Robust Network Design: Formulations, Valid Inequalities, and Computations

Arie M.C.A. Koster* Manuel Kutschka* Christian Raack†

31st August 2011

Abstract

Traffic in communication networks fluctuates heavily over time. Thus, to avoid capacity bottlenecks, operators highly overestimate the traffic volume during network planning. In this paper we consider telecommunication network design under traffic uncertainty, adapting the robust optimization approach of [21]. We present three different mathematical formulations for this problem, provide valid inequalities, study the computational implications, and evaluate the realized robustness.

To enhance the performance of the mixed-integer programming solver we derive robust cutset inequalities generalizing their deterministic counterparts. Instead of a single cutset inequality for every network cut, we derive multiple valid inequalities by exploiting the extra variables available in the robust formulations. We show that these inequalities define facets under certain conditions and that they completely describe a projection of the robust cutset polyhedron if the cutset consists of a single edge.

For realistic networks and live traffic measurements we compare the formulations and report on the speed up by the valid inequalities. We study the "price of robustness" and evaluate the approach by analyzing the real network load. The results show that the robust optimization approach has the potential to support network planners better than present methods.

1 Introduction

Dimensioning or expanding capacity networks is a complex task with many applications in transportation, energy supply, and telecommunications. Doing it carefully with respect to expenditures and expected network demand is crucial for the behavior and flexibility of the resulting network.

In this work we mainly focus on aspects from telecommunication networks but our methods are general enough to be applied also in different contexts. Telecommunication network design typically involves decisions about the network topology, link and node capacities, and traffic routing. It can be considered as a long-term to mid-term strategic planning process. Its goal is to minimize the capital expenditures for network equipment guaranteeing a routing for all considered (data) traffic demands. In the classical combinatorial network design problem integer capacities (corresponding to batches of bit rates, e. g., 40 Gbps) have to be installed on the network links at minimum cost such that all traffic demands can be realized by flow simultaneously without exceeding the link capacities. Given a single traffic matrix, this problem has been studied extensively in the literature, see [8, 10, 23, 24, 37, 47, 59] and the references therein.

In practice, telecommunication networks are typically designed without the knowledge of actual traffic. In most approaches each demand is estimated in the design process, e.g., by using traffic measurements or population statistics [25, 31, 40, 63, 64]. To handle future changes in the traffic volume and distribution and to guarantee robust network designs, these values (and consequently capacities) are (highly) overestimated. Obviously, this approach leads to a wastage of network capacities, investments, and energy. In order to create and operate more resource- and cost-efficient telecommunication networks the uncertainty of future traffic demand has to be already taken into account in the strategic capacity design process.

*RWTH Aachen University, Lehrstuhl II für Mathematik, Wüllnerstr. 5b, D-52062 Aachen, Germany, {koster,kutschka}@math2.rwth-aachen.de

†Zuse Institute Berlin (ZIB), Takustr. 7, D-14195 Berlin, Germany, raack@zib.de

Incorporating uncertainty within the mathematical analysis of operational research has been an effort since its very first beginnings. In the 1950s, Dantzig [28] introduced Stochastic Programming using probabilities for the possible realizations of the uncertain data. The main limitation of such probabilistic approaches is that the distribution of the uncertain data must be known a priori which is often not the case for "real-world" problems. Stochastic programming may also result in extremely hard to solve optimization problems.

In 1973, Soyster [61] suggested another approach based on implicitly describing the uncertain data introducing so-called uncertainty sets and establishing the concept of Robust Optimization. Using this framework we do not need any information about the probabilistic distribution of the uncertainty. Instead a solution is said to be robust if it is feasible for all realizations of the data in the given uncertainty set. In Robust Optimization we aim at finding the cost-optimal robust solution. This approach has been further developed by Ben-Tal and Nemirovski [15, 16, 17], Bertsimas and Sim [21], and others using different convex and bounded uncertainty sets. They introduce the concept of robust counterparts for uncertain linear programs. In [15] it is shown that these can be solved by deterministic linear programs or deterministic conic quadratic programs if the uncertainty set is polyhedral or ellipsoidal, respectively. Bertsimas and Sim [21] introduced a polyhedral uncertainty set that easily allows to control the price of robustness by varying the number Γ of coefficients in a row of the given linear program that are allowed to deviate from its nominal values simultaneously. By changing this parameter Γ the practitioner is enabled to regulate the trade-off between the degree of uncertainty taking into account and the cost of this additional feature.

Robust optimization is also a well known method in telecommunication network design. We distinguish between robust network design using static or dynamic routing which refers to the flexibility of flow to respond to the realization of the demand (while the capacity remains fixed). Static routing means that for every node pair the same paths are used with the same splitting independent of the realization of demand. Contrary, dynamic routing allows for full flexibility in rerouting the traffic if the demand changes. The concept of different routing schemes is strongly related to different levels of recourse in multi-stage stochastic and robust optimization [18]. We refer to Mudchanatongsuk et al. [51] and Poss and Raack [58] for a discussion on how to embed the two classical routing schemes in these more general frameworks. For general two-stage robust network design check [9]. We also note that recently there has been some progress in defining routing schemes in between static and dynamic, see for instance [13, 57, 58, 60].

In multi-period or multi-hour network design [44, 62], an explicit finite set of demand matrices is given, and the network is designed in such a way that each of the demand matrices can be routed non-simultaneously within the installed capacities (potentially expanding the network capacity in every period). In this context, Oriolo [55] introduces the concept of dominated demand matrices. Matrices are dominated if they can be removed from the uncertainty set without changing the problem. Oriolo [55] classifies domination for pairs of demand matrices and static as well as dynamic routing. Ben-Ameur and Kerivin [14] consider static routing and demands that may vary within a polytope given in the space of the commodities. The resulting infinite set of capacity constraints is handled by separation.

For telecommunication network design problems this concept of polyhedral demands has mainly been applied using the *hose model* [30]. In its symmetric version, the hose model defines upper bounds on the sum of the incoming and outgoing node traffic for all network nodes. This model has attracted a lot of attention in recent years, in particular, due to its nice theoretical and algorithmic properties assuming continuous capacities (e.g., polynomial solvable cases, see [26, 34, 38]). For algorithmic and computational studies using the hose model the reader is referred to Altin et al. [4, 6] (static routing)and Mattia [48] (dynamic routing).

Another compact but less studied description of uncertainty of traffic in telecommunication networks is obtained by applying the framework of Bertsimas and Sim [21] to network design problems. As demand uncertainty essentially arises in the coefficients of the capacity constraints of network design formulations, a restriction on the number of coefficients to deviate simultaneously translates to a restriction on the number of individual commodity-demands to deviate simultaneously. The corresponding polyhedral demand uncertainty set, which we call the Γ-model, hence provides a meaningful alternative to the hose model. Given that in realistic traffic scenarios it is unlikely to have all demands at their peak at the same time, the number of simultaneous peaks is restricted to a (small) non-negative value Γ. Adjusting Γ relates to adjusting the robustness and the level of conservatism of the solutions which provides additional flexibility. Altin et al. [4] apply the Γ-model to the classical VPN problem with single-path flows and continuous capacities. Belotti et al. [11] use a simplification of the Γ-model for the special case $\Gamma = 1$ and solve a robust network design problem by Lagrangian relaxation. Klopfenstein and Nace [42] consider bandwidth packing using the Γ-model focusing on the robust knapsack problem given by link capacity constraints. Poss and Raack [58] and Ouorou and Vial [57] study variants of the Γ-model in the context of network design and affine routing as alternative to the static and dynamic routing schemes. Finally Belotti et al. [12] solve real-world network design problems taking into account demand uncertainty using the Γ-model. They computationally compare different layer architectures with respect to equipment cost for different values of Γ.

Contributions of this paper In this paper, we consider the robust network design problem with static routing following Bertsimas and Sim [21]. We enhance in three different ways the classical flow formulation for network design to include demand uncertainties. First, we derive a straightforward exponential-size mixed-integer programming (MIP) formulation. Next, we use duality theory to obtain a compact reformulation, and finally, we project out the flow variables. A computational evaluation reveals that the compact formulation outperforms the other models.

To improve the performance of the MIP solver we study the robust counterpart of the well-known cutset polyhedron for network design. Instead of a single cutset inequality for every network cut, we derive multiple classes of facet-defining cut-based inequalities by exploiting the extra variables available in two of the robust formulations. Computations show that the robust cut-based inequalities significantly reduce the computation times.

Finally, we analyze the robustness for realistic networks and a demand forecast based on live traffic measurements by comparing the cost savings using robust optimization instead of an overestimation. The designs are evaluated on "future" traffic matrices. This process demonstrates how robust optimization can support network planning.

Outline This paper is structured in three parts: formulations, valid inequalities, and computations. In Section 2.2 we introduce three different formulations for robust network design using the Γ-model. Section 3 is devoted to cut-based valid inequalities to improve the formulations. In Section 4, we report on the computational comparison of the formulations and an evaluation of the robust network designs. We close with concluding remarks.

2 Formulations for Robust Network Design

2.1 Modeling Alternatives

We consider the following robust network design problem. We are given an undirected connected graph $G = (V, E)$ representing a potential network topology. On each of the links $e \in E$ capacity can be installed in integral units and costs κ_e per unit. A set of commodities K represents potential traffic demands. More precisely, a commodity $k \in K$ corresponds to node pair (s^k, t^k) and a demand $d^k \geq 0$ for traffic from $s^k \in V$ to $t^k \in V$. The actual demand values are considered to be uncertain. For the moment we assume that the demand vector $d \in \mathbb{R}_+^{|K|}$ corresponding to the demand values $d^k, k \in K$ lies in a given polytope $\mathcal{D} \subset \mathbb{R}_+^{|K|}$ without explicitly specifying $\mathcal{D}$.

The traffic for commodity k is realized by a splittable multi-path flow between s^k and t^k. Of course, the actual multi-commodity flow depends on the realization of the demand $d \in \mathcal{D}$. In this context the literature roughly distinguishes two main routing principles. We either choose an arbitrary flow for every realization of the demand in $\mathcal{D}$, which is known as *dynamic routing* or we fix a *routing template* for every commodity, that is, every realization of a commodity demand has to use the same set of paths between $s^k \in V$ and $t^k \in V$ with the same percentual splitting of flow among the paths. This latter principle is known as *oblivious routing* (or *static routing*) and is considered in this paper. We refer the reader to [48] for solution approaches considering dynamic routing principle.

The robust network design problem using oblivious routing (RND) is to find a minimum-cost installation of integral capacities and a routing template for every commodity such that actual flow does not exceed the link capacities independent of the realization of demands in $\mathcal{D}$. This problem can be formulated as the following integer linear program.

$$\min \quad \sum_{e\in E} \kappa_e x_e \tag{1a}$$

$$s.t. \quad \sum_{j\in N(i)} (y^k_{ij} - y^k_{ji}) = \begin{cases} 1 & i = s^k \\ -1 & i = t^k \\ 0 & \text{else} \end{cases}, \forall i \in V,\, k \in K \tag{1b}$$

$$\sum_{k\in K} d^k y^k_e \leq x_e, \qquad \forall e \in E, d \in \mathcal{D} \tag{1c}$$

$$y, x \geq 0 \tag{1d}$$

$$x \in \mathbb{Z}^{|E|} \tag{1e}$$

Here $N(i)$ denotes the set of neighboring nodes of i. Constraints (1b) describe a multi-commodity flow using a link-flow formulation. The flow for commodity k is directed from (its source) s^k to (its target) t^k without loss of generality. Constraints (1c) are link capacity constraints. Variables x_e denote the number of capacity batches installed on $e \in E$ at cost κ_e per batch.

The flow variables y^k_{ij}, y^k_{ji} denote the fraction of demand (independent of its realization) routed on $e = \{i, j\}$ away from node i, j, respectively. We set $y^k_e := y^k_{ij} + y^k_{ji}$. Since the cost for link capacity is minimized we may ignore cycle flows and hence assume that either $y^k_{ij} = 0$ or $y^k_{ji} = 0$ and $y^k_e \leq 1$ in any optimal solution. The vector $y^k \in \mathbb{R}^E_+$ is called a routing template for commodity $k \in K$. Given a capacity allocation $x \in \mathbb{R}^{|E|}_+$, and routing templates y^k for all $k \in K$, we say that (x, y) supports $d \in \mathcal{D}$ in case (1c) is satisfied for d. Notice that while flow and demands are directed, the actual direction is arbitrary since we sum up the two flows in (1c).

Fixing $d \in \mathcal{D}$, the realized flow $f^k_e(d)$ for commodity k on edge e amounts to

$$f^k_e(d) := d^k y^k_e. \tag{2}$$

This means that we allow the flow to change with the demand fluctuations d but we restrict the flow dynamics to the linear functions given by (2). Notice that (2) is as a special case of so-called *affine recourse* introduced by Ben-Tal et al. [18] in the context of adjustable robust solutions of linear programs with uncertain data. Ouorou and Vial [57] apply affine recourse to network design introducing affine routing. Poss and Raack [58] provide a conceptual discussion of the three routing schemes: oblivious, affine, and dynamic.

Domination In its general form, formulation (1) is semi-infinite. It however by convexity suffices to claim (1c) for every vertex of the polytope $\mathcal{D}$. Moreover, using the concepts of domination of demand vectors introduced by Oriolo [55] we may remove vectors from the uncertainty set that are dominated without changing the problem. A vector $d_1 \in \mathcal{D}$ is said to *totally dominate* a second vector $d_2 \in \mathcal{D}$ if any pair (x, y) supporting d_1 also supports d_2. In this case we may remove the capacity constraints corresponding to d_2 from the formulation which refers to removing d_2 from $\mathcal{D}$. By removing totally dominated demand vectors from $\mathcal{D}$ the set of feasible solutions (x, y) to (1) remains unchanged. Since there is no cost for flow in (1) we can even use a weaker domination concept in order to remove a potentially larger set of demand realizations from $\mathcal{D}$. Following Oriolo [55], a vector $d_1 \in \mathcal{D}$ is said to *strongly dominate* a second vector $d_2 \in \mathcal{D}$ if for every (x, y) supporting d_1 there exists a routing template $y^\star$ such that $(x, y^\star)$ supports both d_1 and d_2. Total domination implies strong domination. Removing strongly dominated $d \in \mathcal{D}$ might change the set of feasible solutions (x, y) but the set of feasible capacity allocations x is not changing. Oriolo [55] shows hat d_1 totally dominates d_2 if and only if $d^k_1 \geq d^k_2$ for all $k \in K$ and provides a similar characterization for strong domination.

Whenever we speak of domination in the sequel we allow to remove demand vectors from the uncertainty set without distinguishing total or strong domination. From the discussion above it follows that (1) can be reduced to a compact formulation if the number of non-dominated vertices of $\mathcal{D}$ is polynomial in the number of nodes and arcs.

Dualization Using linear programing duality it is however easy to show that there exists also a compact formulation for the problem if the number of non-dominated vertices is exponential but the number of facets of $\mathcal{D}$ is polynomial. Dualization of constraints is a central technique in robust optimization, see [16, 17, 19] and [20, 21]. For problem (RND) we observe that the data uncertainty only affects the capacity constraints (1c) which we rewrite as

$$\max_{d \in \mathcal{D}} \sum_{k \in K} d^k y_e^k \ \leq x_e, \forall e \in E. \tag{3}$$

Since $\mathcal{D}$ is a polytope the maximization in (3) refers to a linear program which maximizes the link flow $f_e^k(d)$ over all demand realizations in $\mathcal{D}$ using the linear recourse (2). Dualizing the linear description of $\mathcal{D}$ by introducing a dual variable for every (non-redundant) inequality describing $\mathcal{D}$ and removing the $\min$ in the dualization (which can be done since we are minimizing capacities anyway) we obtain a so-called robust counterpart for (RND). The robust counterpart for general polytopes $\mathcal{D}$ is provided in [6]. It is compact if the linear description for $\mathcal{D} \subset \mathbb{R}^{|K|}$ is compact, that is, the number of facets of $\mathcal{D}$ is polynomial. In the next section we discuss the dualization using a particular uncertainty set.

We may of course exploit domination to derive a formulation for $\mathcal{D}$ that has either polynomial vertices or facets. From the computational point of view it can be even wise to extend $\mathcal{D}$ by adding dominated vectors. Moreover, instead of a formulation for $\mathcal{D}$ in $\mathbb{R}^{|K|}$ it is also possible to use compact extended formulations of $\mathcal{D}$ to derive compact robust counterparts of (RND), see the next section. We denote by DUALIZE any algorithmic approach that is based on solving the robust counterpart given in [6] obtained by dualizing the capacity constraints (1c) as described above.

Separation Instead of dualizing the capacity constraints we can also use the original formulation (1) after reducing it to the (non-dominated) vertices of $\mathcal{D}$. In case this number is exponential the resulting exponential number of capacity constraints (1c) are handled by separation. This approach which we refer to as algorithm SEPARATE has been used by Ben-Ameur and Kerivin [14] to solve (RND) for the hose model. We remove the capacity constraints (1c) from the system (1) (or keep a subset) and add them dynamically. For this we have to solve the separation problem in (3) in every iteration and for every edge. More precisely, given a solution $(x^\star, y^\star)$ to (1) or its linear programming relaxation containing only a subset of the constraints (1c), for every edge $e \in E$, we solve the problem

$$\max \sum_{k \in K} d^k y_e^{\star k} \ \text{ s.t. } \ d \in \mathcal{D}$$

obtaining an optimal solution $d_\star \in \mathcal{D}$. The vector $d_\star$ refers to the worst case demand realization in $\mathcal{D}$ for edge e given flow template $y^\star$. In case $\sum_{k \in K} d_\star^k y_e^{\star k} > x_e^\star$ we add the corresponding capacity constraint $\sum_{k \in K} d_\star^k y_e^k \leq x_e$ to formulation (1) and resolve.

Replacement Altin et al. [6] propose a third way to solve (RND) that is based on projecting out the flow variables in the dualized reformulation which we refer to as the REPLACE approach. The projection results in an exponential formulation in the space of the capacity variables and the variables used to dualize the capacity constraints. The formulation does not rely on general metric inequalities [41, 50, 54] but is solely based on network cuts such that the corresponding separation can be handled by standard max-flow-min-cut [3].

In Section 2.2 we show how to use the three approaches, DUALIZE, SEPARATE, and REPLACE to solve (RND) using a particular demand uncertainty, the Γ-model.

Uncertainty Sets The design of uncertainty sets has a strong impact on the quality and robustness of the solution. There are many criteria for selecting uncertainty sets for a particular application. We believe that the following should receive particular attention. First, any uncertainty set should reflect real data uncertainty as much as possible, that is, in our case, $\mathcal{D}$ should contain the observed (non-dominated) demand fluctuations that happen with high probability, and it should not contain demand realizations that are highly unlikely. Secondly, data to formulate and parametrize the uncertainty set should be available.

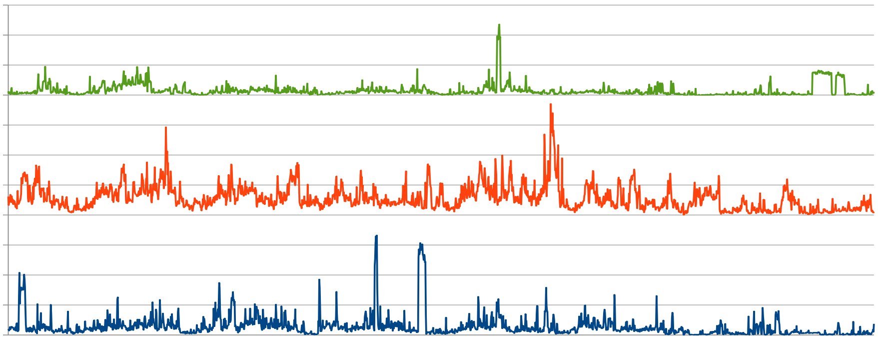


Figure 1: Traffic fluctuations of three source-destination pairs in the US Abilene Internet2 network [1] in time intervals of 5 minutes during one week

Moreover, the uncertainty set with these parameters in a meaningful way. And last but not least, a criterion often neglected, the resulting formulation or reformulation should be computationally tractable.

One such uncertainty set is the already mentioned hose model. We refer to Altin et al. [6] and Ben-Ameur and Kerivin [14] solving network design problems using the hose model and algorithms DUALIZE and SEPARATE, respectively. Theoretical results can be found in [26, 34, 38]. Introduced in the context of virtual private networks (VPNs) [30, 32], the hose model describes feasible traffic realizations by bounds on single nodes only. This data is typically available to the network practitioner [30, 32]. However, describing traffic by node demands strongly simplifies the notion of traffic matrices since the traffic fluctuations are aggregated at the network nodes. Whenever point-to-point traffic data is available or can be calculated from real-life measurements in networks [56, 63, 64] or from population statistics [31], it becomes desirable to work with more flexible uncertainty sets that reflect the observed characteristics and dynamics of the point-to-point traffic to allow for more accurate network designs.

2.2 Γ-uncertainty

For general linear and integer programs, Bertsimas and Sim [20, 21] propose a polyhedral uncertainty set together with a simple way to adjust the price for robustness (i. e., increase of the objective value of a robust solution compared to its non-robust counterpart) by tuning the shape of the set. In their model the coefficients of the constraint matrix may vary around a given nominal value but the number of deviating coefficients is bounded by a (small) number Γ_i for every row i of the matrix. Adjusting Γ_i means to control the price for robustness. Because of its simplicity this concept has been used extensively in robust optimization for many different applications [4, 9, 20, 21, 22, 42].

We already observed that data uncertainty for (RND) only appears in the capacity constraints (1c) with uncertain coefficients $d^k, k \in K$. Applying the framework of Bertsimas and Sim [20] to (RND) means to restrict the number of commodity demands that deviate from a given nominal demand value $\bar{d}^k$, simultaneously. In particular, the number of simultaneous demand peaks is bounded. This assumption has again a strong relation to telecommunications since in typical traffic patterns, in particular in IP (Internet Protocol) networks, traffic peaks do not occur simultaneously, cf. Figure 1. Note also that the main justifications for a meaningful application of the Γ-model to our problem is the huge number of coefficients in the uncertain capacity constraints (1c). The situation can be different for other applications.

In the following we introduce the Γ-model for (RND) in detail. Similar uncertainty models have been used in [4], [5] and [42] applied to different versions of network design and demand packing problems. We assume that the demand for commodity $k \in K$ varies around a given *nominal demand* $\bar{d}^k$ with a maximal possible *deviation* of $0 \leq \hat{d}^k$, that is,

$$d^k \in [\bar{d}^k - \hat{d}^k, \bar{d}^k + \hat{d}^k] \text{ for all } k \in K. \tag{4}$$

Now we limit the number of possible deviations from the nominal value:

$$\sum_{k \in K} \frac{|d^k - \bar{d}^k|}{\hat{d}^k} \leq \Gamma, \tag{5}$$

where $\Gamma \in \{0, \dots, |K|\}$. We use the same Γ for all capacity constraints (1c) since the coefficients (the demand scenarios) are independent of the edges. The corresponding uncertainty polytope $\mathcal{D}$ can be described in $\mathbb{R}^{|K|}$ using exponential many inequalities or alternatively using a compact extended formulation. For the latter, we rewrite $d^k = \bar{d}^k + (\sigma_+^k - \sigma_-^k)\hat{d}^k$ and consider $(\sigma_+, \sigma_-) \geq 0$ satisfying

$$\sigma_+^k + \sigma_-^k \leq 1 \text{ for } k \in K \text{ and } \sum_{k \in K} (\sigma_+^k + \sigma_-^k) \leq \Gamma.$$

However, the Γ-model can also be described in the original space $\mathbb{R}^{|K|}$ with polynomial inequalities if we exploit the concept of domination. Since only the worst-case edge-flow determines the edge-capacity, the problem remains the same if the actual demand is assumed to be in the interval $[\bar{d}^k, \bar{d}^k + \hat{d}^k]$, instead of the interval $[\bar{d}^k - \hat{d}^k, \bar{d}^k + \hat{d}^k]$ for $k \in K$. Demand vectors containing downward deviations from the nominal are totally dominated. We may hence assume that $\sigma_-^k = 0$ for all $k \in K$ without changing the problem. The set of possible deviations scenarios can be defined as

$$\mathcal{D}^\Gamma := \{\sigma \in \mathbb{R}_+^K : \sigma^k \leq 1 \text{ for } k \in K \text{ and } \sum_{k \in K} \sigma^k \leq \Gamma\}$$

and the corresponding Γ-model capacity constraints (1c) reduce to

$$\sum_{k \in K} \bar{d}^k y_e^k + \max_{\sigma \in \mathcal{D}^\Gamma} \sum_{k \in K} \sigma^k \hat{d}^k y_e^k \leq x_e, \forall e \in E. \tag{6}$$

Remark 2.1. By domination we might even force $\sum_{k \in K} \sigma^k = \Gamma$ which removes the all-nominal demand vector (and vertex) $\bar{d}$ from $\mathcal{D}^\Gamma$. This observation however is not improving on the robust counterpart below such that we stick to the full-dimensional decription of $\mathcal{D}^\Gamma$. Setting $\Gamma = 0$, the polytope $\mathcal{D}^\Gamma$ reduces to a singleton, the origin. Hence there is no deviation and (RND) reduces to the deterministic problem of optimizing against the all-nominal vector $\bar{d}$. Similarily, in case $\Gamma = |K|$ (RND) reduces to the problem of optimizing against the worst-case all-peak scenario $\bar{d} + \hat{d}$. By varying Γ in $\{0, \dots, |K|\}$ we may adjust the level of robustness.

We remark that a simple compact alternative to the defined Γ-model which is also described in the original space is to use (4) plus a relaxation of (5):

$$\sum_{k \in K} \frac{d^k - \bar{d}^k}{\hat{d}^k} \leq \Gamma.$$

In this case there might be more than Γ many upward deviations if compensated by an appropriate number of downward deviations and vice versa. This results in a relaxed uncertainty set potentially giving more conservative solutions.

We now continue with the three modelling alternatives in case of Γ-uncertainty. We will assume $0 < \Gamma < |K|$ in the following. In this case the polytope $\mathcal{D}^\Gamma$ is full-dimensional and has $\binom{|K|}{\Gamma} + 1$ many vertices. The capacity of a link has to be determined subject to at most Γ commodities deviating from the nominal demand value. For each deviating commodity, the peak value $\bar{d}^k + \hat{d}^k$ describes the worst-case (capacity-wise). Hence, the complete model reads

SEPARATE:

(1a), (1b), (1d), (1e)

$$\sum_{k \in K} \bar{d}^k y_e^k + \sum_{k \in Q} \hat{d}^k y_e^k \leq x_e \qquad \forall Q \subset K, |Q| = \Gamma. \tag{7a}$$

and we denote its convex hull of feasible solutions by P_Γ^{sepa}. As mentioned above we may handle the exponential number of capacity constraints (related to vertices of $\mathcal{D}^\Gamma$) by separation. In the corresponding SEPARATE approach we have to solve the problem

$$\max \left\{ \sum_{k\in K} \sigma^k \hat{d}^k y_e^{\star k} \quad : \quad 0 \le \sigma^k \le 1 \;\forall k \in K,\; \sum_{k\in K} \sigma^k \le \Gamma \right\} \tag{8}$$

for every edge e given a solution $(x^\star, y^\star)$. This problem can be solved directly by sorting the commodities with respect to the value $\hat{d}^k y_e^{\star k}$. The Γ largest values determine the worst-case commodity subset $Q \subset K$ with $|Q| = \Gamma$ for edge $e \in E$ maximizing the left hand side of the capacity constraint (6). In case violated by $(x^\star, y^\star)$ we add the corresponding capacity constraint.

In our implementation of SEPARATE the initial formulation already contains one capacity constraint per edge corresponding to the all-nominal scenario, that is, we start with (1) where (1c) is used for $d = \bar{d}$ only.

The linear program (8) bounded and (integral) feasible for all vectors y, every edge $e \in E$, and all considered values of Γ. Hence, by strong duality, its optimal objective value coincides with

$$\min \sum_{k\in K} p_e^k + \pi_e \Gamma \quad \text{s.t.} \quad \pi_e, p_e^k \ge 0 \;\forall k \in K,\; p_e^k + \pi_e \ge \hat{d}^k y_e^k \;\forall k \in K \tag{9}$$

for every edge $e \in E$. With this relation we can reformulate the exponential model (1) as:

DUALIZE:

$$\text{(1a), (1b), (1d), (1e)}$$

$$\Gamma \pi_e + \sum_{k\in K} \bar{d}^k y_e^k + \sum_{k\in K} p_e^k \le x_e, \quad \forall e \in E \tag{10a}$$

$$-\pi_e + \hat{d}^k y_e^k - p_e^k \le 0, \quad \forall e \in E, k \in K \tag{10b}$$

$$p, \pi \ge 0 \tag{10c}$$

The DUALIZE approach to solve (RND) using the Γ-model is based on solving the compact model (10). Compared to the (singleton scenario) deterministic network design model obtained by setting $|\mathcal{D}| = 1$ in (1) we have $|E| + |E||K|$ additional variables and $|E||K|$ additional constraints.

We denote by

$$P_\Gamma := \operatorname{conv}\left\{(x, y, \pi, p) \in \mathbb{Z}_+^{|E|} \times \mathbb{R}_+^{2|E||K|} \times \mathbb{R}_+^{|E|} \times \mathbb{R}_+^{|E||K|} \mid (x, y, \pi, p) \text{ satisfies (10)}\right\}$$

the convex hull of all feasible solutions of model (10) and $\operatorname{proj}_{(x,\pi)}(P_\Gamma)$ as the projection on the (x, π) space. Since this model is at the center of this paper, we prove the dimension of its polyhedron.

Proposition 2.2. *The dimension of P_Γ equals $2|E| + 3|E||K| - (|V| - 1)|K|$ whereas $\operatorname{proj}_{(x,\pi)}(P_\Gamma)$ is full-dimensional.*

Proof. For P_Γ, there are $2|E|+3|E||K|$ variables and $(|V|-1)|K|$ linearly independent flow conservation constraints (1b). We show that there are no additional implied equations. Let

$$\sum_{e\in\delta(S)} \alpha_e x_e + \sum_{e\in\delta(S)} \beta_e \pi_e + \sum_{e\in\delta(S)} \sum_{k\in K} \delta_e^k p_e^k + \sum_{e=\{i,j\}\in\delta(S)} \sum_{k\in K} (\mu_{ij}^k y_{ij}^k + \mu_{ji}^k y_{ji}^k) = \gamma \tag{11}$$

be an equation satisfied by all points in P_Γ and let $\hat{p} = (\hat{x}, \hat{y}, \hat{\pi}, \hat{p}) \in P_\Gamma$. For all $e \in E$ we can modify $\hat{p}$ by increasing the capacity without leaving P_Γ. Hence, $\alpha_e = 0$ for all $e \in E$. Once we increased the capacity we can also increase variables π_e and p_e^k for every $e \in E$ and $k \in K$ which gives $\beta_e = \delta_e^k = 0$ for all $e \in E$ and $k \in K$. Now we choose a spanning tree $T \subseteq E$ in G which exists since G is connected. By adding a linear combination of the flow conservation constraints (1b) to (11) we can assume that either μ_{ij}^k or $\mu_{ji}^k = 0$ for all $e = \{i,j\} \in T, k \in K$. Sending a small flow in both direction on every e in T gives $\mu_{ij}^k = \mu_{ji}^k = 0$. Now choosing an arbitrary edge $e \in E$ there is a unique circuit consisting of e and edges in T. Sending small circulation flows (in both direction) on this circuit finally results in $\mu_{ij}^k = \mu_{ji}^k = 0$ for all $e = \{i, j\} \in E, k \in K$. It follows that (11) is a linear combination of flow conservation constraints which gives the desired results. By projecting all constructed points we also conclude that $\mathrm{proj}_{(x,\pi)}(P_\Gamma)$ has dimension $2|E|$. ☐

Altin et al. [6] propose a projection of robust counterparts of the form (10) to the space of the design and dual variables. A similar projection has been studied before by Mirchandani [50] in the context of deterministic network design and can be applied to the compact model (10) as well. For this, we introduce slack variables q_e^k corresponding to inequalities (10b). Hence, flow variables y_e^k in (10a) can be replaced by $\frac{1}{\hat{d}^k}(\pi_e + p_e^k - q_e^k)$. Now the flow for different commodities is no longer bundled. The existence of a flow from s^k to t^k can be guaranteed by a minimum-cut condition replacing the flow formulation (1b) and resulting in the following exponential model

REPLACE:

$$\text{(1a), (1d), (1e)}$$

$$\left(\Gamma + \sum_{k\in K} \frac{\bar{d}^k}{\hat{d}^k}\right)\pi_e + \sum_{k\in K}\left(\frac{\bar{d}^k + \hat{d}^k}{\hat{d}^k} p_e^k - \frac{\bar{d}^k}{\hat{d}^k}\bar{q}_e^k\right) \le x_e, \quad \forall e \in E \tag{12a}$$

$$\sum_{e\in\delta(S)} \left(p_e^k + \pi_e - q_e^k\right) \ge \hat{d}^k, \quad \forall k \in K, S \subset V : s^k \in S, t^k \notin S \tag{12b}$$

$$p, q, \pi \ \ge 0 \tag{12c}$$

with P_Γ^{repl} the convex hull of its feasible solutions. A minimum cut value of at least $\hat{d}^k$ for every $k \in K$ between source s^k and target t^k with respect to the edge weights $\left(\pi_e + p_e^k - q_e^k\right)$ is necessary and sufficient for the exisistence of a flow template y satisfying (1b). The REPLACE approach for solving (RND) using the Γ-model is based on solving this model. The exponential set of inequalities (12b) is handled implicitly by separation using a max-low-min-cut algorithm. Notice that the number of variables remains $O(|K||E|)$.

3 Valid Inequalities

In deterministic network design, cutset inequalities have been proven to be of particular importance [8, 23, 27, 46, 59]. This is true from the theoretical point of view as they define facets but also from the computational point of view as they are known and proven to improve on the performance of branch-and-cut based approaches to solve network design problems. In this section, we generalize the well-known class of cutset inequalities to robust network design.

To obtain strong inequalities based on network cuts we shrink the two shores of a network cut and study the convex hull of solutions to general two-node problems. In this respect we study a structure which is known as a cutset polyhedron, see [8, 59]. As by linear dependency we can save the flow conservation constraints for one of the two nodes, the structure is also referred to as a single node flow set in the literature, mainly in the context of bounded capacity variables and so-called flow-cover inequalities, see [7, 35, 36, 45].

The Γ-robust cutset polyhedron is introduced in Section 3.1. We will however further simplify and project the feasible region to the x and π space. The resulting two-variable set is studied in detail in Section 3.2 providing a complete description and all facet-defining inequalities.

These inequalities are then lifted in Section 3.3 to the space of all variables for the two-node problem and eventually to the original space which establishes facet-defining cutset inequalities for the original problem under certain conditions. These conditions are worked out following the line of projecting and lifting.

3.1 The Γ-Robust Cutset Polyhedron

We consider a proper and nonempty subset S of the nodes V and the corresponding cutset $\delta(S)$ and denote by $Q_S \subseteq K$ the subset of commodities with source s^k and target t^k not in the same shore of the cut. Since we may always reverse single demands without changing the model we may assume in this description $s^k \in S$ for all $k \in Q_S$. We denote by $\bar{d}_S := \sum_{k \in Q_S} \bar{d}^k$ the aggregated *nominal cut-demand* with respect to S. We will throughout assume that $|Q_S| \geq \Gamma \geq 1$. Notice that we can always reduce Γ to $|Q_S|$ without changing the problem on the cut. It follows $\bar{d}_S > 0$. Contracting both shores of the cut $\delta(S)$, we consider the following Γ-robust two-node formulation corresponding to (10):

$$\sum_{\{i,j\} \in \delta(S)} (y_{ij}^k - y_{ji}^k) \;=\; 1 \quad \forall k \in Q_S \tag{13a}$$

$$\sum_{\{i,j\} \in \delta(S)} (y_{ij}^k - y_{ji}^k) \;=\; 0 \quad \forall k \in K \setminus Q_S \tag{13b}$$

$$\Gamma \pi_e + \sum_{k \in K} \bar{d}^k y_e^k + \sum_{k \in K} p_e^k \;\leq\; x_e \quad \forall e \in \delta(S) \tag{13c}$$

$$-\pi_e + \hat{d}^k y_e^k - p_e^k \;\leq\; 0 \quad \forall e \in \delta(S), k \in K \tag{13d}$$

$$x, y, p, \pi \;\geq\; 0 \tag{13e}$$

We define the *robust cutset polyhedron* $P_\Gamma(S)$ with respect to S to be

$$P_\Gamma(S) := \operatorname{conv}\left\{(x, y, \pi, p) \in \mathbb{Z}_+^{|\delta(S)|} \times \mathbb{R}_+^{2|\delta(S)||K|} \times \mathbb{R}_+^{|\delta(S)|} \times \mathbb{R}_+^{|\delta(S)||K|} \mid (x, y, \pi, p) \text{ satisfies (13)}\right\}.$$

such that the following follows from Lemma 2.2 as $P_\Gamma(S)$ defines a two-node robust network design problem. We set $R_\Gamma(S) := \operatorname{proj}_{(x,\pi)}(P_\Gamma(S))$.

Corollary 3.1. *The dimension of $P_\Gamma(S)$ equals $2|\delta(S)| + 3|\delta(S)||K| - |K|$ whereas $R_\Gamma(S)$ is full-dimensional.*

In the sequel we will make use of the well-known mixed integer rounding (MIR) technique several times, see [52]. For some real number d we define $r(d) := d - (\lceil d \rceil - 1)$ as the fractional part of d with $r(d) = 1$ if d is integral.

Lemma 3.2 (Nemhauser and Wolsey [52])**.** *Consider the two variable mixed integer set defined by a single base inequality:*

$$Y = \{(x, \pi) \in \mathbb{Z}_+ \times \mathbb{R}_+ \,:\, cx + i\pi \geq d\}.$$

The inequality

$$rx + \max(0, i)\pi \geq r \left\lceil \tfrac{d}{c} \right\rceil$$

is valid for Y, where $r := cr(d/c)$.

Note that $cr(d/c)$ gives the remainder of the division of d by c if d/c is fractional. Otherwise, if d/c is integral then $r = c$. Given any vector v and a subset of indices I we abbreviate $v(I) := \sum_{e \in I} v_e$.

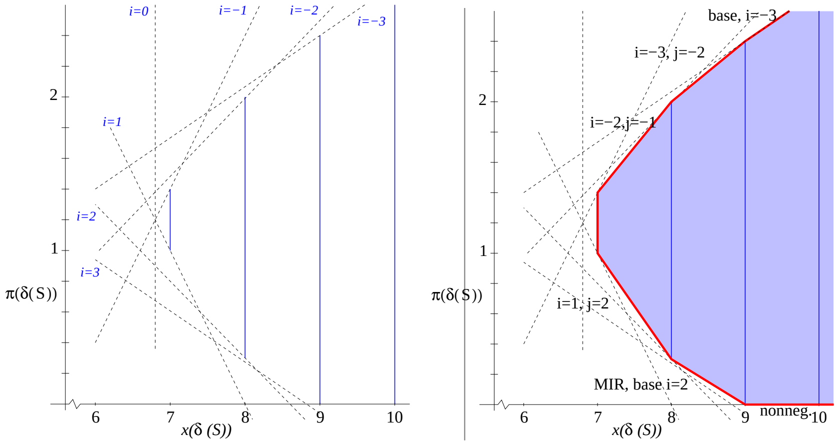
3.2 Robust Cutset Inequalities.

Independent of the realization of demand all cut commodities Q_S have to be realized across the cut $\delta(S)$, that is $y^k(\delta(S)) \geq 1$ for all $k \in Q_S$. It follows that we have to provide sufficient cut capacity $x(\delta(S))$ resulting in the following base cutset inequality to hold:

$$x(\delta(S)) \geq d_0 := \sum_{k \in Q_S} \bar{d}^k + \max_{\sigma_+ \in \mathcal{D}^\Gamma} \sum_{k \in Q_S} \sigma_+^k \hat{d}^k \tag{14}$$

It states that the capacity on the cut should be at least the nominal cut demand plus the Γ largest deviations among Q_S. Notice that the right hand side is independent of the realized flow. The value d_0 only depends on the cut $\delta(S)$ and the value of Γ. As the left hand side is integral we may round up the right hand side giving

$$x(\delta(S)) \geq \lceil d_0 \rceil \tag{15}$$



(a) base inequalities and feasible set (b) convex envelope

Figure 2: Example of X_Γ with $\Gamma = 3$, $|Q_S| = 6$, $\bar{d}_S = \frac{9}{5}$, $\hat{d} = (\frac{11}{5}, \frac{8}{5}, \frac{6}{5}, \frac{6}{5}, \frac{3}{5}, \frac{1}{5})$. The upper convex envelope inequalities (23) for $i = -3$, $j = -2$ ($2x - 5\pi \geq 6$) and for $i = -2$, $j = -1$ ($3x - 5\pi \geq 14$), the lower convex envelope inequalities (22) for $i = 1$, $j = 2$ ($7x + 10\pi \geq 59$), the MIR inequality (20) for $i = 0$ ($x \geq 7$), the MIR inequality (20) for $i = 2$ ($3x + 10\pi \geq 27$), the base inequality $x - \Gamma\pi \geq \frac{9}{5}$, and $\pi \geq 0$, completely describe the convex hull.

This already generalizes the classical cutset inequality for network design [46]. Since no dual variables π_e appear in this inequality, it is also valid for the exponential formulation (1). We use inequality (15) to tighten all three formulations during branch-and-cut. As we will prove in Corollary 3.16, inequality (15) defines a facet of $P_\Gamma(S)$ if $d_0 < \lceil d_0 \rceil$ and either $|\delta(S)| = 1$ or $d_0 > 1$. It also defines a facet of P_Γ if additionally the graphs defined by the two shores S and $V \setminus S$ are connected. In the rest of this section we will generalize this essential result to a more general class of inequalities in the space of the x and π variables.

Let us start by generalizing the base inequality (14). Let Q be an arbitrary but nonempty subset of the cut-commodities Q_S. From the flow-conservation constraints (13a) follows that

$$\sum_{k \in Q} \bar{d}^k y^k(\delta(S)) \geq \bar{d}(Q) \text{ and } \sum_{k \in Q} \hat{d}^k y^k(\delta(S)) \geq \hat{d}(Q). \tag{16}$$

Aggregating all capacity constraints (13c), adding all constraints (13d) for $e \in \delta(S)$ and $k \in Q$, using (16), and relaxing the backward flow variables results in

$$x(\delta(S)) + (|Q| - \Gamma)\pi(\delta(S)) \geq \bar{d}_S + \hat{d}(Q)\ . \tag{17}$$

The left hand side of (17) is not changing as long as the cardinality of the subset Q is constant. Hence among all subsets of Q with cardinality $|Q|$ the one maximizing $\hat{d}(Q)$ gives the strongest inequality (17). To state this inequality we have to sort the commodities nonincreasingly with respect to the maximum deviation $\hat{d}^k$ and define subsets of Q_S corresponding to large deviations. This needs some new notation. Let $\rho : Q_S \mapsto \{0, \ldots, |Q_S|\}$ be a permutation of the commodities in Q_S such that $\hat{d}^{\rho^{-1}(1)} \geq \hat{d}^{\rho^{-1}(2)} \geq \ldots \geq \hat{d}^{\rho^{-1}(|Q_S|)}$ and let $J = \{-\Gamma, \ldots, |Q_S| - \Gamma\}$. Fixing the cut we define $Q_i := \{k \in Q_S : \rho(k) \leq i + \Gamma\}$ for $i \in J$ as the commodities corresponding to the $i + \Gamma$ largest $\hat{d}^k$ values with respect to Q_S. Hence the demand $d_i := \bar{d}_S + \hat{d}(Q_i)$ denotes the total nominal demand plus the $i + \Gamma$ largest peak demands across the cut. This definition is consistent with the definition of d_0 in (14) since $|Q_S| \geq \Gamma$ and hence $\bar{d}_S = \sum_{k \in Q_S} \bar{d}^k$ and $\hat{d}(Q_0) = \max_{\sigma_+ \in \mathcal{D}^\Gamma} \sum_{k \in Q_S} \sigma_+^k \hat{d}^k$.

Using this notation inequality (17) reduces to

$$x(\delta(S)) + i\pi(\delta(S)) \geq d_i. \tag{18}$$

It is valid for all $i \in J$ and by setting $i = 0$ we get inequality (15). In the sequel we consider the polyhedron

$$X_\Gamma(S) = \text{conv}\left\{(x,\pi) \in \mathbb{Z}_+^{|\delta(S)|} \times \mathbb{R}_+^{|\delta(S)|} \mid (x,\pi) \text{ satisfies (18) } \forall i \in J\right\}$$

Every valid inequality for $X_\Gamma(S)$ is also valid for the Γ-robust formulations (10) and (12). In the following we will completely describe $X_\Gamma(S)$ providing all facet-defining inequalities. Since all coefficients in (18) are identical for all edges in $\delta(S)$ it suffices to study the two-dimensional case with base inequalities

$$x + i\pi \geq d_i \tag{19}$$

and the polyhedron

$$X_\Gamma = \text{conv}\left\{(x,\pi) \in \mathbb{Z}_+ \times \mathbb{R}_+ \mid (x,\pi) \text{ satisfy (19) for all } i \in J\right\},$$

also see Figure 3.2.

Notice that $X_\Gamma(S)$ is obtained from X_Γ by copying variables and forcing non-negativity for the copied variables. It follows that every facet for X_Γ translates into a facet for $X_\Gamma(S)$ and vice versa except for the non-negativity constraints. In fact a complete description of X_Γ determines a complete description of $X_\Gamma(S)$ and vice versa.

Lemma 3.3. *Every facet-defining inequality $\alpha x + \beta\pi \geq \gamma$ for X_Γ with $\alpha, \beta, \gamma \in \mathbb{R}$ different from a non-negativity constraint translates into a facet-defining inequality $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$ for $X_\Gamma(S)$. All facets of $X_\Gamma(S)$ defined by inequalities different from non-negativity constraints are of the form $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$ for $\alpha, \beta, \gamma \in \mathbb{R}$ and correspond to a facet-defining inequality $\alpha x + \beta\pi \geq \gamma$ for X_Γ.*

Of course, we have $X_\Gamma = X_\Gamma(S)$ if and only if $|\delta(S)| = 1$.

In the following we will not distinguish facet-defining inequalities of X_Γ and $X_\Gamma(S)$ as long as different from non-negativity constraints.

Let us divide the index set J into the sets $J_- = \{-\Gamma, \ldots, -1\}$ and $J_+ := \{1, \ldots, |Q_S| - \Gamma\}$ such that $J = J_- \cup \{0\} \cup J_+$. Accordingly, the *upper envelope* of X_Γ corresponds to indices in J_- and the *lower envelope* of X_Γ corresponds to indices in J_+, see Figure 3.2. More precisely, the upper envelope is given by $X_\Gamma \cap \{\pi \geq d_0 - d_{-1}\}$ whereas we define the lower envelope of X_Γ as $X_\Gamma \cap \{\pi \leq d_1 - d_0\}$. Notice that the upper envelope is always non-empty. The lower envelope is non-empty if and only if $|Q_S| > \Gamma$ and $d_1 > d_0$.

We call valid inequalities for X_Γ trivial if they are non-negativity constraints or if they are of type (19). In the following we are only interested in non-trivial facets of X_Γ as these will translate to facets of P_Γ. Lower and upper envelope are similar in structure. The lower envelope, however, is cut by $\pi \geq 0$ which leads to one additional type of facet. We will see that besides the vertical facet $x \geq \lceil d_0 \rceil$ there are two classes of non-trivial inequalities describing the lower envelope and one class of non-trivial inequalities describing the upper envelope facets.

Setting $r_i := r(d_i)$ and apply mixed integer rounding to (19) yields

$$r_i x + \max(0, i)\pi \geq r_i \lceil d_i \rceil \tag{20}$$

valid for X_Γ, see Lemma 3.2. In particular, for $i = 0$ this inequality reduces to $x \geq \lceil d_0 \rceil$ which is (15). For $i \in J_-$, inequalities (20) are obvioulsy dominated by (15). For $i \in J_+$ inequality (20) connects the two points $(\lfloor d_i \rfloor, r_i/i)$ and $(\lceil d_i \rceil, 0)$ in case $r_i < 1$. If $r_i = 1$ inequality (20) reduces to the base inequality $x + i\pi \geq d_i$. We get

Lemma 3.4. *Inequality* (20) *defines a facet of X_Γ if $i = 0$ and $r_i < 1$.*

Proof. Consider $\epsilon > 0$ and the two affinely independent points $(\lceil d_0 \rceil, d_0 - d_{-1})$ and $(\lceil d_0 \rceil, d_0 - d_{-1} + \epsilon)$ which both satisfy (20) with equality. To see feasibility notice that $d_0 - d_{-1} \geq 0$ gives the Γ largest deviation demand among Q_S. Setting

$$x_e = \lceil d_0 \rceil, \pi_e = d_0 - d_{-1}, p_e^k = \max(\hat{d}^k - \pi_e, 0), \text{ and } y_{ij}^k = 1 \text{ for } k \in Q_S \tag{21}$$

for some edge $e = \{i, j\} \in \delta(S)$ gives a feasible point for $P_\Gamma(S)$ which has a slack of $1 - r_0$ in the capacity constraint (13c) since $\Gamma\pi_e + \sum_{k\in K} p_e^k \sum_{k\in K} y_e^k = d_0$. Hence $(\lceil d_0 \rceil, d_0 - d_{-1})$ is feasible for X_Γ. As $r_0 < 1$ it follows also that the second point $(\lceil d_0 \rceil, d_0 - d_{-1} + \epsilon)$ is feasible for X_Γ for $\epsilon < 1 - r_0$. □

Lemma 3.5. *Assume* $J_+ \neq \emptyset$ *and* $\lceil d_{|Q_S|-\Gamma} \rceil > \lceil d_0 \rceil$. *Set* $i = \arg\max(r_k/k : k \in J_+$ *with* $\lceil d_k \rceil = \lceil d_{|Q_S|-\Gamma} \rceil)$. *Inequality* (20) *defines a facet of* X_Γ *if* $r_i < 1$.

Proof. Let $\pi' := \max(r_k/k : k \in J_+)$. Since $\lfloor d_i \rfloor \geq \lceil d_0 \rceil$ the two points $(\lfloor d_i \rfloor, \pi')$ and $(\lceil d_i \rceil, 0)$ are feasible. They satisfy (20) with equality and are affinely independent. □

In general, the two inequalities from Lemma 3.4 and Lemma 3.5 do not suffice to provide a complete description of X_Γ. To get a complete description of the lower envelope of X_Γ we have to consider two arbitrary base inequalities $x + i\pi \geq d_i$ and $x + j\pi \geq d_j$ with $i, j \in J_+, i < j$. Its intersection has x-value

$$b_{i,j} := (jd_i - id_j)/(j - i).$$

Now we have to connect the two points $(\lfloor b_{i,j} \rfloor, (d_i - \lfloor b_{i,j} \rfloor)/i)$ and $(\lceil b_{i,j} \rceil, (d_j - \lceil b_{i,j} \rceil)/j)$ Let $r_{i,j} := (j-i)r(b_{i,j})$. Recall that $r_{i,j}$ defined this way is the remainder of the division of $jd_i - id_j$ by $(j - i)$ with $r_{i,j} = (j - i)$ in case $b_{i,j}$ is not fractional, see Lemma 3.2.

Lemma 3.6. *For* $i, j \in J_+$ *with* $i < j$*, the following inequality is valid for* X_Γ*:*

$$(i + r_{i,j})x + ij\pi \;\; \geq \;\; r_{i,j} \lceil b_{i,j} \rceil + id_j \tag{22}$$

Proof. We scale the two base inequalities with j and i, respectively:

$$jx + ji\pi \;\; \geq \;\; jd_i \quad \text{and} \quad ix + ij\pi \;\; \geq \;\; id_j.$$

Introducing the slack $s_j := ix + ij\pi - d_j \geq 0$ of the second constraint and combining the two inequalities gives

$$(j - i)x + s_j \;\; \geq \;\; jd_i - id_j,$$

Applying MIR and re-substituting results in (22). □

In a similar way we combine two base constraints for $i, j \in J_-$ to get valid inequalities for the upper envelope of X_Γ.

Lemma 3.7. *For* $i, j \in J_-$ *with* $i < j$*, the following inequality is valid for* X_Γ*:*

$$(-j + r_{i,j})x - ij\pi \;\; \geq \;\; r_{i,j} \lceil b_{i,j} \rceil - jd_i \tag{23}$$

Proof. We multiply the base constraints for i and j by $-j$ and $-i$, respectively:

$$-jx - ji\pi \;\; \geq \;\; -jd_i \quad \text{and} \quad -ix - ij\pi \;\; \geq \;\; -id_j.$$

Introducing the slack $s_i := -jx - ji\pi + jd_i \geq 0$ for the first constraint and combining gives

$$(j - i)x + s_i \;\; \geq \;\; jd_i - id_j,$$

Applying MIR and resubstituting results in (23). □

In case $b_{i,j}$ is fractional inequalities (22) resp. (23) defined above cut off the fractional intersection point $(b_{i,j}, \pi)$ with $\pi = (d_i - b_{i,j})/i$ of the two base inequalities (19) corresponding to i and j. Note that by construction of the demand values d_i it holds that $b_{i,i+1} \geq b_{i+1,i+2}$ for $0 > i \in J_-$ and $b_{i,i+1} \leq b_{i+1,i+2}$ for $0 < i \in J_+$. Also note that if $(j - i)$ divides $d_{i,j}$ then inequality (22) resp. (23) reduces to the base inequality for i resp. j. Of course not every pair (i, j) results in a facet. In fact, only linearly many of the inequalities (22) and (23) are non-redundant. Let us define the function

$$\pi(k, x) := \frac{d_k - x}{k} \qquad \text{for all } k \in J_- \cup J_+ \text{ and } x \in \mathbb{R}_+.$$

We now consider an arbitrary interval $[a, a + 1]$ with $a \in \mathbb{Z}, a \geq \lceil d_0 \rceil$ and easily determine the indices i, j that yield an inequality of (22) resp. (23) dominating all others of this type on the chosen interval by simply maximizing (resp. minimizing) the value $\pi(k, a)$ and $\pi(k, a + 1)$. Doing so for all relevant values of a we get all (non-trivial) facets of the lower resp. upper envelope:

Lemma 3.8. *Assume $J_+ \neq \emptyset$ and $\lceil d_{|Q_S|-\Gamma} \rceil > \lceil d_0 \rceil + 1$. For $a \in \mathbb{Z}$ with $\lceil d_0 \rceil \leq a \leq \lceil d_{|Q_S|-\Gamma} \rceil - 1$ let $i := \arg\max_{k \in J_+} \pi(k,a)$ and $j := \arg\max_{k \in J_+} \pi(k, a+1)$. If $i \neq j$, then inequality* (22) *defines a facet of X_Γ. If otherwise $i = j$, then the base inequality* (19) *defines a facet of X_Γ.*

Proof. If $i \neq j$ resp. $i = j$ then inequality (22) resp. (19) connects the two affinely independent points $(a, \pi(i,a))$ and $(a+1, \pi(j, a+1))$, that is, they satisfy inequality (22) resp. (19) at equality. To see feasibility of the first point check that for $k \in J_+$ it holds $a + k\pi(i,a) \geq k\pi(k,a) = d_k$ by definition of i. For $k \in J_-$ we have $a + k\pi(i,a) \geq d_0 + k(d_i - d_0)/i \geq d_0 + k(d_k - d_0)/k = d_k$ where the first inequality follows from $a \geq d_0$ and the second inequality follows from $k < 0 < i$ and the definition of the demands d_i. The difference $d_i - d_{i-1}$ is non-increasing with i. Feasibility of the second point can be shown in a similar way. □

Notice that for the lower envelope and $\lceil d_{|Q_S|-\Gamma} \rceil - 1 \leq x \leq \lceil d_{|Q_S|-\Gamma} \rceil$ we get a facet of type (20) by Lemma 3.2. For $x \geq \lceil d_{|Q_S|-\Gamma} \rceil$ we have $\pi \geq 0$ as a facet. These inequalities together completely describe the lower envelope. A complete description of the upper envelope of X_Γ is obtained with the following Lemma which is proved similar to the proof of Lemma 3.8.

Lemma 3.9. *For $a \in \mathbb{Z}$ with $a \geq \lceil d_0 \rceil$ let $i = \arg\min_{k \in J_-} \pi(k, a+1)$ and $j = \arg\min_{k \in J_-} \pi(k,a)$. If $i \neq j$, then inequality* (23) *defines a facet of X_Γ. If otherwise $i = j$ then the base inequality* (19) *defines a facet of X_Γ.*

Notice that for $x \geq \lceil b_{\Gamma-1,\Gamma} \rceil$ the base inequality (19) for $i = -\Gamma$ is the only facet. Also notice that the pairs $\{i, j\}$ in Lemma 3.8 resp. Lemma 3.9 are not unique. However, the resulting facet-defining inequalities are of course unique.

We have established different classes of facet-defining inequalities for X_Γ. It turns out that all these inequalities together with the trivial facets completely describe X_Γ. This essentially follows already from the above since we stated the dominant inequalities for all intervals $[a, a+1]$ with $a \geq \lceil d_0 \rceil$.

Completeness also follows from a result of Miller and Wolsey [49] who study a two-dimensional set (Model W) similar to X_Γ. Applying [49, Theorem 3] for the lower envelope resp. upper envelope (using an appropriate variable transformation) we get

Corollary 3.10.

$$X_\Gamma = \{(x,\pi) \in \mathbb{R} \times \mathbb{R} \mid (x,\pi) \text{ satisfies the constraints (19), (20), (22), (23), and } \pi \geq 0\}.$$

3.3 Lifting

We have provided a complete and non-redundant description of X_Γ and thus of $X_\Gamma(S)$. Next, we show how facets of $X_\Gamma(S)$ translate to facets of the cutset polyhedron $P_\Gamma(S)$ and the original network design polyhedron P_Γ. We also prove that the set X_Γ is identical to the projection of the cutset polyhedron P_Γ to the space of the x and π variables if the cut contains a single edge. Define

$$R_\Gamma(S) := \text{proj}_{(x,\pi)}(P_\Gamma(S)) := \{(x,\pi) \mid \exists y \text{ and } p \text{ such that } (x,y,\pi,p) \in P_\Gamma(S)\}$$

Lemma 3.11. *$R_\Gamma(S) \subseteq X_\Gamma(S)$. Moreover, $R_\Gamma(S) = X_\Gamma(S)$ if and only if $|\delta(S)| = 1$.*

Proof. For $(x,\pi) \in R_\Gamma(S)$ let $(x,y,\pi,p) \in P_\Gamma(S)$. Inequalities (18) are valid for $P_\Gamma(S)$ which gives $(x,\pi) \in X_\Gamma(S)$ and $R_\Gamma(S) \subseteq X_\Gamma(S)$.

Let $\delta(S) = \{e\}$ with $e = \{i,j\}$ for $i,j \in V$. Given $(x,\pi) \in X_\Gamma(S)$ we set $y^k_{ij} := 1, y^k_{ji} := 0$, and $p^k_e := \max(0, \hat{d}^k - \pi_e)$ for all $k \in Q_S$. Now (x,y,π,p) obviously satisfies (13a), (13d), and (13e). Moreover it holds that

$$\begin{aligned} \Gamma\pi_e + \sum_{k \in Q_S} \bar{d}^k y^k_e + \sum_{k \in Q_S} p^k_e &= \Gamma\pi_e + \sum_{k \in Q_S} \bar{d}^k + \sum_{k \in Q_S} \max(0, \hat{d}^k - \pi_e) \\ &= \Gamma\pi_e + \bar{d}_S + \hat{d}(Q_i) - i\pi_e \\ &\leq x_e \end{aligned}$$

for some $i \in \{0, \ldots, |Q_S|\}$ using the introduced ordering of demands and (18). It follows that (x, y, π, p) satisfies (13c) and hence $(x, \pi) \in R_\Gamma(S)$.

It remains to show that $R_\Gamma(S) \neq X_\Gamma(S)$ if $|\delta(S)| > 1$. Let $e_1, e_2 \in \delta(S)$. There is a point (x, y, π, p) in $P_\Gamma(S)$ with $x_{e_1}, \pi_{e_1} > 0$ and $x_{e_2}, \pi_{e_2} = 0$. We simply route all traffic on e_1 and set x_{e_1}, π_{e_1} large enough. For this point it holds $(x, \pi) \in X_\Gamma(S)$ as already shown. We modify this point by shifting the capacity from e_1 to e_2 but keeping the value π_{e_1} such that $x_{e_1} = 0$ and $\pi_{e_1} > 0$. This gives a vector $(x, \pi) \in X_\Gamma(S) \backslash R_\Gamma(S)$ since inequalities (18) are still satisfied but (13c) is violated for e_1. □

Notice that from Lemma 3.11 follows that any point (x, π) which is defined on a single edge, that is, there exists $e \in \delta(S)$ such that $x_f = \pi_f = 0$ for all $f \in \delta(S), f \neq e$, is valid for $X_\Gamma(S)$ if and only if it is valid for $R_\Gamma(S)$. We will use this fact several times below.

Lemma 3.12. *Every facet-defining inequality for $X_\Gamma(S)$ (different from a non-negativity constraint) defines a facet of $R_\Gamma(S)$.*

Proof. We can assume that the facet of $X_\Gamma(S)$ is defined by $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$ for $\alpha, \beta, \gamma \in \mathbb{R}$. Consider $2|\delta(S)|$ affinely independent points $(x^i, \pi^i) \in X_\Gamma(S)$ for $i = 1, \ldots, 2|\delta(S)|$ satisfying $\alpha x^i(\delta(S)) + \beta\pi^i(\delta(S)) = \gamma$. Given an arbitrary edge $f \in \delta(S)$ we construct a point $(\tilde{x}^i, \tilde{\pi}^i)$ for every $i = 1, \ldots, 2|\delta(S)|$ by shifting all entries to edge f, more precisely $\tilde{x}^i_f := \sum_{e\in\delta(S)} x^i_e$ and $\tilde{\pi}^i_f := \sum_{e\in\delta(S)} \pi^i_e$. All other entries are set to zero: $\tilde{x}^i_e := \tilde{\pi}^i_e := 0$ for all$e \in \delta(S)\backslash\{f\}$. The points $(\tilde{x}^i, \tilde{\pi}^i)$ are valid for $X_\Gamma(S)$ and they satisfy $\alpha\tilde{x}^i(\delta(S)) + \beta\tilde{\pi}^i(\delta(S)) = \gamma$. Moreover, since $(\tilde{x}^i, \tilde{\pi}^i)$ is defined on a single edge it holds $(\tilde{x}^i, \tilde{\pi}^i) \in R_\Gamma(S)$. Notice that $(\tilde{x}^i, \tilde{\pi}^i) \neq 0$ as there is at least one cut demand. There must exist at least two affinely independent points among $(\tilde{x}^i, \tilde{\pi}^i)$, otherwise the points (x^i, π^i) cannot be affinely independent. Assume these points are $(\tilde{x}^1, \tilde{\pi}^1)$ and $(\tilde{x}^2, \tilde{\pi}^2)$. The proof is complete for $|\delta(S)| = 1$. In case $|\delta(S)| > 1$ we can assume that either $\tilde{x}^1_f > 0$ or $\tilde{x}^2_f > 0$, and similarly either $\tilde{\pi}^1_f > 0$ or $\tilde{\pi}^2_f > 0$. Otherwise the original points (x^i, π^i) are all contained in the face defined by $x(\delta(S)) \geq 0$ resp. $\pi(\delta(S)) \geq 0$ which is a contradiction as the sum of non-negativity constraints cannot define a facet. Now we vary $f \in \delta(S)$ which gives $2|\delta(S)|$ affinely independent points, both in $R_\Gamma(S)$ and on the face defined by $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$. □

Lemma 3.13. *Let $\alpha, \beta, \gamma \in \mathbb{R}$. If $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$ defines a facet for $R_\Gamma(S)$ then it also defines a facet for $X_\Gamma(S)$.*

Proof. It holds $R_\Gamma(S) \subseteq X_\Gamma(S)$. Since $R_\Gamma(S)$ is full-dimensional we only have to show that $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$ is valid for $X_\Gamma(S)$. Assume the contrary. We take a point in $X_\Gamma(S)$ which violates $\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma$. Now we modify this point by shifting everything to one edge. The constructed point is also valid for $R_\Gamma(S)$ as shown above but violates the facet-defining inequality which is a contradiction. □

We call facet-defining inequalities for $R_\Gamma(S)$ non-trivial if they are non-trivial for $X_\Gamma(S)$, that is, they are different from non-negativity constraints and different from (18).

Theorem 3.14. *Every non-trivial facet-defining inequality*

$$\alpha x(\delta(S)) + \beta\pi(\delta(S)) \geq \gamma \tag{24}$$

for $R_\Gamma(S)$ also defines a facet of $P_\Gamma(S)$ if one of the following conditions hold

- $|\delta(S)| = 1$ *and there exists a feasible point (x, y, π, p) on the face of $P_\Gamma(S)$ defined by* (24) *such that the edge capacity constraint* (10a) *is not tight.*
- $|\delta(S)| \geq 2$ *and there exists a feasible point (x, y, π, p) on the face of $P_\Gamma(S)$ defined by* (24) *such that the capacity constraint* (10a) *is not tight for at least two different edges.*

Proof. We assume that (24) does not define a facet for $P_\Gamma(S)$. Hence every point $(x, y, \pi, p) \in P_\Gamma(S)$ satisfying (24) at equality must be contained in a facet of $P_\Gamma(S)$ defined by

$$\sum_{e\in\delta(S)} \alpha_e x_e + \sum_{e\in\delta(S)} \beta_e \pi_e + \sum_{e\in\delta(S)} \sum_{k\in K} \delta^k_e p^k_e + \sum_{e=\{i,j\}\in\delta(S)} \sum_{k\in K} (\mu^k_{ij} y^k_{ij} + \mu^k_{ji} y^k_{ji}) \geq \gamma \tag{25}$$

By adding flow conservation constraints to (25) we conclude that $\mu_{ij}^k = 0$ for an arbitrary edge $e = ij \in \delta(S)$ and all $k \in K$. We may hence assume that for the same edge the capacity constraint is not tight for the point (x, y, π, p) on the face of $P_\Gamma(S)$ defined by (24). By increasing p_e^k we see that $\delta_e^k = 0$ for all $k \in K$. Similarly, sending a small circulation flow on e, we conclude $\mu_{ji}^k = 0$. Notice that by these perturbations we never leave the face.

Now assume that $|\delta(S)| \geq 2$. There is a second edge $e' \neq e$ such that the corresponding capacity constraint is not tight. Since we may exchange variable values of two different edges without leaving the face (24) edge $e' \neq e$ is in fact arbitrary. By sending circulation flow using edges e and e' and by increasing $p_{e'}^k$ it turns out that $\delta_e^k = \mu_{ij}^k = \mu_{ji}^k = 0$ for all edges $e \in \delta(S)$ and commodities $k \in K$.

Since (24) defines a facet of $R_\Gamma(S)$, $2|\delta(S)|$ affinely independent points exist. These points can be lifted to points in $P_\Gamma(S)$ remaining affinely independent in the (x, π) space and satisfying (24) as well as (25) at equality. We showed that only the $2|\delta(S)|$ coefficients in (25) corresponding to the x and π variables are nonzero. Hence (25) is (24) up to scaling and up to a linear combination of flow conservation constraints. It follows that (24) defines a facet of $P_\Gamma(S)$. □

We call a valid inequality for $P_\Gamma(S)$ non-trivial if it is different from the constraints (13a)-(13e) defining $P_\Gamma(S)$. The following result is a straight-forward generalization of the corresponding result for the deterministic case from [59], also see [2].

Lemma 3.15. *Every non-trivial facet-defining inequality of $P_\Gamma(S)$ defines a facet of P_Γ if both cut shores are connected.*

The proof of Lemma 3.15 is based on the fact that in case both shores are connected, then the flow for commodities in $K \setminus Q_S$ can be routed in the two shores without using cut edges. This means that we can construct feasible points for P_Γ from points valid for $P_\Gamma(S)$ without changing the cut values. This is done by assigning sufficiently large values for x_e, π_e, and p_e^k for edges $e \in E \setminus \delta(S)$ and then decomposing the problem with respect to the two graphs defined by S respectively $V \setminus S$.

Corollary 3.16. *Given a node set $S \subset V$ such that the two shores of the corresponding cut $\delta(S)$ are connected, the cutset inequality* (15) *defines a facet of P_Γ if $r(d_0) < 1$ and either $|\delta(S)| = 1$ or $d_0 > 1$.*

Proof. By Lemma 3.4 and Lemma 3.12 inequality (15) defines a facet of $R_\Gamma(S)$. In this case inequality (15) is also non-trivial for $R_\Gamma(S)$. Fixing edge $e = \{i, j\} \in \delta(S)$ we consider the point (x, y, π, p) on the face of $P_\Gamma(S)$ defined by (15) as defined in (21). All other variables are set to zero. Recall that the capacity constraint of e has a slack of $1 - r(d_0)$. In case $|\delta(S)| \geq 2$ and $d_0 > 1$ and hence $\lceil d_0 \rceil \geq 2$ we can shift one unit of capacity to a second edge. Also a fraction of $1/\lceil d_0 \rceil$ of all other variables is shifted to the second edge. This way we construct a point on the face with two edges not being tight in the capacity constraint. Hence, using Theorem 3.14 and Lemma 3.15 we get the desired result. □

Corollary 3.17. *Given a node set $S \subset V$ such that the two shores of the corresponding cut $\delta(S)$ are connected and $J_+ \neq \emptyset$ as well as $\lceil d_{|Q_S|-\Gamma} \rceil > \lceil d_0 \rceil$. The MIR inequality*

$$r_i x(\delta(S)) + \max(0, i)\pi(\delta(S)) \geq r_i \lceil d_i \rceil \tag{26}$$

defines a facet of P_Γ if $i = \arg\max(r_k/k \,:\, k \in J_+$ with $\lceil d_k \rceil = \lceil d_{|Q_S|-\Gamma} \rceil)$ and $r_i < 1$.

Proof. By Lemma 3.5 and Lemma 3.12 inequality (26) defines a facet of $R_\Gamma(S)$ if $r_i < 1$. In this case inequality (26) is also non-trivial for $R_\Gamma(S)$. Fixing $e \in \delta(S)$ we consider the following point (x, y, π, p) on the face of $P_\Gamma(S)$ defined by (26)

$$x_e = \lceil d_i \rceil = \lceil d_{|Q_S|-\Gamma} \rceil, \pi_e = 0, p_e^k = \hat{d}^k, \text{ and } y_{ij}^k = 1 \text{ for } k \in Q_S.$$

All other variables are set to zero. The point protects against all demands across the cut at their peak. There is a slack of at least $1 - r_i$. It holds $\lceil d_{|Q_S|-\Gamma} \rceil \geq 2$ since $\lceil d_0 \rceil \geq 1$. If $|\delta(S)| \geq 2$ we can hence shift one unit of capacity and a fraction of all other variables to a second edge such that two edges are non-tight in the capacity constraint. Hence, using Theorem 3.14 and Lemma 3.15 we get the desired result. □

Corollary 3.18. *Given a node set $S \subset V$ such that the two shores of the corresponding cut $\delta(S)$ are connected, the upper envelope inequality*

$$(-j + r_{i,j})x(\delta(S)) - ij\pi(\delta(S)) \;\geq\; r_{i,j} \lceil b_{i,j} \rceil - jd_i \tag{27}$$

defines a facet of P_Γ if $i, j \in J_-$, $i < j$, such that $i = \arg\min_{k \in J_-} \pi(k, a{+}1)$ and $j = \arg\min_{k \in J_-} \pi(k, a)$ with $r_{i,j} < 1$ and $a \in \mathbb{Z}$ with $a \geq \lceil d_0 \rceil$ having either $|\delta(S)| = 1$ or $a \geq 2$.

Proof. By Lemma 3.9 and Lemma 3.12 inequality (27) defines a facet of $R_\Gamma(S)$. From $r_{i,j} < 1$ follows $i < j$ and the break point $b_{i,j}$ is fractional and hence (27) is non-trivial for $R_\Gamma(S)$. Let F be the face of $P_\Gamma(S)$ defined by (27). There is a point $(\bar{x}, \bar{\pi})$ with $a < \bar{x} < a + 1$ in the linear relaxation of X_Γ cut off by (27). Using a single edge e we may of course lift this point to a valid point $(\bar{x}, \bar{y}, \bar{\pi}, \bar{p})$ of the linear relaxation of $P_\Gamma(S)$. Set $\alpha = (-j + r_{i,j})$, $\beta = -ij$, and $\gamma = r_{i,j} \lceil b_{i,j} \rceil - jd_i$. The point $(\dot{x}, \bar{\pi})$ with $\dot{x} = \frac{\gamma - \beta\bar{\pi}}{\alpha} > \bar{x}$ is in X_Γ and lies on the facet. Moreover $p_1 := (\dot{x}, \bar{y}, \bar{\pi}, \bar{p}) \in F$ such that for the selected single edge e the capacity constraint is not tight. However p_1 is not feasible as $\dot{x}$ with $a < \dot{x} < a + 1$ is not integral. Consider the two points $(a, \pi(j, a))$ and $(a + 1, \pi(i, a + 1))$ on the facet of X_Γ and denote by p_2 and p_3 the two corresponding points lifted to $P_\Gamma(S)$ on the face F. We can assume that p_2 and p_3 have nonzero values only on edge e and that p_1 is a convex combination of p_2 and p_3. Hence at least one of p_2 or p_3 is not tight in the capacity constraint of e. The proof is complete in case $|\delta(S)| = 1$. Assume $|\delta(S)| \geq 2$. By shifting one unit of capacity to a second edge e_2 we construct points p_4 and p_5 from p_2 and p_3 similar to the proof of Corollary 3.16. As long as $a \geq 2$ at least one these point is not tight in the capacity constraint of at least two edges. By Theorem 3.14 and Lemma 3.15 the claim follows. □

Corollary 3.19. *Let $S \subset E$ be a node set such that the two shores of the corresponding cut $\delta(S)$ are connected and $J_+ \neq \emptyset$ as well as $\lceil d_{|Q_S| - \Gamma} \rceil > \lceil d_0 \rceil + 1$. The lower envelope inequality*

$$(i + r_{i,j})x(\delta(S)) + ij\pi(\delta(S)) \;\geq\; r_{i,j} \lceil b_{i,j} \rceil + id_j \tag{28}$$

defines a facet of P_Γ if $i, j \in J_+$, $i < j$, such that $i := \arg\max_{k \in J_+} \pi(k, a)$ and $j := \arg\max_{k \in J_+} \pi(k, a{+}1)$ with $r_{i,j} < 1$ and $a \in \mathbb{Z}$ with $\lceil d_0 \rceil \leq a \leq \lceil d_{|Q_S| - \Gamma} \rceil - 1$ having either $|\delta(S)| = 1$ or $a \geq 2$.

Proof. Similar to the proof of Corollary 3.18. □

4 Computations

In this section we present the results of three major computational studies. First, we compare the three presented approaches SEPARATE, DUALIZE, and REPLACE by evaluating the computational performance of the corresponding models (7), (10), and (12). Second, we investigate the impact of separating violated inequalities (15), (26)–(28) on the solving process. Finally, we evaluate the costs and realized robustness of the optimal robust network designs for real-life traffic measurements.

Instances. We consider problem instances based on live traffic data from different sources: the U.S. Internet2 Network (ABILENE) [1], the pan-European research backbone network GÉANT, and the national research backbone network operated by the German DFN-Verein [29] mapped on the network (GERMANY17) defined by the NOBEL project [53], and in addition mapped on a larger network (GERMANY50) [56]. For each network the live traffic data is given as a set of measured traffic matrices with a granularity of 5 minutes (ABILENE, GERMANY17, GERMANY50) or 15 minutes (GÉANT). Recently, the live traffic measurements of these networks have also become available in the SNDlib [56]. For ABILENE resp. GÉANT we consider two time periods of one week resulting in two instances ABILENE1 and ABILENE2 resp. GEANT1 and GEANT2. For GERMANY17 and GERMANY50 we consider one day each. In Section 4.3 we evaluate the realized robustness of optimal robust network designs. Therefore, we use additional traffic measurements in the evaluation to simulate uncertain future traffic. In total four weeks of traffic measurements are used for each ABILENE and GÉANT instance. Table 1 summarizes the network and traffic properties of all considered data sets.

Network	ABILENE	GÉANT	GERMANY17	GERMANY50
# nodes	12	22	17	50
# links	15	36	26	89
# demands	66	231	136	1044
available traffic period	6 months	4 months	1 day	1 day
traffic granularity	5 min	15 min	5 min	5 min
# available traffic matrices	48 095	10 737	288	288
# traffic matrices used	$2 \times 8\,064$	$2 \times 2\,688$	288	288
instances	ABILENE1 ABILENE2	GEANT1 GEANT2	GERMANY17	GERMANY50

Table 1: Network and traffic properties of considered data sets

For each data set let T denote the considered time period and let $d^k_{(t)}$ be the demand for commodity $k \in K$ at time step $t \in T$. In a first step we scale the traffic data in such a way that sum of all peak demands $\max_{t \in T}(d^k_{(t)})$ over all commodities $k \in K$ amounts to 1 Tbps. To determine the nominal value $\bar{d}^k$ respectively peak value $\bar{d}^k + \hat{d}^k$ we calculated the arithmetic mean and 95%-percentile of each demand $k \in K$ using the scaled measurements. That is, we set $\bar{d}^k := 1/|T| \sum_{t \in T} d^k_{(t)}$ and $\bar{d}^k + \hat{d}^k$ corresponds to the largest deviation from the nominal value in period T'_k where T'_k is obtained from T by removing the 5% largest demands. The link capacity module size, that is one unit of capacity, is set to 40 Gbps.

General settings. We implemented formulations (7), (10), and (12) of the Γ-robust network design problem in C++ using IBM ILOG CPLEX 12.1 [39] as branch-and-cut framework. We applied the CONCERT framework of CPLEX and callbacks to implement the separation methods. The computations were carried out single-threaded on a Linux machine with 2.93 GHz Intel Xeon W3540 CPU and 12 GB RAM. A time limit of 12 hours was set for solving each problem instance. All other solver settings were left at their defaults if not stated differently.

4.1 Model comparison

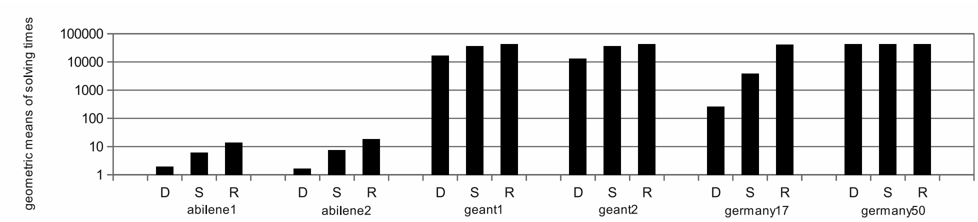
In our first computational study, we compare the models (7), (10), and (12) to evaluate their computational performance. Each of these models follows a different approach: SEPARATE (7), DUALIZE (10), and REPLACE (12). For all networks, $\Gamma \in \{0, 1, \ldots, 10\}$ is considered. This yields 66 realistic test instances.

We complement this study by an extensive evaluation of higher values of Γ for the small ABILENE network: for ABILENE1 and ABILENE2 we consider all possible values of Γ, i. e., values up to the number of commodities ($\Gamma = |K| = 66$).

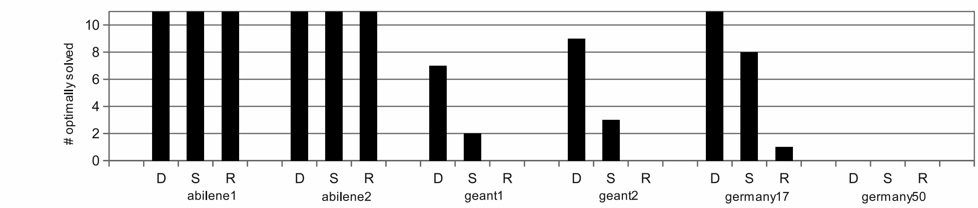
For SEPARATE and REPLACE, the exponential many inequalities (7a) resp. (12b) are treated implicitly: violated ones are separated as so-called lazy constraints during the solving process. Figure 3 visualizes the computational behavior of the three models: First, for each network a comparison of the geometric mean of the solving times for $\Gamma = 0, 1, \ldots, 10$ is shown in Figure 3(a). Second, for each network the number of optimally solved instances out of 11 is shown in Figure 3(b). Third, the solving times for ABILENE1 and ABILENE2 with $\Gamma \in \{0, 1, \ldots, 66\}$ are shown in Figure 3(c) resp. Figure 3(d). Note, the logarithmic scales of the solving time axes in Figure 3(a), 3(c), and 3(d).

Impact on the solving time. First, we investigate the results for $\Gamma = 0, 1, \ldots, 10$: In our computational study, the model following the DUALIZE approach has been the fastest for 64 of 66 test instances. Only for ABILENE1 with $\Gamma = 0$ and ABILENE2 with $\Gamma = 1$ the solving time of the SEPARATE model has been slightly faster by less than a second. The second fastest model (for $\Gamma \leq 10$) in our study is the SEPARATE model approach.

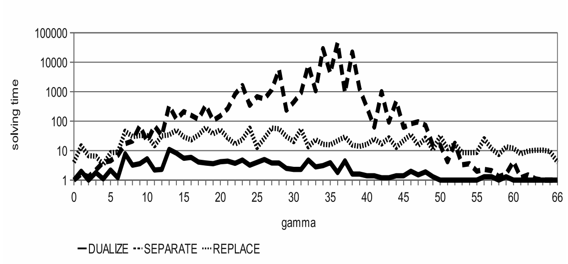
The geometric means of its solving times range from 2.1 (GEANT1) to 14.5 (GERMANY17) times the geometric means obtained by DUALIZE. The model following REPLACE is the slowest (for $\Gamma \leq 10$)



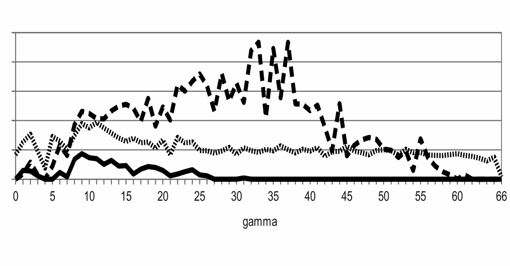
(a) Geometric means of solving times (for $\Gamma \in \{0, 1, \ldots, 10\}$)



(b) Number of optimally solved instances (for $\Gamma \in \{0, 1, \ldots, 10\}$)



(c) Solving times of ABILENE1 and $\Gamma \in \{0, 1, \ldots, 66\}$



(d) Solving times of ABILENE2 and $\Gamma \in \{0, 1, \ldots, 66\}$

Figure 3: Comparison of models for DUALIZE (D), SEPARATE (S), and REPLACE (R) approaches

with geometric means from 2.6 (GEANT1) up to 157 (GERMANY17) times the corresponding one for DUALIZE. For both, SEPARATE and REPLACE, very often the time limit has been reached. In particular, GERMANY50 could not been solved in any case within the time limit of 12 hours. For GERMANY50, the optimality gap has been in the range from 51% to 71% (DUALIZE), 43% to 59% (SEPARATE), and 420% to 736% (REPLACE). For this large instance, the SEPARATE approach yields the smallest optimality gaps.

It turns that out the MIP solver obtains many solutions feasible for the incomplete formulation of the SEPARATE and REPLACE models which are infeasible to the complete problem. Thus, these solutions are separated by additional model inequalities (lazy constraints). In our studies, the amount of these non-redundant cuts slows down the solving processes of the SEPARATE and REPLACE models significantly. These results are in contrast to those in [33] for the set cover problem.

Second, we have a closer look at the supplementary results of our study for ABILENE1 and ABILENE2 with $\Gamma = 0, 1, \ldots, 66$ as visualized in Figure 3(c) and 3(d). It turns out that we cannot generalize our previous evaluation of the geometric means to $\Gamma > 10$: We observe that SEPARATE is faster than REPLACE only for small values Γ or $|K| - \Gamma$. For example, for ABILENE1 and $10 \leq \Gamma \leq 49$ the REPLACE model is up to 1000 times faster than the SEPARATE model (e. g. $\Gamma = 36$). This behavior can be explained easily: The size of the (resulting) formulation for SEPARATE strongly depends on the size of Γ in contrast to DUALIZE and REPLACE. The number of capacity constraints (7a) to be considered for SEPARATE is proportional to the number of vertices of $\mathcal{D}^\Gamma$ which is precisely $\binom{|K|}{\Gamma} + 1$ for $0 < \Gamma < |K|$. That is, both for Γ close to 0 and for Γ close to $|K|$ the SEPARATE formulation is small. Here the SEPARATE approach might even outperform DUALIZE. In all other cases the number of constraints to be considered for separation can be extremely large. Notice that computation times for SEPARATE in Figure 3(c) and 3(d) roughly follow the shape of $\binom{|K|}{\Gamma} + 1$ as a function of Γ.

Impact on solvability. In the following, we consider the number of optimally solved instances within the time limit. First, we observe the ABILENE1 and ABILENE2 instances could be solved optimally for all three models. The instance GERMANY50 could not be solved optimally by any model. For all 33 remaining instances, the DUALIZE approach solved the most instances to optimality (88%), SEPARATE the second most (39%), and the REPLACE model the least (3%). Every instance solved by SEPARATE or REPLACE was also solved by DUALIZE.

Conclusion. The model following the DUALIZE approach outperforms the other two models: more instances can be solved to optimality in less time independent of Γ. Therefore, we focus on the DUALIZE model in the following studies. However, we want to remark that additional special purpose primal heuristics might improve on the computational behavior of SEPARATE at least for small values of Γ or $|K| - \Gamma$.

4.2 Valid inequalities

In our second computational study, we investigate the impact of separating valid inequalities on the overall solving process. Using the callback functionality of CPLEX, we added a separator for the inequalities (15), (26), (27), and (28). The separator is called at the root node of the branch-and-cut tree. Within the separator we implemented three separation algorithms variants: one heuristic (Shrinking) and two exact algorithms (Enumeration, ILP) to separate violated inequalities. The three separation algorithms are described in the following.

Exact separation (Enumeration). To study the effectiveness of the valid inequalities we implemented an exact separation algorithm which enumerates all network cuts explicitly and generates *all* violated inequalities of type (15), (26), (27), and (28). Clearly, this approach is suited for small networks only as the number of network cuts that must be enumerated increases exponentially. Still, for small networks this enumerative algorithm can be used to investigate the maximal effectiveness of these inequalities that can be achieved by separating all existing violated inequalities in terms of improving the root node dual bound.

Exact separation (ILP). Further, we implemented another exact separation algorithm which solves an integer linear program to separate a most violated inequality of type (15). It was introduced in Koster et al. [43] and is computational tractable for larger networks as well.

We define binary variables β_i $(i \in V)$ with $\beta_i = 1$ if and only if $i \in S$ determining the cut, α^k with $\alpha^k = 1$ if and only if $k \in Q_S$ determining the cut-crossing commodities, γ^k with $\gamma^k = 1$ if and only if commodity $k \in Q_S$ deviates from its nominal, and $\bar{\delta}_{ij}$ $(ij \in E)$ with $\bar{\delta}_{ij} = 1$ if and only if $ij \in \delta(S)$ determining the cutset. In addition, let d determine the worst-case total demand value crossing the cut, and let R be the right-hand side value of the corresponding cutset inequality (15). Given an LP solution x^*, we minimize the feasibility (i. e., maximize the violation) of inequality (15) such that a negative objective value yields a violated cut. Then, the ILP formulation of the separation problem reads

$$\min \sum_{ij \in E} x^*_{ij} \bar{\delta}_{ij} - R$$

$$s.t.\ \max\{\delta_i - \delta_j, \delta_j - \delta_i\} \leq \bar{\delta}_{ij} \leq \min\{\delta_i + \delta_j, 2 - \delta_i - \delta_j\} \quad \forall ij \in E \tag{29a}$$

$$\max\{\delta_{s^k} - \delta_{t^k}, \delta_{t^k} - \delta_{s^k}\} \leq \alpha^k \leq \min\{\delta_{s^k} + \delta_{t^k}, 2 - \delta_{s^k} - \delta_{t^k}\} \quad \forall k \in Q \tag{29b}$$

$$\gamma^k \leq \alpha^k \quad \forall k \in Q \tag{29c}$$

$$\sum_{k \in Q} \gamma^k \leq \Gamma \tag{29d}$$

$$\sum_{k \in Q} (\bar{d}^k \alpha^k + \hat{d}^k \gamma^k) = d \tag{29e}$$

$$d \leq R \leq d + 1 - \varepsilon \tag{29f}$$

$$\alpha^k, \delta_i, \gamma^k, \bar{\delta}_{ij} \in \{0,1\}, R \in \mathbb{Z}_+, d \geq 0 \quad \forall k \in Q, \forall ij \in E, \forall i \in V \tag{29g}$$

where constraints (29a), (29b), and (29c) define the logical dependencies between the indicator variables α^k, β_i, and γ^k. Constraint (29d) limits the number of deviating commodities by Γ. The total demand

	robust cutset inequalities (15)			envelope inequalities (26)–(28)	
	enumeration (ex)	ILP (ex)	shrinking (h)	enumeration (ex)	shrinking (h)
strategy (i)	·	·	·	·	·
strategy (ii)	✓	·	·	·	·
strategy (iii)	✓	·	·	✓	·
strategy (iv)	·	·	✓	·	✓
strategy (v)	·	✓	✓	·	✓

Table 2: Considered strategies with exact (ex) and heuristic (h) separation algorithms

d is calculated by (29e). Constraint (29f) guarantees the round-up of the right-hand side variable R using $0 < \varepsilon \ll 1$ to avoid rounding R to $\lceil d \rceil + 1$ or higher. Note, no time limit is set for solving the ILP formulation of the separation problem.

Heuristic separation (Shrinking). Complementing the exact separation algorithms, we propose the following heuristic separation algorithm: Violated inequalities are separated for all single node network cuts (i. e., $\delta(S)$ with $|\delta(S)| = 1$) as well as a set of network cuts resulting from a graph shrinking heuristic. This graph shrinking heuristic generalizes a shrinking heuristic dating back to [24, 37] and used by [59] for the deterministic model (1) Γ-robust network design.

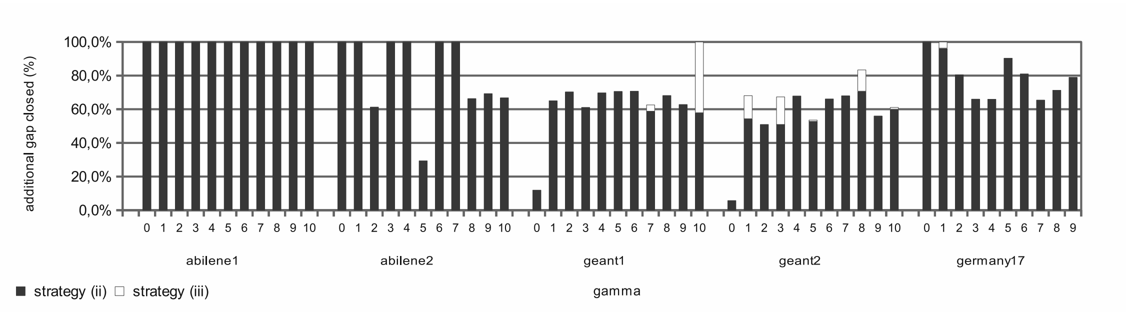
The idea of this extended graph shrinking heuristic is the following: The base inequality of the robust cutset inequality (15) is the sum of flow conservation constraints (1b), capacity constraints (10a), and constraints (10b). For violated cutset inequalities we need (almost) tight base inequalities. Hence we wish to have edges e in the cut $\delta(S)$ that have (almost) no slack in the constraints (10a) and (10b). In the shrinking heuristic we hence shrink edges whose corresponding model constraints have large slacks. Technically, we try to minimize the sum of weights w_e for edges e on the cut: Given the solution of the current LP relaxation, we use $w_e := s_e^{(10a)} + \sum_{k \in K} s_{e,k}^{(10b)}$ where $s_e^{(10a)}$ denotes the slack of the capacity constraint (10a) for edge e and the $s_{e,k}^{(10b)}$ the slack of constraint (10b) for edge e and commodity k. By contracting edges in non-increasing order of $w_e$we shrink the network until only η nodes or no edges with positive weight are left. Based on empirical values of previous computational studies we set $\eta = 5$.

Let $\mathcal{N}(\mathcal{V}, \mathcal{E})$ be the remaining shrunken network with node set $\mathcal{V}$ and edge set $\mathcal{E}$. Then, the set of network cuts returned by the shrinking heuristic consists of all network cuts corresponding to single node network cuts in $\mathcal{N}$ as well as further up to $|\mathcal{V}|^2$ network cuts in $\mathcal{N}$ obtained by enumeration.

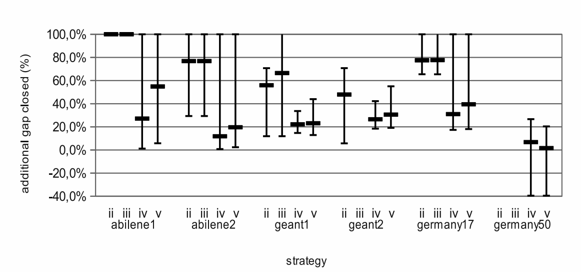
Strategies. We compare five different strategies to solve the Γ-Robust Network Design Problem: (i) solving the compact model (10) with default CPLEX and four further variants (ii)-(v) including the separation of violated inequalities in the root node using CPLEX and separation callbacks. The different strategies and separation algorithms used are summarized in Table 2 referring to the exact (Enumeration, ILP) and heuristic separation algorithm (Shrinking) as introduced above. Note, in strategy (v) the separation problem is solved exactly only after no violated cut has been found by the heuristic.

Impact on the root gap. First, only the root node of the cut-and-branch tree is solved to investigate the effectiveness of separating violated inequalities (15), (26)–(28). We evaluate the additional gap closed in each strategy, i. e., the ratio $(\text{DB} - \text{DB}_{\text{CPLEX}})/(\text{PB} - \text{DB}_{\text{CPLEX}})$ where DB denotes the dual bound of the corresponding strategy after the root node before branching, DB_{CPLEX} the corresponding value of default CPLEX (strategy (i)), and PB the (overall) best known primal bound. Thus, an additional gap closed of 100% means the instance could be solved to optimality in the root node. Note, none of the instances could be solved to optimality in the root node by CPLEX alone (strategy (i)).

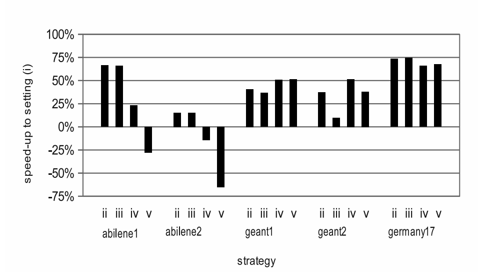
A detailed view on the additional gap closed in the root node for strategies (ii) and (iii) is given in Figure 4(a). Results for GERMANY50 are not shown as strategy (ii) and (iii) follow the enumerative exact separation approach which is not applicable to the GERMANY50 instance because of its size. For ABILENE1 and all values of Γ, we observe that the remaining optimality gap can be closed completely in



(a) additional gap closed in root for strategies (ii) and (iii)



(b) min, max, and geom. mean of additional gap closed in root



(c) speed-up compared to CPLEX (strategy (i))

Figure 4: Effect of separating violated inequalities inequalities (15), (26)–(28)

the root node by separating the considered classes of inequalities. Considering ABILENE2, this is also the case for most values of Γ. In total, for more than 90% of all instances the optimality gap can be closed by at least 50%. For GEANT1 and GEANT2, the least additional gap closed is achieved.

Comparing strategies (ii) and (iii), we observe a difference for only 7 of 66 instances. In these cases, the gap could be closed by additionally 2.5% in the geometric mean in strategy (iii), ranging from 0.1% (GEANT1, $\Gamma = 1$) to 16.4% (GEANT2, $\Gamma = 3$). This is caused by the fact that mainly violated robust cutset inequalities (15) are found and almost no violated envelope inequalities or MIR cuts. We discuss this in the paragraph about the distribution of separated cuts.

Next, we compare all strategies but on a less detailed level: Figure 4(b) illustrates the geometric mean (thick horizontal bar), minimal (lower end of vertical line), and maximal values (upper end of vertical line) of the additional gap closed for each instance and all considered values of Γ. As strategies (iv) and (v) primary follow an heuristic approach, the achieved additional gap closed is less than for strategy (ii) or (iii). But still a reasonable additional gap closed of at least 20% in the geometric mean can be achieved for all instances except ABILENE2 with strategy (iv) and GERMANY50. Although due to memory limits GERMANY50 could not be processed in strategy (ii) and (iii), it has been computational tractable in strategy (iv) and (v). Here, the optimality gap left by CPLEX can be closed by 6.8% (1.64%) in the geometric mean, ranging from -39.5% (-39.5%) to 26.7% (20.4%) for strategy (iv) (resp. (v)) where a negative gap closed means that actually an optimality gap larger than in strategy (i) remained at the time limit (hence, the root node could not be solved in 12h).

Distribution of separated cuts. In the following, we have a closer look on the type of cuts that have been separated in the root node in strategy (iii). In strategy (iii) all existing violated inequalities (15), (26)–(28) are separated by enumeration. Table 3 shows the number of separated cuts for each instance, the value of Γ, and the class of the generated inequalities. For better readability, blanks are printed instead of zeros and completely blank columns are omitted. Column RC states the number of separated robust cutset inequalities (15), column L the lower envelope (28), and M the MIR inequalities (26). No violated upper envelope inequalities (27) have been separated in our computational studies. Hence these columns are missing.

	ABILENE1		ABILENE2		GEANT1			GEANT2			GERMANY17		
gamma	RC	M	RC	M	RC	L	M	RC	L	M	RC	L	M
0	45		46		49910			63732			2664		
1	36		25		6668		5	7408		2	726	5	1
2	28		16		3606		7	3896	1	2	414	1	1
3	29		21		2946		3	3837		12	353		
4	24		23		2428	1	7	3355		2	308		
5	30		14		2697		4	3738		4	314		
6	23		13		2345		4	4462		1	324		1
7	20		8		2351		4	4877		8	296		1
8	19		18		2530		4	5224		27	335		2
9	16		20	2	2367		1	3424		6	335		1
10	21	1	19		2154		4	5255		14	386		

Table 3: Cut distribution. Robust cutset (RC), Lower envelope (L), and MIR (M)

We observe that mostly violated robust cutset inequalities are separated. In particular for $\Gamma = 0$, where the robust cutset inequalities corresponds to the well-known and effective classical cutset inequalities, many violated cuts are found. From $\Gamma = 0$ to $\Gamma = 2$, we observe a significant decrease in the number of separated robust cutset inequalities which we cannot explain. For $\Gamma > 2$, the number of separated robust cutset inequalities does not fluctuate that much. No violated upper envelope inequalities (27) have been found which might be explained as follows: the capacity constraints (10b) act as (variable) upper bounds on π_e because of the positive objective function coefficients of x_e. Hence, the value of π_e is set as small as possible such that the other constraints are still satisfied. But violated upper envelope inequalities only occur for high values of π_e (cf. Figure 3.2). Only a few violated lower envelope inequalities (28) have been found and only in 4 of 60 instances. Some more violated MIR inequalities (26) could be separated for the larger networks GERMANY17, GEANT1, and GEANT2.

Impact on the solving time. Figure 4(c) shows the speed-up of the total solving process (up to the time limit of 12h, no node limit) as geometric mean of the individual speed-ups for $\Gamma = 0, 1, \ldots, 10$. Here speed-up refers to $100 \cdot (1 - t/t_{\text{CPLEX}})$ in percent where t is the computation time for the considered strategy and t_{CPLEX} refers to the computation time of CPLEX (strategy (i)). We observe in the geometric means that the total solving process is always accelerated except for ABILENE1 with strategy (v) and ABILENE2 with strategy (iv) or (v). These instances are relatively small and solved within seconds. The overhead caused by strategy (iv) (graph shrinking etc.) and in particular strategy (v) (solving an ILP) slows down the total solving process for both instances. In general, strategies (ii) and (iii) yield large speed-ups of up to 75% (GERMANY17). For larger networks, the heuristic approaches used in strategies (iv) and (v) perform better. For example, while for the small ABILENE1 instance the speed-up is 66% in strategy (ii) and 23% in strategy (iv), for the larger GEANT2 instance we have a speedup of 38% in strategy (ii) but 51% in strategy (iv).

Impact on solvability. Finally, we address the number of instances that could be solved within the 12 hours time limit. The large GERMANY50 instance could not be solved to optimality in any strategy. But using the heuristic approaches in strategies (iv) and (v) the remaining optimality gap at the time limit could be decreased by 11.6%. Further, any other instances not solved within the time limit by CPLEX (6 out of 55 instances) could be solved to optimality in strategies (iv) and (v).

Conclusion. Separating violated inequalities (15), (26)–(28) speeds up the total solving process by significantly decreasing the optimality gap in the root node already. The very vast majority of separated inequalities are of type (15).

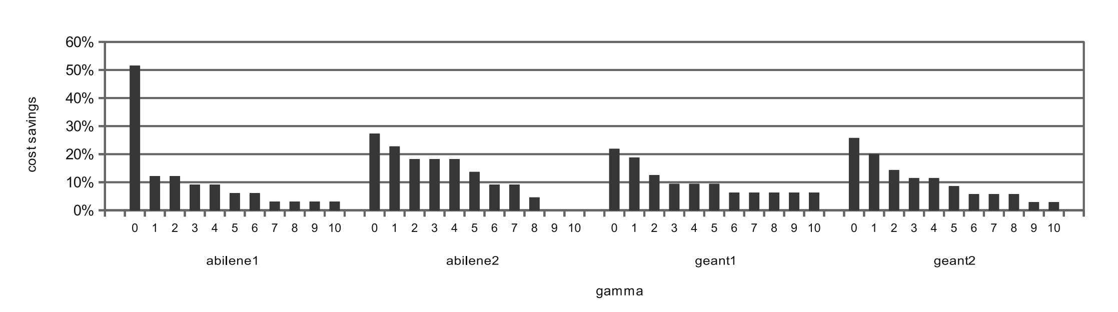


Figure 5: Cost savings of robust ABILENE and GÉANT network design compared to classical network design with peak demand values (i. e., corresponds to Γ-Robust Network Design with $\Gamma = |K|$).

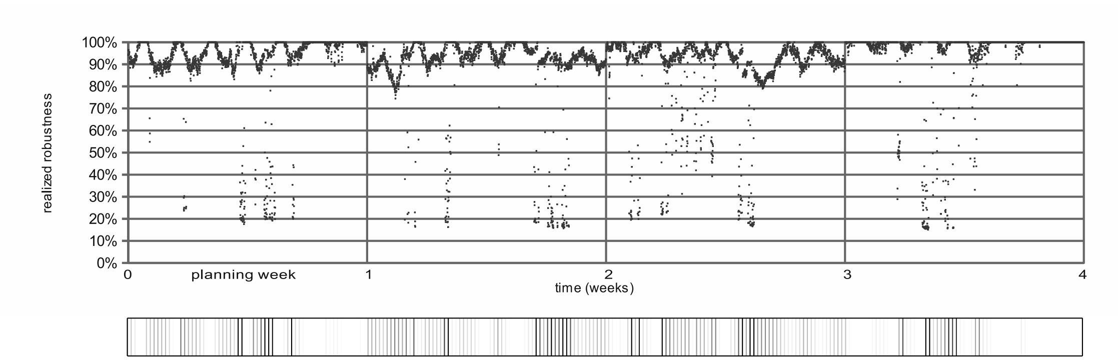


Figure 6: Realized robustness of ABILENE1 and $\Gamma = 0$. Additionally, the corresponding traffic loss profile is shown below the realized robustness diagram.

4.3 Quality of optimal robust network designs

Our third computational study focuses on the quality of optimal robust network designs. We investigate two aspects as quality criteria: the cost of an optimal robust network design and the realized robustness with respect to a given set of traffic matrices.

For the latter, we only consider the ABILENE and GÉANT networks because only for these networks traffic measurements spanning several weeks are available. Given the traffic measurements of one week as input data (as described above) we include additional weeks of traffic measurements in our evaluation of the realized robustness to simulate uncertain future traffic.

Cost savings. The cost is determined as the relative *cost savings* of the considered robust design compared to the cost of the optimal network design obtained by setting Γ to $|K|$. This corresponds to the (conservative) classical network design where capacities are optimized against the worst peak scenario. Note, considering cost savings is just a different view on the so-called price of robustness [21] and emphases the potential cost savings by using Γ-robustness compared to the most-conservative classical approach.

Figure 5 shows the relative cost savings of ABILENE1, ABILENE2, GEANT1, and GEANT2 for $\Gamma = 0, 1, \ldots, 10$. Clearly, the cost savings (compared to $\Gamma = |K|$) decrease with increasing value of Γ as the costly additional installment of link capacity modules is implied. Still, for $\Gamma = 5$ about 10% costs can be saved in all considered networks. We also see that advantage of a robust design in terms of cost is relatively small already for $\Gamma = 10$. That is, allowing for 10 commodities being at the peak simultaneously gives capacity designs at a cost similar to networks that are designed against the all-peak scenario.

Realized Robustness. Given a traffic matrix d, a capacity design x, and a static routing y of all commodities, the *realized robustness* is determined as the maximal fraction of the total demand $\sum_{k \in K} d^k$ that

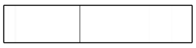
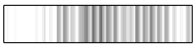
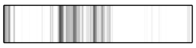
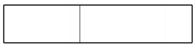
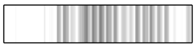
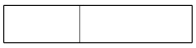
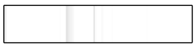
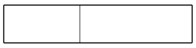
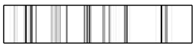
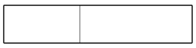
Γ	ABILENE1	ABILENE2	GEANT1	GEANT2
0				
1				
5				
10				
$\|K\|$				

Table 4: Traffic loss profiles of ABILENE and GÉANT networks and selected values of Γ. Note, the traffic loss profile of ABILENE1 and $\Gamma = 0$ is the same as in Figure 6.

can be realized as flow within the given capacities x and using the routing defined by y. To calculate this value we solve a linear program that takes d, x, and y as input and maximizes the fraction of total demand that can be realized. Traffic measurements of four consecutive weeks are used to evaluate the realized robustness in every time step where only the measurements of the first week (so-called planning week) have been analyzed and used as input to the Γ-robust network design problem.

Figure 6 shows the result for ABILENE1 and $\Gamma = 0$. The geometric mean of the realized robustness in this case is 88.2%, that is, on average over the considered time period we are able to realize 88.2% of the demand in the given capacities using the given static routing. Clearly, such a value does not catch the change of the realized robustness over time. We observe most of the time a realized robustness of 85–100%. But there exist several traffic matrices for which the realized robustness is as worse as 15%. To capture this temporal aspect of robustness, we propose a different visualization which we call the *traffic loss profile*. The corresponding traffic loss profile of ABILENE1 and $\Gamma = 0$ is shown below the diagram in Figure 6. This profile visualizes each traffic matrix by a vertical line whose gray scale value corresponds to the relative traffic amount that cannot be routed (i. e., 100% minus the realized robustness of the considered traffic matrix). The darker the line, the more traffic is lost, i. e., the less robustness is realized. Hence, a profile without lines is best and corresponds to a totally robust network design.

Table 4 shows the traffic loss profiles of optimal robust network designs for ABILENE1, ABILENE2, GEANT1, and GEANT2 and selected values of Γ. Notice that the traffic profiles for $\Gamma = |K|$ correspond to the best robustness that can be realized for the given nominal and peak demand values $\bar{d}$ and $\bar{d} + \hat{d}$. None of these profiles are totally robust since only the 95% percentile is used as peak demand $\bar{d}^k + \hat{d}^k$ value for each commodity $k \in K$.

Fixing $\Gamma = 0$, we observe that the realized robustness of optimal network designs of the four instances are quite different: 88.2% (ABILENE1), 99.9% (ABILENE2), 93.2% (GEANT1), and 96.3% (GEANT2). Comparing $\Gamma = 0$ to $\Gamma = 1$ already shows a significant improvement for ABILENE1 (88.2% to 94.9%) and GEANT2 (96.3% to 98.6%). By trend, the realized robustness of a network design increases when Γ increases, i. e., the corresponding traffic loss profile has less vertical lines or the Gray scales of the lines are brighter. A decrease can only occur due to a different and disadvantageous traffic routing. For example, this can be observed for GEANT2 where the realized robustness decreases from 99.8% to 99.7%, compare the traffic loss profiles. Note, for ABILENE2 the classical network design ($\Gamma = 0$) achieves already a realized robustness of almost 100%.

For $\Gamma = |K|$, the realized robustness is 95.8% (ABILENE1), 99.9% (ABILENE2), 98.0% (GEANT1), and 99.9% (GEANT2). This is best for the given choice of nominal and peak demand values. The evaluation of the realized robustnesses for all instances and considered values of Γ yields that for $\Gamma \geq 1$ (ABILENE1), $\Gamma \geq 0$ (ABILENE2), $\Gamma \geq 10$ (GEANT1), and $\Gamma \geq 5$ (GEANT2), at least 99% of the corresponding realized robustness value for $\Gamma = |K|$ is achieved. The traffic loss profiles for these cases hence basically coincide with the corresponding profiles for $\Gamma = |K|$, compare with Table 4.

Further, by comparing the first quarter of each traffic loss profile to its remaining part, we can evaluate the realized robustness of the planning week (the first week of the four-week time period) compared to the remaining three weeks representing uncertain future traffic. For example, we observe that the network design of GEANT1 realizes high robustness during the planning week but is significantly less robust in the following weeks. Clearly, the network load has been larger in the three weeks following the planning week.

Conclusion. Optimal robust network designs provide high potential for significant cost savings compared to the conservative setting where only peak demand values are considered. Further, the traffic loss due to peaks drops significantly already for relatively small values of Γ. In particular, for $\Gamma \leq 5$ a remarkable increase in the realized robustness can be achieved. A good value for Γ seems to depend on the size of the instance. For ABILENE a value $\Gamma = 1$ is sufficient for high robustness while for GÉANT choosing $\Gamma = 5$ gives a good trade-off between cost and robustness. With these values we get almost totally robust networks at a cost of roughly 10-20% less the cost for a network designed for the all-peak scenario.

5 Concluding Remarks

In this paper we considered a robust network design problem with static routing using a polyhedral uncertainty set going back to Bertsimas and Sim [20, 21]. This model allows to adjust the number of point-to-point demands that deviate from a given nominal value simultaneously by changing a parameter $\Gamma > 0$.

We presented mathematical formulations that enhance the classical flow formulation for network design to include demand uncertainties in different ways. A computational evaluation revealed that the compact formulation based on dualizing the capacity constraints outperforms other models based on decomposition and separation.

In a polyhedral study, we derived strong valid inequalities based on network cuts thereby generalizing the well-known cutset inequalities for the deterministic case. Instead of a single cutset inequality for every network cut, we derived multiple classes of facet-defining cut-based inequalities by exploiting the extra variables available in dualized robust counterparts. We are able to completely describe a projection of the robust cutset polyhedron of a single edge. We generalized the corresponding facet-defining inequalities and we developed conditions for the generalized cutset inequalities to define facets of the original model.

The separation of the developed robust cutset inequalities turns out to speed-up the solving process for the compact model significantly: We save up to 66% of computation time compared to default CPLEX. Many instances could only be solved to optimality withing the time limit with the new robust cutset inequalities enabled. In some cases we close the gap already in the root node.

In our computational studies we used realistic networks together with life traffic measurements in IP networks. This allowed us to parametrize the models in a meaningful way but also to evaluate the realized robustness of the resulting solutions using real-life dynamics of the demand. We also compared the possible cost savings compared to overestimating the number of simultaneous peaks in deterministic approaches. Using these results we studied the trade-off between the level of robustness and the needed capital expenditures. It turns out that already small values of Γ between 1 and 5 suffice to get capacity designs that are almost totally robust against all traffic fluctuations. This results in cost savings of 10–20% compared to conservative deterministic designs. This also shows that in real-life traffic dynamics the number of simultaneous peaks is small.

Acknowledgement

This work was supported by the Federal Ministry of Education and Research (BMBF grant 03MS616A, project ROBUKOM, `www.robukom.de`) and the DFG research center MATHEON (`www.matheon.de`).

References

[1] Abilene Internet2 Network. `www.internet2.edu/network`.

[2] Y. K. Agarwal. k-Partition-based facets of the network design problem. *Networks*, 47(3):123–139, 2006.

[3] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[4] A. Altin, E. Amaldi, P. Belotti, and M. Ç. Pinar. Provisioning virtual private networks under traffic uncertainty. *Networks*, 49(1):100–155, 2007.

[5] A. Altin, H. Yaman, and M. C. Pinar. A Hybrid Polyhedral Uncertainty Model for the Robust Network Loading Problem. Technical report, Bilkent University, 2009.

[6] A. Altin, H. Yaman, and M. C. Pinar. The Robust Network Loading Problem under Polyhedral Demand Uncertainty: Formulation, Polyhedral Analysis, and Computations. *INFORMS Journal on Computing*, Articles in Advance, 2010.

[7] A. Atamtürk. Flow pack facets of the single node fixed charge flow polytope. *Operations Research*, 29:107–114, 2001.

[8] A. Atamtürk. On Capacitated Network Design Cut-Set Polyhedra. *Mathematical Programming*, 92: 425–437, 2002.

[9] A. Atamtürk and M. Zhang. Two-Stage Robust Network Flow and Design Under Demand Uncertainty. *Operations Research*, 55(4):662–673, 2007.

[10] P. Avella, S. Mattia, and A. Sassano. Metric inequalities and the network loading problem. *Discrete Optimization*, 4:103–114, 2006.

[11] P. Belotti, A. Capone, G. Carello, and F. Malucelli. Multi-layer MPLS network design: The impact of statistical multiplexing. *Computer Networks*, 52(6):1291–1307, 2008.

[12] P. Belotti, K. Kompella, and L. Noronha. A comparison of OTN and MPLS networks under traffic uncertainty. *Working paper*, 2011. `http://myweb.clemson.edu/~pbelott/papers/robust-opt-network-design.pdf`.

[13] W. Ben-Ameur. Between fully dynamic routing and robust stable routing. In *Design and Reliable Communication Networks, 2007. DRCN 2007. 6th International Workshop on*, pages 1–6. IEEE, 2007.

[14] W. Ben-Ameur and H. Kerivin. Routing of uncertain demands. *Optimization and Engineering*, 3: 283—-313, 2005.

[15] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–14, 1999.

[16] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[17] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and application. *Mathematical Programming*, 92:453–480, 2002.

[18] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[19] A. Ben-Tal, L. El Ghaoui, and A.S. Nemirovski. *Robust optimization*. Princeton University Press, 2009.

[20] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98(1):49–71, 2003.

[21] D. Bertsimas and M. Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004.

[22] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[23] D. Bienstock and O. Günlük. Capacitated Network Design – Polyhedral Structure and Computation. *INFORMS Journal on Computing*, 8:243–259, 1996.

[24] D. Bienstock, S. Chopra, Oktay Günlük, and C. Y. Tsai. Minimum Cost Capacity Installation for Multicommodity Network Flows. *Mathematical Programming*, 81:177–199, 1998.

[25] A. Bley, U. Menne, R. Klähne, C. Raack, and R. Wessäly. Multi-layer network design – A model-based optimization approach. In *Proc. of the Polish-German Teletraffic Symposium (PGTS) 2008*, pages 107–116, 2008.

[26] C. Chekuri, G. Oriolo, M. G. Scutellá, and F. B. Shepherd. Hardness of robust network design. *Networks*, 50(1):50–54, 2007.

[27] S. Chopra, I. Gilboa, and S. T. Sastry. Source sink flows with capacity installation in batches. *Discrete Applied Mathematics*, 86:165–192, 1998.

[28] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, pages 197–206, 1955.

[29] Deutsche Forschungsnetz (DFN). `http://www.dfn.de`.

[30] N. G. Duffield, P. Goyal, A. G. Greenberg, P. P. Mishra, K. K. Ramakrishnan, and J. E. van der Merive. A Flexible Model for Resource Management in Virtual Private Networks. In *Proc. of the ACM SIGCOMM*, pages 95–108, 1999.

[31] A. Dwivedi and R. E. Wagner. Traffic model for USA long distance optimal network. In *Proc. of the Optical Fiber Communication Conference*, pages 156–158, 2000.

[32] J.A. Fingerhut, S. Suri, and J.S. Turner. Designing least-cost nonblocking broadband networks. *Journal of Algorithms*, 24(2):287–309, 1997.

[33] M. Fischetti and M. Monaci. Robustness by cutting planes and the uncertain set covering problem. `www.dei.unipd.it/ fisch/papers/Uncertain_SCP.pdf` (2011-08-26), 2009.

[34] F. Grandonia, V. Kaibel, G. Oriolo, and M. Skutella. A short proof of the VPN Tree Routing Conjecture on ring networks. *Operations Research Letters*, 36:361–365, 2008.

[35] Z. Gu, G. L. Nemhauser, and M. W. P. Savelsbergh. Lifted flow cover inequalities for mixed 0-1 integer programs. *Mathematical Programming*, 85:436–467, 1999.

[36] Z. Gu, G. L. Nemhauser, and M. W. P. Savelsbergh. Sequence independent lifting in mixed integer programming. *INFORMS Journal on Computing*, pages 109–129, 2000.

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

[38] C. A. J. Hurkens, J. C. M. Keijsper, and L. Stougie. Virtual private network design: A proof of the tree routing conjecture on ring networks. *SIAM Journal on Discrete Mathematics*, 21:482–503, 2007.

[39] IBM – ILOG. CPLEX version 12.1, 2009. `http://www.ilog.com/products/cplex`.

[40] F. Idzikowski, S. Orlowski, C. Raack, H. Woesner, and A. Wolisz. Dynamic routing at different layers in ip-over-wdm networks – maximizing energy savings. *Optical Switching and Networking, Special Issue on Green Communications*, 8:181–2000, 2011.

[41] M. Iri. On an Extension of the Maximum-flow Minimum-cut Theorem to Multicommodity Flows. *Journal of the Operations Research Society of Japan*, 13(3):129–135, 1971.

[42] O. Klopfenstein and D. Nace. Valid inequalities for a robust knapsack polyhedron – Application to the robust bandwidth packing problem. In *Proc. of International Network Optimization Conference INOC*, 2009.

[43] A. M.C.A. Koster, M. Kutschka, and C. Raack. Cutset inequalities for robust network design. In *Proceedings of INOC 2011*, pages 118–123. International Network Optimization Conference, 2011.

[44] B. Lardeux, D. Nace, and J. Geffard. Multiperiod network design with incremental routing. *Networks*, 50(1):109–117, 2007.

[45] Q. Louveaux and L. A. Wolsey. Lifting, superadditivity, mixed integer rounding and single node flow sets revisited. *4OR*, 1(3):173–207, 2003.

[46] T. L. Magnanti and P. Mirchandani. Shortest paths, single origin-destination network design and associated polyhedra. *Networks*, 33:103–121, 1993.

[47] T. L. Magnanti, P. Mirchandani, and R. Vachani. Modelling and Solving the Two-Facility Capacitated Network Loading Problem. *Operations Research*, 43:142–157, 1995.

[48] S. Mattia. The Robust Network Loading Problem with Dynamic Routing. Technical Report 3, Università di Roma la Sapienza, 2010.

[49] A. J. Miller and L. A. Wolsey. Tight formulations for some simple mixed integer programs and convex objective integer programs. *Mathematical Programming*, 98:73–88, 2003.

[50] P. Mirchandani. Projections of the capacitated network loading problem. *European Journal of Operational Research*, 122(3):534–560, 2000. ISSN 0377-2217.

[51] S. Mudchanatongsuk, F. Ordóñez, and J. Liu. Robust solutions for network design under transportation cost and demand uncertainty. *Journal of the Operational Research Society*, 59:652–662, 2008.

[52] G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 1988.

[53] NOBEL. NOBEL – Next generation Optical networks for Broadband European Leadership , 2004–2007. `http://www.ist-nobel.org`.

[54] K. Onaga and O. Kakusho. On feasibility conditions of multicommodity flows in networks. *Transactions on circuit theory*, 18(4):425–429, 1971.

[55] G. Oriolo. Domination Between Traffic Matrices. *Mathematics of Operations Research*, 33(1):91–96, 2008.

[56] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski. SNDlib 1.0–Survivable Network Design Library. *Networks*, published online, 2009. `http://sndlib.zib.de`.

[57] A. Ouorou and J.-P. Vial. A model for robust capacity planning for telecommunications networks under demand uncertainty. In *6th International Workshop on Design and Reliable Communication Networks, 2007. DRCN 2007*, pages 1–4, 2007.

[58] M. Poss and C. Raack. Affine recourse for the robust network design problem: between static and dynamic routing. ZIB Report 11-03, Zuse Institute Berlin, February 2011.

[59] C. Raack, A. M. C. A. Koster, S. Orlowski, and R. Wessäly. On cut-based inequalities for capacitated network design polyhedra. *Networks*, 57(2):141–156, March 2011.

[60] M. G. Scutellà. On improving optimal oblivious routing. *Operations Research Letters*, 37(3):197–200, 2009.

[61] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[62] S. E. Terblanche, R. Wessäly, and J. M. Hattingh. Lagrangian Relaxation as a solution approach to solving the Survivable Multi-Hour Network Design Problem. In *Proc. of the South African Telecommuniations and Networks Applications Conference - SATNAC 2007*, 2007.

[63] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon. Providing public intradomain traffic matrices to the research community. *ACM SIGCOMM Computer Communication Review*, 36(1):83–86, 2006.

[64] Y. Zhang. 6 months of abilene traffic matrices. `http://www.cs.utexas.edu/~yzhang/research/AbileneTM/`, August 2009.