problem formulation by introducing some probability space $(\Omega, \mathcal{F}, P)$ and allowing the parameters in question to be dependent on the outcome of a random event $\omega \in \Omega$. Here, the probability distribution P is meant to reflect information about uncertain parameters coming from the above-mentioned forecasts. Thus the demand for bandwidth on edges and VLR-capacity at nodes will be described by the following sets of parameters:

- $D_{kr}(\omega)$ The net demand for commodity k at BSC r. $(k \in K, r \in W)$
- $L_r(\omega)$ The load of BSC r on the VLR in the MSC to which it is connected. $(r \in W)$

We emphasize that $D_{kr}(\omega)$ is the *net demand* for commodity k at BSC r and hence, in particular, that it is negative if and only if BSC r is the origin of commodity k and that $\sum_{r \in W} D_{kr}(\omega) = 0$. The parameter $D_{kr}(\omega)$ is directly related to the traffic between the origin of commodity k and BSC r, whereas the load $L_r(\omega)$ should rather be thought of as depending on the number of customers in the area served by BSC r.

Corresponding to the two types of demand, we have two types of existing capacity in the network - capacity restricting flow on edges of the network and capacity restricting the number of customers served by nodes in the network. These are summarized in the following sets of parameters:

- C_{ij} Flow-capacity on edge $\{i, j\}$. ($\{i, j\} \in E$)
- M_i VLR-capacity of the MSC located at node i. ($i \in V$)

The cost structure is described by the following sets of parameters some of which are treated as exogenous, while others are assumed to be uncertain at the point in time at which the decision has to be made, thus depending on the random event ω:

- c_i The cost of deploying an MSC at node i. ($i \in V_2$)
- $p_{ij}(\omega)$ The cost of adding one unit of capacity on edge $\{i, j\}$. ($\{i, j\} \in E$)
- $q_{ri}(\omega)$ The cost of connecting BSC r to node i. ($r \in W,\ i \in V$)
- $h_{rt}(\omega)$ The penalty cost (for supporting handovers) incurred if BSC r and t are connected to different MSCs. ($r, t \in W$)

Note that we assume the cost of expanding the capacity of a connection to be linear and that we do not include a fixed cost for establishing the connection. The reason for this is the fact that the company, in cooperation with which this research project was engaged upon, had already available a physical network with sufficient link capacities. In order to utilize this capacity, however, it may be necessary to install additional equipment at the end-points of the connection, and this cost is assumed to be linear with respect to the capacity provided.

The main decisions to be taken are deployment of new MSCs and allocation of BSCs to MSCs. These decisions are represented by the following two sets of binary variables:

- $x_i = \begin{cases} 1 & \text{if an MSC is deployed in node } i.\ (i \in V_2) \\ 0 & \text{otherwise} \end{cases}$
- $y_{ri}(\omega) = \begin{cases} 1 & \text{if BSC } r \text{ is connected to MSC } i.\ (r \in W,\ i \in V) \\ 0 & \text{otherwise} \end{cases}$

As indicated by the dependency of the variables y_{ri} on the random event ω, the allocation of BSCs to MSCs is allowed to depend on the outcome of the random parameters. That is, the decision on allocation of BSCs to MSCs is postponed to the second stage to take full advantage of the additional information which is available at this point.

Finally, the following sets of variables are used to describe flow in the network, and the capacity expansion of links needed to carry this flow. Since flow does not occur until demand is realized, these variables all belong in the second stage.

· $f_{ijk}(\omega)$ Flow of commodity k on edge $\{i,j\}$ in direction from i to j. $(k \in K, \{i,j\} \in E)$

· $f_{jik}(\omega)$ Flow of commodity k on edge $\{i,j\}$ in direction from j to i. $(k \in K, \{i,j\} \in E)$

· $v_{ij}(\omega)$ Aggregate flow on edge $\{i,j\}$ in excess of current capacity C_{ij}. $(\{i,j\} \in E)$

To be capable of handling the model computationally, we will assume that there is only a finite number of possible outcomes of random parameters.

(A1) The probability distribution P is discrete and has finite support, say $\Omega = \{\omega^1, \dots, \omega^S\}$ with corresponding probabilities $P(\{\omega^1\}) = \pi^1, \dots, P(\{\omega^S\}) = \pi^S$.

A possible outcome of random parameters $(p(\omega^s), q(\omega^s), h(\omega^s), D(\omega^s), L(\omega^s))$ corresponding to some elementary event $\omega^s \in \Omega$ will be referred to as a scenario. For notational convenience we will refer to such a scenario simply by $(p^s, q^s, h^s, D^s, L^s)$. Likewise, we will use a superscript s on second-stage variables to indicate that these decisions are allowed to differ for different scenarios.

We are now ready to formulate the problem of optimally deploying a number of new MSCs and allocating BSCs to MSCs as a two-stage stochastic program. The first-stage objective is to minimize the sum of the cost of new MSCs and the expected value of the cost incurred in the second stage,

$$z = \min \sum_{i \in V_2} c_i x_i + \sum_{s=1}^{S} \pi^s Q^s(x) \tag{1}$$

$$\text{s.t. } x \in \mathbb{B}^{|V_2|}. \tag{2}$$

Here, the second-stage value function $Q^s(x)$ is given by

$$Q^s(x) = \min \sum_{\{i,j\} \in E} p_{ij}^s v_{ij}^s + \sum_{r \in W} \sum_{i \in V} q_{ri}^s y_{ri}^s + \sum_{\substack{r,t \in W \\ r<t}} h_{rt}^s \sum_{i \in V} (y_{ri}^s - y_{ti}^s)^+ \tag{3}$$

$$\text{s.t. } \sum_{r \in W} L_r^s y_{ri}^s \leq M_i \qquad \forall i \in V_1, \tag{4}$$

$$\sum_{r \in W} L_r^s y_{ri}^s \leq M_i x_i \qquad \forall i \in V_2, \tag{5}$$

$$\sum_{i \in V} y_{ri}^s = 1 \qquad \forall r \in W, \tag{6}$$

$$\sum_{j:\{i,j\} \in E} f_{jik}^s - \sum_{j:\{i,j\} \in E} f_{ijk}^s = \sum_{r \in W} D_{kr}^s y_{ri}^s \qquad \forall i \in V,\ k \in K, \tag{7}$$

$$\sum_{k \in K} \left(f_{ijk}^s + f_{jik}^s \right) \leq C_{ij} + v_{ij}^s \qquad \forall \{i,j\} \in E, \tag{8}$$

$$y^s \in \mathbb{B}^{|W||V|},\ f^s \in \mathbb{R}_+^{2|E||K|},\ v^s \in \mathbb{R}_+^{|E|}. \tag{9}$$

We have used the notation x^+ to denote $\max\{0, x\}$ for $x \in \mathbb{R}$, and hence the third term of the second-stage objective (3) includes the handover cost between BSCs r and t if and only if these BSCs are allocated to different MSCs. The constraints (4) and (5) ensure that the total load from the BSCs connected to an MSC does not exceed the capacity of the VLR. Moreover, the constraint (5) ensures that a BSC can only be connected to an MSC if this is actually deployed (x_i=1) while the constraint (6) ensures that all BSCs are connected to exactly one MSC. The constraint (7) is a flow conservation constraint stating that the net flow of commodity k into MSC i should equal the aggregate net demand for commodity k from BSCs connected to MSC i. Finally, the constraint (8) states that the aggregate flow on an edge $\{i, j\} \in E$ cannot exceed the total capacity installed on the edge.

We note that the nonlinear term in the second-stage objective may easily be replaced by a linear one. Hence let H^s_{rt} be a variable representing the handover cost incurred between BSCs r and t under scenario s. Then H^s_{rt} may be defined using V linear constraints,

$$H^s_{rt} \geq h^s_{rt}(y^s_{ri} - y^s_{ti}) \qquad \forall i \in V, \tag{10}$$

and the nonlinear term may be replaced by a simple summation of the variables H^s_{rt}. Thus if the constraints (10) are added, the third objective term may be replaced by

$$\sum_{\substack{r,t \in W \\ r<t}} H^s_{rt}.$$

3 Survivability

There is an entirely different side to the issue of designing a telecommunications network under uncertainty besides the one we have considered this far. Thus it is possible that not only the parameters of the model, such as demand and prices, are subject to uncertainty. To be specific, we will consider a situation in which nodes and/or edges are subject to potential failures. This forces us to impose different kinds of *survivability constraints* to ensure that the network is not too vulnerable in case of such failures. The concept of survivability has previously been considered in the context of telecommunication networks by numerous authors. (See e.g. Dahl and Stoer [5] and Rios, Marianov and Gutierrez [13].) In general survivability may be achieved either by *diversification* or by *reservation* depending on the assurance required and the ability to restructure the solution in case of failures. In this section we discuss some possible formulations in the context of problem (1)-(9).

By diversification we mean routing demand using two or more edge- and/or node-disjoint paths. Diversification constraints are easily imposed if we are working with the disaggregate formulation in which each commodity $k \in K$ corresponds to a unique point-to-point demand.

Hence we may let $O(k)$ and $D(k)$ denote the origin and destination of commodity k, and d_k^s the demand for commodity k under some scenario s so that D_{kr}^s equals d_k^s for $r = D(k)$, $-d_k^s$ for $r = O(k)$ and zero otherwise. If σ_k is a parameter equal to the maximum fraction of demand for commodity k that is allowed to flow through any given node or edge of the network, we may impose the following diversification constraints:

$$f_{ijk}^s + f_{jik}^s \leq \sigma_k d_k^s \qquad \forall \{i,j\} \in E,\ k \in K \tag{11}$$

$$\sum_{j:\{i,j\} \in E} f_{ijk}^s \leq \sigma_k d_k^s + (1 - \sigma_k) d_k^s y_{O(k),i}^s \qquad \forall i \in V,\ k \in K \tag{12}$$

If paths are not required to be node disjoint the constraints defined by (12) are ignored.

When working with the aggregate formulation on the other hand, we cannot impose such exact diversification constraints. One possibility is to use the following constraint, stating that at most a fraction of σ_k of the aggregate net flow of a commodity into a given MSC can arrive through one connection.

$$f_{jik}^s \leq \sigma_k \sum_{r \in W} D_{kr}^s y_{ri}^s + \sum_{h:\{i,h\} \in E} f_{ihk}^s \qquad \forall \{i,j\} \in E,\ k \in K$$

As mentioned, another way to achieve survivability is by reservation. That is, to ensure the possibility of rerouting a given fraction of demand in the network resulting after a node or edge failure. To include reservation in the problem formulation each scenario should correspond not only to an outcome of the random parameters, but also to a specific failure state (possibly no failure). If all second-stage decisions may be modified in the light of a failure such an extension is easily included in the formulation, simply by modifying the node and/or edge set for each scenario according to the corresponding failure. It is more realistic, however, to assume that only rerouting of traffic is possible, whereas a swift reallocation of BSCs to MSCs or capacity expansion is not practicable. Such a situation would correspond to a three-stage stochastic program. In the first stage, as before, the deployment of MSCs is decided upon. In the second stage the outcome of random parameters is revealed and allocation of BSCs to MSCs and appropriate capacity expansion is carried out. Finally, in the third stage a failure possibly occurs and traffic is rerouted accordingly. Note that a node (MSC) failure in this situation would result in the loss of some demand, since BSCs allocated to the MSC in question would be cut off from the rest. We do not pursue this issue further in the present paper. It should be noted, however, that in theory such a three-stage problem could be solved by the solution procedure presented in the subsequent sections, but in practice the computational overhead involved would render such an approach intractable even for networks of moderate size.

4 Dual Decomposition

In this section we briefly outline the dual decomposition procedure which we are going to apply to problem (1)-(9). Dual decomposition, or scenario decomposition, exploits the fact that the vast majority of variables and constraints in the stochastic program are scenario dependent. In fact the only thing tying the scenarios together are the first-stage decisions on deployment of MSCs. Hence, if we use variable splitting on the first-stage variables, defining a deployment of MSCs for each scenario $x^1, \ldots, x^S$, problem (1)-(9) becomes separable into independent scenario subproblems. The fact that the deployment of MSCs cannot be scenario dependent may now be represented by a *non-anticipativity constraint* stating the problem as

$$\begin{aligned} z = \min \sum_{s=1}^{S} \pi^s \Big(\sum_{i \in V_2} c_i x_i^s + Q^s(x^s) \Big) & \\ \text{s.t. } x^1 = \ldots = x^S, & \\ x^s \in \mathbb{B}^{|V_2|} \qquad \forall s \in \{1, \ldots, S\}. & \end{aligned} \tag{13}$$

Relaxing the non-anticipativity constraint we obtain a problem which is completely separable into independent scenario subproblems. These subproblems are solved to obtain an optimal deployment of MSCs for each scenario. Next non-anticipativity is reinforced by branching on components of these solutions which differ among scenarios. To be specific, we introduce a branching tree initially consisting of only the root node corresponding to the original problem (13). In a given iteration we select a problem from the branching tree and solve the corresponding scenario subproblems obtaining scenario solutions $x^1, \ldots, x^S$. If MSC i is to be deployed in some scenario solutions and not in others, we add two problems to the branching tree imposing for $s = 1, \ldots, S$ the constraints $x_i^s = 0$ and $x_i^s = 1$ respectively. Otherwise, if all scenario solutions are equal, we have a feasible solution of the original problem and may update the upper bound if appropriate. For a thorough description of such a procedure, including a Lagrangian relaxation of the non-anticipativity constraints, we refer to Carøe and Schultz [3].

Clearly, if the scenario subproblems are solved by means of some branch and bound procedure, some effort should be taken to put information from previous iterations in the above procedure to use. Thus a node which is fathomed in a given subproblem in some iteration of the main procedure may be reconsidered in subsequent iterations since more variables are fixed as the main procedure progresses. In fact, for the problem instance considered in Section 6, the number of first-stage variables was so small (less than 20) that an enumeration tree could be created a priori and used for all scenarios, thus precluding any re-evalutions of nodes.

5 Valid Inequalities

In order to solve problem (1)-(9) using the dual decomposition procedure outlined in the previous section we need an efficient procedure for solving the scenario subproblems. To this end we will apply the concept of branch and cut which have proven to be a powerful tool for the solution of (mixed-) integer programming problems. As in ordinary branch and bound we start with the LP-relaxation of the mixed-integer programming problem and build a partitioning of the solution space in order to obtain an integral solution. The crucial idea in branch and cut is to combine this approach with a continuous generation of cutting planes tightening the formulation and thus reducing the size of the branching tree. For a thorough discussion of the branch and cut approach we refer to Padberg and Rinaldi [10] and Günlük [7]. As cutting planes we will use valid inequalities derived through simple polyhedral considerations.

First, we consider an inequality based on the total VLR-capacity installed through deployment of new MSCs. The inequality simply states that the total capacity of all VLRs in the resulting network should exceed the total demand from all BSCs. Formally the inequality is derived by summing the constraints (4)-(5), rearranging and rounding.

$$\sum_{i \in V_2} x_i^s \geq \left\lceil \frac{1}{M} \Big(\sum_{r \in W} L_r^s - \sum_{i \in V_1} M_i \Big) \right\rceil \qquad \forall s \in \{1, \dots, S\}.$$

Here we have defined $M := \max_{i \in V_2} M_i$. Since the deployment of MSCs is not allowed to be scenario dependent this inequality may be strengthened further:

Proposition 1 *The following inequality is valid for the feasible region of all scenario subproblems, $s = 1, \dots, S$.*

$$\sum_{i \in V_2} x_i^s \geq \max_{\tau \in \{1,\dots,S\}} \left\lceil \frac{1}{M} \Big(\sum_{r \in W} L_r^\tau - \sum_{i \in V_1} M_i \Big) \right\rceil .$$

This inequality may be viewed as a global constraint in the sense that it is valid for all scenarios. As mentioned in the previous section we used an enumeration tree to solve subproblems for the instance considered in Section 6. Hence the above inequality was not actually included in the formulation but was merely used to reduce the size of the enumeration tree.

Next we consider a local constraint which is only guaranteed to be valid for the particular scenario from which it was derived. This inequality is based on the VLR-capacity of the individual MSCs and is used to enforce the fact that each BSC must be allocated to a unique MSC. Once again the underlying idea is simple. If the total demand from a group of BSCs exceeds the VLR-capacity of an MSC, we cannot allocate all of these BSCs to the MSC in question. This is formalized in the following proposition.

Proposition 2 *Let U be a subset of W such that $\sum_{r \in U} L_r^s > M_i$ for some MSC $i \in V$ and some scenario $s \in \{1, \ldots, S\}$. Then the following inequality is valid for the feasible region of the s'th scenario subproblem.*

$$\sum_{r \in U} y_{ri}^s \leq |U| - 1.$$

Naturally, this inequality will only be useful when the subset U of W is minimal in the sense that $\sum_{r \in U \setminus \{t\}} L_r^s \leq M_i$ for all $t \in U$, since it is otherwise dominated by other inequalities of the same type.

6 Numerical Results

In this section we will describe the practical application of our model. We have implemented our model on a real problem provided by Sonofon, a Danish mobile communication network provider. In this section we briefly describe the problem instance, the structure of costs and demand, and the practical collection and estimation of data. Due to competitive conditions, however, we cannot be too specific about the problem size and the input data. Finally, we report our computational results.

The problem under consideration has between 5 and 10 existing MSCs, less than 20 potential locations for new MSCs and less than 50 BSCs. The network interconnecting the MSCs is complete. The number of binary variables were reduced by dividing the area of interest into a number of regions and precluding from consideration certain allocations of BSCs to MSCs across regions. In the resulting formulation each scenario subproblem has 707 binary variables, 14598 continuous variables and 12045 constraints.

The cost of a new MSC is orders of magnitude higher than any other cost parameter. The cost of connecting a BSC to an MSC was set to zero if the BSC is currently connected to this particular MSC, and otherwise the total cost of a movement was estimated. Furthermore, the cost of expanding link capacities is given by the total cost of installing new equipment. The issue of determining an appropriate level for the artificial penalty cost for handovers, however, is a more complicated matter. Setting this level too low, may result in solutions with a large number of handovers which are not acceptable from a practical viewpoint. A high level, on the other hand, may result in configurations for which the gained practicability obtained by reducing the number of handovers is not sufficient to justify the increased cost. As a side effect computation time is likely to be increased in this case due to the large number of movements of BSCs required to reduce the number of handovers. In practice we chose to adjust the handover costs, observing their effect on solutions, so as to create geographically connected BSC areas.

The current demand for bandwidth and VLR-capacity was estimated from observations of traffic and the number of customers respectively. Future demand was then calculated using the estimates of current demand scaled by different scenario dependent growth factors. We have used the following procedure to generate demand for VLR-capacity at BSC r under scenario s,

$$L_r^s = \mu^s \cdot \rho_r^s \cdot L_r.$$

Here L_r is the current demand for VLR-capacity at BSC r, μ^s is a parameter reflecting the average growth in the number of customers, while ρ_r^s is a parameter reflecting regional fluctuations from this average growth. To capture the correlation between the demand for VLR-capacity and the demand for bandwidth, the net demand for commodity k at BSC r under scenario s was calculated using current demand D_{kr}, the above-mentioned parameters reflecting growth in the number of customers, and a third parameter σ^s reflecting growth in the demand for bandwidth per customer,

$$D_{kr}^s = \mu^s \cdot \sqrt{\rho_k^s \cdot \rho_r^s} \cdot \sigma^s \cdot D_{kr}.$$

Note that we have used the geometric average of the regional fluctuations ρ_k^s and ρ_r^s. Likewise the different cost terms were made scenario dependent by introducing stochastic fluctuations on future prices. The growth factors were all sampled from uniform distributions reflecting the expectations of Sonofon for the time horizon under consideration. As pointed out in Section 1, the second-stage decision of allocation of BSCs to MSCs is to be made after one year, and this was the time horizon used when estimating growth factors for the cost terms. As for customer demand, however, we have used a four-year time horizon when estimating the appropriate growth factors. This was done to ensure a somewhat stable solution guaranteeing sufficient network capacity for three additional years beyond the completed deployment of new MSCs. This means that demand is in fact only partially revealed at the time the second-stage decisions are to be made, but since the additional information obtained at this point will provide an improved estimate of the true rate of growth in demand, the gain of postponing some decisions to the second stage is likely to be considerable.

The algorithm was implemented in C++ using procedures from the callable library from CPLEX 6.6. Considering 100 scenarios, the solution times were about 3.5 hours CPU-time on a 700 MHz Linux PC. The solution suggested the deployment of one new MSC. Due to the complexity of the problem, the survivability constraints of Section 3 have not been implemented in the application. The valid inequalities of Section 5, however, have speeded up the solution times considerably.

7 Conclusions

In this paper we have set up a model for the optimal deployment of new MSCs in a mobile communications network. The model takes into account the cost of new MSCs, the cost of allocating BSCs to MSCs, and the cost of expanding capacities of links connecting the MSCs. Furthermore, a penalty cost was introduced to limit the number of handovers, inducing connected BSC areas. Since the deployment of MSCs involves a planning horizon of about a year, a number of important parameters of the model are not known with certainty at the point of decision. This lead us to a two-stage stochastic programming formulation of the problem. Considering 100 possible scenarios for the random parameters, the resulting formulation of a real-life problem contained more than a million variables and constraints and hence decompostion methods were called for. We chose to solve the problem using a dual decomposition procedure, solving scenario subproblems by means of branch and cut. The algorithm was implemented in C++ and the problem could be solved to optimality within a few hours of CPU time. We conclude that our model has been successfully implemented, and that it incorporates the most important details of the problem. We also conclude that the stochastic programming model is an important tool in the decision process, giving insight of the dynamics of the expansion problem.

References

[1] A. Balakrishnan, T.L. Magnanti, and R.T. Wong. A dual-ascent procedure for large-scale uncapacitated network design. *Operations Research*, 37:716–740, 1989.

[2] D. Bienstock and O. Günlük. Capacitated network design. Polyhedral structure and computation. *INFORMS Journal on Computing*, 8(3):243–259, 1996.

[3] C.C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24:37–45, 1999.

[4] S.-G. Chang and B. Gavish. Telecommunications network topological design and capacity expansion: Formulations and algorithms. *Telecommunication Systems*, 1:99–131, 1993.

[5] G. Dahl and M. Stoer. A cutting plane algorithm for multicommodity survivable network design problems. *INFORMS Journal on Computing*, 10(1):1–11, 1998.

[6] M.A.H. Dempster, E.A. Medova, and R.T. Thompson. A stochastic programming approach to network planning. *Teletraffic Contributions for the Information Age. Proceedings of the 15th International Teletraffic Congress - ITC 15*, 1:329–339, 1997.

[7] O. Günlük. A branch-and-cut algorithm for capacitated network design. *Mathematical Programming*, 86:17–39, 1999.

[8] A. Merchant and B. Sengupta. Assignment of cells to switches in PCS networks. *IEEE/ACM Transactions on networking*, 3(5):521–526, 1995.

[9] J.M. Mulvey and A. Ruszczynski. A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43:477–490, 1995.

[10] M. Padberg and G. Rinaldi. A branch-and-cut algorithm for the resolution of large-scale symmetric travelling salesman problems. *SIAM Review*, 33(1):60–100, 1991.

[11] M. Riis and K.A. Andersen. Capacitated network design with uncertain demand. Working Paper 2000/5, University of Aarhus, Department of Operations Research, 2000.

[12] M. Riis and K.A. Andersen. On using stochastic programming to plan the multiperiod capacity expansion of one connection in telecommunications. Working Paper 2000/2, University of Aarhus, Department of Operations Research, 2000.

[13] M. Rios, V. Marianov, and M. Gutierrez. Survivable capacitated network design problem: new formulation and Lagrangean relaxation. *Journal of the Operational Research Society*, 51:574–582, 2000.

[14] R.T. Rockafellar and R.J-B Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.

[15] D. Saha, A. Mukherjee, and P.S. Bhattacharya. A simple heuristic for assignment of cells to switches in a PCS network. *Wireless Personal Communications*, 12:209–224, 2000.

[16] S. Sen, R.D. Doverspike, and S. Cosares. Network planning with random demand. *Telecommunication Systems*, 3:11–30, 1994.

[17] E.C. Tzifa, V.P. Demestichas, M.E. Theologou, and M.E. Anagnostou. Design of the access network segment of future mobile communications systems. *Wireless Personal Communications*, 11:247–268, 1999.