Exact Approaches to the Single-Source Network Loading Problem

Ivana Ljubić* Peter Putz* Juan-José Salazar-González†

August 17, 2009

Abstract

This article considers the network design problem that searches for a minimum-cost way of installing capacities on the edges of a network in order to simultaneously route a flow from a given access point to a subset of nodes representing customers with positive demands. We first consider compact and exponential-sized MIP formulations of the problem and provide their theoretical and computational comparison. We also propose a new strong disaggregated flow formulation. To solve the problem in practice, we project out the flow variables and generate Benders cuts within a branch-and-cut framework.

In an extensive computational study we compare the performance of compact MIP models against a textbook implementation and several normalization variants of Benders decomposition. We introduce a set of 32 real-world instances and use these, together with 64 other instances from the literature, to test our approaches. The results show that our branch-and-cut approach beats the best-performing compact formulation leading to the best algorithm today for solving the considered data set.

Keywords: Local Access Network Design, Network Loading, Capacitated Network Design, Benders Decomposition.

1 Introduction

We consider the problem of deploying a broadband telecommunication system that lays optical fiber cable from a *central office* to a number of *end-customers*. In case of the *fiber to the home* technology (FTTH) the end-customers represent houses, whereas when deploying *fiber to the curb* technology (FTTC) the end-customers are usually multiplexor devices. In both cases, we are dealing with a capacitated network design problem that requires an installation of optical fiber cables with sufficient capacity to carry the traffic from the central office to the end-customers. We start with a network without capacities, or

*Department of Statistics and Decision Support Systems, University of Vienna, Austria
†DEIOC, Universidad de La Laguna, Tenerife, Spain

with some pre-installed capacities, and search for the installation of cable types on links that enables simultaneous routing of traffic so that the whole demand in the network is satisfied at minimum cost.

This article provides an exact solution method based on a branch-and-cut approach. A new mixed integer programming formulation (MIP) of the problem, developed by the disaggregation of continuous flow variables, is proposed. Although the network loading problem has been intensively studied in the literature, to the best of our knowledge this specific disaggregation has not been considered so far. From theoretical point of view, the new formulation provides stronger lower bounds when compared to existing MIP models. To make the new model computationally tractable, we project out flow variables by using strengthening, rounded Benders inequalities incorporated into a branch-and-cut framework.

When compared to the best-performing compact formulation, we show that the average gap on a set of benchmark instances from the literature can be improved from 5.5% to 2.5%. When testing instances derived from a real-world telecommunication example, we report 8 optimal solutions, whereas the compact formulation did not solve a single one to optimality.

Problem Definition The *Single-Source Network Loading Problem* (SSNLP) is also known as the *local access network design problem* arising in the design of telecommunication networks. It can be defined as follows.

Let us consider an undirected and connected graph $G^u(V, E)$ with a designated root node $r \in V$ (representing the central office or access point to the backbone network) and a set of customers $D \subseteq V \setminus \{r\}$. Each customer $k \in D$ is associated with a positive demand $d_k \in \mathbb{R}_{>0}$. Each edge $e \in E$ is associated with a length $l_e \in \mathbb{R}_{>0}$. It is allowed to install combinations of different cable types with positive costs and capacities on every edge. Salman et al. [36] have shown that, by using dynamic programming, one can precompute the optimal combination of cable types for each level of flow and for every edge. This provides an increasing non-linear step cost function of flow for every edge of the network. We consider the optimization problem after this transformation, i.e., instead of searching for optimal *combination of cable types*, we are looking for the optimal *module* of the step cost function to be installed on every edge. Thereby, we assume that modules $\mathcal{N}_e = \{n_1, n_2, \ldots, n_{|\mathcal{N}_e|}\}$ are given for each edge $e \in E$, with capacities $u_{e,n} \in \mathbb{R}_{>0}$ and costs $c_{e,n} \in \mathbb{R}_{>0}$ for each $1 \leq n \leq |\mathcal{N}_e|$. We denote $|\mathcal{N}| := \max_{e \in E} |\mathcal{N}_e|$. Then SSNLP looks for a single-source multiple-sink routing and a link capacity assignment, with an installation of *at most one module* on every edge, to satisfy all customer demands.

Since we allow the flow between the access point and some customers to be split apart, we are speaking about a *bifurcated routing* formulation. The optimal solution of SSNLP is not necessarily a tree. Obviously, if there is only one module per edge providing sufficient capacity to route the total flow through it, then the optimal solution will be a tree, and the problem is equivalent to the Steiner tree problem on the graph by considering all customers with positive demand as *terminals* while minimizing the sum of edge lengths ($l_e \cdot c_{e,1}$) taken into the solution. Furthermore, in case of multiple modules obeying economies of scales and with sufficient capacity, the optimal solution will be a tree.

Our definition of SSNLP works for the general setting: Economies of scale may not be given over all modules. We allow the number of modules and their costs and capacities to differ from edge to edge, in contrast to the previous approaches where the modules were considered to be uniform. In addition, capacities on an edge may be limited.

Previous Work Due to its importance in telecommunications, transportation, computer and energy supply networks, network loading problems have been widely studied in the literature. Many authors consider a more general variant in which a routing from multiple sources to multiple sinks is required. Polyhedral structures of the general network design problem with multiple sources and multiple sinks are studied in [4, 7, 12, 17, 23, 29, 38]. Benders decomposition approaches have been studied as well: for the multiple-source multiple-sink case, an exact algorithm based on the *expansion step cost model* from [17] was given in [21]; the latter approach has been improved recently in [19]. In [14] the authors study the relationship between metric and Benders inequalities for the general capacitated network design problem. In [34] the authors look into speeding up Benders decomposition by combining it with local branching. Metric inequalities for the network loading problem have been studied in [3]. The authors work on the multi-commodity flow problem and propose several variants for separating metric inequalities. We build some of our normalization models by extending their ideas, see Section 4.

SSNLP has been studied only under the assumption that costs and capacities satisfy the concept of *economies of scale*, i.e., that the cost per unit capacity of a thick (high capacity) cable is considerably cheaper than that of a thin (low capacity) cable, thus buying capacities in bulks becomes more economical when the traffic increases. For that reason, in the computer science literature, SSNLP is also known as the *single-sink buy-at-bulk* network design problem (see, e.g., [35]). In terms of approximation algorithms, currently the best provable worst-case approximation ratio of 76.5 has been obtained by Gupta et al. [24]. Chopra et al. [10] have shown that the network loading problem with only two cable types and with a single-source and a single-sink node remains NP-hard. Berger et al. [6] have proposed a tabu-search heuristic that relies on the computation of k shortest paths, in order to find alternative paths from the root to each customer node.

Salman et al. [36] proposed the *search by objective relaxation* (SOR) approach for solving SSNLP. The authors solve SSNLP by considering the flow problem with a non-linear step cost function. The step cost function is first approximated by its lower convex envelope. The obtained relaxed problem is solved by a combinatorial algorithm in polynomial time. The process is repeated in every node of the branch-and-bound tree in which branching is done by dividing the interval of possible flow values on an edge into subintervals. In Section 5 we refer to their results where we also use their benchmark instances to test our approach.

Raghavan and Stanojević [33] pointed out that the linear programming (LP) relaxation of a single-commodity flow model for SSNLP (see Section 2.2) also approximates the step cost function by its lower convex envelope. Therefore the SOR approach can also be seen as a stylized branch-and-bound on a

single-commodity flow model. The authors compared their stylized branch-and-bound approach against two MIP models based on the aggregated single-flow formulation. While the authors were able to beat the explicit cost model, their results were always worse than those obtained by the incremental cost MIP model.

The rest of the paper is organized as follows. In Section 2 we recall existing MIP formulations and propose a new disaggregated compact model. A hierarchy of MIP formulations is also given. Section 3 explains how to project out flow variables of the new compact model and how to generate stronger cutting planes. Algorithmic aspects of our approach are discussed in Section 4. We implemented several different approaches for solving the Benders subproblem. An extensive computational comparison of compact vs. Benders approaches is provided in Section 5.

2 MIP Formulations

Network loading problems are often modeled using compact flow-based MIP formulations (see, e.g., [14, 23, 33, 36]) involving integer design variables and continuous flow variables. When disaggregating flow variables, we face the trade-off problem between the increasing quality of lower bounds and the growing size of the underlying linear program. In this section, we first recall three "natural" formulations for SSNLP: two compact and an exponential-sized one. We then propose a new disaggregated flow-based formulation. We finally provide a hierarchy of different MIP formulations with respect to the quality of lower bounds from their LP relaxations.

2.1 Transformation into Directed Problem

It is well known that, in general, the MIP formulations of uncapacitated network design problems on directed graphs provide better lower bounds than their undirected counterparts (see e.g., [11]). However, the MIP approaches to SSNLP up to now, considered in [33, 36], involve undirected graphs. We propose to work with directed graphs and for that purpose we transform our input graph $G^u = (V, E)$ into a directed graph $G = (V, A)$ where $A := \{(i,j),(j,i) \mid \{i,j\} \in E; i,j \neq r\} \cup \{(r,j) \mid \{r,j\} \in E\}$. The available modules on the arcs remain symmetric, i.e., $c_{ij,n} = c_{ji,n} = c_{\{i,j\},n}$, $u_{ij,n} = u_{ji,n} = u_{\{i,j\},n}$ for all $n \in \mathcal{N}_{\{i,j\}}$. To solve SSNLP, we now search for the *directed solution*, i.e., for the installation of at most one module on every *arc* so that there is enough capacity to route the flow from r to every $k \in D$.

Lemma 2.1. *Let G be the graph as described above. Then, any optimal solution of* directed SSNLP *on G contains no directed cycle.*

Proof. Let x be an optimal solution of the directed SSNLP problem such that it contains a directed cycle C. Let f be a feasible flow sent from r toward all customers $k \in D$ using the capacities installed in x. Denote by $\tilde{a}$ the arc from C such that $\tilde{a} = \arg\min_{a \in C} f_a$. Let us construct a new flow f' such that $f'_a = f_a$, if $a \notin C$, and $f'_a = f_a - f_{\tilde{a}}$, if $a \in C$. The flow f' is feasible and corresponds to an integral

solution x' with $x'_{\tilde{a},n} = 0$, for all $n \in \mathcal{N}_{\tilde{a}}$. The cost of x' is strictly less than the cost of x, which is a contradiction. □

Corollary 2.2. *Given the graph G^u, demands d_k, for all $k \in D$ and the cost and capacity functions c and u as described above, any optimal solution of SSNLP on G^u can be transformed into an equivalent* directed solution *on G with the same objective value, and vice versa.*

2.2 Single-Commodity Flow Formulation (SCF)

The single-commodity flow formulation, SCF, models the flow on every arc as the total amount of flow routed from the root toward the customers. To model a non-decreasing step cost function on every arc, binary variables need to be used. There is a possibility to model the problem using the *explicit cost* (also called the *multiple choice model* in [15]), or the *incremental cost model*. The *incremental cost* model for general multi-source multi-sink network design problem was introduced by Dahl and Stoer [17], while for SSNLP it was tested in [33, 36]. Raghavan and Stanojević [33] proved that for SSNLP both models are equivalent in terms of quality of lower bounds, and their LP relaxations both approximate the monotonically increasing step cost function by its lower convex envelope. A more general equivalence result for generic minimization problems with separable non-convex piecewise linear costs is given by Croxton et al. [15], see also Keha et al. [26].

Binary variables $x_{ij,n} \in \{0,1\}$ decide whether the module n shall be installed on the arc (i,j), whereas flow variables $f_{ij} \geq 0$ describe the amount of flow on arc $(i,j) \in A$. Then SCF model is:

$$\text{SCF}: \quad \min \sum_{(i,j)\in A} l_{ij} \sum_{n\in\mathcal{N}_{ij}} c_{ij,n} x_{ij,n} \tag{1}$$

$$\text{s.t.} \quad \sum_{(i,j)\in A} f_{ij} - \sum_{(j,i)\in A} f_{ji} = \begin{cases} -d_i, & i \in D \\ \sum_{k\in D} d_k, & i = r \\ 0, & otherwise \end{cases} \qquad \forall i \in V \tag{2}$$

$$\sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \leq 1 \qquad \forall (i,j) \in A \tag{3}$$

$$0 \leq f_{ij} \leq \sum_{n\in N} u_{ij,n} x_{ij,n} \qquad \forall (i,j) \in A \tag{4}$$

$$x_{ij,n} \in \{0,1\} \qquad \forall (i,j) \in A,\ \forall n \in \mathcal{N}_{ij}. \tag{5}$$

The *flow conservation constraints* (2) ensure that every customer receives desired amount of flow, while the *capacity constraints* (4) ensure that enough capacity is installed on every arc. The *disjunction constraints* (3) are typical for the *explicit cost model*, i.e., on every arc at most one module may be installed.

Observation 2.3. *Given an optimal solution (x', f') of an LP relaxation of* SCF*, the subgraph G' of G obtained by taking all arcs $(i,j) \in A$ such that $\sum_{n\in\mathcal{N}_{ij}} x'_{ij,n} > 0$ contains no directed cycle.*

Therefore, the subtour elimination constraints $x_{ij,n} + x_{ji,n} \leq 1$, for all $(i,j) \in A$, and all $n \in \mathcal{N}_{ij}$, are redundant for both, the SCF model and its LP relaxation. However, summed up over all modules, the following inequalities can be alternatively used to replace disjunction constraints (3):

$$\sum_{n \in \mathcal{N}_{ij}} (x_{ij,n} + x_{ji,n}) \leq 1 \qquad \forall (i,j) \in A.$$

The SCF model contains $O(|A| \cdot |\mathcal{N}|)$ variables and constraints, but due to "big-M" constraints (4), it provides arbitrarily bad lower bounds. In case of economies of scale, the LP relaxation of the SCF model has an optimal solution in which at most one of $x_{ij,n}$ variables (the one with the lowest $c_{ij,n}/u_{ij,n}$ ratio) on every arc is non-zero (see also [36]).

2.3 Multi-Commodity Flow Formulation (MCF)

Disaggregation by commodities is commonly used for the multiple-source multiple-sink network design problems (see, e.g., [1, 29]). In this model, f_{ij}^k describes the amount of flow of commodity $k \in D$ routed through the arc (i,j). Commodities in our case are source-sink pairs $(r,k), k \in D$, i.e., they are directly associated to customers $k \in D$. The MCF model then reads as follows:

$$\text{MCF} : \quad \min \sum_{(i,j) \in A} l_{ij} \sum_{n \in \mathcal{N}_{ij}} c_{ij,n} x_{ij,n} \tag{6}$$

$$\text{s.t.} \quad \sum_{(i,j) \in A} f_{ij}^k - \sum_{(j,i) \in A} f_{ji}^k = \begin{cases} -d_k, & i = k \\ d_k, & i = r \\ 0, & \text{otherwise} \end{cases} \qquad \forall i \in V,\ \forall k \in D \tag{7}$$

$$\sum_{n \in \mathcal{N}_{ij}} x_{ij,n} \leq 1 \qquad \forall (i,j) \in A \tag{8}$$

$$\sum_{k \in D} f_{ij}^k \leq \sum_{n \in \mathcal{N}_{ij}} u_{ij,n} x_{ij,n} \qquad \forall (i,j) \in A \tag{9}$$

$$0 \leq f_{ij}^k \leq d_k \sum_{n \in \mathcal{N}_{ij}} x_{ij,n} \qquad \forall (i,j) \in A,\ \forall k \in D \tag{10}$$

$$x_{ij,n} \in \{0,1\} \qquad \forall (i,j) \in A,\ \forall n \in \mathcal{N}_{ij}. \tag{11}$$

The *flow conservation constraints* (7) and the *capacity constraints* (9) have the same meaning as for the SCF model. The *coupling constraints* (10) ensure that if there is a flow in any module n on the arc (i,j), then the corresponding design variable need to be set up. Obviously, these constraints are redundant for the MIP formulation, but they improve the lower bound of the LP relaxation.

The MCF model contains $O(|A| \cdot |\mathcal{N}| + |A| \cdot |D|)$ variables and $O(|V| \cdot |D| + |A| \cdot |\mathcal{N}| + |A| \cdot |D|)$ constraints. The LP relaxations of the SCF model and MCF model without coupling constraints (10) produce the same lower bound.

2.4 Disaggregated Multi-Commodity Flow Formulation (DMCF)

For the multi-commodity capacitated network design problem, Croxton et al. [16] and Frangioni and Gendron [20] proposed a disaggregation by integer values in a MIP based on the multi-commodity flow formulation. Similarly, in this new MIP model for SSNLP, we disaggregate flow variables with respect to modules. Beside the binary design variables, $x_{ij,n} \in \{0,1\}$, we use the disaggregated flow variables $f^k_{ij,n}$ that define the amount of flow of commodity $k \in D$, routed through the arc (i,j) using the module $n \in \mathcal{N}_{ij}$. The DMCF model reads then as follows:

$$\text{DMCF} : \quad \min \sum_{(i,j)\in A} l_{ij} \sum_{n\in\mathcal{N}_{ij}} c_{ij,n} x_{ij,n} \tag{12}$$

$$\text{s.t.} \quad \sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} f^k_{ij,n} - \sum_{(j,i)\in A} \sum_{n\in\mathcal{N}_{ji}} f^k_{ji,n} = \begin{cases} -d_k, & i = k \\ d_k, & i = r \\ 0, & \text{otherwise} \end{cases} \qquad \forall i \in V,\ \forall k \in D \tag{13}$$

$$\sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \le 1 \qquad \forall (i,j) \in A \tag{14}$$

$$\sum_{k\in D} f^k_{ij,n} \le u_{ij,n} x_{ij,n} \qquad \forall (i,j) \in A, \forall n \in \mathcal{N}_{ij} \tag{15}$$

$$0 \le f^k_{ij,n} \le d_k x_{ij,n} \qquad \forall (i,j) \in A, \forall k \in D, \forall n \in \mathcal{N}_{ij} \tag{16}$$

$$x_{ij,n} \in \{0,1\} \qquad \forall (i,j) \in A, \forall n \in \mathcal{N}_{ij}. \tag{17}$$

The *capacity constraints* (15) ensure that the total flow in module n on arc (i,j) must not exceed the capacity of the given module n. Constraints (16) are redundant for the MIP formulation, but they improve the optimal value of the LP relaxation.

The DMCF model contains $O(|A| \cdot |\mathcal{N}| \cdot |D|)$ constraints and $O(|A| \cdot |\mathcal{N}| \cdot |D|)$ variables, and it is very unlikely that even the most sophisticated MIP solvers may solve instances of moderate size using the DMCF formulation. Our computational experiments with the DMCF model confirmed this claim (see Section 5). To use the advantage of this strong model, we propose to project out the flow variables and to introduce Benders inequalities instead, keeping the quality of lower bounds, and even improving them by rounding techniques.

Denote by

$$\begin{aligned}
\mathcal{P}_{\text{SCF}} &:= \{(x,f) \in [0,1]^{|A|} \times \mathbb{R}^{|A|}_{\ge 0} \mid (x,f) \text{ satisfy } (2)-(4)\} \\
\mathcal{P}^-_{\text{MCF}} &:= \{(x,f) \in [0,1]^{|A|} \times \mathbb{R}^{|A||D|}_{\ge 0} \mid (x,f) \text{ satisfy } (7)-(9)\} \\
\mathcal{P}_{\text{MCF}} &:= \{(x,f) \in [0,1]^{|A|} \times \mathbb{R}^{|A||D|}_{\ge 0} \mid (x,f) \text{ satisfy } (7)-(10)\} \\
\mathcal{P}^-_{\text{DMCF}} &:= \{(x,f) \in [0,1]^{|A|} \times \mathbb{R}^{|A||D||\mathcal{N}|}_{\ge 0} \mid (x,f) \text{ satisfy } (13)-(15)\} \\
\mathcal{P}_{\text{DMCF}} &:= \{(x,f) \in [0,1]^{|A|} \times \mathbb{R}^{|A||D||\mathcal{N}|}_{\ge 0} \mid (x,f) \text{ satisfy } (13)-(16)\}
\end{aligned}$$

the polytopes of the LP relaxations of the above MIP models for SSNLP. Denote by $\text{proj}_x(\mathcal{P}) = \{x \in$

$[0,1]^{|A|} \mid (x,f) \in \mathcal{P}\}$ the natural projection on the space of x variables, for any of the polyhedra $\mathcal{P}$ defined above. It is not difficult to see that the following result holds:

Lemma 2.4.

$$\text{proj}_x(\mathcal{P}_{\text{DMCF}}) \subset \text{proj}_x(\mathcal{P}_{\text{MCF}}) \subset \text{proj}_x(\mathcal{P}_{\text{SCF}}).$$

Section 5 provides computational evidence of this theoretical result. Furthermore, we can also show the following equivalence.

Lemma 2.5.

$$\text{proj}_x(\mathcal{P}^-_{\text{DMCF}}) = \text{proj}_x(\mathcal{P}^-_{\text{MCF}}) = \text{proj}_x(\mathcal{P}_{\text{SCF}}).$$

Proof. Obviously, both disaggregated formulations MCF$^-$ and DMCF$^-$ can be simply aggregated into the SCF model by setting $f_{ij} := \sum_{k\in D} f^k_{ij}$ and $f_{ij} := \sum_{k\in D}\sum_{n\in\mathcal{N}_{ij}} f^k_{ij,n}$, respectively. Furthermore, it is easy to see that $\text{proj}_x(\mathcal{P}_{\text{SCF}}) \subseteq \text{proj}_x(\mathcal{P}^-_{\text{MCF}})$.
Let us now show that every feasible solution $(x',f') \in \mathcal{P}_{\text{SCF}}$ can be projected into a feasible solution $(x,f) \in \mathcal{P}^-_{\text{DMCF}}$. For that purpose, we define another module-based disaggregated formulation, equivalent to the SCF model, using variables $f_{ij,n}$ such that $f_{ij} = \sum_{n\in\mathcal{N}_{ij}} f_{ij,n}$, for all $(i,j) \in A$. The corresponding flow-conservation constraints and the capacity constraints are given as:

$$\sum_{(i,j)\in A}\sum_{n\in\mathcal{N}_{ij}} f_{ij,n} - \sum_{(j,i)\in A}\sum_{n\in\mathcal{N}_{ji}} f_{ji,n} = \begin{cases} -d_k, & i = k \\ \sum_{k\in D} d_k, & i = r \\ 0, & \text{otherwise} \end{cases} \qquad \forall i \in V, \tag{18}$$

$$f_{ij,n} \le x_{ij,n}u_{ij,n} \qquad \forall (i,j) \in A,\ \forall n \in \mathcal{N}_{ij}. \tag{19}$$

Given $(x',f') \in \mathcal{P}_{\text{SCF}}$, we simply set $f_{ij,n} := f'_{ij}\frac{u_{ij,n}x'_{ij,n}}{\sum_{n\in\mathcal{N}_{ij}} u_{ij,n}x'_{ij,n}}$ and $x_{ij,n} = x'_{ij,n}$, for all $(i,j) \in A$, for all $n \in \mathcal{N}_{ij}$. It is easy to see that the obtained flow $f_{ij,n}$ and $x_{ij,n}$ satisfy both (18) and (19). We now disaggregate this flow by every single commodity f^k, for all $k \in D$. What we get is a flow $f^k_{ij,n}$, such that $(x,f) \in \mathcal{P}^-_{\text{DMCF}}$. Hence, $\text{proj}_x(\mathcal{P}_{\text{SCF}}) \subseteq \text{proj}_x(\mathcal{P}^-_{\text{DMCF}})$. □

2.5 Directed Cut-Set Formulation (DCut)

We now recall the cut-set formulation for SSNLP on directed graphs. For each subset $S \subset V$, we will denote by $\delta^+(S) := \{(i,j) \in A : i \in S, j \in V \setminus S\}$ and $\delta^-(S) := \{(i,j) \in A : i \in V \setminus S, j \in S\}$, outgoing and ingoing cuts, respectively. For any $S \subseteq V$, $A(S)$ denotes the induced arc set, i.e., $A(S) := \{(i,j) \in A : i \in S, j \in S\}$.

Projecting out aggregated flow variables $f_{ij} = \sum_{k\in D}\sum_{n\in\mathcal{N}_{ij}} f^k_{ij,n}$ for all $(i,j) \in A$ leads to the following cut-set inequalities:

$$\sum_{(i,j)\in\delta^+(S)}\sum_{n\in\mathcal{N}_{ij}} u_{ij,n}x_{ij,n} \ge \sum_{k\in D\setminus S} d_k \qquad \forall S \subset V : r \in S, S \cap D \neq D. \tag{20}$$

The separation problem of cut-set inequalities in general multiple-source multiple-sink case is NP-hard and can be reduced to the max-cut problem [4]. However, inequalities (20) for SSNLP can be separated in polynomial time as follows. For a given fractional solution x', we define the directed *support graph* $G' = (V', A')$ where $V' := V \cup \{t\}$ with an additional sink t, and $A' := A_1 \cup A_2$ being $A_1 := \{(i,j) \in A : \sum_{n \in N_{ij}} u_{ij,n} x'_{ij,n} > 0\}$ and $A_2 := \{(k,t) : k \in D\}$. The cost associated to each arc $a = (i,j) \in A_1$ is set to $\sum_{n \in N} u_{ij,n} x'_{ij,n}$, and the cost of each arc $a = (k,t) \in A_2$ is set to d_k. If the minimum cut between r and t in G' is less than $\sum_{k \in D} d_k$, there is a violated inequality (20).

Since $x_{ij,n}$ variables are binary, the cut-set inequalities can also be strengthened as follows:

$$\sum_{(i,j) \in \delta^+(S)} \sum_{n \in \mathcal{N}_{ij}} \min(u_{ij,n}, \sum_{k \in D \setminus S} d_k) x_{ij,n} \geq \sum_{k \in D \setminus S} d_k.$$

Observe that, due to "big-M" constants referring to capacities, the LP relaxation obtained by solving the DCut formulation can be arbitrarily bad. In fact, constraints (20) correspond to Benders inequalities obtained by projecting out flow variables from the SCF model and therefore the DCut model is equivalent to the SCF model with respect to the quality of lower bounds of their LP relaxations.
The DCut model can be improved with additional *connectivity constraints*, i.e.:

$$\sum_{(i,j) \in \delta^+(S)} \sum_{n \in \mathcal{N}} x_{ij,n} \geq 1 \quad \forall S \subset V : r \in S, S \cap D \neq D, \tag{21}$$

leading to a new model DCut$^+$. Constraints (21) can be derived from the MCF model by dualizing flow-constraints (7) and (10).

Denote by

$$\begin{aligned} \mathcal{P}_{\text{DCut}} &:= \{x \in [0,1]^{|A|} \mid x \text{ satisfy (20), (3)}\}, \\ \mathcal{P}_{\text{DCut}^+} &:= \{x \in \mathcal{P}_{\text{DCut}} \mid x \text{ satisfy (21) }\}. \end{aligned}$$

Model DCut$^+$ is weaker than the MCF model because the flow variables in the MCF model simultaneously need to satisfy the capacity constraints (9) and (10), while the cut-set inequalities (20) and (21) ensure the existence of two independent flows. Thus, we have:

Lemma 2.6.

$$\text{proj}_x(\mathcal{P}_{\text{MCF}}) \subset \mathcal{P}_{\text{DCut}^+} \subset \mathcal{P}_{\text{DCut}} = \text{proj}_x(\mathcal{P}_{\text{SCF}}).$$

2.6 Further Strengthening Inequalities

We now address two families of valid inequalities that improve the lower bound obtained by solving the LP relaxations of the above MIP models for SSNLP.

2.6.1 Degree-Balance Constraints

Non-customer nodes $V \setminus (D \cup \{r\})$ cannot have incoming (or outgoing) arcs only. Therefore, we can add the following *degree-balance constraints* that only work for single source case:

$$\sum_{(l,i)\in A, l\neq j} \sum_{n\in\mathcal{N}_{li}} x_{li,n} \geq \sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \qquad \forall (i,j) \in A, i \notin D, i \neq r \tag{22}$$

$$\sum_{(j,l)\in A, l\neq i} \sum_{n\in\mathcal{N}_{jl}} x_{jl,n} \geq \sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \qquad \forall (i,j) \in A, j \notin D, j \neq r. \tag{23}$$

Inequality (22) states that if an arc (i,j) emanating from a non-customer node i is being used in the solution, there must be at least one arc entering i. Thanks to Lemma 2.1 the opposite arc (j,i) can be excluded from the summation on the left hand side. Inequality (23) states the opposite case for an arc (i,j) entering a non-customer node j.

Observe that capacitated versions of these cuts, i.e.:

$$\sum_{(l,i)\in A, l\neq j} \sum_{n\in\mathcal{N}_{li}} u_{li,n} x_{li,n} \geq \sum_{n\in\mathcal{N}_{ij}} u_{ij,n} x_{ij,n} \qquad \forall (i,j) \in A, i \notin D, i \neq r$$

$$\sum_{(j,l)\in A, l\neq i} \sum_{n\in\mathcal{N}_{jl}} u_{jl,n} x_{jl,n} \geq \sum_{n\in\mathcal{N}_{ij}} u_{ij,n} x_{ij,n} \qquad \forall (i,j) \in A, j \notin D, j \neq r.$$

are *not valid* in general, but only if there is a uniform capacity/cost structure on edges.

2.6.2 Cover Inequalities

Given a cut-set inequality (20) defined by $S \subset V, r \in S$, define the index set $I(S) := \{(i,j,n) \mid (i,j) \in \delta^+(S), n \in \mathcal{N}_{ij}\}$ and $B := \sum_{k\in D\setminus S} d_k$. Set $M \subset I(S)$ is called a *cover* with respect to $I(S)$ if $\sum_{(i,j,n)\in M} u_{ij,n} < B$ and a *maximal cover* if, in addition, for all M', such that $I(S) \supseteq M' \supset M$: $\sum_{(i,j,n)\in S'} u_{ij,n} \geq B$. If M is a maximal cover with respect to $I(S)$, then the following *cover inequalities* are valid:

$$\sum_{(i,j,n)\in I(S)\setminus M} x_{ij,n} \geq 1. \tag{24}$$

In general, the separation problem of cover inequalities is NP-hard. We show that the problem of finding the most violated cover inequality (24) is equivalent to solving the *precedence constrained knapsack problem*. Assume that indices $n \in \mathcal{N}_{ij}$ are sorted according to increasing arc capacities. To model any cover M with respect to $I(S)$, we introduce the binary variables $z_{ij,n}$ that are equal to one if and only if $(i,j,n) \in M$. For every arc $(i,j) \in \delta^+(S)$, we define $u_{ij,0} = 0$.

For given fractional solution x' and an index set $I(S)$ induced by a cut-set inequality, the *most violated cover inequality* can be found by solving the following model:

$$\begin{aligned}
\text{KNAP}: \quad & \max \sum_{(i,j,n)\in I(S)} x'_{ij,n} z_{ij,n} \\
& \sum_{(i,j,n)\in I(S)} (u_{ij,n} - u_{ij,n-1}) z_{ij,n} < B \\
& z_{ij,n} \geq z_{ij,n+1}, \qquad \forall (i,j,n) \in I(S),\ n < |\mathcal{N}_{ij}| \\
& z_{ij,n} \in \{0,1\}, \qquad \forall (i,j,n) \in I(S).
\end{aligned} \tag{25}$$

Let z' be an optimal solution of model KNAP. The corresponding cover inequality reads then as follows:

$$\sum_{(i,j,n)\in I(S)} (1 - z'_{ij,n})x_{ij,n} \geq 1.$$

If all capacities and demands are integers, (25) can be replaced by $\sum_{(i,j,n)\in I(S)}(u_{ij,n} - u_{ij,n-1})z_{ij,n} \leq B - 1$. The cover inequalities are similar to the *band inequalities* for the incremental cost model in [17].

2.7 Hierarchy of Formulations

The hierarchical scheme given in Figure 1 summarizes the relationships between the LP relaxations of the MIP models considered throughout this paper for SSNLP. A filled arrow specifies that the target formulation is strictly stronger than the tail formulation. An empty arrow specifies that the target formulation is at least as strong as the tail formulation. Thereby, DMCF$^+$ denotes the DMCF formulation extended by degree balance and cover inequalities, and DCut (DCut$^+$) denotes the cut-set formulation (i.e., DCut model extended by connectivity inequalities). *Benders*$^+$ denotes the model with rounded Benders cuts (see next section) extended by degree balance and cover inequalities.

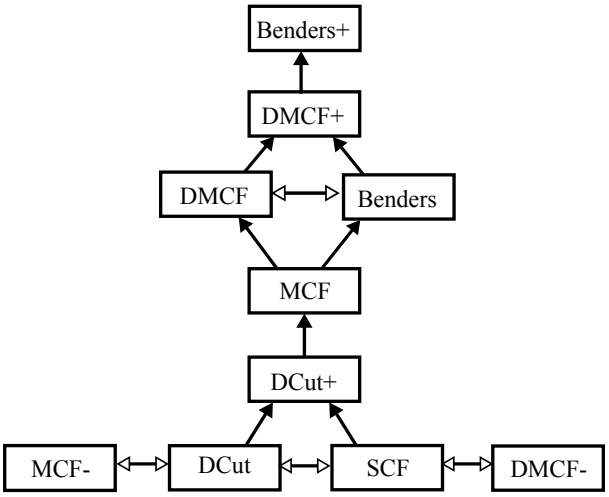


Figure 1: Hierarchy of LP relaxations.

3 Benders Decomposition for DMCF

Magnanti and Wong [31] emphasize that for any MIP, the tighter the LP relaxation of a model, the better Benders cuts can be produced. Therefore, we propose to solve the disaggregated model DMCF by projecting out flow variables by dynamically generating the corresponding violated Benders inequalities. Similar ideas has been applied to the weaker MCF model for several related problems, see e.g. [13, 14, 21].

3.1 The Benders Subproblem

Let the master problem be the one given by the objective function (12) subject to constraints (14) and (17). A solution $x' = [0 \leq x'_{ij,n} \leq 1 : (i,j) \in A, n \in \mathcal{N}_{ij}]$ of the master problem defines a feasible solution for the LP relaxation of the DMCF model if and only if there exist flow variables $[f^k_{ij,n} : (i,j) \in A, k \in D, n \in \mathcal{N}]$ satisfying the linear system of inequalities given by (13), (15) and (16), where $x = x'$.

Farkas' lemma states that a linear system of equations $\{Ax \leq b : x \geq 0\}$ has a solution if and only if $u^T b \geq 0$ for all $u \geq 0$ such that $u^T A \geq 0$. To apply Farkas' lemma to the system (13), (15), (16), we define a dual variable α^k_i associated to each equation (13), a dual variable $\beta^k_{ij,n}$ associated with each inequality (16), and a dual variable $\gamma_{ij,n}$ associated with each equation in (15). Then, the polyhedron defined by this system is non-empty if and only if the following Benders decomposition subproblem SUB is bounded (i.e., its optimal value is equal to zero):

$$\text{SUB} \ : \ \min \quad z(\alpha,\beta,\gamma,x') = \sum_{k\in D} d_k(\alpha^k_r - \alpha^k_k) + \sum_{(i,j)\in A}\sum_{n\in N_{ij}} \left(\sum_{k\in D} d_k\beta^k_{ij,n} + u_{ij,n}\gamma_{ij,n}\right) x'_{ij,n} \tag{26}$$

$$\text{s.t.} \ \ \alpha^k_i - \alpha^k_j + \beta^k_{ij,n} + \gamma_{ij,n} \geq 0 \qquad \forall (i,j) \in A, \ \forall k \in D, \ \forall n \in \mathcal{N}_{ij} \tag{27}$$

$$(\alpha,\beta,\gamma) \geq 0. \tag{28}$$

Violated Benders inequalities can be found and used within a branch-and-cut framework as follows. We first solve the linear relaxation of the master problem, obtaining a fractional solution $x' = [x'_{ij,n} : (i,j) \in A, n \in \mathcal{N}_{ij}]$. With these values x' we define the corresponding subproblem SUB. Note that this subproblem can be either bounded or unbounded. In the latter case, there is an unboundedness direction (α',β',γ') for which $z(\alpha',\beta',\gamma',x') < 0$. To avoid this situation the *Benders cut*

$$\sum_{(i,j)\in A}\sum_{n\in N_{ij}} \left(\sum_{k\in D} d_k\beta'^k_{ij,n} + u_{ij,n}\gamma'_{ij,n}\right) x_{ij,n} \geq \sum_{k\in D} (\alpha'^k_k - \alpha'^k_r) d_k \tag{29}$$

must be added to the master problem. Observe that we can round down the coefficients multiplying $x_{ij,n}$ by setting them to $\min(\sum_{k\in D} d_k\beta'^k_{ij,n} + u_{ij,n}\gamma'_{ij,n}, \sum_{k\in D} d_k(\alpha'^k_k - \alpha'^k_r))$.

The process is iterated until the linear relaxation of the master problem has been solved to optimality, that is, until the subproblem SUB is unable to find more violated Benders cuts. If $x_{ij,n}$ variables are all integer, the SSNLP is solved. Otherwise, we resort to branching.

We refer to this implementation of the separation of violated Benders inequalities as the *textbook implementation*. In Section 5, we compare this basic implementation with more elaborated variants proposed in Section 4.

3.2 Strengthening Benders Inequalities with Metric Inequalities

Let us describe two procedures for generating *metric inequalities* (see, e.g., [23]) as proposed in [3, 14]:

Procedure 1. First consider the following optimization problem for a vector $\mu \geq 0$:

$$\min\{ \sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} \mu_{ij,n} x_{ij,n} \mid (x,f) \in [0,1]^{|A||N|} \times \mathbb{R}^{|A||D|}_{\geq 0}, (x,f) \text{ satisfy } (7),(9)\}.$$

Assigning dual variables α_i^k to (7), and $\gamma_{ij,n}$ to (9), by LP-duality, we have that the subproblem defined by (7) and (9) is feasible if and only if for any $\mu \geq 0$

$$\sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} \mu_{ij,n} x_{ij,n} \geq \max\{\sum_{k\in D} (\alpha_k^k - \alpha_r^k) d_k \mid \alpha_j^k - \alpha_i^k \leq \frac{\mu_{ij,n}}{u_{ij,n}},\ (i,j)\in A, n\in\mathcal{N}_{ij}, k\in D,\ \alpha \geq 0\}.$$

The problem we get on the right hand side is decomposable into k dual versions of shortest path problems on an auxiliary graph $\hat{G}$ whose arc lengths are defined as $\hat{w}_{ij} = \min_{n\in\mathcal{N}_{ij}} \mu_{ij,n}/u_{ij,n}$ for all $(i,j) \in A$.

This property can be used to strengthen any valid inequality of the form: $\sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} \mu_{ij,n} x_{ij,n} \geq M$. In particular, for cut-set inequalities (20) and (21) defined by a subset $S \subset V, r \in S$, the arc lengths for the shortest path problem are given as

$$\hat{w}_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \delta^+(S), \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad \hat{w}_{ij} = \begin{cases} \min_{n\in\mathcal{N}_{ij}} \frac{1}{u_{ij,n}}, & \text{if } (i,j) \in \delta^+(S), \\ 0, & \text{otherwise,} \end{cases}$$

respectively. Denote by $SP(k,\hat{w})$ the length of the shortest path between r and k in G, with arc lengths defined by $\hat{w}$. If $\sum_{k\in D} SP(k,\hat{w}) d_k > M$, we obtain a metric inequality that dominates the original valid inequality.

Procedure 2. Consider now the following network flow problem with two kinds of capacity constraints:

$$\min \Bigg\{ \sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} \mu_{ij,n} u_{ij,n} x_{ij,n} + \sum_{(i,j)\in A} \sum_{k\in D} d_k \nu_{ij,n}^k \sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \mid$$
$$(x,f) \in [0,1]^{|A||N|} \times \mathbb{R}^{|A||D||N|}_{\geq 0},\ (x,f) \text{ satisfy } (13),(15),(16) \Bigg\}$$

After assigning dual variables α_i^k to (13), $\beta_{ij,n}^k$ to (16) and $\gamma_{ij,n}$ to (15), again by LP-duality, we have that the network flow problem (13), (15), (16) is feasible if and only if for any $\mu \geq 0$ and $\nu \geq 0$ we have:

$$\sum_{(i,j)\in A} \sum_{n\in\mathcal{N}_{ij}} \mu_{ij,n} u_{ij,n} x_{ij,n} + \sum_{(i,j)\in A} \sum_{k\in D} d_k \nu_{ij,n}^k \sum_{n\in\mathcal{N}_{ij}} x_{ij,n} \ \geq$$
$$\max \Bigg\{ \sum_{k\in D} (\alpha_k^k - \alpha_r^k) d_k \mid \alpha_j^k - \alpha_i^k \leq \nu_{ij,n}^k + \mu_{ij,n},\ (i,j)\in A, n\in\mathcal{N}_{ij}, k\in D, \quad \alpha \geq 0 \Bigg\}.$$

The problem on the right hand side is again decomposable into k dual versions of shortest path problems on auxiliary graphs G^k whose arc lengths are defined as $w_{ij}^k = \min_{n\in\mathcal{N}_{ij}} \{\nu_{ij,n}^k + \mu_{ij,n}\}$ for all $(i,j) \in A$. Therefore, any Benders inequality associated to a non-extreme ray can be strengthened to a metric inequality as follows. For any fixed Benders cut given by $(\alpha', \beta', \gamma')$ satisfying (27)-(28), one can look for $(\alpha(\beta',\gamma'), \beta', \gamma')$ that might improve the right-hand side of (29) by solving:

$$\begin{aligned} \text{SPDUAL}: \quad & \max \sum_{k\in D} (\alpha_k^k - \alpha_r^k) d_k & & \\ \text{s.t.} \quad & \alpha_j^k - \alpha_i^k \leq \beta_{ij,n}'^k + \gamma_{ij,n}' & & \forall (i,j)\in A,\ \forall k\in D,\ \forall n\in\mathcal{N}_{ij} \\ & \alpha_i^k \geq 0 & & \forall i\in V,\ \forall k\in D. \end{aligned}$$

If $\sum_{k \in D}(\alpha_k^k - \alpha_r^k)d_k > \sum_{k \in D}(\alpha_k'^k - \alpha_r'^k)d_k$, we obtain a metric inequality that dominates the original Benders cut.

4 Branch-and-Cut Algorithm

The overall approach is a branch-and-cut algorithm in which the LP relaxation of the compact model SCF is solved first, followed by the separation of further strengthening inequalities:

1. Initialization:
   (a) The initial LP master consists only of the design variables x and flow variables f.
   (b) The LP relaxation of the model SCF extended by some inequalities (see Section 4.1) is solved.
2. In every node of the branch-and-bound tree:
   (a) As long as there are violated connectivity inequalities (21), add them to the master LP and resolve it.
   (b) Based on the current fractional solution x', create the Benders subproblem.
   (c) Solve the subproblem. If this results in a violated Benders cut, add it to the master LP and resolve it.

As an alternative approach to the initialization of lower bounds by the SCF model, we also tried the branch-and-cut with cut-set inequalities (20). These cut-sets can be separated very fast. However, our preliminary results have shown that using SCF directly, results in a clearly preferable approach in practice.

4.1 Initialization of Lower Bounds

In order to obtain good, yet easily computable, lower bounds that will avoid expensive computation of "trivial" Benders cuts, we extend the SCF model by constraints (22), (23) and the following in-degree inequalities and root-out-degree inequalities:

$$\sum_{(i,j) \in A} \sum_{n \in \mathcal{N}_{ij}} x_{ij,n} \geq 1 \qquad \forall j \in D$$
$$\sum_{(r,j) \in A} \sum_{n \in \mathcal{N}_{rj}} x_{rj,n} \geq 1.$$

Obviously, these inequalities are special cases of connectivity cut-set inequalities (21) for $S = V \setminus \{j\}$ and $S = \{r\}$, respectively.

4.2 Branch-and-Cut Parameters

To improve the overall performance and to avoid numerical difficulties we consider the following two standard ingredients:

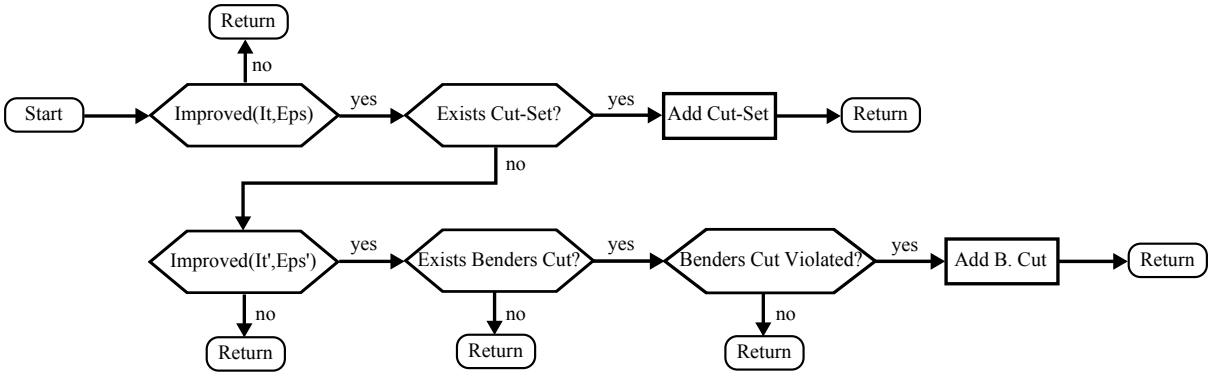


Figure 2: Separation of cuts in the branch-and-cut framework.

- Tailing Off: If the relative improvement of the lower bound is less than `Eps`% in the last `It` iterations of the separation procedure, we stop the separation and resort to branching. The general setting of (`It`,`Eps`) is $(20, 10^{-3})$. However, if only the computationally more expensive Benders cuts were separated in recent iterations, a stricter setting of (`It'`,`Eps'`) $= (10, 10^{-3})$ is applied.

- Degree of Violation: Assume that after solving the Benders subproblem for a given fractional value x', we obtain a violated cut defined by a vector $(\alpha', \beta', \gamma')$. Before inserting the corresponding cut into the master LP, we normalize it by dividing it with its right-hand side (which is always positive) and calculate its violation by the current fractional solution x' as follows:

$$\text{violation}(\alpha', \beta', \gamma', x') = 1 - \sum_{(i,j)\in A} \sum_{n\in N_{ij}} \frac{\sum_{k\in D} d_k \beta'^k_{ij,n} + u_{ij,n}\gamma'_{ij,n}}{\sum_{k\in D} d_k(\alpha'^k_k - \alpha'^k_r)} x'_{ij,n} \tag{30}$$

  If $\text{violation}(\alpha', \beta', \gamma', x') < 10^{-4}$, the cut will not be considered as violated and will not be inserted into the system.

The flowchart in Figure 2 explains how we implemented our cut separation procedure.

4.3 Separation of Benders Cuts

Any implementation of Benders cut's separation heavily affects the overall performance of a MIP approach. In our work, we considered the straight-forward implementation of solving the dual subproblem SUB as defined in Section 3, and three other *normalization approaches* obtained by closing the dual unbounded cone (see Table 1 for an illustration). For each of these models, we explicitly solved the dual or the primal version of the subproblem.

4.3.1 Separation Models

SUB **Model:** In order to get a violated Benders inequality, we search for an extreme ray of the unbounded subproblem SUB. As already observed in [5, 18], this approach has a significant drawback: it returns a randomly chosen extreme ray without having any positive influence on the quality of the

Dual	Primal	Explanation
SUB	-	see Section 3
SUBc	PSUBc	SUB extended by $(\alpha, \beta, \gamma)^t \cdot \mathbf{1} = 1$
SUBn	PSUBn	SUB extended by $\sum_{k \in D} d_k(\alpha_k^k - \alpha_r^k) = 1$
SUBf	PSUBf	SUB extended by $\alpha^t \cdot \mathbf{1} = 1$

Table 1: Different normalization approaches for separating Benders cuts.

violated cut found. An advantage of this method is that it returns a violated constraint much faster than the corresponding more sophisticated methods described below.

SUBc **and** PSUBc **Models:** Instead of solving the subproblem on the pointed unbounded cone, one can close it by adding the following hyperplane:

$$\sum_{(i,j) \in A} \sum_{n \in N_{ij}} \sum_{k \in D} \beta_{ij,n}^k + \sum_{(i,j) \in A} \sum_{n \in N_{ij}} \gamma_{ij,n} + \sum_{i \in V} \sum_{k \in D} \alpha_i^k = 1. \tag{31}$$

Obviously, the Benders cut generated by solving SUB extended by (31) is violated if and only if the objective value is strictly less than zero. Furthermore, each vertex of such obtained polyhedron (except the origin) corresponds to an extreme ray of the unbounded subproblem.

One easily observes that the model SUBc is equivalent to the similar problem of maximizing the value of $\Theta \leq 0$ subject to constraints (13), (15) and (16) in which Θ is added to the left-hand side of each of them. This primal model is denoted by PSUBc.

SUBn **and** PSUBn **Models:** Recall that before inserting a cut into the master problem we check its violation according to (30). Since we are interested in looking for a maximally violated Benders cut, one can normalize the subproblem by fixing the right-hand side to one. The SUB model is therefore extended by the following constraint:

$$\sum_{k \in D} d_k(\alpha_k^k - \alpha_r^k) = 1.$$

The resulting subproblem SUBn is bounded and its solution (if negative) always corresponds to the most violated Benders cut according to (30). Again, the master solution x' is infeasible if and only if the solution of SUBn is strictly less than zero. The primal of SUBn, denoted by PSUBn, is known as the *maximum concurrent flow model* (see, e.g., [8]), and it has been used by Avella et al. [3] for separation of metric inequalities.

SUBf **and** PSUBf **Models:** One can also consider the following flow feasibility subproblem:

$$\begin{aligned}
&\text{PSUBf}: \qquad \max \Theta \\
\sum_{(i,j)\in A}\sum_{n\in\mathcal{N}_{ij}} f^k_{ij,n} - \sum_{(j,i)\in A}\sum_{n\in\mathcal{N}_{ji}} f^k_{ji,n} + \Theta &= \begin{cases} d_k, & i = r \\ -d_k, & i = k \\ 0, & \text{otherwise} \end{cases} && \forall i \in V,\ \forall k \in D \\
\sum_{k\in D} f^k_{ij,n} &\le u_{ij,n} x'_{ij,n} && \forall (ij) \in A,\ \forall n \in \mathcal{N}_{ij} \\
0 \le f^k_{ij,n} &\le d_k x'_{ij,n} && \forall (i,j) \in A,\ \forall k \in D,\ \forall n \in \mathcal{N}_{ij} \\
\Theta &\le 0.
\end{aligned}$$

This problem has a nice flow structure that can easily be recognized by an LP solver (like Cplex), therefore we consider it as another alternative normalization approach for finding a violated Benders inequality. In the corresponding dual variant of the model, denoted by SUBf, we extend SUB with $\sum_{k\in D}\sum_{i\in V}\alpha^k_i = 1$.

SUBcap **and** PSUBcap **Models:** In our preliminary computational experiments we also tried the variant in which Θ is added only to capacity and coupling constraints (15) and (16), respectively. The latter model is a generalization of the *capacity reduction problem*, used by Avella et al. [3] to generate the so-called *strong metric inequalities* for the MCF model of NLP. However, in our preliminary results, both SUBcap and PSUBcap were significantly outperformed by other models mentioned above. In particular, these submodels could not be solved to optimality within our default time limit for solving Benders subproblems. Therefore, we do not report results for these models in Section 5.

4.3.2 Further Potential Enhancements

We implemented the following additional techniques that are known to significantly improve the performance of (Benders) separation algorithms, in general.

Connectivity Cuts: We separate *nested*, *backward* and *minimum cardinality* connectivity cuts (21), that are used as a standard procedure for accelerating cutting plane methods (see, e.g. [27]). The idea of separating nested cuts is to look for violated inequalities whose coefficient vector is orthogonal to an already detected cut. Using backward cuts one can generate two violated inequalities per a single max-flow computation. Minimum cardinality cuts search for a cut set with the smallest number of arcs having the same max-flow value. We separate up to 100 connectivity cuts per iteration. For finding the maximum flow in a directed graph, we used an adaptation of Cherkassky and Goldberg's maximum flow algorithm [9].

Nested Benders Cuts: We also tried to apply the separation of nested cuts for adding several disjoint Benders cuts per iteration. Assume that $(\alpha', \beta', \gamma')$ is a solution to the current Benders subproblem.

Before solving the LP relaxation of the master, we fix to zero the variables β and γ with positive value in the current dual solution. In contrast to the separation of connectivity inequalities, there is a drawback of separating nested Benders cuts: For the instances we tested, solving a single LP relaxation of Benders subproblem is computationally more expensive than resolving the primal master LP.

Magnanti-Wong Implementation: The ideas of Magnanti and Wong [31] has been widely used for accelerating separation of Benders cuts (see, e.g., [28, 34]). The authors proposed to accelerate the convergence of the basic Benders algorithm by adding Pareto-optimal Benders cuts. A Pareto-optimal Benders cut is given by the following definition: a cut $z(\alpha'', \beta'', \gamma'', x) \geq 0$ *dominates* another cut $z(\alpha', \beta', \gamma', x) \geq 0$ if and only if $z(\alpha', \beta', \gamma', x) \geq z(\alpha'', \beta'', \gamma'', x)$ for all $x \in \{0,1\}^{|A||\mathcal{N}|}$ satisfying (3), and the strict inequality holds for at least one x. A Benders cut is said to be *Pareto-optimal* if no other cut dominates it. In case that there are multiple optimal solutions to the Benders subproblem, Magnanti and Wong have proposed an approach to search for a Pareto-optimal cut by solving an additional subproblem in the separation phase:

1. Given a fractional solution x', solve the Benders subproblem SUBx to get a violated cut defined by $(\alpha', \beta', \gamma')$. If $z(\alpha', \beta', \gamma', x') = 0$, no violated cut exists. Stop.
2. Set $z' := z(\alpha', \beta', \gamma', x')$.
3. Solve the new subproblem defined as:

$$\min\{z(\alpha, \beta, \gamma, x_0) \mid (\alpha, \beta, \gamma) \text{ satisfy the constrains in SUBx and } z(\alpha, \beta, \gamma, x') = z'\}.$$

4. Denote by $(\alpha'', \beta'', \gamma'')$ the solution to this subproblem. Then, the cut $z(\alpha'', \beta'', \gamma'', x) \geq 0$ is inserted into the master problem.

In this procedure, SUBx denotes any of the normalized variants (bounded subproblems) described above and x_0 is a given fractional solution called *core point*, i.e., a point that belongs to the relative interior of the convex hull of all binary vectors x satisfying (3). As already observed by Papadakos [32], for the above procedure to work efficient, one needs to start it with a different core point every time the procedure is applied. For that purpose, we start with a randomly chosen point from the interior, and later we generate a random convex combination of two incumbent solutions.

The obvious drawback of this procedure is, that we have to solve two time-consuming subproblems within each separation. Furthermore, solving the Magnanti-Wong subproblem is computationally more expensive that solving the master problem.

4.4 Primal Heuristic

We employ the following rounding heuristic based on min-cost-flow. Denote the total installed capacity on an arc by $X_{ij} = \sum_{n \in \mathcal{N}_{ij}} u_{ij,n} x_{ij,n}$. The cheapest fitting module to support a certain capacity U

is denoted by $n(U) = \arg\min_{\{n \in \mathcal{N}_{ij} | u_{ij,n} \geq U\}} c_{ij,n}$. Starting from a fractional solution x, we create a binary solution x' and subsequently a cheaper binary solution x''. Initialize $x' := 0$. Now for every arc (i,j) install the cheapest fitting module, i.e. $x'_{ij,n(U)} := 1$. The resulting x' is integer feasible, but also typically overly generous and can be improved. To this end we use an augmented graph with an additional sink t, similar to the one from Section 2.5: Let $G' = (V', A')$ where $V' = V \cup \{t\}$ and $A' = A \cup \{(k,t) : k \in D\}$). The arc capacities are set to X'_{ij} for all $(i,j) \in A$ and d_k for all (k,t). Arc costs are defined as $C_{ij} := \sum_{n \in \mathcal{N}_{ij}} c_{ij,n} x'_{ij,n}$. Initialize $x'' := 0$. We now compute the min-cost-flow $f \in \mathbb{R}^{|A|}$ in G'. This induces our new incumbent candidate $x'' : x''_{ij,n(f_{ij})} := 1$.

We use the min-cost-flow implementation based on capacity scaling and successive shortest path computation found in the commercial library LEDA, 5.2 (see [1, 2]). This algorithm only works for integer capacity and cost. Therefore we round these values to the nearest integer prior to the min-cost-flow computation. A result of this rounding is that x'' will, on rare occasions, be infeasible. This is easily detected by a subsequent computation of max-flow/min-cut and an infeasible x'' is discarded.

5 Computational Results

This section reports on our computational experience with the proposed branch-and-cut framework. We implemented our algorithms using C++ and CPLEX 11.1 [25]. An Intel Core 2 personal computer with 1.8 GHz and 3.25 GB of RAM was used for testing purposes. If not mentioned otherwise, the default Cplex settings are used.

We set a time limit of 1000 seconds for solving benchmark instances. For all instances that cannot be solved to optimality because that limit was reached, we report the gap between the best known upper bound (UB) and the lower bound (LB) obtained, calculated as $\frac{UB-LB}{UB} \cdot 100\%$. The time limit for each single separation of Benders cuts was set to 45 seconds. In our preliminary computational experiments we did not see any advantage of the procedure for strengthening Benders by metric inequalities (as long as the SCF model and undirected cut-sets are used for initialization), so the technique described in Section 3.2 is not used in the results presented below.

5.1 Preprocessing

All reported instances are preprocessed according to the following rules:

(i) Each customer node $k \in D$ with degree one is joined with his neighbor using the cheapest module that allows the flow of d_k to be routed.

(ii) Each non-customer node with degree one is simply deleted.

(iii) Each non-customer-node with degree two and with the same modules on both sides is replaced by a single edge, with the same modules.

(iv) Each non-customer-node with degree two and different modules on both size is replaced by an edge with all possible module combinations.

(v) Rules (iii) and (iv) may result in parallel edges. Parallel edges are replaced by a single one with all possible module combinations.

(vi) Rules (iv) and (v) may lead to dispensable modules. A module $n \in \mathcal{N}_e$ is dispensable if there exists another module $n' \in \mathcal{N}_e$ with $u_{e,n'} \geq u_{e,n}$ and $c_{e,n'} \leq c_{e,n}$. Dispensable modules are deleted.

(vii) Sets of excess modules $N'_e = \{n \in \mathcal{N}_e \mid u_{e,n} \geq \sum_{k \in D} d_k\}$ are replaced by a single module n' with $c_{e,n'} = min_{n \in N'_e} c_{e,n}$ and $u_{e,n'} = \sum_{k \in D} d_k$.

Observe that rules (iv) and (v) may generate instances with non-uniform modules, even if the modules of the original instance were uniform.

5.2 Benchmark Instances

Instances from Salman [37]: Salman instances include four problems originally defined in [22] (problems ARPA, OCT, USA, and RING) and 60 problems randomly generated by Salman [37], also used in [36]. For the latter ones, there are 12 groups with 20, 30 and 40 nodes. There are 9 cable types obeying economies of scale. The cheapest cable type has capacity of 6 – see [6, 36] for a detailed description. The convex combinations of these generate up to $\lceil \frac{\sum_k d_k}{6} \rceil$ modules. The notation `e(n)(s)(d)` provides summary information on the instances: `n` denotes the number of nodes, `s` explains the location of the root node (`c` stands for *central*, `r` stands for *random position*), `d` explains the level of demand (`l` stands for *low demand*, which is randomly generated between 0 and 30; `h` stands for *high demand*, randomly generated between 0 and 60). In [33, 36] two kinds of experiments were performed: using all 9 cable types and using only 4 of them. Since our method does not depend on the number of cable types, but on the number of modules, we only performed the more challenging variant involving all 9 cable types. Table 3 provides input information on Salman instances: each of twelve `e(n)(s)(d)` groups contains 5 instances, and the average values per group are reported. The number of nodes ($|V|$), the number of edges ($|E|$), the number of customers ($|D|$), and the number of modules ($|\mathcal{N}|$) represent the averaged values obtained after preprocessing.

Real-World Instances (Bregenz): We used the street map of the Austrian city Bregenz with 1014 nodes and 1191 edges as underlying network. As customers we considered 4 different sets of nodes with cardinalities $|D| \in \{29, 36, 45, 67\}$. We classified the instances into two groups: Group `H` contains graphs with *higher demands*, i.e., each customer has a demand randomly chosen from $\{4, 8, 12, 16, 20\}$; Group `L`, in contrast, contains graphs with *lower* demand, i.e., each customer is assigned a demand of 4 units.

We employed up to four different modules as displayed in Table 2. These modules imitate real-world data we obtained from an Austrian telecommunication company. In particular, not all of these modules obey economies of scale: it is possible that there are empty conduits with limited modular capacity available at low costs, but if higher capacities need to be installed, new trenches need to be pre-paired, which involves very high investment costs.

Type	$\|\mathcal{N}\|$	(capacity $u_{ij,n}$, cost $c_{ij,n}$) ...
A	2	(120, 7.0), (1020, 146.0)
B	2	(30, 2.2), (1020, 146.0)
C	3	(30, 2.2), (60, 4.0), (1020, 146.0)
D	4	(30, 2.2), (60, 4.0), (120, 7.0), (1020, 146.0)

Table 2: The four different sets of modules used for Bregenz instances.

The preprocessing greatly reduces the size of the graph and the number of customers goes down to 28, 33, 41 and 61. Furthermore, although we start with uniform modules, we end up with non-uniform ones. Table 5 illustrates that: the number of minimal, maximal and average number of modules per arc is given in columns $|\mathcal{N}_{\min}|$, $|\mathcal{N}_{\max}|$ and $|\mathcal{N}_{\text{avg}}|$, respectively.

5.3 Solving Salman Instances

We first report on the results with the three compact MIP models presented in Section 2: DMCF, MCF and SCF. We also compared the branch-and-cut approaches based on different separation models as explained in Section 4.3.1. The main goal of this study was: a) to compare the qualities of lower bounds obtained by solving compact models versus branch-and-cut approaches, and b) to determine whether there is a difference in the performance of the branch-and-cut approach when the textbook implementation is compared against normalized separation approaches. For that purpose, we wanted to ensure that the obtained results are not biased by the quality of incumbent solutions found by the MIP solver. Therefore, we initialized all the models with the best known upper bounds (BKUB) found during previous extensive computations and we turned all heuristics off. For the same reason, for this particular test, we turned Cplex cuts and the presolver off. Benders cuts are separated at the root and in each 10th node of the branch-and-bound tree.

Table 3 provides values averaged over 5 instances per group, for `e(n)(s)(d)` instances, and the values for additional four instances from [22].

Gap at the root node: In Table 3 we first report on the quality of LP relaxations of three compact models and the corresponding value of the LP relaxation at the root node of the branch-and-bound tree for the SUBc approach. The gaps between obtained lower bounds and the best known upper bound (provided in column *UB*) are shown. The SUBc approach was the one among all branch-and-cut approaches to provide the tightest lower bounds at the root node. The average (median) gap over

all 64 instances of the SUBc approach is 6.0% (5.9%). The worst LP relaxation gap among Benders approaches is obtained by solving the SUB model: the average (median) gap is 7.5% (7.5%). These results are coherent with our theoretical discussion provided in Section 4.

Comparing compact formulations, we observe that the average (median) gap of the SCF model of 21.3% (20.4%) can be improved to 9.4% (9.4%) by solving the MCF model, which can further be improved to 6.3% (5.8%) by solving the DMCF formulation. Looking at gaps of the SUBc approach and the DMCF model, we can observe two different effects. In some cases SUBc produces better gaps. This results from tightening Benders cuts by rounding down the coefficients (see groups `e20_c_l`, `e40_c_h` and `e40_r_h` in Table 3). In other cases the gap of SUBc is slightly worse than the one of DMCF. This is explained by tailing-off and violation checks. Particularly, if at some point the current Benders cut does not satisfy the violation test (30) and we decide not to add this particular cut and instead resort to branching, the lower bound at the root node will be slightly smaller than the value of the LP relaxation of DMCF.

Gap after the time limit: For the SCF model and for Benders separation approaches Table 3 also reports the lower-bound gap after the time limit was reached. Every single variant of our branch-and-cut approach beats the compact SCF model. The best results are obtained by solving the SUBf approach: the average (median) gap after 1000 seconds is 2.5% (2.5%), while SCF terminates at 5.5% (5.6%).

SUBf solves 14 out of 20 instances of group `e20` to optimality, while SCF finds optimum only in 7 out of 20 cases. Despite the bad quality of gaps of the LP relaxation, the model SCF succeeds to improve the final gap by drawing the advantage of branching. The average number of branch-and-bound nodes when solving SCF is close to 750 000, while the number of nodes processed by our Benders implementations varies between 212 (SUBn) and 6043 (SUBf). Due to the huge number of branch-and-bound nodes, in 12 out of 64 cases SCF terminates abnormally, due to the terminates abnormally, due to the memory overload.

The rightmost column in Table 3 shows the average gaps reported by Salman et al. [36] obtained by solving SORb2 approach. The average gaps obtained by Raghavan and Stanojević [33] were always worse than those obtained in [36], therefore we report only on the latter ones. However, we stress that these gaps are not directly comparable with ours, because of different software and hardware settings.

Table 4 reports on the correlations between the average time needed to solve the subproblem, the number of branch-and-bound nodes and the tightness of the bounds at the root node of the branch-and-bound tree. The average values over all 64 Salman instances for the following parameters are provided: Time_0 and Gap_0 denote the running time and the gap at the root node of the branch-and-bound tree, respectively; Benders_0 denotes the number of Benders cuts separated at the root node; $\text{Time}_0/\text{Benders}_0$ provides the ratio between the total time spent and the number of Benders cuts. The values $\text{Gap}_{\text{total}}$ and $\text{Time}_{\text{total}}/\text{Benders}_{\text{total}}$ are the corresponding values provided for the total running time $\text{Time}_{\text{total}}$ of 1000 seconds. The last row shows how many branch-and-bound nodes have been processed within the time limit. For the results after 1000 seconds, the two best performing approaches are shown in bold

Problem								Gap at the Root Node				Gap after 1000s								Gap
	s	d	$\|V\|$	$\|E\|$	$\|D\|$	$\|N\|$	UB	DMCF	MCF	SCF	SUBc	SCF	SUB	SUBc	SUBn	SUBf	PSUBc	PSUBn	PSUBf	[Salman et al.]
e20	c	l	18.2	37.8	8.6	12.6	111,9	5.5	11.2	32.3	**5.4**	0.3	**0.0**	**0.0**	0.4	**0.0**	**0.0**	**0.0**	**0.0**	0.0
e20	r	l	17.4	35.4	10.0	13.0	143,3	**6.6**	10.9	33.2	6.8	4.7	1.6	2.0	2.5	**1.2**	1.7	1.5	1.5	1.9
e20	c	h	18.2	38.0	8.6	27.6	194,2	**6.0**	9.8	22.1	**6.0**	1.8	2.4	2.6	4.3 [1]	1.3	2.0	2.0	**1.2**	0.7
e20	r	h	17.6	36.2	9.2	23.2	239,8	**5.8**	9.4	19.7	**5.8**	2.9	1.6	1.7	3.8 [2]	**0.4**	0.9	1.1	0.9	1.0
e30	c	l	26.6	54.4	14.6	24.2	323,5	**6.6**	10.8	22.7	7.1 [1]	7.2 (2)	4.0	4.3 [1]	6.1 [2]	**2.7**	3.5	3.2	3.4	7.1
e30	r	l	26.8	55.2	13.8	22.2	290,1	**5.9**	9.1	22.9	**5.9**	7.8	4.0	4.4	6.0 [2]	**2.8**	4.0	3.3	3.5	7.6
e30	c	h	26.2	53.6	14.2	47.4	529,1	**5.1**	8.8	15.6	**5.1** [1]	6.8 (1)	4.4	4.4 [1]	7.5 [5]	**3.0**	3.7	3.5	3.5	6.2
e30	r	h	26.6	54.8	12.2	33.4	469,0	**4.9**	6.9	13.1	**4.9**	4.7 (2)	3.7	4.1	5.5 [3]	**2.8**	3.6	3.1	3.3	4.5
e40	c	l	36.6	75.2	19.0	27.4	409,1	**6.8**	11.0	25.4	7.1 [2]	14.6 (4)	6.5	6.5 [2]	8.6 [1]	**4.5**	6.4 [2]	5.2	5.6	14.7
e40	r	l	35.6	74.6	19.4	31.4	572,4	**5.8** [1]	9.0	17.7	6.1 [2]	8.9 (2)	5.2	5.7 [2]	6.8 [2]	**3.9**	5.2	4.4	5.0	10.4
e40	c	h	35.8	73.8	19.0	49.4	674,8	9.8 [2]	8.2	15.5	**5.8** [4]	9.3 (1)	5.1	5.8 [4]	6.6 [2]	**3.7**	5.3 [1]	4.2	4.7 [1]	7.9
e40	r	h	33.8	71.8	16.8	46.2	745,6	6.4 [3]	7.2	13.4	**5.2** [4]	7.0 (1)	4.8 [1]	5.0 [4]	6.3 [5]	**3.4**	4.5	4.0	4.4	6.3
oct			16	20	14	39	2432,3	**6.1**	8.8	17.5	6.3	4.0	2.8	4.0	5.2	1.4	3.0	**1.2**	2.5	0.0
ring			26	54	17	47	1391,3	**7.2**	11.5	25.6	7.3	11.1	5.6	5.9	7.9	**3.8**	5.4	4.4	4.6	6.5
usa			26	39	16	44	2233,2	**7.4**	10.4	19.5	**7.4**	8.4	6.2	6.2	8.2 [1]	**4.6**	5.6	5.1	5.4	4.8
arpa			16	21	12	35	2571,4	**5.0**	9.4	15.0	5.1	1.7	1.5	2.2	4.5	**0.5**	1.7	1.7	1.3	0.0

Table 3: Results for Salman's instances. The upper part of the table shows average values over 5 instances in each class `e(n)(s)(d)`. The lower part shows results for the four instances from Gavish and Altinkemer [22]. We report the gaps to the best known upper bound obtained at the root node and after the time limit. For the SCF model the numbers in round brackets show how many (out of 5 cases) experiments terminated because Cplex ran out of memory. The corresponding values are not included in the average gap. The numbers in squared brackets denote in how many (out of 5) cases the separation at the root node was not finished within the time limit.

face.

The normalized Benders subproblems have a complicated flow structure with two kinds of capacity constraints. Therefore, the problem of solving a normalized subproblem by closing the unbounded cone with an additional constraint may become a difficult task. Row $\text{Time}_0/\text{Benders}_0$ of Table 4 provides an estimate of an average time (in seconds) needed to solve each Benders subproblem. The fastest subproblems are SUBf and PSUBn (followed by the separation of extreme rays with the SUB approach). Correspondingly, these two variants are first to be finished at the root node of the branch-and-bound tree. Therefore, they are also separating the most Benders cuts and traversing the most nodes of the branch-and-bound tree. However, the SUBf bounds obtained at the root node are tighter than the corresponding bounds of the PSUBn model, which makes the SUBf approach the winner, when solving this data set.

This study shows that:

- Strong Benders cuts derived from the DMCF formulation beat the compact SCF model.
- Two important aspects decide on the quality of our Benders approach: a) the running time needed to solve the Benders subproblem, and b) the quality of the derived Benders cuts. The model that succeeds to balance the trade-off between these two aspects is the most desirable one.

Average	DMCF	MCF	SCF	SUB	SUBc	SUBn	SUBf	PSUBc	PSUBn	PSUBf
Time_0	211.6	1.2	0.1	68.3	422.5	603.5	22.0	247.0	35.3	170.0
Benders_0	-	-	-	40.9	38.3	67.8	57.2	36.0	115.5	52.9
$\text{Time}_0/\text{Benders}_0$	-	-	-	1.7	11.0	8.9	0.4	6.9	0.3	3.2
Gap_0	6.3	9.4	21.3	7.5	6.0	6.9	6.4	6.1	6.9	6.2
$\text{Benders}_{\text{total}}$	-	-	-	740.8	164.7	159.2	**1281.5**	303.4	**1498.6**	556.0
$\text{Time}_{\text{total}}/\text{Benders}_{\text{total}}$	-	-	-	1.3	6.1	6.3	**0.8**	3.3	**0.7**	1.8
$\text{Gap}_{\text{total}}$	-	-	5.5	3.6	3.9	5.4	**2.5**	3.5	**3.0**	3.1
Nodes	-	-	748 245	2152	492	212	6043	1326	3655	2501

Table 4: Average values over all 64 Salman's instances.

5.4 Solving Real-World Instances

We now show the comparison of results obtained for the set of real-world instances derived from Bregenz, a city in Austria.

LP relaxations: We first compare the gap of LP relaxations for three compact models, SCF, MCF and DMCF, whose values are given in Table 5. We turned Cplex cuts and the presolver off, to be able to compare the gaps achieved with the proposed models and to avoid distortion due to "cleverness" of

the MIP/LP solver. Therefore, the reported results should be solver independent. For all 32 instances, the LP relaxation of the SCF model was solved within 1 or 2 seconds, but the average (median) gap over all Bregenz instances is 46.6% (42.0%). As expected, lower bounds obtained by solving the MCF model are significantly better: 19.1% (7.1%), but the LP relaxation of only 13 out of 32 instances was solved to optimality in less than 1000 seconds (within 383.8 seconds, on average). Finally, the average (median) gap obtained by solving the DMCF model is 13.9% (7.8%), but only in 3 out of 32 cases the LP relaxation was solved to optimality within the given time limit (in 55 seconds, on average). This also explains why some of the presented gaps of the MCF model are better than the corresponding DMCF ones (LP relaxations are solved by dual simplex method).

Therefore, the only compact model that can be practically incorporated into a general branch-and-bound framework is the SCF model.

Solving MIPs: For the SCF model and for the seven branch-and-cut variants described above, we ran the code for 1000 seconds, with default Cplex settings. Only when solving the SUB model, the Cplex presolver needs to be turned off. Since the separation of Benders cuts may become a time-consuming task for instances of that size, we separate them only at the root node of the branch-and-bound tree. The results given in Table 5 show that even on this challenging set of instances we are able to beat the compact model. For 8 out of 32 instances, our branch-and-cut approach (PSUBf) is able to find the optimal solution within the given time limit, while the SCF model did not solve a single instance to optimality. Furthermore, for 18 out of the remaining 24 instances we found better gaps than SCF.

Box-plots in Figure 3 provide an overview of the obtained gaps at the root node of the branch-and-bound tree. We observe that the huge gaps of the SCF model (46.6% average and 42.0% median value), Figure 3(b)) can be reduced down to an average (median) value of 4.2% (3.4%), by involving Cplex cuts and the presolver. From Figure 3(a), we see that the lower bounds obtained by our Benders approaches are even better, although it is not a trivial task to improve the general purpose cuts produced by Cplex.

Figure 4 shows the gaps after the time limit was reached. Looking at the overall gaps after the given time limit, we observe that is is difficult to point out the differences between particular normalization approaches when default Cplex settings are used (see Figure 4(a)). Therefore, we cannot say that there is a clear winner among different Benders approaches.

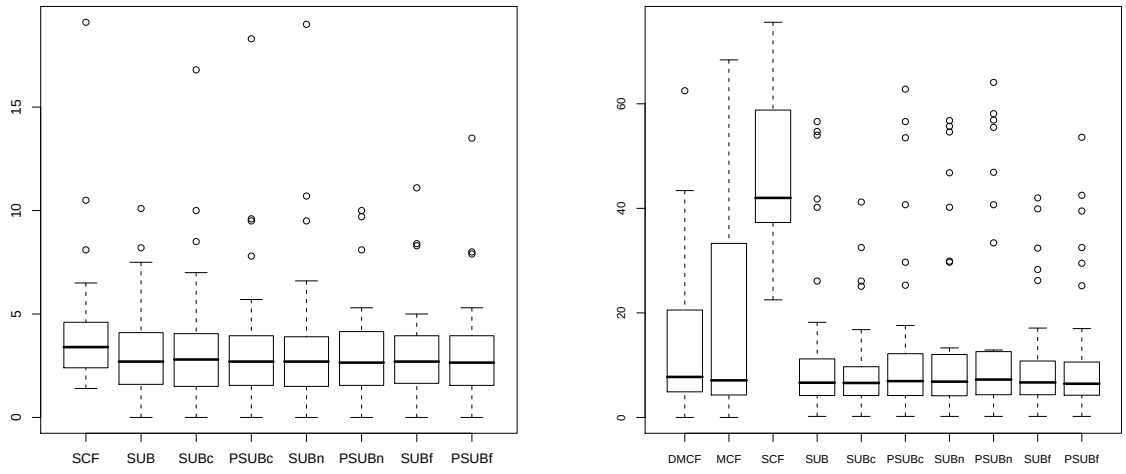
Although the Benders cuts obtained by solving the SUBc model are among the tightest ones (see, e.g., Figure 3), the separation was not finished at the root node of the branch-and-bound tree in 24 out of 32 cases. Figure 6 illustrates a typical situation in which SUBc gets stuck in the separation phase, while the SUB approach, for example, can draw an advantage out of branching.

5.5 Testing Potential Enhancements

We report on negative results when trying to enhance the Benders decomposition with nested cuts and by using Magnanti-Wong (MW) approach. In particular, solving the LP subproblem related to nested

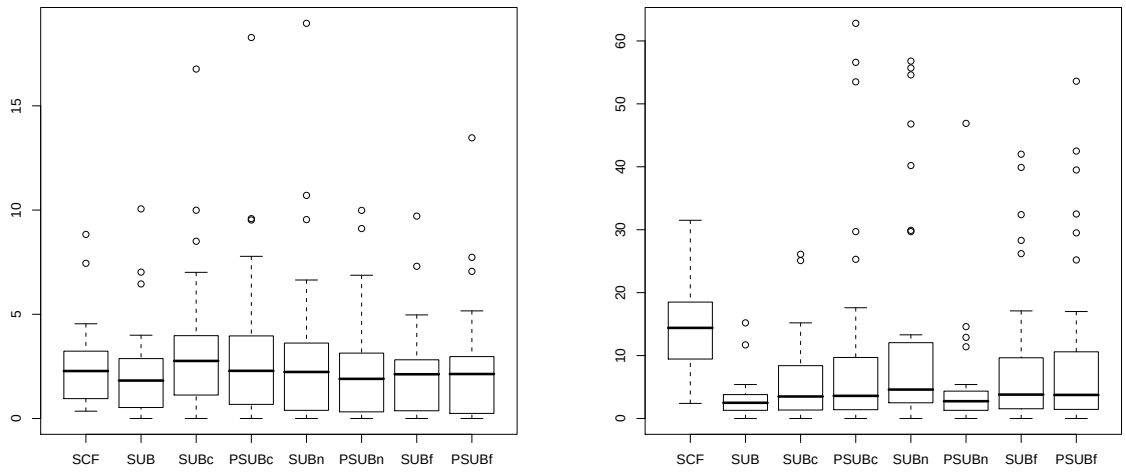
	Problem						best	best	Gap at the root			Gap after 1000s							
Instance	V	E	D	$\|\mathcal{N}_{\min}\|$	$\|\mathcal{N}_{\text{avg}}\|$	$\|\mathcal{N}_{\max}\|$	**LB**	**UB**	DMCF	MCF	SCF	SCF	SUB	SUBc	SUBn	SUBf	PSUBc	PSUBn	PSUBf
29_A_H	322	488	28	2	2.0	3	**110998.28**	**110998.28**	3.8	3.9	58.5	1.1	**0.0**	1.1	0.3	0.2	1.2	**0.0**	**0.0**
29_B_H	322	488	28	2	2.0	4	227009.39	228146.84	30.6	68.4	71.1	**0.5**	0.6	7.0	6.6	2.1	7.8	0.7	2.1
29_C_H	322	488	28	3	3.0	7	55166.67	56779.45	24.6	10.7	25.7	3.3	3.1	3.3	2.9	2.9	4.0	**2.8**	2.9
29_D_H	322	488	28	4	4.1	10	50447.35	52546.87	21.7	11.6	33.1	4.5	**4.0**	5.5	5.0	4.1	5.0	4.6	4.3
36_A_H	325	490	33	2	2.0	4	**174491.64**	**174491.64**	3.5	3.7	42.0	1.0	0.5	0.4	0.4	0.2	0.5	0.2	**0.0**
36_B_H	325	490	33	2	2.0	4	807815.62	819877.83	12.8	37.1	37.4	**1.5**	2.1	4.5	5.8	2.8	5.7	1.9	2.1
36_C_H	325	490	33	3	3.0	7	192314.14	193821.25	14.5	56.1	60.8	**0.8**	1.0	1.8	1.8	**0.8**	2.4	0.9	**0.8**
36_D_H	325	490	33	4	4.1	11	82432.96	84676.01	16.0	10.6	26.0	2.8	2.6	3.2	3.0	2.7	**2.6**	3.1	2.7
45_A_H	333	498	41	2	2.0	4	**206863.16**	**206863.16**	3.1	2.7	43.7	1.1	0.2	1.0	**0.0**	**0.0**	0.1	**0.0**	**0.0**
45_B_H	333	498	41	2	2.0	4	851564.22	915891.12	33.6	40.0	39.9	7.4	**7.0**	10.0	10.7	7.3	9.6	9.1	7.7
45_C_H	333	498	41	3	3.0	7	224778.79	230380.49	22.3	58.9	59.9	**2.6**	6.5	8.5	9.5	5.0	9.5	6.9	7.1
45_D_H	333	498	41	4	4.1	11	100670.85	104114.25	19.4	19.6	27.3	3.3	3.3	3.7	3.4	**3.0**	3.7	3.5	3.6
67_A_H	351	516	61	2	2.0	4	236879.59	238483.81	6.4	5.9	37.2	1.2	0.8	1.8	0.7	0.8	1.2	**0.7**	**0.7**
67_B_H	351	516	61	2	2.0	4	1677023.89	1839517.90	43.4	67.7	64.8	**8.8**	10.1	16.8	19.0	9.7	18.3	10.0	13.5
67_C_H	351	516	61	3	3.0	7	611825.02	625948.67	14.2	46.8	42.0	**2.4**	3.7	5.0	5.1	3.8	5.1	4.9	5.2
67_D_H	351	516	61	4	4.1	11	138159.10	141382.31	26.4	29.5	22.5	**2.3**	2.6	3.4	3.1	2.4	3.4	3.2	3.1
29_A_L	322	488	28	1	1.0	1	**101951.52**	**101951.52**	**0.0**	**0.0**	75.6	0.4	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
29_B_L	322	488	28	2	2.0	4	38318.98	38472.82	7.7	7.4	55.7	1.6	0.7	1.2	0.7	2.5	0.8	**0.4**	0.5
29_C_L	322	488	28	3	3.0	4	34927.38	35758.16	5.1	4.6	59.1	3.2	2.3	3.3	2.3	**2.2**	2.4	2.3	2.3
29_D_L	322	488	28	3	3.0	4	34734.91	35533.78	4.7	4.0	60.5	3.8	**2.2**	3.1	2.4	2.4	2.3	2.3	2.3
36_A_L	325	490	33	2	2.0	2	**163386.07**	**163386.07**	0.6	0.6	53.4	0.4	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
36_B_L	325	490	33	2	2.0	4	**59201.52**	**59201.52**	6.4	6.0	36.0	0.7	**0.0**	1.5	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
36_C_L	325	490	33	3	3.0	5	57067.61	57797.48	7.2	6.3	39.7	2.3	1.6	2.2	1.2	1.4	**1.1**	1.3	1.5
36_D_L	325	490	33	4	4.0	5	54881.13	56167.45	5.8	4.6	41.9	3.1	2.3	2.7	2.3	**2.1**	2.2	2.4	2.6
45_A_L	333	498	41	2	2.0	2	**193052.96**	**193052.96**	0.1	0.1	56.7	0.6	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
45_B_L	333	498	41	2	2.0	4	70211.46	70546.13	7.8	6.8	37.7	1.3	0.7	0.6	**0.4**	0.5	0.6	0.5	0.7
45_C_L	333	498	41	3	3.0	6	67156.12	68224.05	7.1	5.7	40.7	2.3	1.6	2.5	**1.5**	1.9	1.9	1.9	1.7
45_D_L	333	498	41	4	4.0	6	64901.08	66471.38	6.4	4.8	43.8	3.3	2.4	2.8	**2.3**	**2.3**	**2.3**	2.5	2.6
67_A_L	351	516	61	2	2.0	3	**218267.74**	**218267.74**	1.1	1.1	54.6	0.6	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
67_B_L	351	516	61	2	2.0	4	200000.95	201352.36	62.5	61.1	67.9	0.9	1.0	2.8	**0.7**	0.9	2.0	**0.7**	1.5
67_C_L	351	516	61	3	3.0	7	80926.47	82804.18	12.2	12.0	35.5	2.5	2.3	2.8	2.2	**2.1**	2.7	2.3	2.5
67_D_L	351	516	61	4	4.1	8	79031.51	81906.07	12.8	13.7	41.0	4.0	3.5	4.2	3.8	**3.4**	3.9	4.2	3.8

Table 5: Lower bounds obtained by solving LP relaxations of three compact approaches (with Cplex presolver turned off) are compared. The overall gaps (obtained with default Cplex settings, after 1000 seconds) of the compact model SCF and seven branch-and-cut variants are provided.



(a) (b)

Figure 3: Box-plots over 32 `Bregenz`-instances: the gaps of lower bounds at the root node of the branch-and-bound tree obtained a) with Cplex default settings; b) by turning off Cplex cuts and the presolver. In the latter case, gaps for MCF and DMCF models are also given.



(a) (b)

Figure 4: Box-plots over 32 `Bregenz`-instances: the overall gaps obtained after 1000 seconds a) with Cplex default settings; b) by turning off Cplex cuts and the presolver.

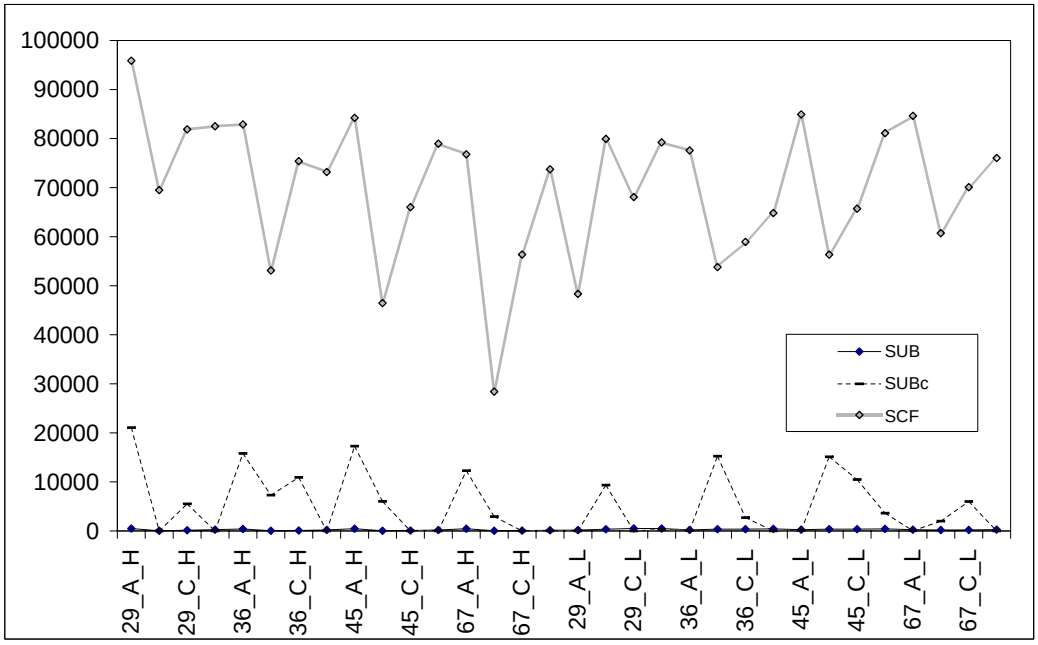
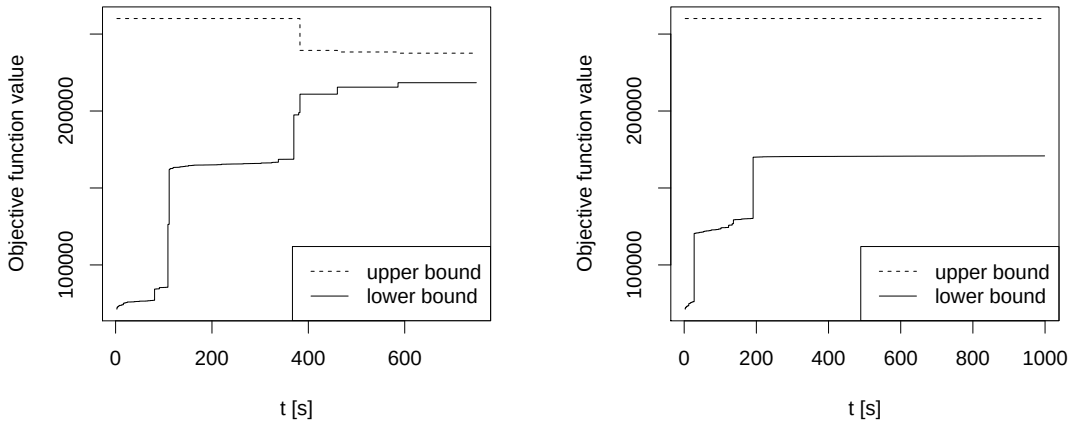


Figure 5: Comparing the number of branch-and-bound nodes traversed by three different approaches within a time limit of 1000 seconds (Cplex cuts and the presolver turned off).



(a) (SUB)

(b) (SUBc)

Figure 6: Lower bound growth vs. time (CPU seconds) with models SUB and SUBc for instance `Bregenz29_B_H`. (a) The first huge increase of lower bound is due to two subsequently found Benders cuts, the second increase is due to branching. (b) The separation at the root node is not finished when solving SUBc.

cuts is a time-consuming task which takes longer than resolving the primal problem. Since the only purpose of nested cuts is to speed up the separation without resolving the master problem, we did not obtain better results by using them.

As already observed above, if Cplex general purpose cuts are turned on, it is difficult to point out the differences between different variants of Benders separation models. Therefore, to test the effects of applying the Magnanti-Wong approach, we turned the Cplex cuts off. The MW approach generates most improving cuts when applied to the SUBn approach, for which we report the gaps obtained within the time limit of 1000 seconds (see Figure 7). We observe that the MW approach slows down the performance: the overall number of included Benders cuts is reduced while there is no significant improvement in the quality of lower bound obtained per iteration.

6 Conclusions

We have presented a new disaggregated flow formulation DMCF that produces tighter gaps than the MCF model which is typically used for network loading problems. Using Benders decomposition, we solve 8 of our 32 new single-source instances to optimality within a reasonable time limit. For 18 out of the remaining 24 instances, we report better gaps than the best performing compact formulation.

Comparing normalization strategies for the Benders decomposition, we see that depending on the structure of the inputs, different normalizations are preferable. However, in contrast to a common belief,

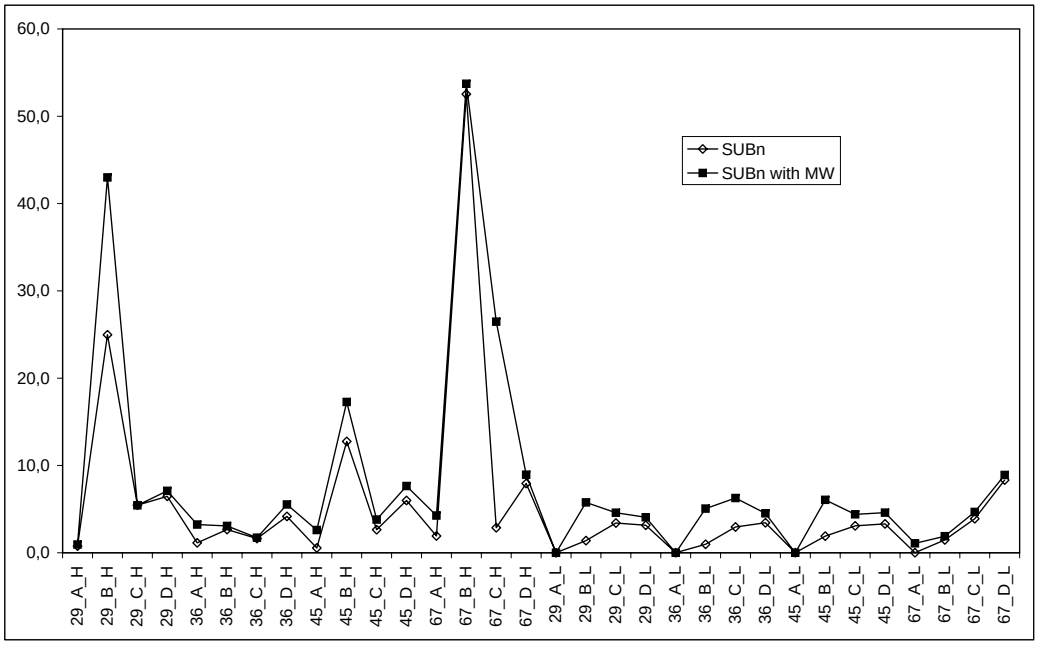


Figure 7: Comparing the gaps obtained within the time limit of 1000 seconds (Cplex cuts turned off): SUBn and SUBn extended by Magnanti-Wong cuts.

the separation of extreme rays, which is also called the *textbook implementation*, provides relatively good results across all instances.

There are several arguments explaining this observation:

1. We solve the problem starting from a compact formulation (the SCF model) and we use Benders cuts only in order to improve the quality of lower bounds, i.e., they are *not necessary* for SSNLP to have a complete MIP formulation. This is a first difference with respect to known approaches for solving the multiple-source multiple-sink network loading, where metric inequalities are separated in a similar way.

2. In opposite to our objective function, many related problems consider settings with flow-dependent objective values. In such cases, Benders cuts are used to separate both, feasibility and optimality cuts. The quality of optimality cuts is essential for such problems and therefore enhancing approaches (like those given in e.g. [18, 30, 31, 34]) play a crucial role to make Benders decomposition work.

3. We confirm the claim of Magnanti and Wong [30], that the crucial role in the generation of efficient Benders separation approaches is played by the size (see our Lemma 2.4) of the convex hull of the relaxed master problem. We show that the textbook implementation of Benders separation is not the worst possible choice, if a "good" LP-model is used to generate the corresponding cuts. Typically there is a trade-off problem in Benders decomposition approaches between the strength of the subproblem and the running time needed to solve it. To overcome this problem, the separation of extreme rays turns out to be a good compromise: an extreme ray is usually found much faster than an extreme point of a bounded subproblem.

Acknowledgments

The first author is supported by the Hertha-Firnberg Fellowship of the Austrian Science Foundation (FWF). The second author is supported by the Austrian Research Promotion Agency, FFG, within Bridge program (812973). The third author thanks the support of "Ministerio de Ciencia e Innovación" through the research project MTM2006-14961-C05-03.

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows*. Prentice Hall, 1993.

[2] Algorithmic Solutions Software Gmbh. LEDA Library for Efficient Data types and Algorithms. `http://www.algorithmic-solutions.com/leda/index.htm`; visited on June 1st 2009.

[3] P. Avella, S. Mattia, and A. Sassano. Metric inequalities and the network loading problem. *Discrete Optimization*, 4(1):103–114, 2007.

[4] F. Barahona. Network design using cut inequalities. *SIAM Journal on Optimization*, 6(3):823–837, 1996.

[5] J. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[6] D. Berger, B. Gendron, J.-Y. Potvin, S. Raghavan, and P. Soriano. Tabu search for a network loading problem with multiple facilities. *Journal of Heuristics*, 6(2):253–267, 2000.

[7] D. Bienstock, S. Chopra, O. Günlük, and C.-Y. Tsai. Mininum cost capacity installation for multicommodity flows. *Mathematical Programming*, 81(2):177–199, 1998.

[8] D. Bienstock and O. Raskina. Asymptotic analysis of the flow deviation method for the maximum concurrent flow problem. *Mathematical Programming*, 91(3):479–492, 2002.

[9] B. V. Cherkassky and A. V. Goldberg. On implementing push-relabel method for the maximum flow problem. *Algorithmica*, 19:390–410, 1997.

[10] S. Chopra, I. Gilboa, and S. T. Sastry. Source sink flows with capacity installation in batches. *Discrete Applied Mathematics*, 85(3):165–192, 1998.

[11] S. Chopra and M. R. Rao. The Steiner tree problem I: Formulations, compositions and extension of facets. *Mathematical Programming*, 64(1–3):209–229, 1994.

[12] M. Chouman, T. G. Crainic, and B. Gendron. A cutting plane algorithm for multicommodity capacitated fixed-charge network design. Technical Report CIRRELT-2009-20, CIRRELT, 2009.

[13] A. M. Costa. A survey on Benders decomposition applied to fixed-charge network design problems. *Comput. Oper. Res.*, 32(6):1429–1450, 2005.

[14] A. M. Costa, J.-F. Cordeau, and B. Gendron. Benders, metric and cutset inequalities for multicommodity capacitated network design. *Computational Optimization and Applications*, 42:371–392, 2009.

[15] K. L. Croxton, B. Gendron, and T. L. Magnanti. A comparison of mixed-integer programming models for non-convex piecewise linear cost minimization problems. *Management Science*, 49:1268–1273, 2003.

[16] K. L. Croxton, B. Gendron, and T. L. Magnanti. Variable disaggregation in network flow problems with piecewise linear costs. *Operations Research*, 55(1):146–157, 2007.

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

[18] M. Fischetti, D. Salvagnin, and A. Zanette. Minimal infeasible subsystems and Benders cuts. Technical report, University of Padua, Italy, 2009.

[19] B. Fortz and M. Poss. An improved Benders decomposition applied to a multi-layer network design problem. Technical report, GOM, Université Libre de Bruxelles, Belgium, 2008.

[20] A. Frangioni and B. Gendron. 0-1 reformulations of the multicommodity capacitated network design problem. *Discrete Applied Mathematics*, 157(6):1229 – 1241, 2009. Reformulation Techniques and Mathematical Programming.

[21] V. Gabrel, A. Knippel, and M. Minoux. Exact solution of multicommodity network optimization problems with general step cost functions. *Operations Research Letters*, 25(1):15–23, August 1999.

[22] B. Gavish and K. Altinkemer. Backbone network design tools with economic tradeoffs. *ORSA Journal on Computing*, 2(3):236–252, 1990.

[23] O. Günlük. A branch-and-cut algorithm for capacitated network design problems. *Mathematical Programming*, 86(1):17–39, 1999.

[24] A. Gupta, A. Kumar, M. Pál, and T. Roughgarden. Approximation via cost sharing: Simpler and better approximation algorithms for network design. *Journal of the ACM*, 54(3):11, 2007.

[25] IBM ILOG. CPLEX. `http://www.ilog.com/products/cplex/`; visited on August 1st 2009.

[26] A. B. Keha, I. R. de Farias Jr., and G. L. Nemhauser. Models for representing piecewise linear cost functions. *Operations Research Letters*, 32(1):44–48, 2004.

[27] I. Ljubić, R. Weiskircher, U. Pferschy, G. Klau, P. Mutzel, and M. Fischetti. An algorithmic framework for the exact solution of the prize-collecting Steiner tree problem. *Mathematical Programming*, 105(427–449), 2006.

[28] T. Magnanti, P. Mireault, and R.T. Wong. Tailoring Benders decomposition for uncapacitated network design. *Mathematical Programming Study*, 18(1):112–154, 1984.

[29] T. L. Magnanti, P. Mirchandani, and R. Vachani. Modeling and solving the two-facility capacitated network loading problem. *Operations Research*, 43(1):142–157, 1995.

[30] T. L. Magnanti and R. T. Wong. Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484, 1981.

[31] T. L. Magnanti and R. T. Wong. Network design and transportation planning: Models and algorithms. *Transportation Science*, 18(1):1–55, 1984.

[32] N. Papadakos. Practical enhancements to the Magnanti-Wong method. *Operations Research Letters*, 36(4):444 – 449, 2008.

[33] S. Raghavan and D. Stanojević. A note on search by objective relaxation. In *Telecommunications Planning: Innovations in Pricing, Network Design and Management*, volume 33 of *Operations Research/Computer Science Interfaces Series*, pages 181–201. Springer US, 2006.

[34] W. Rei, J.-F. Cordeau, M. Gendreau, and P. Soriano. Accelerating Benders decomposition by local branching. *INFORMS Journal on Computing*, 21(2):333–345, 2009.

[35] F. S. Salman, J. Cheriyan, R. Ravi, and S. Subramanian. Approximating the single-sink link-installation problem in network design. *SIAM Journal on Optimization*, 11(3):595–610, 2000.

[36] F. S. Salman, R. Ravi, and J. N. Hooker. Solving the capacitated local access network design problem. *INFORMS Journal on Computing*, 20(2):243–254, 2008.

[37] S. Salman. *Selected Problems in Network Design: Exact and Approximate Solution Methods*. PhD thesis, Carnegie Mellon University, Pittsburgh, 2000.

[38] S.P.M. van Hoesel, A.M.C.A. Koster, R. L. M. J. van de Leensel, and M. W. P. Savelsbergh. Bidirected and unidirected capacity installation in telecommunication networks. *Discrete Applied Mathematics*, 133(1-3):103–121, 2003.