Hierarchical Infrastructure Planning in Networks

Thorsten Koch and Roland Wessäly

Zuse Institute Berlin, Takustr. 7, D-14195 Berlin, Germany,
tel: +49-30-84185-{213,245}, fax: +49-30-84185-269, e-mail: {koch,wessaely}@zib.de
atesio GmbH, Rubensstr. 126, D-12157 Berlin, Germany,
tel: +49-30-79786846, fax: +49-30-79786843, e-mail: {koch,wessaely}@atesio.de

Abstract

In this article, strategical infrastructure planning problems in the design of large-scale telecommunication networks are discussed based on experiences from three projects with industrial partners: The access network planning of the German Gigabit-Wissenschaftsnetz (G-WiN) for DFN (Verein zur Förderung eines Deutschen Forschungsnetzes e.V.), the mobile network switching center location planning project for e·plus Mobilfunk, and the fixed network switching center location planning project for TELEKOM AUSTRIA.

We introduce a mathematical model for a hierachical multi-commodity capacitated facility location problem, present adaptions of this basic model to the specific requirements within the different projects and discuss the individual peculiarities and model decisions made. Eventually, we present and discuss computational results of three associated case studies, illustrating “how we did the job” with mathematical methods.

1 Introduction

This article is a digest of experiences from three projects: The access network planning for the German Gigabit-Wissenschaftsnetz (G-WiN) conducted together with the DFN (Verein zur Förderung eines Deutschen Forschungsnetzes e.V.), the mobile switching center location planning project conducted together with e·plus, and the fixed network switching center location planning project conducted together with TELEKOM AUSTRIA. The main purpose of this article is to show how the mathematical toolbox can be applied to real world planning problems. After describing the background of the three planning tasks in Section 2, a mathematical model for the hierarchical multi-commodity capacitated facility location problem is presented in Section 3. Though these projects are technically different, we will see how they can be mapped to essentially the same mathematical model. We will present for each project how we adapted the model to its specific requirements, how we dealt with peculiarities, and note special problems that result from the decisions made in the projects. Since these are case studies, we do not try to completely cover the subjects, but give illustrated “how did we do it” stories with some notes on details that need attention.

2 Three network hierarchy planning tasks

We investigated the problem to define a network hierarchy for three different providers: (i) DFN as IP-network provider, (ii) e·plus as mobile phone network provider, (iii) TELEKOM AUSTRIA as (incumbent) fixed network provider.

2.1 Planning an Access Router Network (G-WiN)

In 1998 we got involved into the planning of what should become Germany's largest IP network, the Gigabit Research Network (G-WiN) operated by the DFN. All major universities and research facilities were to be connected. The network was planned to handle up to 220 TB traffic per hour in its first year. An increase rate of 2.2 annually was anticipated, leading to a planned capacity of about 5,000 TB in 2003.

DFN is not a carrier owning the physical infrastructure to connect locations. Therefore, a call for bids had to be issued to find a carrier for the network. European law requires that any call for bids exactly specifies what the participants are bidding on. This means the DFN had to come up with a network design before calling for bids. As a result, it was decided to design some kind of sensible network and hope the participants of the bidding were able to implement it cheaply. The network should consist of 30 backbone nodes. Ten of these backbone nodes should become interconnected *core* nodes, while the other 20 backbone nodes should be connected pairwise to a core node. We will see in Section 4.2 that the decision to have ten core nodes was probably the most important one in the whole process. For more information on the design of the network connecting the core nodes see Bley and Koch (2000), Bley et al. (2004).

For G-WiN, no distinction between transport network and switching network was necessary, as the bid was for the logical or virtual network as specified by the DFN. The mapping of logical to physical connections was left to the provider of the link. As a result no pricing information for installing links between the nodes was available before the bidding. It was decided to use costs according to those of the predecessor network B-WiN (Bley et al., 1998), but scale them by some factor to anticipate declining prices. Bee-line distances between the locations were used as link distances. Since the hardware to be installed at the nodes was either unknown, not yet available from the vendors, or dependent on the carrier, no real costs or capacities were known. The initial problem for the access network was given as follows: Having 337 nodes from which 224 are potential backbone nodes, select 30 backbone nodes and connect each of the remaining nodes to them.

Connections to backbone nodes had to have one of the following discrete capacities: 128 kbit/s, 2 Mbit/s, 34 Mbit/s, 622 Mbit/s, 2.4 Kbit/s, or 10 GBit/s. Initially clients demanding 128 kbit/s were not considered and none of the clients needed more than 622 Mbit/s. The associated cost function is shown in Figure 1. Additionally Figure 2 visualizes for each location the demand for the peak traffic hour. Since the hardware installed at the backbone nodes has to operate 24 hours, seven days a week, places with suitable maintenance, air conditioning and uninterruptible power supplies are required.

While 224 of the sites were capable in principle to host the hardware, some were preferred. We modeled this by decreasing the cost for the preferred nodes by a small margin.

But even for the preferred locations, the conditions for hosting the equipment had to be negotiated. This led to iterated solutions with consecutively more and more fixed sites. In the end the problem degenerated to a pure assignment problem. For the same reasons the selection of the core nodes was done at DFN. Finally we also added the 128 kbit/s clients to the problem, bringing the total number of locations to 761.

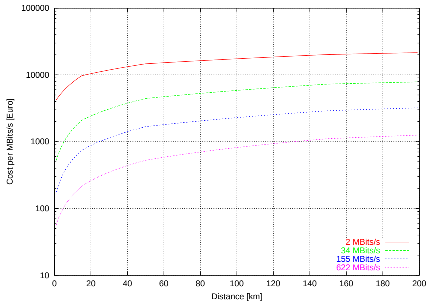


Figure 1. Cost rel. to distance per Mbit/s

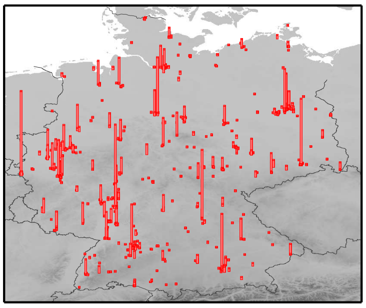


Figure 2. Location demands of G-WiN

We said in the beginning that the annual increase in traffic was to be taken into account. Since the increase was given as a linear factor on all traffic, the only change could have resulted from the discretization of the link capacities, but it turned out that the resulting difference were neglegtible.

2.2 Planning Mobile Switching Center Locations (e·plus)

In the project conducted together with e·plus, the logical layout of a part of a GSM network had to be examined. In a GSM mobile network, the signal from a mobile is transmitted to an antenna that is located at a *Base Transceiver Station* (BTS). The BTS is connected to a *Base Station Controllers* (BSC). The BSC manages the transmitter and receiver resources for the connected base stations and controls the traffic between the base station and the *Mobile Switching Center* (MSC).

The MSCs are at the core of the network and connect all the BSCs with each other via connections to the other MSCs in the network. MSCs are switches, that can be build with different capacities. One resource limiting the capacity of a MSC is the number of *subscribers* which is determined by the database attached to the MSC, the so-called *Visitor Location Register* (VLR). Depending on the traffic it manages, each BSC takes up a number of subscribers. The installation cost of a MSC depends on its subscriber capacity. The connection costs between a BSC and a MSC de-

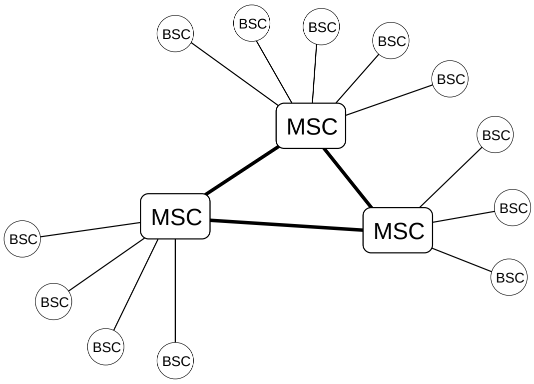


Figure 3. Logical GSM backbone

pend on the data rate of the link. Since e·plus owned only part of its transport network and leased links on demand, it was difficult to associate costs to links in a combinatorial way. The price for each new link had to be individually investigated. As a result we tried different costs functions within the project, either similar in appearance to the one given in Figure 1, or just linear depend on the capacity and the distance.

We can state the problem as follows: Given a list of BSCs, a list of potential MSC locations, and a list of possible MSC configurations, decide where to place MSCs and for each BSC to which MSC it should be connected. Choose a suitable configuration for each MSC.

2.3 Planning Fixed Network Switching Center Locations (Telekom Austria)

Telephone networks are so-called *circuit switched* networks, i. e., if one terminal is calling another terminal, the request is transmitted first to the appropriate switching center, which, depending on the location of the destination terminal, selects (switches) the route to the next switching center. This is repeated until the destination terminal is reached. In a circuit switched network this route between the two terminals is created at the beginning of the transmission and stays alive until the call is finished. The required bandwidth remains reserved all of the time.

The switching network of the Telekom Austria has a hierarchical design. Seven *Main Switching Centers* (HV, german: Hauptvermittlungsstellen) are the backbone of the network. On the level below are about 100 *Network Switching Centers* (NV, german: Netzvermittlungsstellen). Next are about 140 *City Switching Centers* (OV, german: Ortsvermittlungsstellen) and finally, at the bottom level, are about 1,200 *Passive Switching Centers* (UV, german: Unselbstständige Vermittlungsstellen).

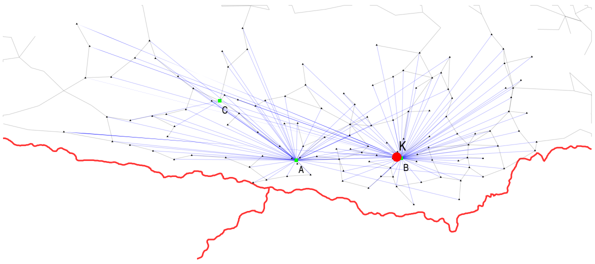


Figure 4. Switching network

The topology of the network is basically a tree apart from the HVs which are linked directly to each other. All other switching centers have to be connected to a center on a higher level than themselves. HVs, NVs, and OVs are called *Full Switching Centers* (VV, german: Vollvermittlungsstellen) because they are able to handle internal traffic themselves, i. e., traffic that does not need to be routed higher up in the hierarchy. In contrast UVs transfer all traffic to their associated VV. In Figure 4, the large (red) circle is a HV, the (green) squares are VVs and the (black) triangles mark UVs. The (blue) lines are the logical connections between switching centers, while the (gray) lines show the physical transport network.

Due to technical advances the capacity of a single switching center has been vastly increased in the last years. At the same time the cost for the transport network has steadily declined. Since the operation costs for maintaining a location is high, a smaller number of switching centers is desirable. In consequence, the goal of the project was to develop planning scenarios for the reduction of the number of full switching centers, that is, to decide for each switching center whether it should be up- or downgraded and to which other center it should be connected.

There are a few things to note: (i) changing a switching center either way induces some cost since this is not a green-field scenario, (ii) the distinction in the hierarchy between OVs and NVs has been dropped and only a three level network, i. e., UV, VV, and HV, was considered, (iii) the switching centers are build by two different manufacturers which are not compatible below HV level. In consequence, only switching centers of the same manufacturer can be connected to each other.

2.3.1 Demands and Capacities

The capacities of the switching centers are limited by the number of users connected and by the amount of traffic to be switched. There was an other restriction called *Zoning Origins* (Verzonende Ursprünge) of which each full switching center could only serve a limited number. Since the whole subject was rather archaic and the numbers were somewhat hard to compute precisely, it was decided later in the project to drop the restriction, even though it fitted easily into the model.

There are three possible terminals connected to an UV: (i) POTS (Plain Old Telephone Service), (ii) ISDN-basic-rate (Integrated Services Digital Network), and (iii) ISDN-primary-rate. Only POTS and ISDN-basic-rate draw from the users restrictions of the switching centers. Assigned to each terminal type is a typical amount of traffic in Erlangs. This is converted along the Erlang-B formula (Erlang, 1917) to 64 kbit/s (voice) channels. All traffic is assumed to be symmetric between the terminals.

As noted before, all non-passive switching centers can route internal traffic directly. This can reduce the amount of traffic to the HVs and within the backbone. Finding an optimal partitioning of the network to minimize the external traffic is $\mathcal{NP}$-hard (see Garey and Johnson (1979), Grötschel and Wakabayashi (1990), Chopra and Rao (1993)) and requires a complete end-to-end traffic matrix. Since end-to-end traffic demands were not available in the project, it was not possible to precisely model this effect. But there is quite accurate empirical knowledge on the percentage of external traffic for each region. So it is possible to attach a fixed traffic reduction factor β (see Section 3.1) to each VV to approximate the effect.

To give an idea about the scenario, here are some numbers: More than three million users are distributed about 300:30:1 onto POTS, ISDN-basic-rate, and ISDN-primary-rate terminals. The traffic demand per terminal is about 0.06, 0.12, and 0.9 Erlangs, respectively. The total traffic is more than 220,000 Erlangs. A VV can serve about 120,000 users and 80,000 channels. The capacity of a HV is about ten times as big. These are just general numbers as switching centers can be configured in various ways.

2.3.2 Costs

Hardware costs for installing, changing, and operating switching centers are relatively easy to determine. The biggest problems are:

- How to assess hardware that is already deployed and payed for.
- How to assess the depreciation of the purchase cost, if it is to be included at all.
- If different configurations for switching centers are possible, usually only a complete setup can be assigned a price tag.

The switching network is a logical network that creates the circuits between the terminals by building paths of logical connections between switching centers. The transport network is the physical network below that transmits the data. It is difficult to estimate the cost of links in the switching network based on the transport network since the transport network is already existing and the relation between traffic demands and routing of such links in the transport network is not clear. Assuming that the transport network is able to cope with the traffic demand as induced by the current switching network and assumed further that our "optimized" switching network does not require an extension of the transport network, no real cost will occur. It follows, that the cost optimal switching network has the minimum number of switching centers possible according to the capacity restrictions. And the question where a switching center should be connected to, could be mostly neglected.

Nevertheless the transport network has a cost associated with it. Assuming that the amount of voice calls is rather static, any excess capacity can be used for other services, e. g. packet data services. As a result some cost function is needed, but can be arbitrarily defined.

After some discussions TELEKOM AUSTRIA supplied the cost function shown in Figure 5. The basic idea is to pay a base price per channel for the existing fiber optic cables. Since these cables need repeaters after every 45 km which induce operating cost the price is raised after 45 and 90 km. The reason for the higher cost of the VV to HV connections results from the higher infrastructure demands of these links due to the higher capacities needed. Note that since we usually assume about 30% internal traffic, the price for the VV to HV connection is multiplied with $\beta = 0.7$, making it in total cheaper than the UV to VV connection for the same number of channels.

Given the cost function, the question which distances to used arises. In the former projects we always used bee-line distances, since the transport network was not owned by the network operator and not much was known about it. In this project we had the possibility to compute distances in the transport network. Regarding the rationale for the cost function, which involved repeaters in the fiber network, this seemed to allow a much better estimation of the involved costs.

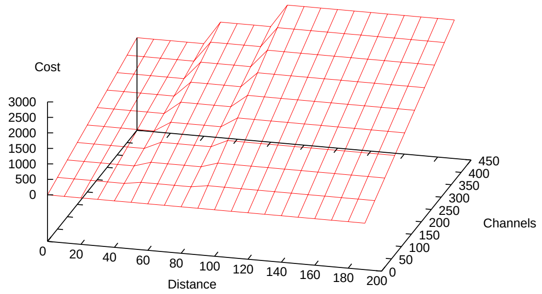


Figure 5. Cost function w.r.t distance and channels

3 Mathematical models for network hierarchy planning

In this section we present a linear mixed-integer model for the hierarchical multi-commodity capacitated facility location problem, which serves as basic formulation for all three network hierarchy planning problems. This basic formulation is then specialized to fulfill particular project requirements.

3.1 The basic model

Given is a layered directed graph $G = (V, A)$ with N hierarchy-levels $L = \{1, \dots, N\}$. The nodes are partitioned into layers as $V_1, \dots, V_N$, with $V_m \cap V_n = \emptyset$ for all $m, n \in L$ and $V = \bigcup_{n \in L} V_n$. Without loss of generality $|V_N| = 1$ is assumed and $r \in V_N$ denotes the *root*. The nodes are connected with arcs $A \subseteq \{(u, v) \in A \mid u \in V_n, v \in V_{n+1}, n, n+1 \in L\}$.

We are looking for a tree that connects all level one nodes with the root. Since G is layered this means each level one node has to be connected to exactly one level two node. These in turn have to be connected to exactly one level three node and so on. This is essentially a Steiner arborescence problem (see, for example, Koch and Martin, 1998) with $V_1 \cup \{r\}$ as terminal set. For each node $v \in V$ and each arc $(u, v) \in A$, binary variables y_v and x_{uv} are introduced, respectively. Each y_v and each x_{uv} is equal to one if and only if the node or arc is active, i. e., is part of the solution. This leads to the following formulation:

$$y_v = 1 \qquad \text{for all } v \in V_1 \tag{1}$$

$$x_{uv} \le y_v \qquad \text{for all } (u, v) \in A \tag{2}$$

$$\sum_{(v,w) \in A} x_{vw} = y_v \qquad \text{for all } n \in L \setminus \{N\}, v \in V_n \tag{3}$$

Note that for $r \in V_N$ the above system implies $y_r = 1$.

Commodities For each node $v \in V$ and each $d \in D$ of a set of commodities (resources), a demand $\delta_v^d \ge 0$ is given, specifing the demand which must be routed from node $v \in V$ to the root node $r \in V_N$. This can be modeled by introducing a non-negative continuous variable f_{uv}^d, $(u, v) \in A$, with upper bound $\rho_{uv}^d \ge 0$, denoting the amount of flow of commodity d from node u to v.

$$\delta_v^d + \beta_v^d \sum_{(u,v) \in A} f_{uv}^d = \sum_{(v,w) \in A} f_{vw}^d \qquad \text{for all } v \in V \setminus \{r\}, d \in D \tag{4}$$

$\beta_v^d > 0$ is a "compression" factor, i. e., all incoming flow into node v of commodity d can be compressed (or enlarged) by β_v^d. Applications for this factor are, for example, data compression in higher hierachie levels, or heuristic approximation of traffic handled locally. If we assume all β_v^d equal within each layer, the total amount of flow of each commodity reaching the root will be constant. Note that for any $v \in V_1$ inequality (4) reduces to $\delta_v^d = \sum_{(v,w) \in A} f_{vw}^d$. Only active arcs can carry flow,

i. e.,

$$\rho^d_{uv} x_{uv} \geq f^d_{uv} \quad \text{for all } (u,v) \in A, d \in D \tag{5}$$

Note that equation (1) is redundant for any $d \in D$ with $\delta^d_v > 0$ as a result of (5).

Capacities For each node $v \in V$ a set S_v of configurations is defined. Associated with each configuration $s \in S_v$ is a capacity κ^d_s for each commodity $d \in D$. We introduce binary variables z_s for each $v \in V$ and each $s \in S_v$. The variable z_s is one if and only if configuration s is active for node v. For each active node, a configuration with sufficient capacity to handle the incoming flow is required, i. e.,

$$\sum_{s \in S_v} z_s = y_v \qquad \text{for all } v \in V \tag{6}$$

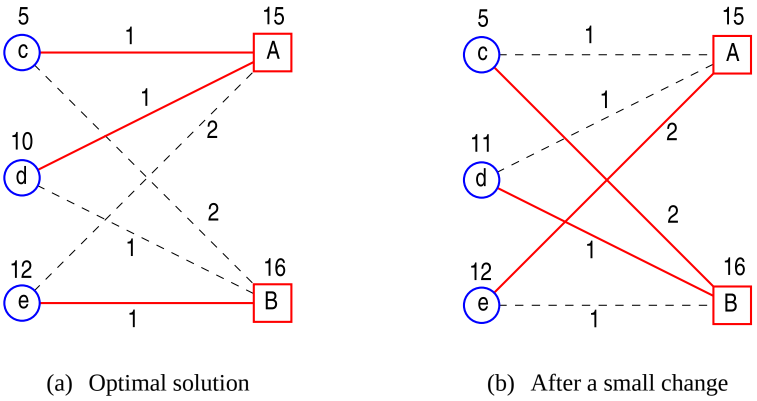
$$\sum_{(u,v) \in A} f^d_{uv} \leq \sum_{s \in S_v} \kappa^d_s z_s \qquad \text{for all } v \in V, d \in D \tag{7}$$

Notice that the configuration of all level one nodes can be predetermined, since there is no incoming flow apart from δ^d_v. For each link $(u,v) \in A$ and each commodity $d \in D$, a set K_d of discrete capacities is given. Which one is selected is modeled with binary variables $\bar{x}^{dk}_{uv}$ by the following constraints

$$\sum_{k \in K_d} \bar{x}^{dk}_{uv} = x_{uv} \qquad \text{for all } (u,v) \in A, d \in D \tag{8}$$

$$\sum_{k \in K_d} k \bar{x}^{dk}_{uv} \geq f^d_{uv} \qquad \text{for all } (u,v) \in A, d \in D \tag{9}$$

Equation (8) ensures that for each active arc one of the possible capacities is chosen. Inequality (9) makes sure the link has sufficient capacity. Note that depending on the particular problem simplifications are possible, especially regarding (3) and (8), and (5), (7), and (9).



(a) Optimal solution (b) After a small change

Figure 6. Instable solution

Configurations, as all types of (hard) capacity constraints, can lead to instable solutions, i. e., solutions that vary considerable upon small changes of the input data. Figure 6 shows an example. Given are three nodes c, d, e with demands $\delta_v = 5, 10, 12$, respectively. The two serving nodes A and B have a fixed capacity of 15 and 16, respectively. The costs for connecting the demand nodes with the serving nodes is given in the figures along the connections. Figure 6a shows the optimal solution when minimizing connection costs. Now increasing the demand of node d by one, results in the solution shown in Figure 6b, that is, changing a single demand by a small amount leads to a completely different solution. This is undesirable, since input data is often inaccurate.

Apart from being unstable, solutions where nodes are not connected to the cheapest available higher level node just look wrong. To prevent this, inequalities like

$$x_{uv} \leq 1 - y_w \quad \forall (u,v) \in A, w \in V \text{ with } c_{uv} > c_{uw}$$

can be introduced, where c_{uv} for $(u,v) \in A$ denotes the costs associated with a connection between node u and node v.

Objective function The objective is to minimize the total cost of the solution, i. e.,

$$\min \sum_{v \in V} (y_v + \sum_{s \in S_v} z_s) + \sum_{(u,v) \in A} \Big(x_{uv} + \sum_{d \in D} (f^d_{uv} + \sum_{k \in K_d} \bar{x}^{dk}_{uv}) \Big)$$

with appropriate (defined for each project) objective coefficients for all variables.

Literature The capacitated facility location problem is well-studied and of considerable importance in practice. As we mentioned before, it can be seen as a capacitated Steiner arborescence problem, or as a partitioning or clustering problem. Many variations are possible. As a result a vast amount of literature on the problem, variations, subproblems, and relaxations has been published. See, for example, Balakrishnan et al. (1995), Hall (1996), Mirchandani (1996), Bienstock and Günlück (1996), Aadal et al. (1996), Ferreira et al. (1996, 1998), Park et al. (2000), Holmberg and Yuan (2000), Ortega and Wolsey (2003), Gamvros and Golden (2003), Bley (2003). It should be noted though that the majority of the publications is not related to real-world projects.

3.2 Modeling G-WiN

We modeled the problem with three layers and discrete link capacities between the backbone and the core nodes. Using this model, we will make a comparison between our original solution and a less restricted one, were the optimization can decide were to place the core nodes.

V_1 is the set of all nodes. Potential backbone nodes are split into a client part, carrying the demand and belonging to V_1 and a backbone part belonging to V_2. The set of potential core nodes is denoted V_3. While the three sets are disjunctive, the elements might refere to the same physical locations. The function $\sigma(V_3) \to V_2$ maps core nodes to the corresponding co-located backbone node. The set of arcs is defined as $A \subseteq (V_1 \times V_2) \cup (V_2 \times V_3)$. The variables are defined similar to Section 3.1. In particular $x_{uv}, (u,v) \in A \cap (V_1 \times V_2)$ are binary variables denoting which connections are active, $\bar{x}_{vwk}, (v,w) \in A \cap (V_2 \times V_3), k \in K$ are binary variables denoting which capacity is used for a link between a backbone and a core node. In addition to the binary variables $y_v, v \in V_2$, denoting the active backbone nodes, a second set of binary variables $\bar{y}_w, w \in V_3$, denoting the active core nodes is introduced. Since only a single commodity is present, no indexing of variables with $d \in D$ is necessary. The following model describes the problem setting:

$$\sum_{(u,v)\in A} x_{uv} = 1 \qquad \text{for all } u \in V_1 \tag{10}$$

$$x_{uv} \le y_v \qquad \text{for all } u \in V_1, (u,v) \in A \tag{11}$$

$$\sum_{(v,w)\in A} \sum_{k\in K} \bar{x}^k_{vw} = y_v \qquad \text{for all } v \in V_2 \tag{12}$$

$$\sum_{k\in K} \bar{x}^k_{vw} \le \bar{y}_w \qquad \text{for all } v \in V_2, (v,w) \in A \tag{13}$$

$$\sum_{(v,w)\in A} \sum_{k\in K} k\bar{x}^k_{vw} \ge \sum_{(u,v)\in A} \delta_u x_{uv} \qquad \text{for all } v \in V_2 \tag{14}$$

$$\bar{y}_w \le y_{\sigma(w)} \qquad \text{for all } w \in V_3 \tag{15}$$

$$\sum_{w\in V_3} \bar{y}_w = 10 \tag{16}$$

$$\sum_{(v,w)\in A} \sum_{k\in K} \bar{x}^k_{vw} = 3 * \bar{y}_w \qquad \text{for all } w \in V_3 \tag{17}$$

Note that (10) results from combining (1) with (3). Inequality (11) corresponds to (2), while (12) is a combination of (3) and (8), and (13) is a combination of (2) and (8). Inequality (14) is a simplified form of (9). Inequality (15) ensures that only backbone nodes can become core nodes. The number of core nodes is fixed to ten by equation (16). Finally, equation (17) fixes the number of backbone nodes to 30, with the additional requirement that one of every three backbone nodes has to become a core node with two other backbone nodes attached to it. We call this the *normal* scenario. We also examined a *relaxed* scenario, were equations (16) and (17) are removed. If we refer to the *original* scenario, this means the solution used in the project.

3.3 Modeling e·plus

This problem can be formulated using a simplification of the model given in Section 3.1.

$$\sum_{(v,w)\in A} x_{vw} = 1 \qquad \text{for all } v \in V_1 \tag{18}$$

$$\sum_{s\in S_v} z_s = 1 \qquad \text{for all } v \in V_2 \tag{19}$$

$$\sum_{(u,v)\in A} \delta_u x_{uv} \leq \sum_{s\in S_v} \kappa_s z_s \qquad \text{for all } v \in V_2 \tag{20}$$

Let V_1 be the set of BSCs and V_2 be the set of potential MSCs. The parameter δ_u, denotes for each BSC $u \in V_1$ the number of associated subscribers. For each MSC $v \in V_2$, the parameter $\kappa_s, s \in S_v$, denotes the number of subscribers which can be handled with configuration s.

Note that (19) enforces a "zero" configuration, i. e., there has to be exactly one $s \in S_v$ with $\kappa_s = 0$ for each $v \in V_2$. This has the nice property that already existing configurations can be modeled this way. Instead of assigning the "building cost" to a configuration, the cost involved with a particular change is used.

For the computational study in Section 4.2 the following objective function was used:

$$\min \sum_{(u,v)\in A} \mu\, 10\, d_{uv}\, l_{uv}\, x_{uv} + \sum_{v\in V_2} \sum_{s\in S_v} c^z_s z_s$$

with μ being a predefined scaling factor, d_{uv} denoting the bee-line distance in kilometers between BSC u and MSC v, and l_{uv} denoting the number of 64 kbit/s channels needed for the traffic between u and v. The building cost for a configuration s is denoted by c^z_s.

3.4 Modeling Telekom Austria

The model can again be derived from the one described in Section 3.1. The nodes in the model are the switching centers. We call the set of all UV's U, the set of all potential VV's V and the set of all potential HV's H. We denote the set of all switching centers by $W = U \cup V \cup H$. Regarding the notation in Section 3.1, $V_1 = U$, $V_2 = V$, and $V_3 = H$. While the sets U, V, and H are pairwise disjoint, the locations associated with the members of the sets may be the same. We introduce a function $\sigma(w), w \in W$, that maps a switching center to its location. If $\sigma(u) = \sigma(v)$ for $v, w \in W$ we call u and v *co-located*. $A_{UV} \subseteq U \times V$ denotes the set of all possible links between UV's and VV's, $A_{VH} \subseteq V \times H$ the set of all possible links between VV's and HV's. The set of all possible links is denoted by $A = A_{UV} \cup A_{VH}$.

Two types of commodities are used, i. e., $D := \{\text{users}, \text{channels}\}$. Demands $\delta^d_u, u \in U, d \in D$ are only given for UV's. For each VV with demands, a co-located UV with a zero-cost link to the VV is generated.

We introduce binary variables x_{ij} for each $(i, j) \in A$ to indicate active links and the binary variables $y_w, w \in V \cup H$, indicate active VV's in case $w \in V$, and active HV's in case $w \in H$. Finally continuous variables $f^d_{vh} \leq \rho^d_v, d \in D, (v,h) \in A_{VH}$ are used to represent the amount of commodity d requested by VV v from HV h. The parameter ρ^d_v denotes the maximum capacity of commodity d that can be handled by VV v. Similarly parameter $\rho^d_h, h \in H$ represents the maximum capacity of commodity d that can be handled by HV h. This leads to the following model:

$$\sum_{(u,v)\in A_{UV}} x_{uv} = 1 \qquad \text{for all } u \in U \tag{21}$$

$$\sum_{(v,h)\in A_{VH}} x_{vh} = y_v \qquad \text{for all } v \in V \tag{22}$$

$$x_{uv} \leq y_v \qquad \text{for all } (u,v) \in A_{UV} \tag{23}$$

$$x_{vh} \leq y_h \qquad \text{for all } (v,h) \in A_{VH} \tag{24}$$

Constraints (21) to (24) are equivalent to (1) to (3).

$$\sum_{(u,v)\in A_{UV}} \delta^d_u x_{uv} = \sum_{(v,h)\in A_{VH}} f^d_{vh} \qquad \text{for all } v \in V \tag{25}$$

$$\rho^d_h x_{vh} \geq f^d_{vh} \qquad \text{for all } (v,h) \in A_{VH}, d \in D \tag{26}$$

$$\beta_v \sum_{vh\in A_{VH}} f^d_{vh} \leq \rho^d_h \qquad \text{for all } h \in H, d \in D \tag{27}$$

Equation (25) is a simplification of (4) and (26) is similar to (5). Inequality (27) limits the capacity of the VV's. Since the utilization of a VV is dependent on the incoming demands, we have not applied β_v to (25), but to inequality (27) as it reduces the outgoing demands. It is not necessary to explicitly limit the capacity of the VV, since

$$\sum_{uv\in A_{UV}} \delta^d_u x_{uv} \leq \rho^d_v \quad \text{for all } v \in V$$

is implied by (22), (26) and (27). Regarding co-locationed switching centers two special requirements have to be ensured: If a VV is active, any co-located UV has to be connected to it:

$$y_v = x_{uv} \quad \text{for all } (u,v) \in A_{UV} \text{ with } \sigma(u) = \sigma(v)$$

Co-locating a VV and a HV is not allowed:

$$y_v + y_h \leq 1 \quad \text{for all } v \in V, h \in H \text{ with } \sigma(v) = \sigma(h)$$

It should be noted that in the investigated scenarios all $y_h, h \in H$ were fixed to one, since a reduction of the number of HV was not considered.

4 Computing Network Hierarchies

In this section, selected computational results for the three network hierarchy planning problems are presented. All computational results were optained on a 3.2 GHz Pentium-4-EE PC with 2 GB RAM using CPLEX 9.0.

4.1 Results for G-WiN

In this scenario, 337 nodes were considered, comprising 224 potential backbone nodes. 30 nodes were preferred backbone nodes, giving them a small bonus in the objective function. Only connections between demand nodes and backbone nodes of less than 300 kilometers were allowed. Backbone nodes that were attached to core nodes had to be at least 50 kilometer apart from the core node. The cost for opening a backbone node was set to 600. Opening a core node again involved a cost of 600, or 599 for a preferred node. The resulting integer program for the normal scenario has 137,581 binary variables, 70,669 constraints and 581,806 non-zero entries in the constraint matrix.

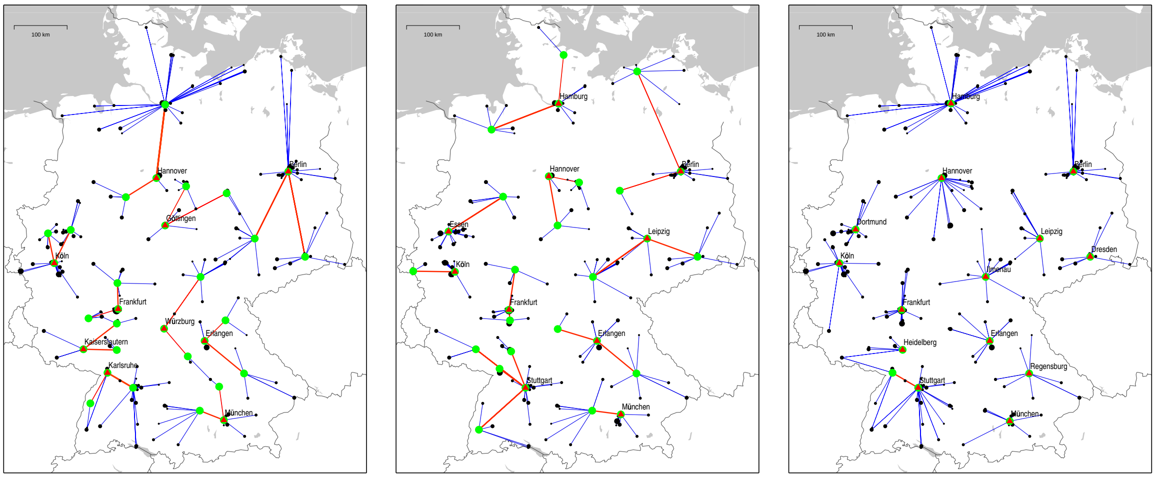
Scenario	Gap [%]	Time [h]	BB	Core	Objective
Normal	7.12	18	30	10	67,593
Relaxed	0.28	3	16	15	58,022
Original	—	—	29	10	67,741

Table 1. G-WiN solution

Table 1 lists the results for the different scenarios. *Gap* shows the optimality gap of the solution. *Time* is the approximate CPU time spend by CPLEX for solving the instance. *BB* and *Core* give the number of backbone and core nodes, respectively. *Objective* list the objective function value for the scenarios. The cost for the *original* scenario is almost equal to the cost for the *normal* scenario, indicating, that given the number of backbone and core nodes is fixed in advance, the original solution is less than 10% off the optimum.

Figure 7 shows images of the results. Backbone nodes are marked as thick (green) circles, core nodes are drawn as (red) triangles. The picture indicates that the cost for the backbone to core node links was set too high to make them pay off. While the relaxed scenario seem to incur the least cost, keep in mind that we have not included any costs for the core network whatsoever and the objective value for the relaxed scenario is only 17% smaller than for the original one.

The bid for the carrier was won by the German Telekom. By now the G-WiN is running very successful for more than four years and has been reconfigured and upgraded several times. Between 2001 and 2002 we investigated the profitability of introducing a third layer of backbone nodes and discovered some potential candidates.



(a) Normal (b) Original (c) Relaxed

Figure 7. Result for G-WiN access network planning.

4.2 Results for e·plus

We computed solutions for ten different scenarios. Table 2 list the parameters that are equal in all cases. The scenarios are partitioned into two groups. The number of subscribers in the first group was 3.4 million, and 6.8 million in the second group. For each group, five solutions with different connection costs were computed, using $\mu = 0.1, 0.5, 1, 2, 10$. All computations were conducted with CPLEX, default settings[1].

Number of BSCs $\lvert V_1 \rvert$	212
Number of potential MSCs $\lvert V_2 \rvert$	85
Maximum allowed BSC/MSC distance [km]	300
Minimum MSC capacity (subscribers)	50,000
Maximum MSC capacity (subscribers)	2,000,000
Number of different configurations $\lvert S_v \rvert$	14
Binary variables	10,255
Constraints	382
Non-zeros entries in constraint matrix	20,425
CPLEX time limit	1 h

Table 2. Scenario parameters

Table 3 and Figure 8 show the result for the scenarios. *Gap* list the gap between the primal solution and the best dual bound when either the time limit of one hour was reached, or CPLEX ran out of memory. *MSC* is the number of MSCs that serve BSCs. ϕ *util.* is the geometric mean of the utilization

---

[1] In one case it was necessary to lower the *integrality tolerance* from 10^{-5} to 10^{-8}.

of the MSCs. *Hardw. cost* is the sum of the cost of the MSC configurations $(=\sum_{v\in V_2}\sum_{s\in S_v} c^z_s z_s)$. *Chan.*×*km* is the sum of the number of 64 kbit/s-channels times the distance in kilometers needed to connect all BSCs $(=\sum_{(u,v)\in A} d_{uv} l_{uv} x_{uv})$. *Total cost* is the objective function value. All cost figures given are divided by 1,000 and rounded to improve clarity.

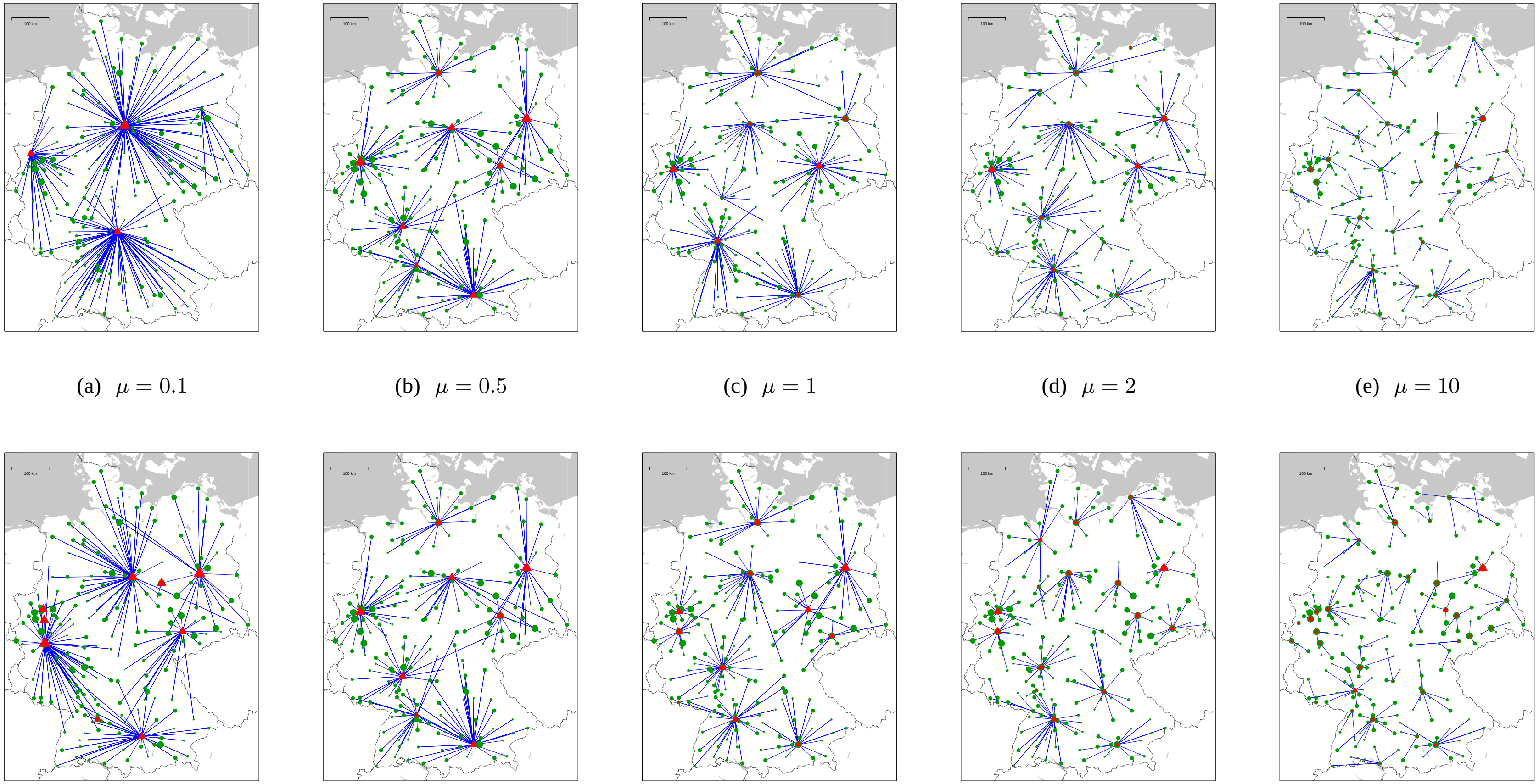
The results are hardly surprising. The higher the connections costs, the more MSCs are opened. Most of the results show the problem we mentioned above, that BSCs got connected to farther away MSCs to circumvent upgrading the nearer ones. This can also happen as a result of not having solved the instances to optimality. It is therefore necessary to post-process these solutions before presenting them, making sure the solutions are two-optimal regarding connection changes.

μ	Fig.	Gap [%]	MSC	ϕ util. [%]	Hardw. cost	Chan. × km	Total cost
			3.4 million users				
0.1	8a	4.66	4	99.4	23,850	1,612	25,462
0.5	8b	3.74	7	99.0	25,884	772	29,744
1	8c	1.96	8	98.7	26,716	659	33,309
2	8d	0.00	12	98.6	29,335	486	39,059
10	8e	0.00	32	93.9	43,199	191	62,265
			6.8 million users				
0.1	8f	2.78	5	99.8	46,202	1,987	48,189
0.5	8g	1.49	8	99.7	47,952	1,179	53,846
1	8h	0.30	11	99.6	49,636	926	58,897
2	8i	1.12	19	97.5	55,473	570	66,873
10	8j	0.13	40	96.7	72,200	250	97,199

Table 3. Results for MSC location planning

In an earlier similar study, DISCNET (Wessäly, 2000) was used in a subsequent step to design and dimension the inter-MSC transport network for each of the solutions. As it turned out, the costs for the inter-MSC network varied highly between the scenarios and dominated the costs for the BSC-MSC network by a huge amount.

From this result we concluded that some interaction between the location planning and the planning of the inter-MSC backbone network is necessary. A possible solution might be to assign costs to the connections between the V_2 nodes (MSCs) and a virtual root. But since this corresponds to a star shaped backbone network it is not clear if it is possible to find suitable costs that resemble a real backbone network somehow. An integrated model, as presented in Bley et al. (2004), seems to be more promising here.



(a) $\mu = 0.1$ (b) $\mu = 0.5$ (c) $\mu = 1$ (d) $\mu = 2$ (e) $\mu = 10$

(f) $\mu = 0.1$ (g) $\mu = 0.5$ (h) $\mu = 1$ (i) $\mu = 2$ (j) $\mu = 10$

Figure 8. Result for MSC location planning. Upper row 3.4 million users, lower row 6.8 million users

4.3 Results for Telekom Austria

Austria has nine federal states Burgenland, Carinthia, Lower Austria, Upper Austria, Salzburg, Styria, Tyrol, Vorarlberg, and Vienna. This is reflected in the telecommunication network, since all equipment within each state is from the same manufacturer. An exception is Vienna, which has two main switching centers, one from each manufacturer.

Region	HV	VV	UV
Salzburg / Upper Austria	1	12	358
Tyrol / Vorarlberg	1	7	181
Carinthia / Styria	1	9	362
Vienna / Burgenland / Lower Austria	2	15	522

Table 4. Size of computational regions

The problem can be "naturally" decomposed into four regions, which consist of Salzburg and Upper Austria, Tyrol and Vorarlberg, Carinthia and Styria, and as the biggest one Vienna, Burgenland, and Lower Austria. Table 4 shows the number of switchings centers for each region. Figures 9, 10, 11, and 12 show a graphical representation of the results for the respective regions.

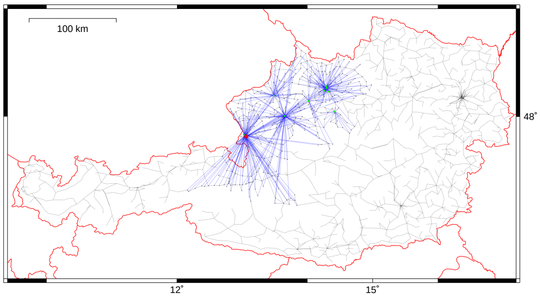


Figure 9. Solution for regions Salzburg and Upper Austria

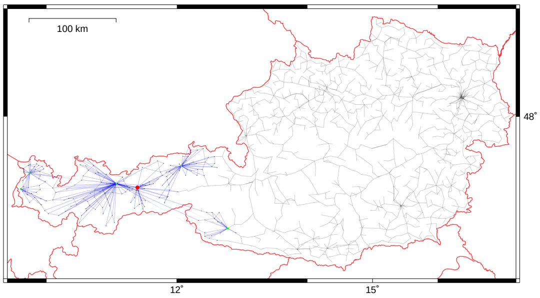


Figure 10. Solution for regions Tyrol and Vorarlberg

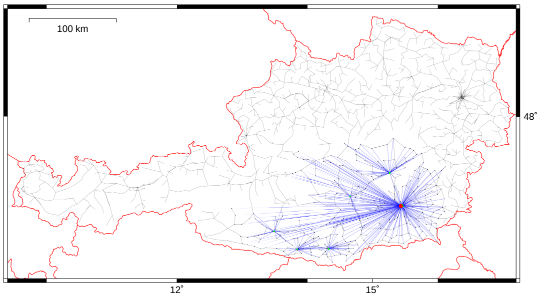


Figure 11. Solution for regions Carinthia and Styria

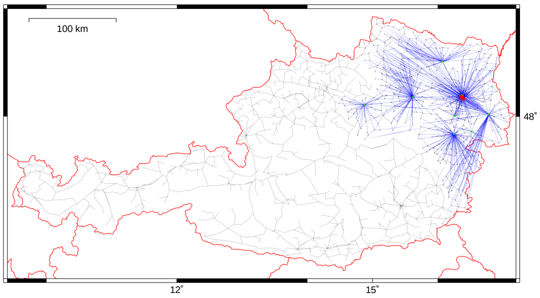


Figure 12. Solution for regions Vienna, Burgenland, and Lower Austria

CPLEX is usually able to solve all scenarios to optimality in reasonable time. (Solving time for facility location problems depends very much on the cost ratio between connections and facilities. If the ratio is balanced, the problem can get very hard to solve computationally.) The only exception was the Vienna, Burgenland and Lower Austria scenario which had an optimality gap of 0.835%, which is far below the accuracy of the data.

5 Conclusions

We have shown in this article how to uniformly handle seemingly different problems. Table 5 gives a summary of the diverse objects that were cast into the same model. As can be seen, none of the projects stretched the abilities of the model to the limit. In fact, most of the time in the projects was spend on assembling, checking, and correcting data, to compile coherent datasets that fit into the model.

	DFN	e·plus	TELEKOM AUSTRIA
V_1 nodes	Client	BSC	UV
V_2 nodes	Backbone	MSC	VV
V_3 nodes	Core		HV
Commodities		subscribers	channels users
Configurations		10 per MSC	
Link capacities	discrete		

Table 5. Different names, same mathematics

We mentioned in the beginning “changing our attitude”. It took some time to understand that our task was not to design networks, but to give decision support. In all the projects, the networks in question were virtual networks in an existing (mature) infrastructure. It became more and more clear that the precise cost of changes can not be determined in general and for all possible combinations of changes. Therefore, such models and methods should be used to

- assess the correctness of assumptions (how many switching centers are needed?)
- compare different scenarios (what is the impact of capacity changes?)
- make qualitative cost decisions (which switching network is likely cheaper?)
- verify the quality of the model (are the assumptions and rules sound?)
- estimate the potential for savings (what would a green-field solution cost?)

but *not* to

- compute quantitative cost results or
- use of the results without further consideration

In our experience regarding facility location problems, real-world data and requirements produce rather restricted problem instances, in the sense that the results are often predictable and that it is hard to find any realistic feasible solution that is much worse than the optimum.

While this sounds, as if our work was unnecessary, the opposite is the case. Precisely the fact that the solutions are so inertial shows there usefulness. Given that the data we based our computations on are often only predictions or forecasts and the cost functions are only virtual approximations, highly fluxionary solutions indicate that any decisions based on the results are questionable, because they depend largely on the assumptions made.

The other experience we gained from these projects was that the ability to quickly adapt the model is far more important than to solve the resulting instances to optimality. This insight triggered the use of modeling languages and the development of ZIMPL.

Authors

Thorsten Koch graduated at the Technical University of Berlin, Germany, with the Master's degree in Mathematics in 1996. Starting 1986, he works as a freelance software developer. He is a member of the optimization group at ZIB, where he is concerned a project on Steiner trees and their industrial application in the design of telecommunication networks. He participated in the network design of the G-WiN, the next generation backbone network of the Deutsches Forschungsnetz e.V. and the EU-project MOMENTUM.

Roland Wessäly graduated at the Technical University of Berlin with a Master's Degree in Computer Science. Since 1994 he is a member of the optimisation group at the Zuse Institute Berlin (ZIB). He developed optimisation methods for the design of survivable capacitated telecommunication networks as part of his PhD Thesis in Mathematics (finished April 2000). In 2001 he received the Vodafone Innovations award for his scientific work on network design. Since 2000 he is managing director of the ZIB spin-off atesio GmbH, a company specialized on planning and optimization algorithms for telecommunication network operators.

References

K. Aadal, M. Labbé, J. Leung, and M. Queyranne. On the two-level uncapacitated facility location problem. *INFORMS Journal on Computing*, 8(3):289–301, 1996.

A. Balakrishnan, T. L. Magnanti, and R. T. Wong. A decomposition algorithm for local access telecommunication network expansion planning. *Operations Research*, 43(1):58–76, 1995.

D. Bienstock and O. Günlück. Capacitated network design—polyhedral structure and computation. *INFORMS Journal on Computing*, 8(3):243–259, 1996.

A. Bley. A lagrangian approach for integrated network design and routing in IP networks. In *Proceedings of International Network Optimization Conference (INOC 2003), Evry/Paris, France, 2003*, pages 107–113. 2003.

A. Bley, M. Grötschel, and R. Wessäly. Design of broadband virtual private networks: Model and heuristics for the B-WiN. In N. Dean, D. Hsu, and R. Ravi, editors, *Robust Communication Networks: Interconnection and Survivability*, volume 53 of *DIMACS*, pages 1–16. AMS, 1998.

A. Bley and T. Koch. Optimierung des G-WiN. *DFN-Mitteilungen*, pages 13–15, 2000.

A. Bley, T. Koch, and R. Wessäly. Large-scale hierarchical networks: How to compute an optimal hierarchy? In H. Kaindl, editor, *Networks 2004: 11th International Telecommunications Network Strategy and Planning Symposium, June 13-16, 2004, Vienna, Austria - Proceedings*, pages 429–434. VDE Verlag, 2004.

S. Chopra and M. R. Rao. The partition problem. *Mathematical Programming*, 59(1):87–115, 1993.

A. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektrotkeknikeren*, 13, 1917.

C. E. Ferreira, A. Martin, C. C. de Souza, R. Weismantel, and L. A. Wolsey. Formulations and valid inequalities for the node capacitated graph partitioning problem. *Mathematical Programming*, 74:247–266, 1996.

C. E. Ferreira, A. Martin, C. C. de Souza, R. Weismantel, and L. A. Wolsey. The node capacitated graph partitioning problem: A computational study. *Mathematical Programming*, 81:229–256, 1998.

I. Gamvros and S. R. B. Golden. An evolutionary approach to the multi-level capacitated minimum spanning tree problem. In G. Anandalingam and S. Raghavan, editors, *Telecommunications Network Design and Management*. Kluver, 2003.

M. R. Garey and D. S. Johnson. *Computers and Intractability*. Freeman, 1979. ISBN 0-7167-1045-5.

M. Grötschel and Y. Wakabayashi. Facets of the clique partitioning polytope. *Mathematical Programming*, 47(3):367–387, 1990.

L. Hall. Experience with a cutting plane algorithm for the capacitated spannang tree problem. *INFORMS Journal on Computing*, 8(3):219–234, 1996.

K. Holmberg and D. Yuan. A lagrangian heuristic based branch-and-bound approach for the capacitated network design problem. *Operations Research*, 48(3):461–481, 2000.

T. Koch and A. Martin. Solving Steiner tree problems in graphs to optimality. *Networks*, 32:207–232, 1998. ZIB-Report SC 96-42.

P. Mirchandani. The multi-tier tree problem. *INFORMS Journal on Computing*, 8(3):202–218, 1996.

F. Ortega and L. A. Wolsey. A branch-and-cut algorithm for the single-commodity, uncapacitated, fixed-charge network flow problem. *Networks*, 41(3):143–158, 2003.

K. Park, K. Lee, S. Park, and H. Lee. Telecommunication node clustering with node compatibility and network survivability requirements. *Managment Science*, 46(3):363–374, 2000.

R. Wessäly. *DImensioning Survivable Capacitated NETworks*. Ph.D. thesis, Technische Universität Berlin, 2000.