Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ANDREAS BLEY
THORSTEN KOCH
ROLAND WESSÄLY

Large-scale hierarchical networks: How to compute an optimal architecture?

ZIB-Report 04–04 (March 2004)

Large-scale hierarchical networks: How to compute an optimal architecture?*

Andreas Bley
Thorsten Koch
Roland Wessäly

March, 2004

Abstract

In this article, we present a mathematical model and an algorithm to support one of the central strategic planning decisions of network operators: How to organize a large number of locations into a hierarchical network? We propose a solution approach that is based on mixed-integer programming and Lagrangian relaxation techniques. As major advantage, our approach provides not only solutions but also worst-case quality guarantees. Real-world scenarios with more than 750 locations have been solved within 30 minutes to less than 1% off optimality.

1 Introduction

The structure of communication networks is influenced by many technological and organizational restrictions. Besides cost, geographical issues must be taken into account as well as technical and functional aspects. Larger networks are usually partitioned into backbone and access networks. These smaller sub-networks are easier to operate than the entire network as a whole. Figure 1 sketches a typical three-level hierarchy.

As the central strategic planning task, operators of large networks (with several hundreds of locations) must define an appropriate hierarchy of manageable sub-networks. In the process of defining this hierarchy, a number of questions must be answered:

- How to optimize the trade-off between connection and equipment cost?
- What is the optimal number of backbone locations?
- Which are the backbone locations?
- How to dimension the equipment?
- Which access locations should be attached to which backbone location?

This type of planning task has to be solved by any operator of a larger communication network, independent of the technology (e.g., IP/OSPF, MPLS, ATM, SDH, WDM). We have been working on several variants of this task with various project partners, e.g., Telekom Austria as a fixed-network operator, E-Plus Mobilfunk as a mobile-phone network operator, and the DFN-Verein as the provider of the largest private IP network

*This work was supported by the DFG Research Center "Mathematics for key technologies" (FZT86), Berlin, www.fzt86.de, Project B3: Integrated planning of multi-level/multi-layer networks.

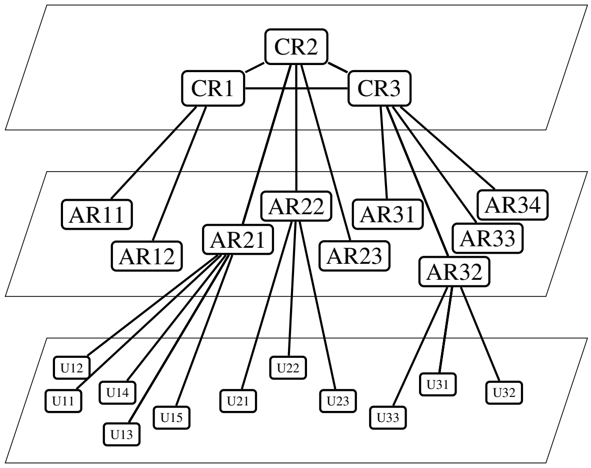

Figure 1: Network hierarchy with three levels. Core routers (CR) form the backbone, access routers (AR) the second level, and users (U) the third level.

in Germany [6, 7]. Recently, similar planning issues also arise in the design of UMTS networks, where decisions about new equipment such as Media Gateways (MGWs) or Remote Network Controllers (RNCs) have to be taken. The answers to the questions above heavily influence all subsequent operational planning issues and, thereby, the cost of the overall network.

With this article, we aim at indicating that today's mathematical optimization technology provides the means to solve problems of such complexity very efficiently and to support automated strategic planning. In Section 2, we describe the practical background of the planning task, focusing on the design of the German Research Network as an example. In Sections 3 and 4, we formulate the problem mathematically and show how to solve the resulting mathematical model by an algorithm that combines Lagrangian relaxation and mixed-integer programming techniques. Finally, in Section 5, we demonstrate the power of this algorithm, reporting results for real-world planning instances with more than 750 locations.

2 Planning task

The DFN-Verein [8] operates the German Research Network, which connects universities and research institutes from all over Germany and serves as a platform to develop and test new applications. It connects more than 750 locations. The backbone of this network, the so-called Gigabit-Wissenschaftsnetz G-WiN, has been launched in 2000 as a replacement for the Breitband-Wissenschaftsnetz B-WiN, which was no longer expandable to carry the continuously increasing traffic volume. The G-WiN has been dimensioned for a monthly traffic of 220 Terabyte initially.

The design of an appropriate structure for such a large network is a complex task. In general, this planning problem can be described as follows: (i) for each location, decide which hierarchy level it belongs to, (ii) connect each location (except for the backbone) to the next higher hierarchy level, (iii) decide about the topology, the hardware configuration, and the capacities of the backbone network, and (iv) identify a routing of the communication demands which respects the installed capacities. The objective is to minimize the sum of connection and infrastructure cost.

For operational reasons, the DFN aimed at a high-capacity backbone with a relatively small number of locations. All other locations needed to be attached to this backbone using either one or two access network levels.

2.1 Locations, hardware and capacities

Starting point is the set of locations, comprising all user locations as sources or destinations of communication traffic. For each location, it is specified to which hierarchy levels it might belong, e.g., a location may be used in the backbone only if sufficient space for the hardware is available as well as the necessary maintenance and service personnel. The number of nodes in a particular sub-network may be restricted.

Each location must be equipped appropriately according to its hierarchy level. This includes infrastructure (building, electricity, air conditioning, etc.), technical devices (switches, routers, line and tributary cards, accounting, etc.), as well as maintenance and operations personnel. In principle, it is possible to consider these requirements at a very detailed level. However, if the planning horizon is too long to obtain an accurate traffic forecast, precise equipment cost, or even the technological specifications of future equipment, it is advisable to aggregate known information about the network. We consider a small set of admissible configurations of which one is selected for each location in a solution.

Similarly, alternative link types can be chosen to connect the network locations. What link type is available to connect a particular pair of locations typically depends on their hierarchy levels and their chosen configurations. For the G-WiN, for instance, high-speed connections of 2.4 Gbit/s or above could be installed only between backbone locations. However, there also may be other restrictions like, for example, the distance or geographical obstacles, that make particular link types or entire connections between certain locations unavailable.

The access networks are supposed to be strongly hierarchical organized, that is, each location of a lower level is connected to exactly one location of the next higher level.

2.2 Routing and survivability

The central role of the backbone usually requires a design that is able to survive single link or single node failures. In such a case, the backbone network must be at least biconnected. In the access networks, a star- or tree-like topology often is sufficient.

The capacities selected for the links must permit a feasible routing of the forecasted communication demands. Within the access networks, the tree-like network structure also implies the routing. Only within the backbone network, the routing depends on the particular routing protocol.

The OSPF routing protocol is used in the G-WiN. With this protocol, each communication demand is sent from its source to its destination along a shortest path with respect to the links' routing weights specified by the network administrator. For operational reasons, traffic splitting extensions of the OSPF protocol are not used in the G-WiN. Each communication demand must be sent unsplit on a single path from its source to its destination. Hence, the OSPF routing weights need to be chosen such that there is a single and uniquely determined shortest path between each pair of backbone locations, see also [5, 7].

2.3 Optimization target

The objective is to identify a network structure which minimizes the overall cost, comprising the cost of setting up (or renting) the locations' infrastructure and equipment, the cost of operation and maintenance as well as the connection costs.

3 Mathematical model

In this section, we introduce the integer programming model for the planning task described in the previous section. For simplicity of presentation, we assume that there are only two hierarchy levels, i.e., each location becomes either a backbone or an access node, and that only backbone nodes may be dimensioned. The model generalizes straightforward to include more than two hierarchy levels and different node configurations at the access level.

3.1 Topology, hardware and link capacities

The potential network is modeled by an undirected *supply graph* $G = (V, E)$, where V is the set of all node locations and E the set of all links which may be possibly included in the final topology. At each location in $W \subseteq V$, a backbone node may be set up. All other locations in $V \setminus W$ must become access nodes.

Links can be installed only between two backbone nodes or between a backbone and an access node, that is, $u \in W$ or $v \in W$ for all $uv \in E$. Each backbone node must be appropriately equipped. The list T_W specifies the potential *node configurations*. Similarly, T_B denotes all potential *link configurations* between two backbone nodes and T_A all potential link configurations between an access and a backbone node. The set of all potential link configurations is $T := T_A \cup T_B$. For a particular link $uv \in E$, we have $T_{uv} := T_A \cup T_B$ if $u, v \in W$, and $T_{uv} := T_A$ otherwise. Each node configuration $t \in T_W$ provides a capacity c^t and each link configuration t provides a bidirected routing capacity c^t. For each node, the node capacity must be larger than the sum of the capacities of all attached links. Furthermore, at most M access nodes can be attached to a single backbone node.

In order to model the decision whether a location is a backbone node and how it is configured then, we introduce a variable $x_v^t \in \{0, 1\}$ for each potential backbone node $v \in W$ and each possible node configurations $t \in T_W$. Similarly, the variables $x_{uv}^t \in \{0, 1\}$ for all $uv \in E$ and all $t \in T_{uv}$ model which links are installed with what configuration. For notational simplicity, we use artificial variables $z_v \in \{0, 1\}$, $v \in W$, and $y_{uv} \in \{0, 1\}$, $uv \in E$, to express that v is a backbone node and that an access node u is linked to a backbone node v, respectively.

$$z_v := \sum_{t \in T_W} x_v^t \leq 1 \qquad v \in W \tag{1}$$

$$y_{uv} := \sum_{t \in T_A} x_{uv}^t \leq z_v \qquad \begin{array}{l} v \in W \\ u \in V \setminus \{v\} \\ uv \in E \end{array} \tag{2}$$

$$\sum_{uv \in \delta(v)} y_{uv} = 1 \qquad v \in V \setminus W \tag{3}$$

$$\sum_{uv \in \delta(v)} y_{uv} \geq 1 - z_v \qquad v \in W \tag{4}$$

$$y_{uv} + 2 \sum_{t \in T_B} x_{uv}^t \leq z_u + z_v \qquad \begin{array}{l} u, v \in W \\ uv \in E \end{array} \tag{5}$$

$$2 - y_{uv} \geq z_u + z_v \qquad \begin{array}{l} u, v \in W \\ uv \in E \end{array} \tag{6}$$

$$\sum_{uv \in \delta(v)} y_{uv} \leq 1 + M z_v \qquad v \in W \tag{7}$$

$$\sum_{u \in \delta(v)} \sum_{t \in T_{uv}} c^t x_{uv}^t \leq \sum_{t \in T_W} c^t x_v^t \qquad v \in W \tag{8}$$

Inequalities (1) ensures that at most one configuration is chosen for each node. Note that v is a backbone node iff $z_v = 1$, i.e., inequality (1) holds with equality. Inequalities (2), (3), (4), and (5) imply that each access node is connected to a backbone node, and that for each installed access link at least one and for each installed backbone link both terminal nodes are indeed chosen as backbone nodes. Inequalities (6) ensures that access link technologies cannot be chosen between backbone nodes. Finally, inequalities (7) and (8) guarantee that no backbone nodes serves more than M access nodes and that the capacities of the access links do not exceed the backbone node's capacity.

Notice that inequalities (1)–(8) only guarantee that each access node is connected to exactly one backbone node and that the chosen link technologies match the type of the terminal nodes. This system does not guarantee that the (backbone-) network is connected. Appropriate connectivity is enforced by the flows via the flow formulation in the next section and via additional metric inequalities [11].

Let $\mathcal{C}$ be the space of all node and link configuration variables introduced above. We define

$$X := \{x \in \{0,1\}^{\mathcal{C}} : x \text{ satisfies (1)–(8) }\}.$$

3.2 Routing

For the solution approach presented in the next section, we do not need explicit variables describing the routing paths of the communication demands. The following model of a non-bifurcated shortest path routing implicitly describes the relation between the routing weights and the resulting link flows.

For each pair of nodes $u, v \in V$, let $d^{u,v} \in \mathbb{R}_+$, denote the directed traffic demand from u to v. For each undirected link $uv \in E$, we denote by (u, v) and (v, u) its two associated directed arcs and set $A := \{(u, v), (v, u) : uv \in E\}$. We assume that a link configuration t chosen for an undirected link $uv \in E$ provides the same routing capacity c^t for both directions of the link. The routing weights for the arcs $(u, v) \in A$ are modeled by variables $w_{(u,v)} \in \mathbb{N}$. The traffic flows induced for the given communication demands and the chosen routing weights are expressed by edge flow variables $f_{(u,v)} \in \mathbb{R}_+$, $(u, v) \in A$. We let

$$\begin{aligned} F := \{(f, w) \in (\mathbb{R}^A, \mathbb{N}^E) : \; & w \text{ induces unique shortest} \\ & \text{paths between all } u, v \in V\text{, and } f \text{ are induced} \\ & \text{arc flows of demands } d \text{ in shortest path routing} \\ & \text{for routing weights } w\} \;. \end{aligned}$$

With this notation, the shortest path routings that satisfy the installed link capacities can be described as follows:

$$(f, w) \in F \;, \tag{9}$$

$$f_{(u,v)} \leq \sum_{t \in T_{uv}} c^t x^t_{uv} \qquad (u, v) \in A. \tag{10}$$

Condition (9), which involves the implicitly defined set F, ensures that w induces unique shortest paths for all communication demands and that f is the corresponding flow on the directed links. Inequality (10) guarantees that these arc flows do not exceed the provided link capacities.

3.3 Cost minimization

The objective of the network design problem is to minimize the total network cost. These comprise the cost k^t_v of setting up a backbone node at $v \in W$ with configuration t and the cost k^t_{uv} of link configuration

t between u and v:

$$\min \ k^T x := \sum_{v \in W} \sum_{t \in T_W} k_v^t x_v^t + \sum_{uv \in E} \sum_{t \in T_{uv}} k_{uv}^t x_{uv}^t \ . \tag{11}$$

4 Algorithmic approach

In this section, we present our solution approach for the integrated network design and shortest path routing problem and discuss some implementation details [4].

The complete model (1)–(11) can be written as

$$\left\{ \begin{array}{rl} & \min \ k^T x \\ \text{s.t.} & f_{(u,v)} \leq \displaystyle\sum_{t \in T_{uv}} c^t x_{uv}^t \quad \text{for all } (u,v) \in A \\ & x \in X \\ & (f,w) \in F \end{array} \right\} \tag{12}$$

If we relax the capacity constraints (10) and let $\mu \in \mathbb{R}_+^A$ be the associated Lagrangian dual multipliers, the resulting Lagrangian function is

$$\begin{aligned} L(\mu) &= L^X(\mu) + L^F(\mu) \text{ , with} \\ L^X(\mu) &:= \min \left\{ k^T x - \sum_{\substack{(u,v) \in A \\ t \in T_{uv}}} \mu_{(u,v)} c^t x_{uv}^t \ : \ x \in X \right\} \\ L^F(\mu) &:= \min \left\{ \sum_{(u,v) \in A} \mu_{(u,v)} f_{(u,v)} \ : \ (f,w) \in F \right\} \ . \end{aligned}$$

For each $\mu \in \mathbb{R}_+^A$, the value $L(\mu)$ is a lower bound for the optimal value of the original problem (12). Hence,

$$L^* := \max_{\mu \in \mathbb{R}_+^A} L(\mu) \leq k^T x^*,$$

where (x^*, f^*, w^*) is an optimal solution of (12). As there are only finitely many different (basic) solutions x and f with $x \in X$ and $(f,w) \in F$, both functions $-L^X(\mu)$ and $-L^F(\mu)$ are convex in μ and the optimal dual multipliers μ can be found by a general convex function optimization algorithm.

This Lagrangian approach is very attractive for practical computations. One reason is that the Lagrangian function $L(\mu)$ decomposes into the sum of two functions $L^X(\mu)$ and $L^F(\mu)$, which both can be evaluated efficiently for real-world size networks.

Evaluating the first function $L^X(\mu)$ corresponds to the problem of finding a valid network structure and hardware installation that minimizes a linear objective function. Although this problem is $\mathcal{NP}$-hard in general, its integer programming formulation can be solved very efficiently for real-world problems by state-of-the-art integer programming solvers. Note that this formulation contains only the variables x and the inequalities (1)–(8) . The traffic demands and flows only affect the objective function coefficients via the Lagrangian dual multipliers. In our implementation, we tighten the formulation X of this subproblem with additional band inequalities and strengthened metric inequalities [11], which improves the overall performance of this approach significantly.

The problem associated with evaluating the second function $L^F(\mu)$ is to find a non-bifurcated shortest path routing that minimizes the total flow costs for μ. This can be solved by any shortest path algorithm.

Clearly, choosing for each pair of nodes the shortest path with respect to μ yields the optimal solution. Ties between equally long shortest paths are broken by using an arbitrary numbering of the nodes or links as a secondary length function. This guarantees the existence of integer routing weights that induce the same set of shortest paths as μ. These weights then can be computed in a post-processing step by solving an integer inverse shortest path problem [3].

Another reason for the Lagrangian approach being computationally attractive is the easiness to include other heuristics. After each iteration of the convex optimization algorithm, the current duals μ can be interpreted as routing weights. In practice, these weights often provide good starting points for heuristics that are based on the modification of routing weights. From this perspective, the Lagrangian approach can also be seen as a primal heuristic that modifies the current routing weights according to the dual information and, as a byproduct, produces a lower bound for the optimum solution value. For the computations discussed in the next section, we used a simple three-step heuristic: In the first step, we solve the integer linear program associated with $L^X(\mu)$ for the current duals. Then we interpret the current duals as routing weights, penalize those links that are not installed in the integer program's solution, and compute the shortest path routing. Finally, in the third step, we compute a cost-minimal hardware installation that can accommodate the resulting arc flows, if one exists.

5 Computational results

In this section, we present results of applying the algorithm described in Section 4 to a scenario we have investigated for the planning of the Gigabit-Wissenschaftsnetz G-WiN.

5.1 Planning scenario

The G-WiN comprises over 750 locations. In the presented planning scenario, these needed to be organized in a 2-level hierarchy: backbone and access. The backbone network must be biconnected, and each access location must be attached to exactly one backbone location. There are 30 potential backbone nodes, but for operational reasons at most 20 of them can be selected as backbone location. Each of these 30 nodes can be connected with any other potential backbone node. For all other nodes, 10 possible links are considered for the connection to the backbone. These links are pre-selected according to geographical criteria.

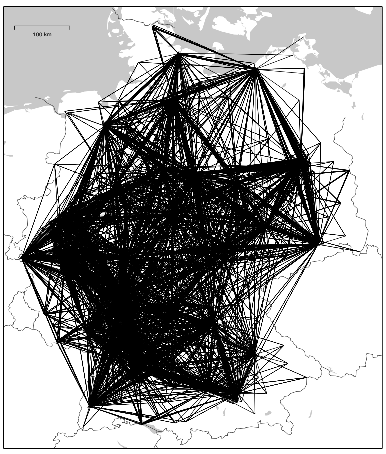

Figure 2: Potential network

At each location, there is an access router. Additionally, a backbone router must be installed at each chosen backbone location. Because the cost for access routers is unavoidable, it can be neglected during the optimization. It is only necessary to consider the additional cost for setting-up a backbone node, which is 1,000,000 per node in this scenario.

The available link capacities to connect access to backbone routers or to connect two backbone routers are (in MBit/s):

access–backbone:	0.128 , 2 , 34 , 155 , 622 , and
backbone–backbone:	155 , 622 , 2,400 , 10,000 .

The cost associated with installing these capacities on the available links depend on the link lengths and are:

capacity	fix-cost	km-length	km-cost
0.128	3278	15-50-999	131-32-19
2	6093	15-50-999	401-111-53
34	45691	15-50-999	1523-623-429
155	52015	15-50-999	2109-1054-436
622	104030	15-50-999	4218-2108-872
2,400	208060	15-50-999	8436-4216-1744
10,000	416120	15-50-999	16872-8432-3488

Table 1: Link capacities and cost.

For example, installing a link capacity of 2 Mbit/s incurs fixed cost 6093 and additional cost depending on the length of the link: Each kilometer up to 15 costs 401, each kilometer between 15 and 50 costs 111 and each kilometer above 50 costs 53.

The DFN continuously performs traffic measurements in its network and has therefore means to provide reasonably accurate traffic distributions. As indicated in Figure 3, the structure of the communication demands is very heterogeneous. Only few locations handle large fractions of the overall traffic.

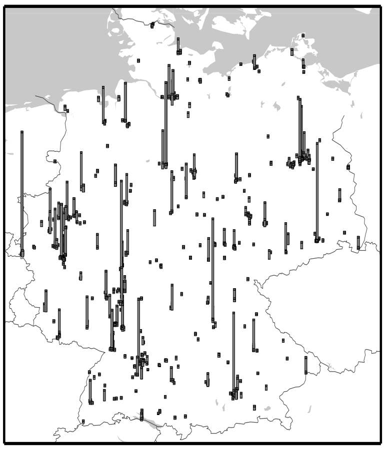

Figure 3: Source-demands

5.2 The results

This planning task was solved with the Lagrangian approach presented in the previous section. The implementation is part of the DISCNET network optimization library [2]. The data structures and algorithms are based on the standard C++ library and LEDA [1]. The CONICBUNDLE algorithm of Helmberg [9] is used to solve the convex optimization problem of finding the optimal dual multipliers. All linear and integer linear programs are solved by CPLEX [10].

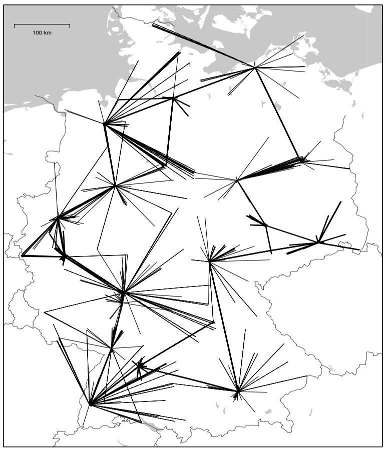

Figure 4: Best known solution

We could solve the above planning scenario within 15 minutes on a 3 GHz Pentium 4 PC. The best solution of the scenario is shown in Figure 4. The cost of this solution is 64,159,949.

This is a particular solution which satisfies *all* requirements of the planning scenario. However, even if an operator has a tool to compute a feasible network configuration, there is still a number of open questions to answer:

- What is the quality of this solution?
- Could another tool compute a solution that is substantially better?
- Is it worth spending more time to find a better solution?

With the lower bound provided by the Lagrangian approach, it is possible to seriously answer these questions. For the above scenario, a lower bound of 63,788,411 on the optimum solution value has been computed. Therefore, the optimality gap $\frac{sol-lb}{lb}$ is less than 0.583 percent. In other words, it might be that an optimal solution has been computed by our algorithm, but even if not, it has inherently been proven at run-time that no other solution is more than 0.583 percent better than the one found. Usually, already the inaccuracy in the input data is larger than that. Consequently, operators can decide on a reliable basis and it becomes possible to make conceptual comparisons in contrast to comparisons which are based on heuristic results of undetermined quality.

For the G-WiN, for example, we also studied variants of the above planning scenario where the traffic demands were projected farther into the future. Our computations revealed that if the initial network topology is well-chosen, it is not necessary to restructure the network if the traffic increases. Figure 5 shows that even though larger link and node configurations are necessary to cope with the larger traffic, the set of chosen backbone locations and the assignment of the access locations in the best solution is nearly the same as in the original scenario.

6 Conclusions

In this article, we presented a mathematical model and an algorithm to support one of the central strategic planning decision of network operators: How to structurally organize a large number of locations in

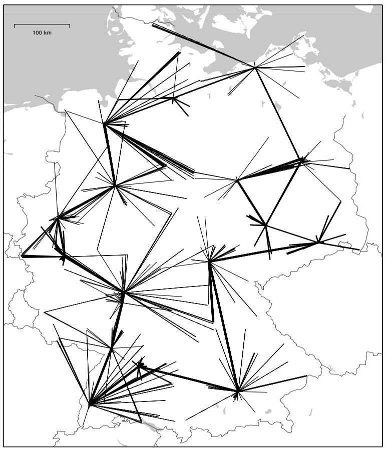

Figure 5: Best solution for doubled traffic

hierarchical levels? One advantage of our approach is that not only solutions but also worst-case quality guarantees are provided. We have been able to compute near-optimal solutions for real-world planning scenarios with more than 750 locations.

References

[1] Algorithmic Solutions Software GmbH. *LEDA: Library of Efficient Data types and Algorithms*, 1998–2003. www.algorithmic-solutions.com.

[2] atesio GmbH. DISCNET – *Network optimization software package*, 2000–2003. www.atesio.de.

[3] W. Ben-Ameur and E. Gourdin. Internet routing and related topology issues. *SIAM Journal on Discrete Mathematics*, (to appear).

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

[5] A. Bley, M. Grötschel, and R. Wessäly. Design of broadband virtual private networks: Model and heuristics for the B-WiN. In N. Dean, D.F. Hsu, and R. Ravi, editors, *Robust Communication Networks: Interconnection and Survivability*, volume 53 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 1–16. American Mathematical Society, 1998.

[6] A. Bley and T. Koch. Optimierung des G-WiN. *DFN-Mitteilungen*, 54:13–15, 2000.

[7] A. Bley and T. Koch. Integer programming approaches to access and backbone IP-network planning. Technical Report ZR-02-41, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2002.

[8] DFN. *Deutsche Forschungsnetz Verein*. www.dfn.de.

[9] C. Helmberg and K.C. Kiwiel. A spectral bundle method with bounds. *Mathematical Programming*, 93(2):173–194, 2002.

[10] ILOG CPLEX Division. *CPLEX 8.0 Reference Manual*, 2002. www.cplex.com.

[11] R. Wessäly. *Dimensioning Survivable Capacitated NETworks*. PhD thesis, Technische Universität Berlin, April 2000.