Technische Universität München

Lehrstuhl für Logistik und Supply Chain Management

Methodological Advances and New Formulations for Bilevel Network Design Problems

Dipl.-Math. Univ. Pirmin Stefan Reinhold Fontaine

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Andreas Schulz
Prüfer der Dissertation: :	1.	Univ.-Prof. Dr. Stefan Minner
	2.	Prof. Teodor Gabriel Crainic, Ph.D., FRSC Université du Québec á Montréal / Kanada

Die Dissertation wurde am 30.05.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.07.2016 angenommen.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Stefan Minner for the continuous support, very constructive criticism and motivation throughout my PhD studies. Giving me the opportunity to work on several different research topics also besides this thesis, was always very interesting for me and I enjoyed it very much. I also want to express my gratitude to Prof. Teodor Gabriel Crainic, Ph.D., for the invitation to CIRRELT in Montreal, for his useful ideas and feedback on the risk equilibration problem for hazardous material and for being part of the examination committee. I also want to thank Prof. Dr. Andreas Schulz for being the chairman of the examination committee.

Further, I am very grateful for the support of my former and current colleagues of the chair of Logistics and Supply Chain Management at the Technical University of Munich: Szymon Albinski, Christian Bohner, Dr. Maximilian Budde, Miray Közen, Christian Mandl, Jun.-Prof. Dr. Anna-Lena Sachs, Dr. Martin Stößlein, Florian Taube and Dariush Tavaghof Gigloo. Thank you for all the discussions and sharing your ideas. I also want to thank Evelyn Gemkow for her support and proofreading of my work.

Moreover, I would like to thank all the colleagues I met during my stay at CIRRELT in Montreal for the interesting discussions. Particularly, I want to thank Prof. Michel Gendreau for sharing his ideas on the risk equilibration problem for hazardous material.

Finally, I would like to thank my parents and my brother Severin for their constant support, encouragement and critical comments over the last years. A special thank goes to Andrea for her patience and motivation during the last years.

Abstract

Bilevel problems are mathematical programming problems where the decision takers are devided into a leader and a follower. The objective function of the leader depends on the decision of the follower, who pursues a different objective and whose feasible region depends on the leader's decision. Bilevel problems appear in hierachical or decentralized decision structures and have applications in many areas like Stackelberg games, production planning or network design. However, exact solution methods for bilevel problems are still scarce in the literature. This thesis proposes a Benders decomposition algorithm to solve discrete-continuous bilevel problems to optimality. The efficiency of the method is shown on existing problems from the literature - namely the Discrete Network Design Problem, the Decentralized Facility Selection Problem and the Hazmat Transport Network Design Problem. Depending on the problem structure, the convergence of Benders decomposition is improved by using the multi-cut version or pareto-optimal cuts. For the Discrete Network Design Problem, a linearization of the convex objective function without introducing binary decision variables is shown and the run time is impoved by more than 60% compared to the mixed-integer linear program. Moreover, the Discrete Network Design Problem is extended to a multi-period model for planning maintenance work in traffic networks. Further, we show on the Decentralized Facility Selection Problem and on the Hazmat Transport Network Design Problem run time improvements of more than 90% compared to the mixed-integer linear program. To include risk equity in hazardous material shipment, a population-based risk definition is introduced and the Hazmat Transport Network Design Problem is extended to a multi-mode model. The numerical results show a better distribution of risk compared to classical models in the literature and a convex relation between risk equilibration and risk minimization.

Contents

List of Tables ix

List of Figures xi

Acronyms xiii

1 Introduction 1

1.1 Motivation . . . 1
1.2 Problem statement . . . 3
1.3 Structure of the thesis . . . 4

2 Fundamentals and literature 7

2.1 Bilevel programming . . . 7
2.2 Benders decomposition . . . 9
2.3 Discrete Network Design Problem . . . 11
2.4 Decentralized Capacitated Facility Selection Problem . . . 13
2.5 Hazardous material transportation . . . 14
2.5.1 Risk assessment . . . 14
2.5.2 Network design . . . 16

3 Benders decomposition for discrete-continuous linear bilevel problems with application to traffic network design 19

3.1 Introduction . . . 19
3.2 Bilevel problem and algorithm . . . 20
3.2.1 Transformation to a single-level problem . . . 20
3.2.2 Benders decomposition . . . 22
3.3 Discrete Network Design Problem . . . 26
3.3.1 Problem definition . . . 26
3.3.2 Linearization of non-linear convex functions . . . 27

3.4 Numerical study 29
3.5 Conclusion 34

4 A dynamic Discrete Network Design Problem for maintenance planning in traffic networks 37
4.1 Introduction 37
4.2 Mathematical model 39
4.3 Solution algorithm 41
4.3.1 Transformation to a linear bilevel problem 41
4.3.2 Benders decomposition 43
4.4 Numerical study 47
4.4.1 Test setup 47
4.4.2 Sioux-Fall network 48
4.4.3 Extended Sioux-Fall network 53
4.5 Conclusion 54

5 Benders Decomposition for the Decentralized Facility Selection Problem 55
5.1 Introduction 55
5.2 Decentralized Facility Selection Problem 56
5.3 Solution method 57
5.3.1 Mixed-integer linear program 57
5.3.2 Benders decomposition 58
5.4 Numerical study 61
5.5 Conclusion 64

6 Benders decomposition for the Hazmat Transport Network Design Problem 65
6.1 Introduction 65
6.2 Bilevel formulation 66
6.3 Solution approach 68
6.3.1 Mixed-integer linear program 68
6.3.2 Multi-cut Benders decomposition 69
6.4 Numerical study 72
6.5 Conclusion 76

7 Population-based Risk Equilibration for the Capacitated Multi-Mode Hazmat Transport Network Design Problem 79
7.1 Introduction . . . 80
7.2 Problem definition . . . 81
7.3 Population-based risk definition and evaluation . . . 83
7.3.1 Risk definition . . . 83
7.3.2 Risk equilibration measures . . . 84
7.4 Capacitated Multi-Mode Hazmat Transport Network Design Problem . . . 85
7.4.1 Bilevel formulation . . . 86
7.4.2 Transformation to a mixed-integer linear program . . . 87
7.5 Numerical study . . . 88
7.5.1 Risk equilibration . . . 91
7.5.2 Trade-off between risk equilibration and risk minimization . . . 92
7.5.3 Comparison to one mode decision model . . . 95
7.5.4 Comparison to maximum arc risk equilibration . . . 96
7.6 Conclusions . . . 97

8 Conclusion 99
8.1 Summary . . . 99
8.2 Limitations and future research . . . 100

Bibliography 101

List of Tables

3.1 Key information of examples 29
3.2 Results for example (S) 31
3.3 Results for example (M) with 10 new links 32
3.4 Results for example (M) with 15 new links 33
3.5 Results for example (L) 34

4.1 Congestion in European cities 2013 (TomTom, 2014) 38
4.2 Genetic algorithm: Parameters 49
4.3 Objective values for Sioux-Fall network (in 10^9) 50
4.4 Objective values for the extended Sioux-Fall network (in 10^8) 54

5.1 Computational results for 300×1500 instances of Avella and Boccia (2009) 62
5.2 Computational results for 500×500 instances of Avella and Boccia (2009) 63

6.1 Computational results for $|K| = 60$ 73
6.2 Computational results for $|K| = 80$ 74
6.3 Computational results for $|K| = 100$ 74
6.4 Computational results for $|K| = 120$ 75
6.5 Computational results for $|K| = 140$ 75
6.6 Average computational results 76

7.1 Risk evalutation for demand scenario 1 91
7.2 Results for 4 demand scenarios using the biobjective risk function 92
7.3 Trade-off between Trad and AdA for 4 demand scenarios using the constrained approach 93
7.4 Comparison of the multi-mode model with single-level decisions 95
7.5 Comparison to maximum arc risk model 96

List of Figures

2.1 Possible impact shapes along a route segment 16

3.1 Example of an approximation of $f(x)$ with 4 data points 28

4.1 Example of modelling an arc of A_2 39
4.2 Example of an approximation of $f(x)$ with 4 data points 42
4.3 Objective values of the Sioux Fall network for $B = 50$ 51
4.4 Objective values of the Sioux Fall network for $B = 75$ and $B = 100$ . . . 52

7.1 Example of the population-based risk definition 84
7.2 Sioux Fall network with population centers and population density 89
7.3 Run time analysis for AdA 90
7.4 Trade-off between Trad and AdA for 4 demand scenarios using the constrained approach 94

Acronyms

B&B	Branch-and-Bound.
BD	Benders Decomposition.
cmHTNDP	Capacitated Multi-Mode Hazmat Transport Network Design Problem.
CNDP	Continuous Network Design Problem.
DCFSP	Decentralized Capacitated Facility Selection Problem.
DCLBP	Discrete-Continuous Linear Bilevel Problem.
DDNDP	Dynamic Discrete Network Design Problem.
DNDP	Discrete Network Design Problem.
DS	Demand Scenario.
GA	Genetic Algorithm.
HCF	Highest Congestion First.
HTNDP	Hazmat Transport Network Design Problem.
KKT	Karush-Kuhn-Tucker.
LBP	Linear Bilevel Problem.
LP	Linear Program.
LRF	Lowest Reliability First.
MILP	Mixed-Integer Linear Program.
MIP	Mixed-Integer Program.
MNDP	Mixed Network Design Problem.

OD	Origin-Destination.
TAP	Traffic Assignment Problem.

Chapter 1

Introduction

1.1 Motivation

Transportation has a huge impact on everyone's daily life: Many commuters are using public transportation systems or their own car and goods are shipped all over the world to satisfy customer demand. This leads to congested cities and highways as well as a deterioration of the network. Therefore, the improvement and regulation of traffic networks is very important for society.

Recently, the Federal Ministry of Transport and Digital Infrastructure of Germany presented the first draft of the 2030 Federal Transport Infrastructure Plan (BMVI, 2016). This plan aims at selecting and prioritizing projects for the federal trunk roads, the federal railway infrastructure and the federal waterway sector to improve the mobility in passenger traffic and to guarantee freight transportation. Besides these major goals, the improvement of safety, the reduction of emissions, the limitation of use of nature, the conservation of nature and the improvement of the quality of life are also covered. For the realization, very restrictive budget constraints with a total investment of 264.5 billion Euro apply. About 69% of this budget will be used for maintenance to preserve the current network. The rest is used to build new and upgrade existing infrastructure to avoid and reduce bottlenecks until 2030.

The study is based on the forecast of transport interconnectivity 2030 (BMVI, 2014b). The transport performance of passenger transport will increase about 12.2% from 1,184 billion pkm (passenger-kilometer) in 2010 to 1,329 billion pkm in 2030. This growth is mostly driven by the increase of long-distance travelling, since the traffic volume will only increase by 1.2% to 103 billion passengers.

In freight transportation, the expected growth of demand is even higher. The transport performance within the territory of Germany will rise by 38% (from 607.1 billion tkm (ton-kilometer) to 837.6 billion tkm) and the traffic volume by 18% (3,704.7 million tons to 4,358.4 million tons) between 2010 and 2030. In particular, the transport performance of rail transportation will increase by 42.9%. (BMVI, 2014b)

One consequence of the huge amount of traffic is congested cities. TomTom (2016) showed in its annual traffic index that in the most congested cities in the world (Mexico City, Bangkok and Istanbul), the average travel time is 59%, 57% and 50% higher than in a traffic free network. In the morning and evening peak, the number even goes up to 97%, 85% and 62%. But also in Europe this so-called congestion level goes up to 44% (Moscow) while in the four most congested German cities (Hamburg, Cologne, Munich, Berlin) the additional travel time is on average between 30% and 28%.

The predicted increase in traffic shows that the federal transport infrastructure needs to be adjusted and the most improving projects should be selected. The Federal Transport Infrastructure Plan includes more than 2,000 projects for expanding the different transportation modes. But also the local networks, which are not part of the Federal Transport Infrastructure Plan, need to be expanded. In 2015 and 2016, Munich planned with 600 construction zones in the traffic network each year (Schmidt and Karowski, 2016; Völklein, 2015). In Berlin in 2015, more than 2,500 construction zones were planned (Huth, 2015). This number is supposed to increase in 2016 and 2017 (Neumann, 2016).

These construction sites will further reduce the capacity of the network during maintenance phases, which can lead to even more congestion. This will especially be true if too many projects are scheduled in one area and a city struggles with congestion in general.

One special case in freight transportation is hazardous material (hazmat) shipment. In 2013, 14% of the transported volume in Germany were dangerous goods of which 47% were transported on the road, 20% by train and about 16.5% each were shipped on inland waterways and the sea (Statistisches Bundesamt Wiesbaden, 2015). Even though technical advances can reduce the risk of hazardous accidents, the risk should additionally be reduced by choosing transportation paths with a lower risk, since the consequences of hazardous accidents are truly fear-inspiring. Moreover, citizens demand more and more often that the risk is also fairly distributed and not some specific parts of the population take all the risk.

1.2 Problem statement

Because of the high increase in traffic and since most authorities have a tight budget for improving their network, it is even more important to use this budget efficiently when expanding the network and, further, to regulate the transport of dangerous goods in order to reduce risks. These problems are called network design problems. They have a hierarchical structure in common where the users of the network only want to minimize their own travel time or costs. But the authorities want to regulate the traffic to minimize the overall congestion (e.g. LeBlanc, 1975) or to reduce the risk of dangerous good accidents (e.g. Kara and Verter, 2004). Since these two objectives do not necessarily correspond, the decision taking authorities (also called leader) have to anticipate the reaction of the user (also called follower). These problems are formulated as bilevel problems (Schneeweiß, 2003). In bilevel programming, the leader optimizes his objective function subject to a nested optimization problem: the optimization problem of the follower. However, optimal solution methods for bilevel problems are scarce. Most of the optimal approaches date back to the beginning of bilevel programming (e.g. Bard and Moore, 1990; Hansen et al., 1992) and so far the literature had a stronger focus on heuristics and metaheuristics.

The literature on traffic network design problems still mainly focusses on the Discrete Network Design Problem (e.g. LeBlanc, 1975), which deals with possible network extensions, even though maintenance is becoming more and more important. This is also reflected in the budget share for maintenance in the current Federal Transport Infrastructure Plan (BMVI, 2016).

In hazmat shipments, risk minimization is the main objective of most of the models so far (e.g. Kara and Verter, 2004). In the Hazmat Transport Network Design Problem, the leader decides which roads of a network allow dangerous goods shipments to reduce the total risk in the network. Even though some researchers already pointed out that risk equilibration needs to be considered (e.g. Erkut et al., 2007), so far only a few publications in this direction exist (Bianco et al., 2009, 2015). These ideas distribute the risk fair among arcs, but neglect that different transportation modes need to be considered. The resulting distribution, however, is generally not fair with respect to the population.

This thesis contributes to the literature by addressing these methodological challenges:

1. How to solve linear bilevel problems with discrete leader variables efficiently to optimality?

2. How to approximate the non-linear Discrete Network Design Problem to a linear bilevel problem without additional binary variables?

3. How to model a maintenance problem as a bilevel problem and solve it?

4. How to use the multiple follower structure in the Hazmat Transport Network Design Problem to solve it efficiently?

5. How to model hazardous material risk for fair distribution?

Using these methodological advances, the following research questions will be addressed in this thesis and shall support the responsible authorities in their decision:

1. How can a bilevel model for maintenance planning improve the use of the budget to reduce congestion compared to practical heuristics?

2. Why is it important to consider different modes in the Hazmat Transport Network Design Problem?

3. What is the trade-off between risk minimization and risk equilibration?

1.3 Structure of the thesis

The remainder of this thesis is structured as follows.

In Chapter 2, we first review the literature on bilevel programming and Benders decomposition (BD). After that, we summarize related work on traffic network design, decentralized production planning and hazardous material shipment.

We will then introduce a new method for solving linear bilevel problems (LBPs) with binary leader variables and continuous follower variables in Chapter 3. The bilevel formulation is transformed into a mixed-integer linear program (MILP) by using the Karush-Kuhn-Tucker (KKT) conditions and by linearizing bilinear terms. The resulting MILP is further solved with BD. Moreover, we introduce a new formulation of the Discrete Network Design Problem (DNDP), which approximates the non-linear objective functions by piecewise linear terms without introducing binary auxiliary variables. The BD approach is tested on the DNDP and the computational benefits are discussed in the results. This chapter is based on (Fontaine and Minner, 2014).

In Chapter 4, we introduce a new model for traffic maintenance planning by extending the model of the previous chapter to a multi-period model. The problem is solved with a multi-cut version of the BD of Chapter 3. In a numerical study, we show that this method finds good solutions faster than a genetic algorithm and simple greedy heuristics that might be used in practice. This chapter is based on (Fontaine and Minner, 2016a).

Chapter 5 shows the efficiency of the BD algorithm of Chapter 3 for the Decentralized Capacitated Facility Selection Problem (DCFSP). This chapter is based on (Fontaine and Minner, 2016b).

In Chapter 6, we apply the multi-cut BD to the Hazmat Transport Network Design Problem (HTNDP). In the results, we show run time improvements of 90% and the importance of using a bilevel formulation. This chapter is based on (Fontaine and Minner, 2016c).

In Chapter 7, we fairly distribute the risk of hazardous accidents among the population. Therefore, we introduce a population-based risk definition and equilibration objective functions. Moreover, the HTNDP is extended to a multi-mode network design problem. The numerical results show that the new definitions are necessary and different transportation modes need to be considered whenever risk should be spread fairly. This is joined work with Stefan Minner (Technical Univerisity of Munich), Teodor Gabriel Crainic (University of Quebec in Montreal) and Michel Gendreau (University of Montreal) and is based on (Fontaine et al., 2016).

The thesis ends with a conclusion and an outlook for possible future research.

Chapter 2

Fundamentals and literature

In this chapter, we will first summarize the methodological fundamentals and the related literature of bilevel programming and Benders decomposition, which are the two main methodological concepts of this thesis. Subsequently, we will review the literature of the Discrete Network Design Problems, decentralized production environments and hazardous material shipment - the applications of the newly introduced methods.

2.1 Bilevel programming

The idea of bilevel programming goes back to the introduction of Stackelberg games in Stackelberg (1934) and defines hierarchical or decentralized decision problems (Schneeweiß, 2003). The decision makers are divided into a so-called leader and a so-called follower while both are optimizing their own objective function. Because of the hierarchical decision structure, the leader decides subject to a nested optimization problem - the so-called follower problem. The feasible region of this follower problem depends on the leader decision and the leader objective depends on the decision of the follower. They can cover a broad range of applications, not only in traffic network design (e.g. LeBlanc, 1975) and the HTNDP (e.g. Kara and Verter, 2004), where our research is located, but also, for example, in finding chemical equilibria (Bard, 1998), defining tolls in a network (Labbé et al., 1998) or competitive facility location problems (Eiselt and Laporte, 1997). For a detailed summary of applications and an introduction into bilevel programming, the reader is referred to the books of Bard (1998), Dempe (2002), Dempe et al. (2015), Migdalas et al. (2013) and Talbi (2013) and surveys by Colson et al. (2005) and Colson et al. (2007).

Bilevel problems are nondifferentiable and nonconvex optimization problems (Dempe et al., 2015) and difficult to solve. Ben-Ayed and Blair (1990) showed that even the linear case with continuous decision variables is $\mathcal{NP}$-hard. The first mathematical formulation was introduced by Bracken and McGill (1973) and since then many applications and solution methods were developed in the literature. Depending on the solution space of the follower problem, assumptions about the cooperation level between follower and leader are needed. If the follower problem has more than one optimal solution, these solutions can have different impacts on the objective value of the leader. The two mostly used definitions are the optimistic bilevel problem and the pessimistic bilevel problem. In the optimistic case, the leader assumes that the follower chooses the solution among all optimal follower solutions, which is the best for the leader. This is also called the weak bilevel programming problem. On the other hand, in the pessimistic case, the leader assumes that the follower takes the worst case decision: the one follower solution out of all the optimal follower decisions that results in the worst objective value of the leader. This is called the strong bilevel programming problem. If no cooperation can be assumed, the pessimistic approach gives a solution that limits the damage (Dempe, 2002). In case of a unique lower level solution for all leader decisions, the optimistic and the pessimistic case return the same solution.

In this thesis, we will focus on LBPs with discrete leader variables and continuous follower variables. But since many solution methods for solving LBPs for continuous leader variables can be easily adapted for solving discrete-continuous linear bilevel problems (DCLBPs), we will review the methods for the general case. The first methods (Bard and Falk, 1982; Candler and Townsley, 1982) were based on enumeration strategies. Later, Bialas and Karwan (1984) proposed the Kth best method for solving the optimistic LBPs. This method starts with the solution of the single level problem that ignores the objective of the follower. From this vertex, neighbor vertices are explored until the global optimal solution is found.

Other approaches are based on the KKT conditions. If the follower problem has a unique solution or an optimistic problem formulation is used, these conditions are not only necessary optimality conditions, but also sufficient (Dempe, 2002). In 1990, Bard and Moore replaced the follower problem with the KKT conditions. The resulting problem was solved in a branch-and bound framework without the bilinear term in the complementary slackness conditions. As long as the complementary slackness conditions were not satisfied, a new branch was evaluated. A modification for the mixed-integer LBP was shown in Moore and Bard (1990). A second branch-and-bound algorithm, which branches on binding follower constraints, was proposed by Hansen et al. (1992). Besides that, also penalty function methods, which add the duality gap to the objective function, were introduced (White and Anandalingam, 1993).

Saharidis and Ierapetritou (2009) proposed an algorithm based on BD (see Section 2.2) for the DCLBP. Their algorithm solves a MILP both in the master problem and in the slave problem as an LBP is solved in the slave problem by using the Active Constraint Strategy by Grossmann and Floudas (1987).

2.2 Benders decomposition

BD was introduced in 1962 by Benders and is nowadays one of the most used solution methods for solving large-scale mixed-integer programs (MIPs). However, BD also has a wide range of applications in stochastic programming, where it is more common under the L-Shaped method (Van Slyke and Wets, 1969), and in non-linear optimization (Geoffrion, 1972).

The general idea of BD is to decompose the problem into two easier problems - the master problem and the slave problem - and solve these problems repeatedly until convergence is reached. Let y and x be two sets of decision variables. From now on, we will refer to y as complicating variables and to x as easy variables. The complicating variables can be integer or continuous and further restricted with a set of constraints. Both will be included in the constraint set $y \in Y$. The easy variales x need to be continuous and w.l.o.g. $x \geq 0$. With A and B being constraint matrices, b the right hand side and c and f the objective function coefficients, a generic MILP can be defined as follows (Martin, 1999):

$$\min \quad f^\top y + c^\top x \tag{2.1}$$

$$\text{s.t.} \quad Ax + By \geq b \tag{2.2}$$

$$y \in Y \tag{2.3}$$

$$x \geq 0 \tag{2.4}$$

By fixing the complicating variables $y = y^*$, the optimal decision for x can be determined by solving the following subproblem $SP(y)$:

$$\min \quad c^\top x \tag{2.5}$$

$$\text{s.t.} \quad Ax \geq b - By^* \tag{2.6}$$

$$x \geq 0 \tag{2.7}$$

With u being the dual variables to the constraints (2.6), the dual subproblem $DSP(y)$ is defined as follows:

$$\max \quad u^\top(b - By^*) \tag{2.8}$$
$$\text{s.t.} \quad u^\top A \leq c \tag{2.9}$$
$$u \geq 0 \tag{2.10}$$

We assume that the feasible region of the $DSP(y)$ is not empty. Otherwise the generic MILP is unbounded or infeasible. With that assumption, the solution space of $DSP(y)$ can be defined by its extreme points p_i, $i = 1,..,I$ and extreme rays q_j, $j = 1,..,J$. As the solution space does not depend on y, the generic MILP can be reformulated as follows:

$$\min \quad f^\top y + z \tag{2.11}$$
$$\text{s.t.} \quad p_i^\top(b - By^*) \geq z \qquad \forall i = 1,...,I \tag{2.12}$$
$$q_j^\top(b - By^*) \geq 0 \qquad \forall j = 1,...,J \tag{2.13}$$
$$y \in Y \tag{2.14}$$
$$z \in \mathbb{R} \tag{2.15}$$

Constraints (2.12) define the Benders optimality cuts and constraints (2.13) the Benders feasibility cuts. Instead of calculating all extreme points and extreme rays of $DSP(y)$, a relaxed master problem (RMP) (2.11), (2.14) and (2.15) is solved in each iteration. With the solution of the master problem, the dual subproblem is solved and, if the dual subproblem is unbounded, a feasibility cut (2.13) is added to the relaxed master problem, whereas, if it is bounded, an optimality cut (2.12) is added. Moreover, if the subproblem is bounded, the objective value gives a new feasible solution and a potential new upper bound z_{up}^{new}. Each objective value of the master problem gives a new lower bound z_{low} and the algorithm ends with the optimal solution as soon as the lower and upper bound z_{up} are equal. The method is outlined in Algorithm 1.

As the convergence of the classical BD can be very slow and solving the RMP can be very difficult, several strategies for improving the algorithm have been proposed. If the RMP is difficult to solve, optimality cuts and feasibility cuts can also be generated by solving the RMP not to optimality during a first phase (Geoffrion and Graves, 1974) or even by solving it heuristically (Cote and Laughton, 1984). Moreover, Rei et al. (2009) proposed local branching in the master problem to generate multiple cuts in one

Algorithm 1: Benders decomposition

Create RMP without constraints (2.12) and (2.13);
Set $z_{low} \leftarrow -\infty, z_{up} \leftarrow \infty$ **while** $z_{up} > z_{low}$ **do**
  solve $RMP \rightarrow y^*, z_{low}$;
  solve $DSP(y^*) \rightarrow z_{up}^{new}$;
  if DSP *bounded* **then**
    add optimality cut to RMP;
    if $z_{up}^{new} < z_{up}$ **then** $z_{up} \leftarrow z_{up}^{new}$;
  else
    add feasibility cut to RMP;
  end
end

iteration and showed significant run time improvements.

A second problem of the classical BD can be the generation of weak cuts in the subproblem. To address this issue, Magnanti and Wong (1981) proposed a method to generate non-dominated optimality cuts which show large improvements in reducing the number of iterations in many cases. Further acceleration ideas for these cuts were proposed by Papadakos (2008) and Sherali and Lunday (2013).

Besides the classical cuts, Codato and Fischetti (2004) proposed combinatorial cuts to reduce the number of iterations.

2.3 Discrete Network Design Problem

Road Network Design Problems are one application of bilevel programming. In the lower level problem, the travelers can decide on their road in a network. The most used lower level model is the Traffic Assignment Problem (Sheffi, 1985), where the resulting flow pattern is a user equilibrium. The travel demand is represented by an origin-destination (OD) matrix that defines how many users want to travel from their origin to their destination. Each traveler wants to minimize its own travel time and no cooperation is assumed between the travelers. In this equilibrium, no traveler can improve its own travel time by switching to another route (Wardrop, 1952).

The upper level is a design problem: The leader wants to minimize - compared to the follower - the total travel time in the network, which is called the system-optimum. Braess (1968) showed that the system-optimum and the follower-optimum are not necessary the same and therefore a bilevel formulation is needed. To improve the network, the leader can add new links or capacity to existing links in a network to minimize the overall congestion in the network. The first one is the DNDP, the latter one the Continuous Network Design Problem (CNDP). The combination of both is the Mixed Network Design Problem (MNDP). Farahani et al. (2013) summarized the existing models of urban transportation planning and gave a survey on existing solution methods.

Even though the focus of this thesis is the DNDP, we also want to give a short summary of the other two cases. Abdulaal and LeBlanc (1979) introduced the CNDP with user equilibrium and presented a direct search algorithm. Suh and Kim (1992) presented a descent algorithm for non-linear bilevel problems. Due to the complexity of the problem, many metaheuristics were proposed in the literature. Friesz et al. (1992, 1993) and Meng and Yang (2002) used simulated annealing procedures and Xu et al. (2009) and Mathew and Sharma (2009) used genetic algorithms for solving the problem. Wang and Lo (2010) transformed the bilevel problem into a MILP by transforming the equilibrium constraints into mixed-integer constraints and linearizing the travel time function.

The DNDP with user equilibrium, which is a more difficult problem, was first introduced by LeBlanc (1975) and solved with a branch-and-bound method. Poorzahedy and Turnquist (1982) solved the Traffic Assignment Problem (TAP) and let the followers decide over the new links in the network. This formulation was solved in a branch-and-bound framework. Moreover, an approximated lower bound was used in a branch-and-bound based heuristic. Gao et al. (2005) proposed a generalized BD approach with the use of support functions for the bilevel formulation. Luathep et al. (2011) used the variational inequality problem of the follower problem to formulate the problem as a MILP. The non-linear travel time function was approximated with piecewise linear terms and bilinear terms were relaxed by introducing binary auxiliary variables. Instead of using variational inequalities, Farvaresh and Sepehri (2011) transformed the problem into a non-linear mixed-integer program by using the KKT conditions and binary auxiliary variables were used to linearize this formulation. Moreover, Wang et al. (2013) used the relation between the leader and the follower objective for a global optimization method.

Metaheuristics were also used to solve larger and real-sized instances for the DNDP. Besides ant colony systems (Poorzahedy and Abulghasemi, 2005), hybrid metaheuristics based on tabu search, simulated annealing and a genetic algorithm (Poorzahedy and Rouhani, 2007) were proposed. The authors showed that the hybrid performs even better than what they introduced in their previous work.

Traffic Network Design Problems with a planning horizon over several periods for maintenance planning are scarce in the literature. Only Ng et al. (2009) proposed a genetic algorithm for maintenance planning which solves a mesoscopic traffic simulation in the follower problem. However, for a network of 13 nodes and 24 arcs, a run time of about 2 days has been reported.

2.4 Decentralized Capacitated Facility Selection Problem

Decentralized production and distribution planning is another application for bilevel programming. These models are tailored to one specific setting and therefore vary significantly with every setting. Cao and Chen (2006) formulated a decentralized production selection problem as a bilevel problem and named it the DCFSP. The principal firm is the leader who decides on opening facilities for production while minimizing plant opening costs and opportunity costs for unused production capacities. The opened plants minimize the production-related operational costs and satisfy the demand restricted by their own capacity. The production plan is done independently from the principal firm. However, coordination among the plants is assumed. The LBP is transformed into a MILP and applied to small instances with up to six facilities and eight products.

Calvete et al. (2011) proposed a bilevel model for production and distribution planning. The leader is represented by a distribution company with several depots. This company wants to minimize the transportation costs and the costs for acquiring a product from the production plants to the depot subject to a multi-depot vehicle routing problem with a homogeneous fleet. After knowing the demand of each depot, the follower – the production company – assigns the demand of the depots to a set of their own plants while minimizing operational costs. The authors used an ant colony metaheuristic to solve the problem.

In 2014, Calvete et al. proposed a bilevel formulation for decentralized distribution network planning. Compared with the previous work, the leader can now decide whether to open a depot or not. However, the distribution is done with direct shipments as in a classical transportation problem. Besides the transportation costs from the depot to the customer, the leader minimizes the depot opening costs and the transportation costs from the plant to the depot. The follower reacts on the demand of each depot and minimizes the operational costs while assigning production quantities from each plant to the depots to satisfy the depots' demand. As the problem is difficult to solve to optimality, they proposed a genetic algorithm for solving larger instances.

2.5 Hazardous material transportation

Research about hazmat transportation started around thirty years ago. Erkut et al. (2007) give a very detailed introduction into the topic and summarize the literature. They identify four main streams in the literature:

- risk assessment
- routing
- combined facility location and routing
- network design

Since we focus on a new solution method for solving the HTNDP and a new risk definition of the HTNDP, we will review the literature of the first and last stream. For details on the other two streams, the interested reader is referred to Erkut et al. (2007).

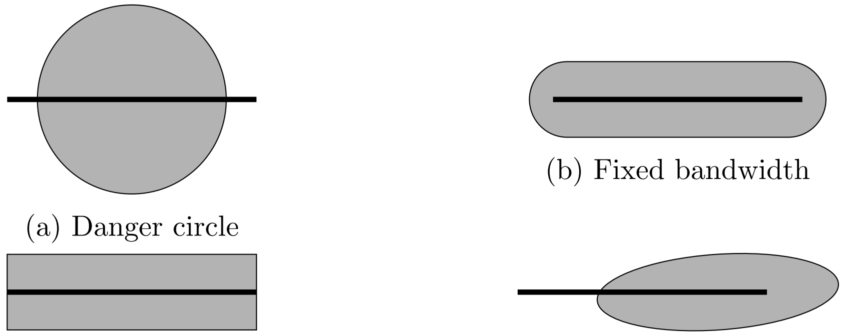
2.5.1 Risk assessment

When transporting hazardous materials (also called dangerous goods), more than just the traditional factors, e.g. transportation time, transportation costs, and environmental aspects like CO_2 emissions have to be considered. One also has to consider the risk of potential accidents and the consequences of these accidents. The economic effects of an hazmat accident can be divided into seven categories (Abkowitz et al., 2001):

- injuries and fatalities
- cleanup costs
- property damage
- evacuation
- product loss
- traffic incident delay
- environmental damage

The costs of injuries and fatalities are often also known as population exposure and used for assessing the risk of hazmat transport. The risk of shipping one unit on an arc is then the product of the probability of an accident on that arc and the exposed population (e.g. Batta and Chiu, 1988; Erkut and Verter, 1998). This is called the traditional risk. Besides the traditional risk, various other risk measures have been proposed over the last years. The population exposure (Batta and Chiu, 1988) ignores the probability of an accident and just sums over the exposed populations. In contrast, the incident probability sums only over the accident probabilities (Saccomanno and Chan, 1985). Especially for evaluating the risk on a path, further advanced risk measures have been proposed (Erkut et al., 2007). To include the fact that the population might not be risk-neutral and favor a higher probability of a low-consequences accident over a lower probability with high consequences, Abkowitz et al. (1992) introduced the perceived risk. The exposed population is exponentiated with a risk preference. If this risk preference is greater than 1, a risk averse population is assumed, if it is 1, the population is risk neutral and if it is smaller than 1, the population is risk prone. All these measures focused on risk minimization. However, a few works in the area of hazardous material routing also focused on a fair distribution of risk. Gopalan et al. (1990) proposed a shortest-path problem for routing the trucks that minimizes the total risk and ensures risk equity between zones by a constraint. However, this model ignores the fact that the carrier's main goal is to minimize the costs. Lindner-Dutton et al. (1991) extended this model and further include the sequencing of the trucks, to have a fair distribution at every point and not over the whole planning horizon. Carotenuto et al. (2007) defined a MILP to find minimal and equitable risk routes for hazardous material shipment. This fair risk distribution is done by distributing it equally among the arcs. The problem is solved by a modified k-shortest path algorithm. All these definitions assume independent risk probabilities per road segment. Kara et al. (2003) proposed a method to evaluate the risk on a path accurately. However, Erkut and Verter (1998) showed that the approximation error of the independent risk assumption is small.

A second factor of this risk definition is the calculation of the exposed population. Depending on the transported dangerous good but also on other factors like topology, weather and wind, different impact shapes were proposed in the literature. Figure 2.1 illustrates four different shapes: The first three figures are very closely related to each other. The danger circle (e.g. Erkut and Verter, 1998) has a hazmat depending radius with the center at the accident location. If this circle is moved along the edge, it yields the fixed bandwidth approximation (e.g. ReVelle et al., 1991). By cutting off the semicircles at the end of the fixed bandwidth, one gets the rectangle approximation (e.g. ALK Associates, 1994). All these models assume that the impact of an accident is the same for every person within the danger area without depending on the distance to the origin of the accident. Since the impact of an airborne hazmat is significantly different from all others, Patel and Horowitz (1994) used the Gaussian plume model to express the consequence of such an incident.



(a) Danger circle
(b) Fixed bandwidth
(c) Rectangle
(d) Gaussian plume

Figure 2.1: Possible impact shapes along a route segment (adapted from Erkut et al., 2007)

2.5.2 Network design

The network design problem is typically represented as a bilevel model to reflect the different interests of authorities and carriers. The government or authorities (the leader) want to regulate the transportation of dangerous goods to reduce the risk for the population. This can be done either by forbidding parts of the transportation network for hazmat transportation (e.g. Kara and Verter, 2004) or by introducing tolls for the transportation of dangerous goods (e.g. Marcotte et al., 2009). The followers are the carriers who need to ship their demand through the network. They want to minimize their transportation costs and therefore the follower problem is modelled as a shortest path problem.

The literature on the HTNDP is rather scarce. The first bilevel formulation was introduced by Kara and Verter (2004). Their model was transformed into a single-level mixed-integer program by using KKT conditions for the follower problem and solved with a commercial solver. In (Verter and Kara, 2008), a path-based formulation is proposed for solving the problem more efficiently. Erkut and Alp (2007) propose a method that creates a hazardous material network out of a tree-structured subset of the existing transportation network. This solution is used in a path-addition heuristic. An authority can decide whether to reduce costs or increase the risk by adding more allowed arcs to the network. All these approaches used the optimistic case of the bilevel problem where the carrier uses the route with the lowest risk if several routes cost the same. These solutions are called unstable and Erkut and Gzara (2008) proposed a heuristic that generates stable solutions for the HTNDP. Further, a stable MILP was introduced by Amaldi et al. (2011). To also distribute the risk fairly, Bianco et al. (2009) introduced a bilevel model that not only minimizes the total risk in the network, but also the maximum risk over all arcs. This idea was also used in a game-theoretic approach for the fair distribution of risk by Bianco et al. (2015). To calculate a Nash-equilibrium, they used a local search heuristic not only to minimize the total risk but also to equilibrate the risk by looking at the maximum risk on an arc. This approach is further restricted to one hazmat type and the authors pointed out that an extension makes the problem much harder as the Nash game is not convex anymore. Recently, Sun et al. (2015) proposed a network design model that includes risk uncertainty in the decision.

Chapter 3

Benders decomposition for discrete-continuous linear bilevel problems with application to traffic network design

We propose a new fast solution method for linear bilevel problems with binary leader and continuous follower variables under the partial cooperation assumption. We reformulate the bilevel problem into a single-level problem by using the Karush-Kuhn-Tucker conditions. This non-linear model can be linearized because of the special structure achieved by the binary leader decision variables and subsequently solved by a Benders decomposition Algorithm to global optimality. We illustrate the capability of the approach on the Discrete Network Design Problem which adds arcs to an existing road network at the leader stage and anticipates the traffic equilibrium for the follower stage. Because of the non-linear objective functions of this problem, we use a linearization method for increasing, convex and non-linear functions based on continuous variables. Numerical tests show that this algorithm can solve even large instances of bilevel problems.

3.1 Introduction

As we are facing increasing population in cities, the demand for transportation increases. This leads to more congested roads and longer travel times. Moreover, congestions lead to air pollution, noise pollution and a lower quality of living. Therefore, traffic networks have to be expanded and an efficient usage of the budget in network expansions should be achieved. In the literature, these problems are addressed as bilevel problems (e.g.

Farahani et al., 2013; Gao et al., 2005; LeBlanc, 1975; Luathep et al., 2011; Poorzahedy and Turnquist, 1982).

Besides the contribution to solve considerably larger instances of DCLBPs, we present a formulation of the Discrete Network Design Problem which approximates the non-linear convex objective functions only by piecewise linear terms (without additional binary variables) and can be solved by our algorithm. Compared to Luathep et al. (2011) and Farvaresh and Sepehri (2011), we avoid introducing binary auxiliary variables and the relaxation of bilinear terms. Because of the very small number of binary variables, the MILP formulation of the DNDP has computational benefits and we can solve even large instances for the DNDP. We further show how to accelerate the run time of the slave problem.

The remainder of this chapter is structured as follows. First, we introduce the general LBP and our algorithm in Section 3.2. Section 3.3 introduces the bilevel formulation of the DNDP and shows the linearization. In Section 3.4, we evaluate the performance of the algorithm on several instances and end with a summary of the proposed procedure, results and outline some future research opportunities.

3.2 Bilevel problem and algorithm

Section 3.2.1 shows the transformation of the DCLBP into a single-level MILP and BD is applied in Section 3.2.2.

3.2.1 Transformation to a single-level problem

In the following, we introduce the general formulation for the DCLBP. The leader variables are given by y_i for all $i \in I$ with I the corresponding set of indices and the follower variables by x_j for all $j \in J$ with J the corresponding set. (3.1) shows the leader objective function with $f_i' \in \mathbb{R}$ for all $i \in I$ and f_j for all $j \in J \in \mathbb{R}$ the objective coefficients. The follower problem is represented by (3.2) - (3.4) with the follower objective function in (3.2) and the follower objective coefficients $c_j \in \mathbb{R}$ for all $j \in J$. K is the set of follower constraints in (3.3), where each constraint $k \in K$ is defined by its coefficients $a_{kj} \in \mathbb{R}$ for all $i \in I$, $a'_{kj} \in \mathbb{R}$ for all $j \in J$ and the right hand side b_k. For simplification, we omit constraints in the leader main problem, but the following transformations can all be applied to the more general formulation. Moreover, we assume the partial cooperation assumption (Bialas and Karwan, 1984; Dempe, 2002) - also called an optimistic

DCLBP. This allows the leader to select an optimal follower decision among all optimal follower decisions if there exists more than one.

$$\min_{y}\ z_L(y,x) = \sum_{i\in I} f'_i y_i + \sum_{j\in J} f_j x_j \tag{3.1}$$

$$\text{s.t.}\ \ \min_{x} z_F(x) = \sum_{j\in J} c_j x_j \tag{3.2}$$

$$\sum_{j\in J} a_{kj} x_j + \sum_{i\in I} a'_{ki} y_i \le b_k \qquad \forall k \in K \tag{3.3}$$

$$x_j \ge 0 \qquad \forall j \in J \tag{3.4}$$

$$y_i \in \{0,1\} \qquad \forall i \in I \tag{3.5}$$

As a first step, we reformulate this problem to an equivalent non-linear MIP, which is derived by substituting the follower problem by its KKT conditions, see Cao and Chen (2006) and Bard (1998). In this single-level model, u_k are the dual variables corresponding to the follower constraints (3.3), and (3.9) are the dual constraints corresponding to the primal follower variables x_j. (3.8) compares the objective value of the primal follower problem on the left side with the objective value of the dual follower problem on the right side. Through the duality theorem, this equation guarantees the optimality of the follower problem while optimizing the leader's objective.

$$\min_{y}\ z_L(y,x) = \sum_{i\in I} f'_i y_i + \sum_{j\in J} f_j x_j \tag{3.6}$$

$$\text{s.t.}\ \ \sum_{j\in J} a_{kj} x_j + \sum_{i\in I} a'_{ki} y_i \le b_k \qquad \forall k \in K \tag{3.7}$$

$$\sum_{j\in J} c_j x_j \le \sum_{k\in K} u_k b_k - \sum_{k\in K}\sum_{i\in I} a'_{ki} u_k y_i \tag{3.8}$$

$$\sum_{k\in K} a_{kj} u_k \le c_j \qquad \forall j \in J \tag{3.9}$$

$$u_k \le 0 \qquad \forall k \in K \tag{3.10}$$

$$x_j \ge 0 \qquad \forall j \in J \tag{3.11}$$

$$y_i \in \{0,1\} \qquad \forall i \in I \tag{3.12}$$

The non-linear term $u_k y_i$ in (3.8) can be linearized using the approach used in Cao and Chen (2006) and Farvaresh and Sepehri (2011) and (3.8) is replaced by the following linear constraints with M being a large positive number:

$$\sum_{j\in J} c_j x_j \leq \sum_{k\in K} u_k b_k - \sum_{k\in K}\sum_{i\in I} a'_{ki}\mu_{ki} \tag{3.13}$$

$$\mu_{ki} \leq u_k + M(1-y_i) \qquad \forall i \in I, k \in K \tag{3.14}$$

$$\mu_{ki} \geq u_k \qquad \forall i \in I, k \in K \tag{3.15}$$

$$\mu_{ki} \geq -My_i \qquad \forall i \in I, k \in K \tag{3.16}$$

$$\mu_{ki} \leq 0 \qquad \forall i \in I, k \in K \tag{3.17}$$

Constraints (3.14) - (3.17) ensure that the newly introduced decision variables μ_{ki} take the value 0 if $y_i = 0$ and u_k if $y_i = 1$.

3.2.2 Benders decomposition

The structure of the MILP derived in the last section allows a solution by BD (Benders, 1962). The basic idea of BD is to decompose the problem into a master problem and a slave problem and to solve these problems repeatedly. The decision variables are divided into complicating variables, which in our case are the binary variables y_i, and a set of easier variables, the continuous variables x_j, u_k, μ_{ki}. In each iteration, the master problem determines one possible leader decision. This solution is used in the slave problem to generate optimality cuts and a feasible solution or feasibility cuts, which are added to the master problem. The main structure of a BD algorithm is shown in Algorithm 2.

Algorithm 2: Benders decomposition

1 Initialization: upper bound $UBD = \infty$; set y^* to any feasible solution of y
2 Solve the slave problem for $y = y^*$. Let z_S be the current optimal value regarding the slave problem and set the upper bound $UBD = \min\{UBD, z_S\}$. If the slave problem is bounded, then add an optimality cut to the master problem, else add a feasibility cut to the master problem.
3 Solve the current master problem and save the solution y^*. Let z_M be the current optimal value regarding the master problem.
4 If $UBD - z_M < tolerance$ then stop else continue with Step 2.

Because of the linearization, we do not have to apply the Generalized Benders decomposition (Geoffrion, 1972) and consequently avoid the convergence problems for bilinear terms. Sahinidis and Grossmann (1991) showed that the Generalized Benders decomposition might end in a local optimum or even not in an optimum at all for different starting points.

Dual slave problem

The dual slave problem is derived by fixing the decision variables y with y^* and dualizing the linearized single-level problem ((3.1) - (3.7), (3.9) - (3.12), (3.13) - (3.17)). In this model, the dual variables α_k, β, γ_j, δ_{1ki}, δ_{2ki} and δ_{3ki} correspond to the constraints (3.7), (3.13), (3.9), (3.14), (3.15) and (3.16) and the dual constraints (3.20), (3.21) and (3.22) to the variables x_j, u_k and μ_{ki}.

$$\max_{\alpha,\beta,\gamma,\delta} \sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki}y_i^*\right)\cdot \alpha_k + \sum_{j\in J} c_j\gamma_j \tag{3.18}$$

$$+\sum_{i\in I}\sum_{k\in K}(M(1-y_i^*)\delta_{1ki} - My_i^*\delta_{3ki}) \tag{3.19}$$

$$\text{s.t.} \sum_{k\in K} a_{kj}\alpha_k + c_j\beta \le f_j \qquad \forall j\in J \tag{3.20}$$

$$-b_k\beta + \sum_{j\in J} a_{kj}\gamma_j - \sum_{i\in I}(\delta_{1ki}+\delta_{2ki}) \ge 0 \qquad \forall k\in K \tag{3.21}$$

$$a'_{ki}\beta + \delta_{1ki} + \delta_{2ki} + \delta_{3ki} \ge 0 \qquad \forall i\in I, k\in K \tag{3.22}$$

$$\alpha_k \le 0 \qquad \forall k\in K \tag{3.23}$$

$$\beta \le 0 \tag{3.24}$$

$$\gamma_j \le 0 \qquad \forall j\in J \tag{3.25}$$

$$\delta_{1ki} \le 0 \qquad \forall i\in I, k\in K \tag{3.26}$$

$$\delta_{2ki}, \delta_{3ki} \ge 0 \qquad \forall i\in I, k\in K \tag{3.27}$$

If this problem is feasible with a solution α^*, β^*, γ^* and δ^*, we add an optimality cut

$$\sum_{i\in I} f'_i y_i + \sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki}y_i\right)\cdot\alpha_k^* + \sum_{j\in J} c_j\gamma_j^* + \sum_{i\in I}\sum_{k\in K}(M(1-y_i)\delta_{1ki}^* - My_i\delta_{3ki}^*) \le z \tag{3.28}$$

to the master problem. If it is unbounded, a new constraint, which bounds the objective function with a Big-M M_2, is added to the slave problem. M_2 has to be large enough such that no extreme point is cut off and only the extreme rays are bounded. The slave problem can then be solved by including this constraint:

$$\sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki} y_i\right)\cdot \alpha_k + \sum_{j\in J} c_j\gamma_j + \sum_{i\in I}\sum_{k\in K}\left(M(1-y_i)\delta_{1ki} - My_i\delta_{3ki}\right) \le M_2 \tag{3.29}$$

The solution of this problem generates the following feasibility cut, which can be added to the master problem:

$$\sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki} y_i\right)\cdot \alpha^*_k + \sum_{j\in J} c_j\gamma^*_j + \sum_{i\in I}\sum_{k\in K}\left(M(1-y_i)\delta^*_{1ki} - My_i\delta^*_{3ki}\right) \le 0 \tag{3.30}$$

This slave problem only contains continuous variables and is easy to solve.

Master problem

Let C_O be the set of solutions $(\alpha^*, \gamma^*, \delta^*)$ of optimality cuts and C_F be the set of solutions $(\alpha^*, \gamma^*, \delta^*)$ of feasibility cuts. In each iteration of the BD, a cut based on the solution of the slave problem is added to the respective set. Then, the corresponding master problem is defined as follows:

$$\min\ z \tag{3.31}$$

$$\text{s.t. } z \ge \sum_{i\in I} f'_i y_i + \sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki} y_i\right)\cdot \alpha^*_k + \sum_{j\in J} c_j\gamma^*_j + \sum_{i\in I}\sum_{k\in K}\left(M(1-y_i)\delta^*_{1ki} - My_i\delta^*_{3ki}\right) \quad \forall(\alpha^*,\gamma^*,\delta^*)\in C_O \tag{3.32}$$

$$0 \ge \sum_{k\in K}\left(b_k - \sum_{i\in I} a'_{ki} y_i\right)\cdot \alpha^*_k + \sum_{j\in J} c_j\gamma^*_j + \sum_{i\in I}\sum_{k\in K}\left(M(1-y_i)\delta^*_{1ki} - My_i\delta^*_{3ki}\right) \quad \forall(\alpha^*,\gamma^*,\delta^*)\in C_F \tag{3.33}$$

$$y_i \in \{0,1\} \quad \forall i\in I \tag{3.34}$$

$$z \in \mathbb{R} \tag{3.35}$$

Through the decomposition we have two smaller subproblems which can be solved much faster: the continuous slave problem and the usually rather small linear mixed-integer master problem.

Acceleration for slave problem

Let z_F* being the optimal objective value of the follower problem (3.2) - (3.4) for a fixed y^*. As the dual variables of the KKT conditions u_k, μ_{ki} only ensure optimality of the follower problem, they don not appear in the leader objective function (3.6) and the primal slave problem can be expressed as follows:

$$\min\ z_L(y,x) = \sum_{i\in I} f'_i y^*_i + \sum_{j\in J} f_j x_j \tag{3.36}$$

$$\text{s.t.}\ \sum_{j\in J} a_{kj}x_j + \sum_{i\in I} a'_{ki}y^*_i \leq b_k \qquad \forall k \in K \tag{3.37}$$

$$\sum_{j\in J} c_j x_j \leq z^*_F \tag{3.38}$$

$$x_j \geq 0 \qquad \forall j \in J \tag{3.39}$$

Having the optimal solution value of the follower objective, (3.38) ensures that the leader objective is minimized under the condition that the follower objective is minimal. The dual variables α_k and β can be calculated by this formulation, as they are not effected by constraint (3.9). Afterwards, the dual of the follower problem is solved:

$$\max \sum_{k\in K} u_k b_k - \sum_{k\in K}\sum_{i\in I} a'_{ki}\mu_{ki} \tag{3.40}$$

$$\text{s.t.}\ \sum_{k\in K} a_{kj}u_k \leq c_j \qquad \forall j \in J \tag{3.41}$$

$$(3.14) - (3.17)$$

$$u_k \leq 0 \qquad \forall k \in K \tag{3.42}$$

Let γ'_j, δ'_{1ki}, δ'_{2ki} and δ'_{3ki} be the dual variables of this problem. As the dual of the follower problem does not influence the leader objective function directly but only (3.40) which is z^*_F in (3.38), the dual variables of the slave problem can be calculated as follows: $\gamma_j = \gamma'_j\beta$, $\delta_{1ki} = \delta_{1ki}\beta$, $\delta_{2ki} = \delta_{2ki}\beta$ and $\delta_{3ki} = \delta_{3ki}\beta$. As γ'_j is the shadow price of constraint (3.41) for (3.40) which is further the shadow price for (3.38) and as β is the shadow price of (3.38) for the leader objective function, $\gamma'_j\beta$ is the shadow price of constraint (3.41) for the leader objective function. Furthermore, if the calculation in the second step ends in $\beta = 0$, the dual follower problem does not have to be solved.

3.3 Discrete Network Design Problem

A general definition of the DNDP is given in Section 3.3.1 and a continuous variable based linearization of a convex function is given in Section 3.3.2.

3.3.1 Problem definition

In the DNDP (Gao et al., 2005; LeBlanc, 1975; Poorzahedy and Turnquist, 1982), an existing transportation network is modeled as a set of nodes N, representing origins and destinations or intersections. The nodes are connected via a set of arcs A, which represents the road system. Every arc $a \in A$ is specified by a travel time function $t_a(x) := T_a \left(1 + B_a \left(\frac{x}{c_a}\right)^4\right)$ (Bureau of Public Roads, 1964). T_a is the free flow travel time, B_a the congestion influence parameter and c_a the capacity limit. The demand in the network is represented by the OD matrix. The set of origins is defined as $R \subseteq N$ and the set of destinations as $S \subseteq N$. The OD matrix is then defined by the values q_{rs}, which is the number of travelers from $r \in R$ to $s \in S$. The set of arcs is divided into two subsets: A_1 is the set of already existing roads, and A_2 the set of possible new roads, which each would cost b_a to build.

The decision maker (leader) has to decide which of the possible new roads of the networks to build subject to a budget B. These decisions will be the binary variables y_a, which are 1 if a new route $a \in A_2$ is built and 0 if not. The leader's objective is to avoid congestion and minimize the total travel time in the network, which is called the system-optimum. The follower, the travellers through the network, minimize their own travel time, which is based on Wardrop's first principle (Wardrop, 1952). This optimum is called the user-optimum. The Paradox of Braess (Braess et al., 2005) showed why these two optima are not necessarily the same and so the formulation as bilevel problem is necessary. The flow on each arc $a = (i, j) \in A$ will be the continuous decision variable x_a and the flow for each destination s on arc a the continuous variables $f_a^s = f_{ij}^s$.

The follower problem, which is an uncapacitated TAP (Nguyen, 1974), can now be formulated for a fixed leader decision y_a^* as follows (Poorzahedy and Turnquist, 1982):

$$\min \sum_{a \in A} \int_0^{x_a} t_a(x)dx = \min \sum_{a \in A} \left(T_a x_a + \frac{T_a B_a}{5c_a^4} x_a^5\right) \tag{3.43}$$

$$\text{s.t.} \sum_{j \in N} f_{kj}^s - \sum_{i \in N} f_{ik}^s = \begin{cases} -\sum\limits_{r \in R} q_{rs}, & k = s \\ q_{rs}, & k = r \\ 0, & \text{o.w.} \end{cases} \qquad \forall s \in S, k \in N \tag{3.44}$$

$$x_a = \sum_{s \in S} f_a^s \qquad \forall a \in A \tag{3.45}$$

$$x_a \leq M_3 y_a^* \qquad \forall a \in A_2 \tag{3.46}$$

$$f_{ij}^s \geq 0 \qquad \forall s \in S, (i,j) \in A \tag{3.47}$$

Constraint (3.44) is the flow conservation constraint for each destination node $s \in S$ and each node $k \in N$. This constraint ensures that all flows with destination s, which flow into node k, and the demand of node k with destination s have to flow out of node k. In (3.45) the aggregated flow on a link is computed and (3.46) ensures that only roads which are built can be used. The Big-M M_3 has to be larger than the maximum possible flow of the network.

The leader problem, which only consists of the objective function with a budget constraint, is:

$$\min \sum_{a \in A} x_a t_a(x_a) \tag{3.48}$$

$$\text{s.t.} \sum_{a \in A_2} b_a y_a \leq B \tag{3.49}$$

$$y_a \in \{0,1\} \qquad \forall a \in A_2 \tag{3.50}$$

The objective functions of both problems each contain a linear term and a non-linear term of the form x^5. For the non-linear terms, we use the following piecewise linear approximation, which only requires continuous auxiliary variables.

3.3.2 Linearization of non-linear convex functions

Let $f(x)$ be an increasing, convex and non-linear function. Assume $m+1$ approximation points (ϑ_0, f_0), $(\vartheta_1, f_1), \ldots, (\vartheta_m, f_m)$ with $f_i := f(\vartheta_i)$. Further, $f(x)$ is a function in the single flow variable x_a and $\vartheta_m \geq \max_{a \in A} x_a$ has to hold. The trivial upper bound is $\sum_{r \in R, s \in S} q_{rs}$. However, as x_a can be much smaller than $\sum_{r \in R, s \in S} q_{rs}$, empirical upper bounds can further improve the quality of the approximation. Define $a_i := \frac{f_i - f_{i-1}}{\vartheta_i - \vartheta_{i-1}}$ and $b_i := -\vartheta_{i-1} a_i + f_{i-1}$. Then $f(x)$ can be approximated by the following piecewise linear function:

$$\bar{f}(x) := \begin{cases} a_i x + b_i, & \text{for } x \in [\vartheta_{i-1}, \vartheta_i)\,, i = 1, \dots, m-1 \\ a_i x + b_i, & \text{for } x \in [\vartheta_{i-1}, \infty)\,, i = m \end{cases} \tag{3.51}$$

It is clear that $a_i - a_{i-1} \geq 0$ and Nemhauser and Wolsey (1988) stated that no binary variables are needed for the approximation.

Instead, $\bar{f}(x)$ can be minimized by the following linear program (LP):

$$\min f_0 + a_1 x_1 + \sum_{i=2}^{m} (a_i - a_{i-1}) x_i \tag{3.52}$$

$$\text{s.t. } x_1 \leq x_i + \vartheta_{i-1} \quad i = 2, \dots, m \tag{3.53}$$

$$x_i \geq 0 \quad i = 1, \dots, m \tag{3.54}$$

As in each (ϑ_i, f_i) a new slope a_i starts, we have to add $(a_i - a_{i-1})x_i$ from that point on with $x_i = x_1 - \vartheta_{i-1}$, but do not subtract anything if $x_1 < \vartheta_{i-1}$. Because of the minimization problem and the definition of the objective function constraints, (3.53) and (3.54) ensure that x_i takes the value of $\min\{0, x_1 - \vartheta_{i-1}\}$ and the defined optimization problem minimizes $f(x)$.

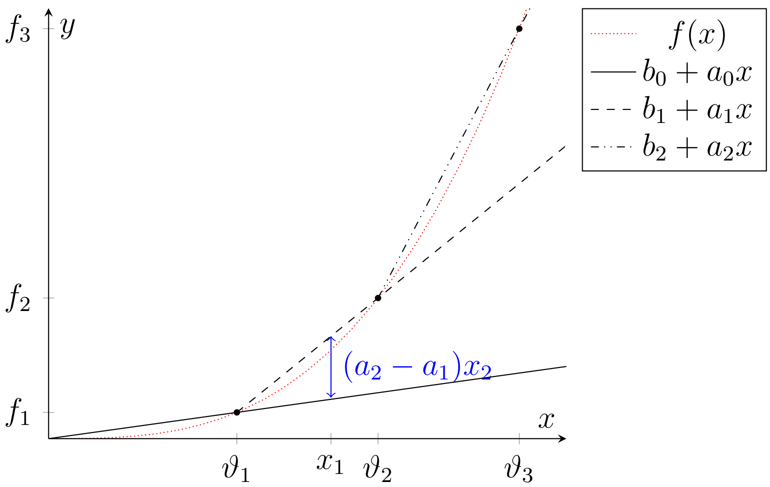


Figure 3.1: Example of an approximation of $f(x)$ with 4 data points

In the example of Figure 3.1, $x_1 > \vartheta_1$ and we have to add $(a_2 - a_1)x_2$ with $x_2 = (x_1 - \vartheta_1)$ (blue line), but $x_3 = 0$.

Imamoto and Tang (2008) proposed a recursive algorithm to find the optimal minimax solution of the piecewise linear approximation of convex functions. In the preprocessing, this algorithm optimizes the x-coordinates such that the maximal approximation error between two adjacent points is minimized.

Applying this piecewise linear approximation to the non-linear objective functions of the DNDP transforms the problem into a LBP without violating the convexity of the objective functions. Therefore, the user equilibrium keeps its unique solution (Sheffi, 1985) and the KKT conditions in Section 3.2 are sufficient (Bard, 1998).

3.4 Numerical study

To show the efficiency of our approach, we used three different examples: a small example (S) taken from Gao et al. (2005), the well-known Sioux-Fall network (M) of LeBlanc (1975) and - as a large scale example - the network of Berlin Mitte Center (L) (Bar-Gera, 2013). In the latter two networks, we used the data of the TAP and added possible new streets to the system in order to have large but also more complex instances, as the complexity also is caused by the number of possible new roads. LeBlanc (1975), Gao et al. (2005) and Luathep et al. (2011) solved the Sioux-Fall network only for 5 possible new roads and (L) was never considered as a DNDP. In the small example, the costs for building all 6 new roads are 70, in (M) we added 10 (15) new roads with total costs of 110 (155) and in the large example, we added 10 potential new links with total building costs of 180. As empirical upper bounds for ϑ_m were not available, we solved the TAP in a preprocessing with several approximations until $\vartheta_m \geq \max_{a \in A} x_a$ was satisfied. The key information on the network sizes is given in Table 3.1.

The reported objective value was calculated by evaluating the non-linear objective function of the leader with the solution obtained from the linearization approximation.

	#nodes	#OD nodes	#arcs	#OD pairs
S	12	2	23	1
M	24	24	86	528
L	398	36	871	1260

	total flow	#new arcs	total costs	ϑ_m
S	20	6	70	11
M	3606	10 (15)	110 (155)	260
L	11481.92	10	180	1400

Table 3.1: Key information of examples

The tests were performed on an Intel(R) Core(TM) i7-2640M CPU, 2.8GHz, 4GB RAM and implemented in Xpress-MP 7.3. Technically, we solved the primal formulations in the slave problem of the BD and got the dual variables through the Xpress functions because the tests showed that the solution time of the primal formulation in our problem is faster than the dual formulation. The calculations were done with several different budgets B and different numbers of approximation points m as introduced in Section 3.3.2. Tables 3.2 - 3.5 show the computational results. The performance of the algorithm was measured by the run time (*time*) and the number of iterations (*iter*) of the BD algorithm. The calculation time of (M) and (L) were distinguished between the total run time (*time total*) and the solution time for the first slave problem (*time step 1*), which shows the complexity of the TAP. Furthermore, the single-level formulation for (M) was solved to compare the run time of the single-level formulation (*SL time*) and the run time for finding the optimal solution in the BD (*opt found BD*) and the single-level formulation (*opt found SL*).

For (S), the leader objective value of the BD algorithm (*obj BD*) was compared with the leader objective value of Gao (*obj Gao*) and the GAP between both values (*GAP*) gives a measure for the quality of the approximation. As no optimal solution for (M) and (L) is reported in the literature, we used the GAP between the approximated objective value and the value of the evaluated non-linear objective function ($\mathrm{GAP_A}$) as a quality measure.

The results show that the run time of the algorithm depends on several factors: More approximation points m and the size of the network increase the run time. Increasing the number of possible new links, which we did in Table 3.4 compared to Table 3.3, increases the number of iterations accompanied with an increased run time. Furthermore, for (M) and (L) the number of iterations and run time increases by increasing the budget B up to 50% of the total costs but starts to decrease again. In the small example, the GAP to the optimal solution is small in all cases and the algorithm ends in the optimal leader decision in all instances. In (M) and (L), the approximation GAP GAP_A decreases by increasing the approximation points but the optimal leader decision does not change for the different approximations.

Furthermore, one can see in Table 3.2 for the small example that the leader objective is oscillating around the objective value by Gao et al. (2005), because we are not only approximating the leader objective, but also the follower objective, which means we are approximating the feasible region of the leader problem. However, further tests showed that this effect is highly dependent on the test instance. Adding additional flows or arcs

B	m	time (sec)	iter	obj BD	obj Gao	GAP (%)
10	10	0.03	3	4088.28	4076.59	0.29
	20	0.03	3	4075.37	4076.59	-0.03
	40	0.06	3	4075.00	4076.59	-0.04
20	10	0.09	8	3959.99	3952.53	0.19
	20	0.12	8	3944.64	3952.53	-0.20
	40	0.17	8	3952.20	3952.53	-0.01
30	10	0.29	12	2754.68	2668.58	3.23
	20	0.30	12	2678.22	2668.58	0.36
	40	0.60	12	2677.94	2668.58	0.35
40	10	0.47	19	2560.18	2524.59	1.41
	20	1.15	19	2520.03	2524.59	-0.18
	40	1.06	19	2527.47	2524.59	0.11
50	10	1.13	23	2397.81	2404.82	-0.29
	20	1.34	23	2406.67	2404.82	0.08
	40	1.42	23	2397.94	2404.82	-0.29
60	10	1.62	28	2314.27	2281.73	1.43
	20	1.85	28	2286.88	2281.73	0.23
	40	2.19	28	2281.58	2281.73	-0.01
70	10	1.68	32	2289.24	2256.96	1.43
	20	1.91	32	2259.99	2256.96	0.13
	40	1.71	32	2255.37	2256.96	-0.07

Table 3.2: Results for example (S)

to the network already reduced this effect for some budgets.

The slower run time for larger networks and more approximation points is related to the larger Traffic Assignment Problem to be solved. As in (L) even the TAP is difficult to solve, the run time is much higher. The number of iterations comes from the complexity of the problem: More possible links enlarge the solution space, but if the budget is small, one only builds 1 or 2 roads and the number of possibilities is small. If the budget is large, almost every road can be built and the decision is easier.

For further acceleration, we applied pareto-optimal cuts as proposed in Magnanti and Wong (1981) to reduce the number of iterations. However, the effort of calculating these cuts could not reduce the total run time. The reason is the increased effort of solving the slave problem.

Besides the comparison with the single level formulation we also compared our algorithm with the algorithm of Bard and Moore (1990). Our algorithm finds the optimal solution significantly faster. Even the case with $m = 10$ and $B = 10$ in (S) could only

B	m	BD time (sec) step 1	BD time (sec) total	iter	SL time (sec)	opt found (sec) BD	opt found (sec) SL	GAP_A (in %)
20	20	0.15	1.63	16	9.15	0.20	3.12	1.739
	40	0.27	2.78	16	21.99	0.35	8.95	1.239
	80	0.64	4.88	15	52.09	0.65	50.99	0.293
	140	1.24	8.89	16	82.35	1.11	78.43	0.098
40	20	0.13	6.27	40	13.46	1.25	6.33	1.461
	40	0.26	11.21	47	19.85	1.91	3.80	0.641
	80	0.63	22.42	51	46.82	4.40	42.43	0.256
	140	1.31	37.93	52	59.38	5.11	56.78	0.055
60	20	0.14	8.41	59	19.22	2.14	19.05	3.086
	40	0.27	14.01	61	16.77	3.45	15.69	1.582
	80	0.69	24.27	60	22.97	6.88	22.49	0.345
	140	1.28	39.88	59	48.70	11.49	47.43	0.073
80	20	0.15	2.24	27	8.05	1.08	1.98	3.097
	40	0.27	4.80	31	14.05	2.94	14.03	1.427
	80	0.64	7.90	26	26.05	5.16	22.91	0.394
	140	1.41	17.80	29	49.10	12.27	48.19	0.083
100	20	0.14	2.51	32	10.46	0.16	10.45	2.974
	40	0.27	4.56	29	6.47	0.31	6.45	1.267
	80	0.65	10.29	34	20.73	0.61	19.98	0.352
	140	1.30	13.88	24	29.72	1.16	28.41	0.125
120	20	0.14	0.26	2	2.05	0.26	2.05	2.745
	40	0.27	0.51	2	8.46	0.51	8.44	0.920
	80	0.65	1.10	2	16.90	1.10	16.85	0.317
	140	1.25	2.86	2	13.07	2.86	13.04	0.127

Table 3.3: Results for example (M) with 10 new links

B	m	BD time (sec) step 1	BD time (sec) total	iter	SL time (sec)	opt found (sec) BD	opt found (sec) SL	GAP_A (in %)
20	20	0.16	3.40	23	8.69	0.30	4.47	1.739
	40	0.32	8.47	27	68.98	0.63	68.45	1.239
	80	1.29	16.64	26	154.04	1.28	150.26	0.293
	140	1.87	19.05	27	417.73	1.41	409.79	0.098
40	20	0.19	24.14	82	14.16	0.59	10.75	2.599
	40	0.37	31.25	82	58.87	0.76	37.73	1.204
	80	0.68	48.34	86	182.49	1.12	180.63	0.165
	140	1.32	76.26	86	375.84	1.77	337.72	0.105
60	20	0.19	19.92	74	14.15	2.15	13.29	2.302
	40	0.38	42.43	86	43.59	4.44	35.49	1.091
	80	1.08	62.50	81	164.84	5.40	156.90	0.280
	140	1.48	92.04	98	399.12	22.54	397.48	0.136
80	20	0.23	20.65	73	19.44	1.70	19.28	5.286
	40	0.45	20.71	59	45.80	4.21	45.62	1.377
	80	0.66	47.49	82	132.34	11.58	130.98	0.488
	140	1.54	59.83	76	326.61	33.85	303.80	0.161
100	20	0.19	8.17	46	15.56	3.38	15.53	5.122
	40	0.33	24.46	89	59.61	2.20	59.58	1.243
	80	0.80	29.10	67	133.41	13.46	133.13	0.519
	140	1.30	72.32	92	339.39	23.58	338.28	0.109
120	20	0.17	6.81	65	16.59	5.66	16.47	4.979
	40	0.30	13.84	66	50.58	9.64	50.32	1.108
	80	0.95	36.21	93	144.01	26.87	141.07	0.543
	140	1.29	61.25	88	453.26	33.41	453.09	0.097
140	20	0.17	3.62	38	12.07	1.33	12.05	5.081
	40	0.31	11.77	55	48.98	3.21	48.93	1.079
	80	0.80	15.11	36	134.50	5.88	132.79	0.473
	140	1.30	49.70	73	389.57	12.25	386.08	0.077
160	20	0.19	0.35	2	3.87	0.35	3.86	5.348
	40	0.34	0.97	2	33.05	0.97	33.03	1.084
	80	0.79	2.01	2	117.43	2.01	117.40	0.467
	140	1.30	42.13	64	323.60	1.32	323.53	0.113

Table 3.4: Results for example (M) with 15 new links

B	m	time step 1 (sec)	time total (min)	iter	GAP_A (in %)
30	40	22	6	17	0.251
	80	85	22	17	0.045
	140	154	43	17	0.017
60	40	25	23	63	0.250
	80	84	109	79	0.053
	140	180	218	79	0.018

Table 3.5: Results for example (L)

be solved in 36 seconds and 881 iterations by the algorithm of Bard and Moore (1990). This is related to the large number of follower variables relative to the number of leader variables. In fact, in the small example, only 2 instances could be solved in less than 1 minute and 50% could not be solved in 1 hour. Tables 3.3 and 3.4 show that the BD reduced the run time on average by 61% and 63%. Moreover, the run time for finding the optimal solution is reduced on average by 84% and 92%. Even though the BD was slightly slower in a few cases, the optimal solution is still found significantly faster.

Finally, we can state that the algorithm always ended in the same decision for building roads without depending on the accuracy of the approximation. This means that the approximation GAP_A of 2% did not influence the leader decision: Moreover, the results of all instances in Tables 3.3 - 3.5 show that as few as 10, 20 resp. 40 approximation points gave a good approximation and the effort of the better approximation does not improve results; but increases run times.

3.5 Conclusion

We proposed a BD method for solving DCLBPs to global optimality. BD is applied to a reformulation of the DCLBP to a single-level MILP and the slave problem solving is further accelerated. The computational results show that this approach, to the best of our knowledge, is the first algorithm solving such large instances for DCLBP and DNDP, even though we just used a basic BD. As the solution time of the slave problem, compared to the master problem, is rather large, the acceleration of solving this problem is of interest for future research.

Our approach enhances computational capabilities of existing (discrete) network design approaches. However, at the same time it has several limitations which are venues for future research. These are given demand matrices rather than elastic demands and road pricing models, discrete potential new roads with given capacity rather than continuous network design problems for large urban networks, and deterministic rather than stochastic user equilibria.

Chapter 4

A dynamic Discrete Network Design Problem for maintenance planning in traffic networks

We propose a dynamic model for network maintenance planning by extending the Discrete Network Design Problem. The leader decides which road in the network is maintained in which period and the follower, as in the Discrete Network Design Problem, optimizes its own path through the network. The non-linear bilevel problem is first linearized and then transformed into a single-level mixed-integer linear program by using the Karush-Kuhn-Tucker conditions. This model is solved with Benders Decomposition. The numerical study shows that this method finds better solutions faster compared to solving the mixed-integer formulation directly and using a genetic algorithm. Furthermore, we show the benefit of this approach compared to simple greedy heuristics.

4.1 Introduction

Many metropolises around the world face the problem of congestion. Moscow and Istanbul in Europe and Rio de Janeiro and Mexico City in America are only some examples of congested cities not only during rush hour. Table 4.1 shows the most congested cities in Europe. The congestion percentage is the average travel time in the city compared to the free flow travel time in 2013 (TomTom, 2014). In 48 European cities the average travel time is 20% or more above a free flow network and in 40 cities the congestion index in the morning peak is 40% or higher (TomTom, 2014).

Moreover, the deterioration of the streets decreases the quality of the whole network and causes even more congestion. Therefore, streets and bridges in a network need to

Rank	City	Country	Congestion	Morning peak	Evening peak
1	Moscow	Russia	74%	111%	141%
2	Istanbul	Turkey	62%	87%	129%
3	Palermo	Italy	39%	60%	64%
4	Warsaw	Poland	39%	71%	75%
5	Rome	Italy	37%	71%	64%

Table 4.1: Congestion in European cities 2013 (TomTom, 2014)

be maintained from time to time. The German interstate and federal highway network consists of roughly 39,000 bridges and the Federal Ministry of Transport and Digital Infrastructure is planning to invest 1.06 billion Euro in bridge maintenance projects with an investment volume of more than 5 million Euro between 2015 and 2017 (BMVI, 2014a). The German railway company DB is going to invest 28 billion Euro over the next 5 years for maintenance: Among other things, 875 bridges will be renewed. This will add up to 850 construction zones per day (DB, 2014). In Munich, 600 construction zones are planned for 2015 (Völklein, 2015). During these maintenance phases, the capacity of streets can be less or streets have to be closed, which leads to congestion and longer travel times for the user during these periods, but decreases travel time afterwards. Furthermore, the repair costs can increase over the years if you postpone the maintenance.

We formulate a non-linear bilevel model for this problem and introduce the Dynamic Discrete Network Design Problem (DDNDP), as for each period in the planning horizon a DNDP is solved. As in the DNDP, which was first introduced by LeBlanc (1975), the congestion in a network is minimized by finding the optimal maintenance plan of the network under the assumption that all travelers in a network minimize their own travel time from origin to destination. Braess et al. (2005) showed that the objective of minimizing congestion and the goal of minimizing travel time are not necessarily the same. Similar to the DNDP, we are dealing with a hierarchical decision process (Schneeweiß, 2003) and different objective functions. Therefore, we need a bilevel formulation for the DDNDP as well.

The contribution of this chapter is first the introduction of a bilevel model for the DDNDP that calculates a long-term schedule for the maintenance work, minimizes the congestion in the network but is restricted to a certain budget per year. Second, we use a BD algorithm for solving the linearized problem by approximating the travel time function. We show that a terminated BD algorithm that is stopped after an iteration limit finds good solutions in reasonable time. Moreover, we show that simple greedy heuristics used in practice, which are based on the road reliability and the current road congestion, cannot compete with our approach especially in fulfilling budget restrictions.

The remainder of this chapter is structured as follows. In Section 4.2, we introduce the bilevel formulation for the Dynamic Discrete Network Design Problem. In the following section, the model is reformulated into a MILP and solved with BD. In Section 4.4, the model is tested in a numerical study and the chapter ends with conclusions and an outlook on further research.

4.2 Mathematical model

The network is defined by a set of nodes N and a set of arcs A, which are the roads of the network. The OD matrix defines the demand q_{rs} for all possible origins $r \in R \subseteq N$ and destinations $s \in S \subseteq N$. We consider a set T of periods (for example years) as planning horizon, in which the maintenance work has to be done. Furthermore, we have a subset of arcs $A_2 \subset A$ which have to be maintained. For modeling purposes, we additionally add a set of arcs $A_2^M \subset A$ and a set $A_2^N \subset A$, which represent the same links as A_2 during and after maintenance. Two mapping functions $f^M : A_2 \to A_2^M$ and $f^N : A_2 \to A_2^N$ link these arcs: $f^M(a)$ represents link a during maintenance and $f^N(a)$ represents link a after maintenance (Figure 4.1). Further, let $\bar{A}_2 = A_2 \cup A_2^M \cup A_2^N$. Therefore, exactly one arc $(i,j) \in A_2$ or $(i,j) \in A_2^M$ or $(i,j) \in A_2^N$ is open for traffic in the network in a specific period.

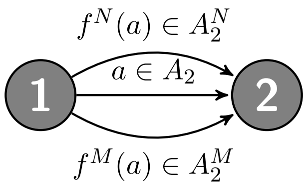


Figure 4.1: Example of modelling an arc of A_2

For each arc $a \in A_2$, there exists a due date l_a until which it has to be maintained. The costs are b_a^t for maintaining link a in period t. Moreover, there is a budget B for every period. The capacity of an arc is defined by an initial value c_a and a reliability index profile r_a^t. As shown in Ng et al. (2009), this reliability gives the percentage of used capacity in each period. This reliability is reduced from period to period, which is associated with longer travel time, and increased after maintenance. Furthermore, T_a is the free flow travel time, K_a the congestion influence parameter and c_a the capacity limit for each arc a. Including the reliability index in the BPR function (Bureau of Public Roads, 1964) gives the following modified travel time function for the travelers on arc x:

$$h_a^t(x) := T_a \left(1 + K_a \left(\frac{x}{c_a r_a^t}\right)^4\right)$$

With the binary decision variables y_a^t, the leader decides on the maintenance period of each arc $a \in A_2$ subject to the budget in each period. $y_a^t = 1$ decides for $a \in A_2$ that a is not yet maintained in period t, for $a \in A_2^M$ that a is maintained in period t and for $a \in A_2^N$ that a was maintained before period t. The follower decides on the continuous variables $f_a^{st} = f_{ij}^{st}$ representing the flow on arc $a = (i,j)$ to destination s in period t and on the cumulative flow variable x_a^t on arc a in period t. The leader minimizes the system-optimum - the total congestion in the network - over all periods and the follower, the travelers in the network, minimizes the sum of the travel times for all origin-destination pairs in each period. This is called the user-optimum.

The leader problem is given by

$$\min \sum_{a\in A, t\in T} x_a^t h_a^t(x_a^t) \tag{4.1}$$

$$s.t. \sum_{a\in A_2} b_a^t y_a^t \leq B \qquad \forall t \in T \tag{4.2}$$

$$\sum_{t\in T, t\leq l_a} y_a^t = 1 \qquad \forall a \in A_2^M \tag{4.3}$$

$$y_a^t + y_{a'}^t + y_{a''}^t = 1 \qquad \forall a \in A_2, a' = f^M(a), a'' = f^N(a), t \in T \tag{4.4}$$

$$y_a^{t-1} \leq y_a^t + y_{a'}^t \qquad \forall a \in A_2, a' = f^M(a), t \in T \setminus \{1\} \tag{4.5}$$

$$y_{a'}^{t-1} + y_{a''}^{t-1} \leq y_{a''}^t \qquad \forall a \in A_2, a' = f^M(a), a'' = f^N(a), t \in T \setminus \{1\} \tag{4.6}$$

$$y_a^t \in \{0,1\} \qquad \forall a \in \bar{A}_2, t \in T \tag{4.7}$$

In (4.2), the budget is limited for each period and (4.3) ensures that all arcs are maintained before a certain deadline. (4.4) ensures that an arc is either unmaintained or currently maintained or has already been renewed. Equation (4.5) ensures that arc a can either be maintained ($a \in A_2^N$) or stay unmaintained ($a \in A_2$) in period t if it is unmaintained in period $t-1$. Equation (4.6) guarantees that a renewed arc $a \in A_2^N$ has to be used in period t if it was maintained in period $t-1$ ($a \in A_2^M$) or earlier ($a \in A_2^N$).

The follower problem is an uncapacitated TAP (Nguyen, 1974). This can be formulated for fixed leader decisions $(y_a^t)^*$ as follows (Poorzahedy and Turnquist, 1982):

$$\min \sum_{a\in A,t\in T} \int_0^{x_a^t} h_a^t(x)dx = \min \sum_{a\in A,t\in T} T_a x_a^t \left(1 + \frac{K_a}{5c_a^4 r_a^t}(x_a^t)^4\right) \tag{4.8}$$

$$\text{s.t.} \sum_{j\in N} f_{kj}^{st} - \sum_{i\in N} f_{ik}^{st} = \begin{cases} -\sum\limits_{r\in R} q_{rs}, & k = s \\ q_{rs}, & k = r \\ 0, & \text{o.w.} \end{cases} \qquad \forall s \in S, k \in N, t \in T \tag{4.9}$$

$$x_a^t = \sum_{s\in S} f_a^{st} \qquad \forall a \in A, t \in T \tag{4.10}$$

$$x_a^t \leq M(y_a^t)^* \qquad \forall a \in \bar{A}_2, t \in T \tag{4.11}$$

$$f_{ij}^{st} \geq 0 \qquad \forall s \in S, (i,j) \in A, t \in T \tag{4.12}$$

Constraint (4.9) is a flow conservation constraint that also satisfies the demand. In (4.10), the aggregated flow on a link is computed and (4.11) ensures with a large positive number M that only links are used which are open during that period. The maximum flow on an arc is restricted by the total demand of the network. Therefore, M can be defined as $\sum_{r\in R,s\in S} q_{rs}$. Obviously, this problem can be decomposed into $|T|$ subproblems for every period.

Both objective functions (4.1) and (4.8) contain a linear term and a non-linear term x^5. In the following section, we linearize the non-linear term with a piecewise approximation.

4.3 Solution algorithm

The non-linear bilevel problem of Section 4.2 is solved with the BD approach by Fontaine and Minner (2014). The idea of this algorithm is to linearize the model first (Section 4.3.1). This LBP is transformed into an equivalent mixed integer program and solved via BD (Section 4.3.2). Further, the slave problem is decomposed into $|T|$ subproblems.

4.3.1 Transformation to a linear bilevel problem

To transform the non-linear bilevel problem into an LBP, the leader and follower objective functions are linearized (Fontaine and Minner, 2014). $f(x) = x^5$, which is an increasing, convex and non-linear function, is approximated by piecewise linear terms. This piecewise linear function is divided into m linear segments and represented by $m+1$

data points $(\vartheta_0, f_0), (\vartheta_1, f_1), \ldots, (\vartheta_m, f_m)$ with $f_i := f(\vartheta_i)$. An upper bound for ensuring $\vartheta_m \geq \max_{a\in A, t\in T} x_a^t$ is given by $\sum_{r\in R, s\in S} q_{rs}$. Nevertheless, empirical upper bounds can reduce the number of approximation points or improve the approximation. In practice, the demand often splits up and the flow on an arc x_a^t can be much smaller than the cumulated demand of the whole network.

With $a_i := \frac{f_i - f_{i-1}}{\vartheta_i - \vartheta_{i-1}}$ and $b_i := -\vartheta_{i-1}a_i + f_{i-1}$, $f(x)$ can be approximated by the following piecewise linear function:

$$\bar{f}(x) := \begin{cases} a_i x + b_i, & \text{for } x \in [\vartheta_{i-1}, \vartheta_i)\,, i = 1, \ldots, m-1 \\ a_i x + b_i, & \text{for } x \in [\vartheta_{i-1}, \infty)\,, i = m \end{cases} \tag{4.13}$$

Let $I = 2, \ldots, m$. Due to the convexity, no binary auxiliary variables are needed (Nemhauser and Wolsey, 1988) for the approximation and we can minimize $\bar{f}(x)$ by the following LP:

$$\min f_0 + a_1 x_1 + \sum_{i\in I}(a_i - a_{i-1})x_i \tag{4.14}$$

$$\text{s.t. } x_1 \leq x_i + \vartheta_{i-1} \quad i \in I \tag{4.15}$$

$$x_i \geq 0 \quad i \in I \tag{4.16}$$

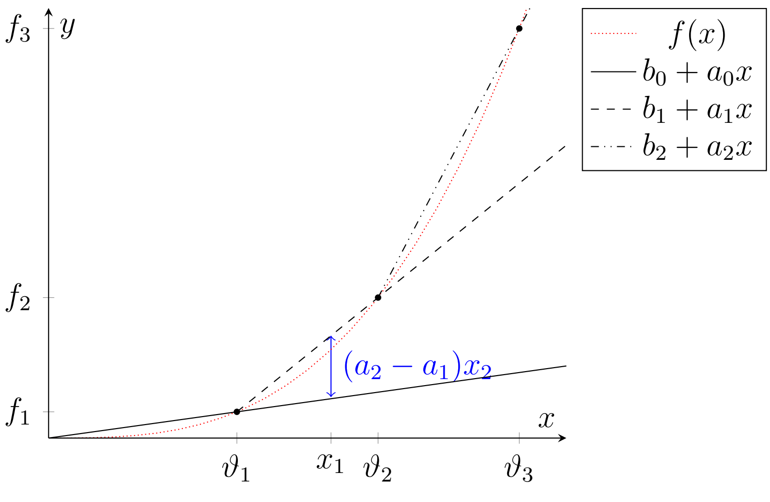


Figure 4.2: Example of an approximation of $f(x)$ with 4 data points

Figure 4.2 illustrates the idea of the approximation. As this is a minimization problem, x_i takes value $\max(x_1 - \vartheta_i, 0)$ and adds the difference between the two slopes a_i and a_{i-1} (blue line) to the objective function and $f(x_1)$ is approximated. For $\vartheta_2 < x_1 \leq \vartheta_3$, both

$(a_2 - a_1)x_2$ and $(a_3 - a_2)x_3$ are added to the objective value.

In the DDNDP, the travel time of each arc $a \in A$ is approximated in every period $t \in T$. Therefore, the continuous auxiliary variable x^t_{ia} is introduced for all $i \in I$, $a \in A$ and $t \in T$. This piecewise linear approximation transforms the DDNDP into an LBP. However, the user equilibrium keeps its unique solution (Sheffi, 1985) as the convexity of the objective functions is not violated.

4.3.2 Benders decomposition

The follower problem is replaced by its KKT conditions. Bard (1998) showed that these conditions are sufficient if the follower solution is unique. For the KKT conditions, we introduce the dual variables u^t_{sk}, v^t_a and w^t_{ia} for the follower constraints (4.20) - (4.22), the dual constraints (4.24) and (4.25) to the primal follower variables f^{st}_{ij} and x^t_{ia} and a large positive number M_2. Because of the duality theorem, equation (4.23) guarantees the optimality of the follower problem through the KKT conditions while the leader's objective is optimized. We further define the indicator function χ_a, which is 1 for all $a \in \bar{A}_2$ and 0 otherwise and

$$C^t_{1a} := T_a\left(1 + \frac{K_a}{c_a r^t_a} a_1\right) \qquad C^t_{ia} := \frac{T_a K_a}{c_a r^t_a}\left(a_i - a_{i-1}\right) \tag{4.17}$$

$$\hat{C}^t_{1a} := T_a\left(1 + \frac{K_a}{5c_a r^t_a} a_1\right) \qquad \hat{C}^t_{ia} := \frac{T_a K_a}{5c_a r^t_a}\left(a_i - a_{i-1}\right) \tag{4.18}$$

for all $a \in A$, $i \in I$ and $t \in T$

$$\min \sum_{t\in T}\left(\sum_{\substack{a\in A\\ s\in S}} C^t_{1a} f^{st}_a + \sum_{\substack{i\in I\\ a\in A}} C^t_{ia} x^t_{ia}\right) \tag{4.19}$$

s.t.

(4.2) – (4.7)

$$\sum_{j\in N} f^{st}_{kj} - \sum_{i\in N} f^{st}_{ik} = q_{ks} \qquad \forall s \in S, k \in N, t \in T \tag{4.20}$$

$$\sum_{s\in S} f^{st}_a \le M y^{t*}_a \qquad \forall a \in \bar{A}_2, t \in T \tag{4.21}$$

$$\sum_{s\in S} f^{st}_a \le x^t_{ia} + \vartheta_{i-1} \qquad \forall a \in A, i \in I, t \in T \tag{4.22}$$

$$\sum_{\substack{a\in A\\ s\in S}} \hat{C}^t_{1a} f^{st}_a + \sum_{\substack{i\in I\\ a\in A}} \hat{C}^t_{ia} x^t_{i,a} \leq \sum_{\substack{s\in S\\ k\in N}} q_{k,s} u^t_{ks} + \sum_{a\in \bar{A}_2} M v^t_a y^t_a + \sum_{\substack{i\in I\\ a\in A}} w^t_{ia} \vartheta_{i-1} \qquad \forall t \in T \tag{4.23}$$

$$u^t_{sk} - u^t_{sj} + \chi_a v^t_a + \sum_{i\in I} w^t_{ia} \leq \hat{C}^t_{1a} \qquad \forall s \in S, a=(k,j)\in A, t\in T \tag{4.24}$$

$$-w^t_{ia} \leq \hat{C}^t_{ia} \qquad \forall a\in A, i\in I, t\in T \tag{4.25}$$

$$f^{st}_a \geq 0 \qquad \forall s\in S, a\in A, t\in T \tag{4.26}$$

$$x^t_{ia} \geq 0 \qquad \forall i\in I, a\in A, t\in T \tag{4.27}$$

$$u^t_{sk} \in \mathbb{R} \qquad \forall s\in S, k\in N, t\in T \tag{4.28}$$

$$v^t_a \leq 0 \qquad \forall a\in \bar{A}_2, t\in T \tag{4.29}$$

$$w^t_{ia} \leq 0 \qquad \forall i\in I, a\in A, t\in T \tag{4.30}$$

The bilinear term (4.23) in the optimality condition following from duality theory is linearized by introducing the auxiliary variables μ^t_a and replacing constraint (4.23) with the following set of equations (Cao and Chen, 2006; Farvaresh and Sepehri, 2011). The newly introduced decision variables μ^t_a are forced to take the value 0 if $y^t_a = 0$ and v^t_a if $y^t_a = 1$.

$$\sum_{\substack{a\in A\\ s\in S}} \hat{C}^t_{1a} f^{st}_a + \sum_{\substack{i\in I\\ a\in A}} \hat{C}^t_{ia} x^t_{i,a} \leq \sum_{\substack{s\in S\\ k\in N}} q_{k,s} u^t_{ks} + \sum_{a\in \bar{A}_2} M \mu^t_a + \sum_{\substack{i\in I\\ a\in A}} w^t_{ia} \vartheta_{i-1} \qquad \forall t \in T \tag{4.31}$$

$$\mu^t_a \leq v^t_a + M_2(1-y^t_a) \qquad \forall a\in \bar{A}_2, t\in T \tag{4.32}$$

$$\mu^t_a \geq v^t_a \qquad \forall a\in \bar{A}_2, t\in T \tag{4.33}$$

$$\mu^t_a \geq -M_2 y^t_a \qquad \forall a\in \bar{A}_2, t\in T \tag{4.34}$$

$$\mu^t_a \leq 0 \qquad \forall a\in \bar{A}_2, t\in T \tag{4.35}$$

This MILP is solved with BD (Benders, 1962). BD devides the decision variables into complicating and non-complicating variables. In the DDNDP, the complicating variables are the binary variables y^t_a and the non-complicating ones are the continuous variables f^{st}_{ij}, x^t_{ia}, u^t_{sk}, v^t_a, μ^t_a and w^t_{ia}. The problem is decomposed into the master problem, which calculates a new feasible solution for the y^t_a in each iteration, and the slave problem, which solves the MILP (4.19), (4.20) - (4.22), (4.24) - (4.30), (4.31) - (4.35) for the y_a^t calculated in the master problem being fixed and generates an optimality cut or a feasibility cut for the master problem. Due to the problem structure of $|T|$ independent TAPs as follower problem, the slave problem can be decomposed into $|T|$ independent subproblems which are solved in parallel for each $t \in T$. Algorithm 3 shows an outline of the BD algorithm.

Algorithm 3: Benders decomposition: Main structure

1 Set upper bound $UBD = \infty$ and y^* to any feasible solution of y.
2 Solve the subproblems of the slave problem for $y = y^*$ in parallel. Let z_S be the optimal solution value of the slave problem. Update the upper bound $UBD = \min\{UBD, z_S\}$. If the slave problem is bounded, then add an optimality cut to the master problem, else add a feasibility cut to the master problem.
3 Solve the master problem and save the solution y^*. Let z_M be the optimal solution value of the solved master problem.
4 If stopping criterion is true: stop, else: go to Step 2.

We used BD as a heuristic, which stops after a predefined time limit. In other words: the algorithm stops after a predefined time limit if the classical stopping criterion $UBD - z_M < tolerance$ is not met before.

Dual slave problem

The slave problem is defined by minimizing (4.19) subject to (4.20) - (4.22), (4.24) - (4.30), (4.31) - (4.35) for fixed decision variables y^*. This problem is decomposed into $|T|$ subproblems and each subproblem t is dualized. The dual variables α_{sk}^t, β_a^t, γ_{ia}^t, ν_{sa}^t, τ_{ia}^t, δ^t, λ_a^{1t}, λ_a^{2t} and λ_a^{3t} correspond to constraints (4.20) - (4.22), (4.24) - (4.25), (4.31) - (4.34) and the dual constraints (4.37) - (4.42) to the primal variables f_a^{st}, x_{ia}^t, u_{sk}^t, v_a^t, μ_a^t and w_{ia}^t.

$$\max \sum_{\substack{s\in S\\k\in N}} q_{ks}\alpha_{sk}^t + \sum_{a\in\bar{A}_2} \beta_a^t + \sum_{\substack{a\in A\\i\in I}} \vartheta_{i-1}\gamma_{ia}^t + \sum_{\substack{a\in A\\s\in S}} \hat{C}_{ia}^t \nu_{sa}^t + \sum_{\substack{a\in A\\i\in I}} \hat{C}_{ia}^t \tau_{ia}^t + \sum_{a\in\bar{A}_2} \left(M_2(1-y_a^{t*})\lambda_a^{1t} - M_2 y_a^{t*}\lambda_a^{3t}\right) \tag{4.36}$$

$$\text{s.t. } \alpha_{sk}^t - \alpha_{sj}^t + \chi_a\beta_a^t + \sum_{i\in I}\gamma_{ia}^t + \hat{C}_{1a}^t\delta^t \le C_{1a}^t \qquad \forall s \in S, a=(k,j)\in A \tag{4.37}$$

$$-\gamma_{ia}^t + \hat{C}_{ia}^t\delta^t \le C_{ia}^t \qquad \forall a\in A, i\in I \tag{4.38}$$

$$- q_{ks}\delta^t + \sum_{j \in N} \nu_{kj}^{st} - \sum_{i \in N} \nu_{ik}^{st} = 0 \qquad \forall s \in S, k \in N \tag{4.39}$$

$$- \lambda_a^{1t} - \lambda_a^{2t} + \sum_{s \in S} \nu_{sa}^t \geq 0 \qquad \forall a \in A \tag{4.40}$$

$$\lambda_a^{1t} + \lambda_a^{2t} + \lambda_a^{3t} - M\delta^t \geq 0 \qquad \forall a \in A \tag{4.41}$$

$$- \vartheta_{i-1}\delta^t + \sum_{s \in S} \nu_{sa}^t - \tau_{ia}^t \geq 0 \qquad \forall i \in I, a \in A \tag{4.42}$$

$$\alpha_{sk}^t \in \mathbb{R} \qquad \forall s \in S, k \in N \tag{4.43}$$

$$\beta_a^t, \lambda_a^{1t} \leq 0 \qquad \forall a \in \bar{A}_2 \tag{4.44}$$

$$\lambda_a^{2t}, \lambda_a^{3t} \geq 0 \qquad \forall a \in \bar{A}_2 \tag{4.45}$$

$$\delta^t \leq 0 \tag{4.46}$$

$$\nu_{sa}^t \leq 0 \qquad \forall a \in A, s \in S \tag{4.47}$$

$$\gamma_{ia}^t, \tau_{ia}^t \leq 0 \qquad \forall i \in I, a \in A \tag{4.48}$$

If this problem is feasible for all $t \in T$ with a solution α^*, β^*, γ^*, δ^*, ν^*, τ^* and λ^*, we add the following optimality cut

$$\sum_{t \in T} \Bigg(\sum_{\substack{s \in S \\ k \in N}} q_{ks}\alpha_{sk}^{t*} + \sum_{a \in \bar{A}_2} \beta_a^{t*} + \sum_{\substack{a \in A \\ i \in I}} \vartheta_{i-1}\gamma_{ia}^{t*} + \sum_{\substack{a \in A \\ s \in S}} \hat{C}_{ia}^t \nu_{sa}^{t*} + \sum_{\substack{a \in A \\ i \in I}} \hat{C}_{ia}^t \tau_{ia}^{t*} + \sum_{a \in \bar{A}_2} \left(M_2(1 - y_a^t)\lambda_a^{1t*} - M_2 y_a^t \lambda_a^{3t*} \right) \Bigg) \leq z \tag{4.49}$$

to the master problem. If the problem is unbounded for $t \in T$, the objective of the dual slave problem is bounded by adding the following constraint:

$$\sum_{\substack{s \in S \\ k \in N}} q_{ks}\alpha_{sk}^t + \sum_{a \in \bar{A}_2} \beta_a^t + \sum_{\substack{a \in A \\ i \in I}} \vartheta_{i-1}\gamma_{ia}^t + \sum_{\substack{a \in A \\ s \in S}} \hat{C}_{ia}^t \nu_{sa}^t + \sum_{\substack{a \in A \\ i \in I}} \hat{C}_{ia}^t \tau_{ia}^t + \sum_{a \in \bar{A}_2} \left(M_2(1 - y_a^{t*})\lambda_a^{1t} - M_2 y_a^{t*} \lambda_a^{3t} \right) \leq M_3 \tag{4.50}$$

with M_3 being a large enough number to bound the extreme rays without cutting off any extreme points. The slave problem is solved and the solution generates a feasibility cut, which is added to the master problem:

$$\sum_{t\in T}\left(\sum_{\substack{s\in S\\k\in N}} q_{ks}\alpha_{sk}^{t*} + \sum_{a\in\bar{A}_2}\beta_a^{t*} + \sum_{\substack{a\in A\\i\in I}}\vartheta_{i-1}\gamma_{ia}^{t*} + \sum_{\substack{a\in A\\s\in S}}\hat{C}_{ia}^{t}\nu_{sa}^{t*} + \sum_{\substack{a\in A\\i\in I}}\hat{C}_{ia}^{t}\tau_{ia}^{t*} + \right.$$
$$\left.\sum_{a\in\bar{A}_2}\left(M_2(1-y_a^t)\lambda_a^{1t*} - M_2 y_a^t\lambda_a^{3t*}\right)\right) \leq 0 \tag{4.51}$$

Master problem

With C_O being the set of optimality cuts and C_F being the set of feasibility cuts, the BD algorithm adds the generated constraint of the slave problem to the respective set in each iteration. The master problem can be defined as follows:

$$
\begin{aligned}
&\min\ z && && (4.52)\\
&\text{s.t.}\ (4.2)-(4.7) && && \\
&\quad c_o && \forall c_o \in C_O && (4.53)\\
&\quad c_f && \forall c_f \in C_F && (4.54)\\
&\quad y_a^t \in \{0,1\} && \forall a \in \bar{A}_2, t \in T && (4.55)\\
&\quad z \in \mathbb{R} && && (4.56)
\end{aligned}
$$

4.4 Numerical study

4.4.1 Test setup

The proposed BD algorithm was tested in a numerical study and compared with the MILP formulation solved with branch-and-bound (B&B) on a commercial solver and with a genetic algorithm (GA). Moreover, two simple greedy heuristics were used as a benchmark from practice. The methods were tested on the well-known Sioux-Fall network of LeBlanc (1975) and an extended Sioux-Fall network originally proposed by Farvaresh and Sepehri (2013) that is four times larger. The BD and the B&B were executed in Xpress-MP 7.6 with default settings. The genetic algorithm was implemented in Matlab R2014a. The chromosomes of the GA were defined as an integer vector of the length $|A_2|$. Each point in the vector takes values from 1 to $|T|$, which defines the period of maintenance, and all violations of the budget constraint were penalized in the objective function. The chromosomes were evaluated by solving $|T|$ TAPs using the Xpress interface for Matlab. The first greedy heuristic selects the arcs with the lowest reliability first (LRF), while the second greedy heuristic selects them according to the highest congestion first (HCF) until the budget of a period is exceeded.

The tests were performed on an Intel(R) Core(TM) i7-3770 CPU, 3.4GHz, 32GB RAM with 8 threads. All models used all 8 threads: The BD solved the subproblems in parallel, the B&B used parallel computing method for solving the MILP and the GA evaluated the chromosomes in parallel. The approximation points are calculated in the preprocessing according to Imamoto and Tang (2008). A recursive method calculates the x-coordinates in such a way that the maximal approximation error between two adjacent points is minimized.

4.4.2 Sioux-Fall network

We used the TAP data of the well-known Sioux-Fall network of LeBlanc (1975) and generated the additionally needed data. The network consists of 24 nodes, which are all origin and destination nodes, and 76 arcs. There are 528 OD-pairs with a total flow of 3,606 in the network. The planning horizon was 10 periods.

We relaxed constraint (4.3) and had no deadlines; the objects only need to be maintained during the planning horizon. The reliability was randomly generated between 60% and 100% in period 1 and decreases in every period. Further, we assume that the maintenance of every road takes one period and the reliability of the link during that period is reduced by 50% ($\frac{r_a^t}{2} = r_{a'}^t$ for $a \in A_2$ and $f^M(a) = a' \in A_2^M$). After maintenance, the reliability increases by 15%. This reliability development was adapted from Ng et al. (2009).

We tested the network with $|A_2| = 38$ and $|P| = 10$. The costs for maintaining all roads in the first period sum up to 420 and increase between 3% and 6% in every period. We used three different budgets to evaluate the algorithm ($B = 50$, $B = 75$ and $B = 100$). For the approximation, we used $m \in \{20, 40, 60\}$ data points. Fontaine and Minner (2014) showed that a good approximation for solving the Sioux-Fall network without maintenance can already be found with 20 and 40 data points. Furthermore, the objective value of an optimal network, in which every road has already been maintained in every period, is $5.89bn$ ($5.71bn$ for $m = 40$ and $5.68bn$ for $m = 60$), which gives a lower bound for the optimization problem. These values give some lower bounds and show the effect of maintenance

We tested the genetic algorithm with different parameter settings. The crossover fraction and the population size had the highest impact on the solution quality. Table 4.2 shows the parameter setting that returned the best results and is used for the numerical study in this chapter.

parameter	value
population size	50
fitness scaling	rank
selection	stochastic uniform
elite count	2
crossover fraction	0.7
crossover	scattered
mutation	uniform (0.01)

Table 4.2: Genetic algorithm: Parameters

Moreover, the time limit was set to one hour for all methods in this setting.

Results

Table 4.3 shows the objective values for the solution methods. The GA solved each problem instance 20 times and Figures 4.3a - 4.4f show the average values of the algorithm. Moreover, the solutions were evaluated in the non-linear objective function and shown in angle brackets under the solution. The approximation gaps are similar to the ones in Fontaine and Minner (2014). For a budget of $B = 50$, the B&B could not find any solution, neither the genetic algorithm nor both heuristics (LRF and HCF) could compete with the BD in both cases. The solution space in these instances is very small, as the budget is close to the minimal required budget. Therefore, the GA and the greedy approaches had problems finding feasible solutions which do not violate the budget constraints. Hence, we further allowed the use of left over budget from previous periods in the greedy heuristic to generate a solution. In Table 4.3, the superscript in round brackets indicate indicates the number of periods where the budget limit was exceeded. The BD is the only method that produced feasible solutions. Moreover, these solutions have better objective values than the other methods even though no constraint is violated.

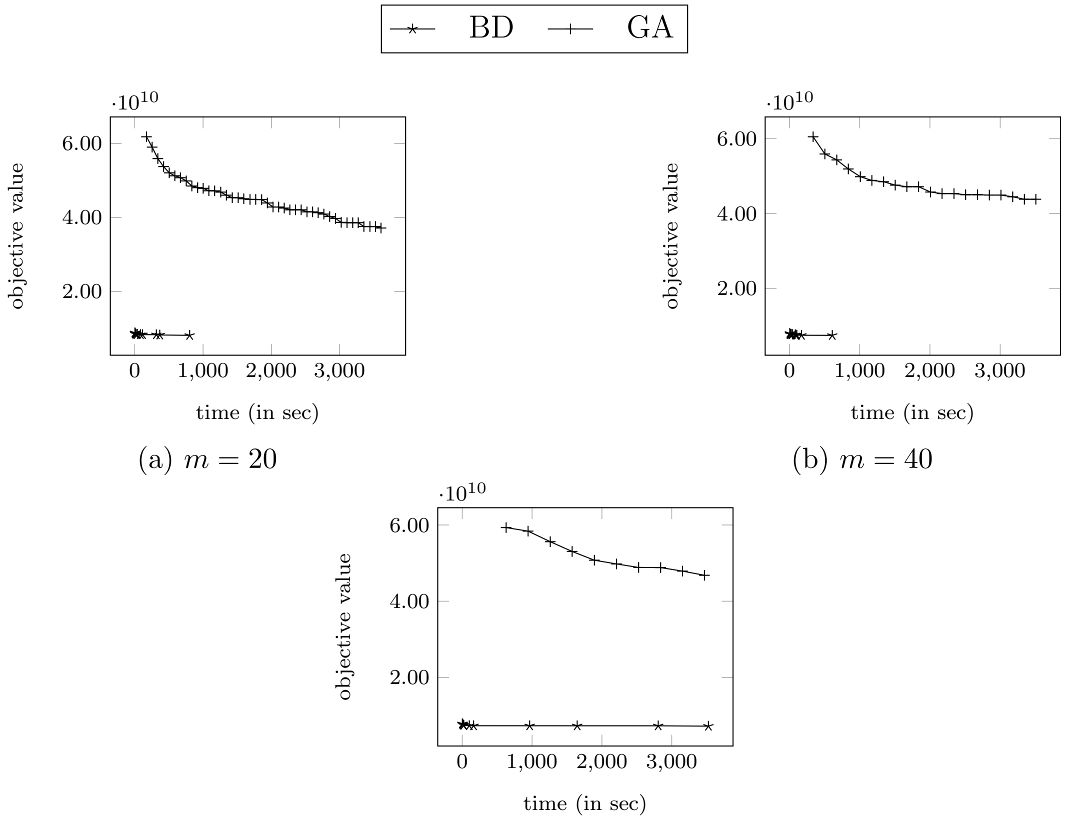
In the other two budget cases, the B&B always performed worst. The LRF heuristic also finds good solutions for a budget of $B = 75$. In fact, those solutions even outperform the GA. However, for a budget of $B = 100$, the order is reversed. The HCF, which had the second best results for a budget of $B = 50$, is the worst for the other two cases. These

B	m	BD	LRF	HCF	B&B	GA best	GA worst	GA mean
50	20	8.07 $\langle 7.60\rangle$	$8.49^{(5)}$ $\langle 7.88\rangle$	$8.42^{(3)}$ $\langle 7.90\rangle$	-	$8.28^{(3)}$ $\langle 7.78\rangle$	$8.76^{(7)}$ $\langle 8.32\rangle$	$8.53^{(4.7)}$ $\langle 8.02\rangle$
	40	7.39 $\langle 7.29\rangle$	$7.63^{(5)}$ $\langle 7.55\rangle$	$7.60^{(3)}$ $\langle 7.54\rangle$	-	$7.52^{(3)}$ $\langle 7.45\rangle$	$8.00^{(7)}$ $\langle 7.93\rangle$	$7.73^{(4.8)}$ $\langle 7.65\rangle$
	60	7.18 $\langle 7.14\rangle$	$7.51^{(5)}$ $\langle 7.49\rangle$	$7.49^{(3)}$ $\langle 7.46\rangle$	-	$7.35^{(3)}$ $\langle 7.31\rangle$	$7.80^{(7)}$ $\langle 7.77\rangle$	$7.59^{(4.9)}$ $\langle 7.55\rangle$
75	20	8.00 $\langle 7.53\rangle$	8.05 $\langle 7.52\rangle$	8.39 $\langle 7.88\rangle$	8.20 $\langle 7.77\rangle$	8.02 $\langle 7.48\rangle$	8.27 $\langle 7.74\rangle$	8.13 $\langle 7.64\rangle$
	40	7.18 $\langle 7.1\rangle$	7.28 $\langle 7.21\rangle$	7.62 $\langle 7.55\rangle$	7.68 $\langle 7.61\rangle$	7.26 $\langle 7.17\rangle$	7.56 $\langle 7.47\rangle$	7.40 $\langle 7.31\rangle$
	60	7.06 $\langle 7.03\rangle$	7.18 $\langle 7.16\rangle$	7.52 $\langle 7.48\rangle$	7.38 $\langle 7.35\rangle$	7.27 $\langle 7.21\rangle$	7.46 $\langle 7.43\rangle$	7.34 $\langle 7.31\rangle$
100	20	7.80 $\langle 7.30\rangle$	8.02 $\langle 7.45\rangle$	8.34 $\langle 7.72\rangle$	8.19 $\langle 7.72\rangle$	7.74 $\langle 7.24\rangle$	8.00 $\langle 7.58\rangle$	7.91 $\langle 7.44\rangle$
	40	7.10 $\langle 7.01\rangle$	7.19 $\langle 7.17\rangle$	7.53 $\langle 7.46\rangle$	7.45 $\langle 7.32\rangle$	7.18 $\langle 7.04\rangle$	7.36 $\langle 7.26\rangle$	7.25 $\langle 7.15\rangle$
	60	6.97 $\langle 6.93\rangle$	7.12 $\langle 7.10\rangle$	7.43 $\langle 7.41\rangle$	7.69 $\langle 7.66\rangle$	7.08 $\langle 7.06\rangle$	7.28 $\langle 7.24\rangle$	7.18 $\langle 7.15\rangle$

Table 4.3: Objective values for Sioux-Fall network (in 10^9)

results indicate that, in general, it is beneficial to maintain links with a low reliability fast. However, the disadvantage of these heuristics is the ineffective usage of the budget and the uncoordinated maintenance plan. Regions with several maintenance works can cause highly congested areas.

Figures 4.3 and 4.4 show the improvement of the objective values for the BD, the GA and the B&B with a Budget of 50, 75 and 100 and an approximation of $m = 20$, $m = 40$ and $m = 60$. In general, the BD has a lot of improvements in the beginning and finds the best solution very fast. These results correspond to the results in Fontaine and Minner (2014). In the DNDP, the BD even found the optimal solution in few iterations, but the proof of optimality took many iterations. The GA also improves continuously; however, the improvements are slower.



(a) $m = 20$

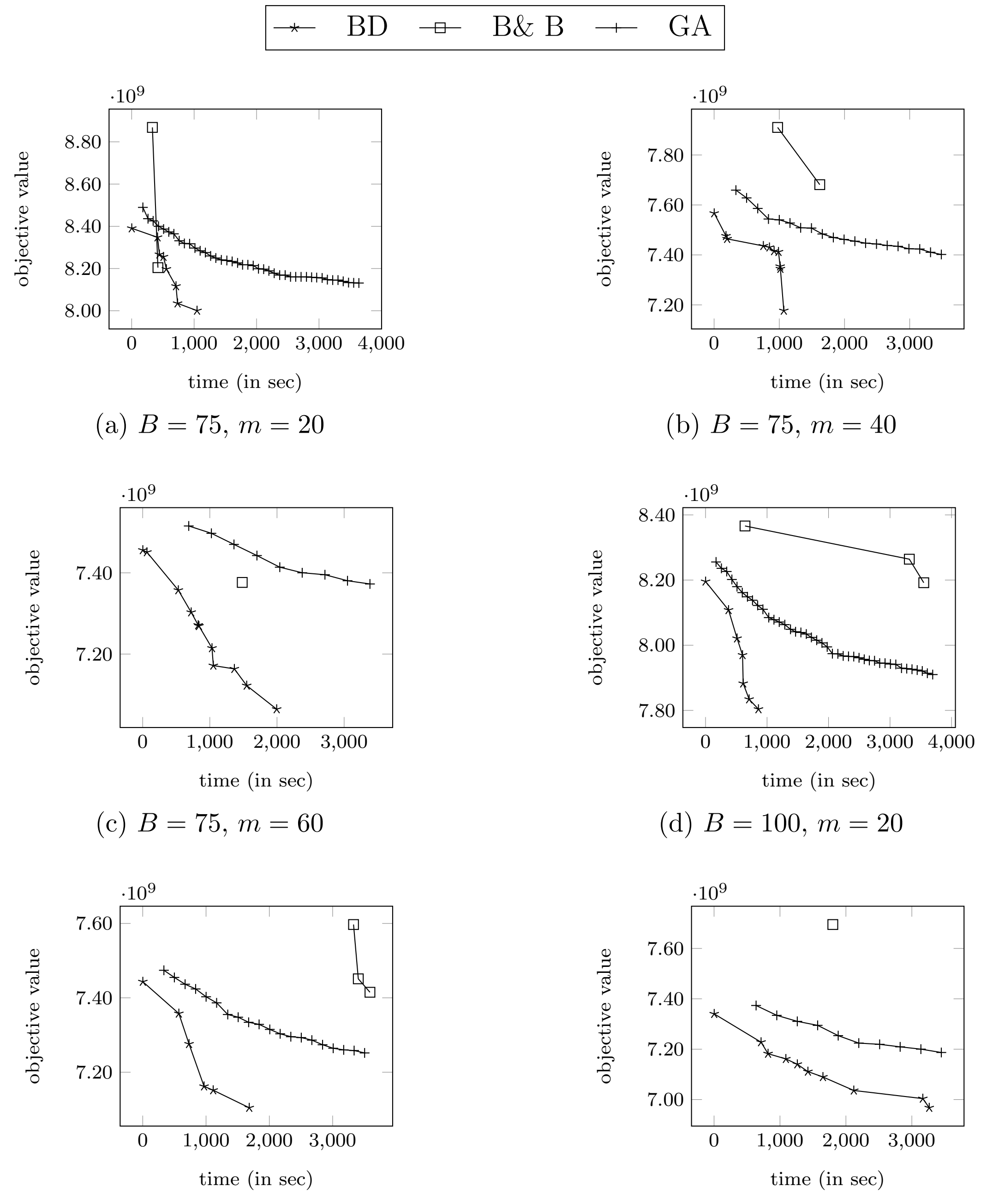
(b) $m = 40$

(c) $m = 60$

Figure 4.3: Objective values of the Sioux Fall network for $B = 50$

For the other two budget cases (Figures 4.4a - 4.4f), the BD and the GA have similar initial solutions, but the BD improves more and much faster than the GA. With increasing m, the gap between the BD and the GA gets even bigger. Even though the B&B

finds several feasible solutions, the initial solution of the B&B is (except of $m = 60$, $B = 75$) the worst solution if compared to all the approaches and the best found is only competitive for $B = 75$ and $m = 20$.



(a) $B = 75$, $m = 20$

(b) $B = 75$, $m = 40$

(c) $B = 75$, $m = 60$

(d) $B = 100$, $m = 20$

(e) $B = 100$, $m = 40$

(f) $B = 100$, $m = 60$

Figure 4.4: Objective values of the Sioux Fall network for $B = 75$ and $B = 100$

4.4.3 Extended Sioux-Fall network

Farvaresh and Sepehri (2013) introduced a network that combines 4 Sioux-Fall networks to one larger network. It consists of 100 nodes and 317 arcs. The possible 30 new arcs are included in the network and chosen as maintenance objects. The planning horizon is set to 8 periods. The original dataset consists of 817 OD-pairs. However, the total demand was only about 1/3 of the demand of the classical Sioux-Fall. As maintenance planning is only necessary in congested networks, we multiply the demand with 12 to get congestion effects similar to those in the classical Sioux-Fall network. The reliability is generated as before and the costs of Farvaresh and Sepehri (2013) are used as maintenance costs.

The total maintenance costs in the first period sum up to 2,062 and the costs increase between 3% and 6% per period. We used two budget cases: 300 and 400. We used two approximation precisions with $m = 20$ and $m = 40$. Due to the smaller solution space in the maintenance plan, the population size was reduced to 25. However, because of the large TAP, the run time was increased to 4 hours for all methods and the MILP was only solved with BD and not with the B&B anymore.

Results

Table 4.4 shows the results for the numerical tests of the larger network. The genetic algorithm was again executed 20 times and best, worst and mean values are shown. As in the low budget case, the solution with the best/worst objective value und the solution with the best/worst budget penalty were not always the same, both solutions are reported in the table. Because of the large number of arcs in relation to the number of maintenance objects, the effects of maintenance works are smaller and the solutions are closer to each other. This is also reflected by the objective value of an optimal network, in which every road has already been maintained in every period, which is $5.08 \cdot 10^8$ for $m = 20$ and $4.95 \cdot 10^8$ for $m = 40$.

The results of the second network show again that BD finds the best solutions. Only with a budget of 400 and 20 approximation points, the GA finds a better solution once; however, on average BD is still better. Surprisingly, the HCF heuristic finds good solutions for the high budget case. Nevertheless, with a low budget, only the BD finds solutions that satisfy the budget as in the first network for the low budget, the scheduling makes the problem more difficult. Moreover, in the case of those solutions where the budget is not satisfied for the other methods, the congestion in the solution of the BD is less. Especially for $m = 20$, if the budget constraint is violated only twice, the congestion

B	m	BD	LRF	HCF	GA best	GA worst	GA mean
300	20	5.38 $\langle 5.17\rangle$	$5.41^{(3)}$ $\langle 5.20\rangle$	$5.37^{(4)}$ $\langle 5.18\rangle$	$5.39^{(4)}$,$5.43^{(2)}$ $\langle 5.18\rangle$,$\langle 5.23\rangle$	$5.45^{(3)}$,$5.40^{(5)}$ $\langle 5.26\rangle$,$\langle 5.21\rangle$	$5.42^{(3.6)}$ $\langle 5.21\rangle$
	40	5.13 $\langle 5.06\rangle$	$5.16^{(3)}$ $\langle 5.09\rangle$	$5.13^{(4)}$ $\langle 5.07\rangle$	$5.14^{(1)}$ $\langle 5.07\rangle$	$5.18^{(3)}$,$5.16^{(6)}$ $\langle 5.13\rangle$,$\langle 5.10\rangle$	$5.16^{(3.6)}$ $\langle 5.10\rangle$
400	20	5.38 $\langle 5.18\rangle$	5.42 $\langle 5.22\rangle$	5.38 $\langle 5.18\rangle$	5.38 $\langle 5.17\rangle$	5.42 $\langle 5.21\rangle$	5.40 $\langle 5.19\rangle$
	40	5.13 $\langle 5.07\rangle$	5.18 $\langle 5.12\rangle$	5.13 $\langle 5.07\rangle$	5.13 $\langle 5.07\rangle$	5.17 $\langle 5.11\rangle$	5.14 $\langle 5.08\rangle$

Table 4.4: Objective values for the extended Sioux-Fall network (in 10^8)

value is among the worst of all 20 runs.

4.5 Conclusion

We proposed a non-linear bilevel formulation for a multi-period maintenance planning problem. The formulation was linearized and a BD algorithm was applied to solve this problem. The terminated algorithm finds good feasible solutions very fast compared to the MILP formulation solved with a B&B and using a genetic algorithm. Especially with a tight budget, which is often the case in practice, the BD is the only algorithm that finds good solutions without violating the budget. As convergence of the BD could not be realized, reducing the number of iterations and finding good lower bounds is of interest for future research. Since the classical genetic algorithms have problems finding feasible solutions for low budgets, specific operators should also be developed.

Chapter 5

Benders Decomposition for the Decentralized Facility Selection Problem

Production environments become more and more complex and monolithic models are no longer appropriate. The Decentralized Capacitated Facility Selection Problem addresses this problem and assumes a hierarchical decision structure where the principal firm only decides which plants to open. The plants are independent from the principal firm and want to minimize their operational costs. The principal firm, on the other hand, wants to minimize plant opening costs and the opportunity costs for unused capacities. This problem is modelled as a bilevel problem and transformed into a mixed-integer linear program using Karush-Kuhn-Tucker conditions. We apply Benders decomposition to this mixed-integer linear program and show that large instances can be solved up to 400 times faster than the mixed-integer linear program.

5.1 Introduction

Over the past decades, many companies have spread their production over several plants stretching over one or more countries. The advantages of these decentralized production systems can be cost reduction, efficiency improvement or also quick reaction to market changes even though more coordination effort is needed. Ertogral and David Wu (2000) showed that monolithic models, which use centralized decision schemes, should not be used in decentralized environments. Therefore, Cao and Chen (2006) proposed a bilevel model for plant selection in decentralized production environments.

The authors replaced the follower problem with its KKT conditions. Since the resulting MILP could only solve small instances, we propose a solution method based on BD. The numerical results show that even large instances can be solved very fast.

The remainder of this chapter is structured as follows: First, we introduce the bilevel formulation of the DCFSP. In Section 5.3, this formulation is transformed into a MILP and solved with BD. The numerical results are shown in Section 5.4 before we finish the chapter with the conclusion.

5.2 Decentralized Facility Selection Problem

We define the DCFSP as introduced in (Cao and Chen, 2006). The DCFSP assumes a decentralized manufacturing environment. A principal firm, the leader, decides which plants out of a set of plants I to open. Each plant $i \in I$ has fixed opening costs f_i and an available production capacity C_i. Unused capacity in plant i is penalized with the opportunity costs p_i if the plant is opened. The principal firm wants to minimize the sum of opening costs and opportunity costs.

The followers are the plants that have to satisfy the demand d_j of a set of products $j \in J$. Each plant has a capacity consumption ratio a_{ij} for processing one unit of product j in plant i. Using one unit of production capacity of plant i costs w_i and transferring one unit of product j to that plant costs r_{ij}. Moreover, $J_i \subseteq J$ defines the set of products that can be produced in plant i and $I_j \subseteq I$ defines the set of plants capable of producing product j. The plants operate independently of the principal firm and want to minimize production and transportation costs while cooperation among the plants exists.

The leader decision is modeled with the binary variable y_i for $i \in I$, which is 1 if the plant i is opened and 0 if not. The continuous follower variables x_{ij} define the fraction of the demand of product j produced in plant i.

The leader problem is then defined as follows:

$$\min \sum_{i\in I} f_i y_i + \sum_{i \in I} p_i \left(C_i y_i - \sum_{j \in J_i} d_j a_{ij} x_{ij} \right) \tag{5.1}$$

$$\text{s.t.} y_i \in \{0,1\} \qquad \forall i \in I \tag{5.2}$$

The opening costs and the opportunity costs are minimized while the binary condition is ensured. This problem is optimized subject to the follower problem:

$$\begin{aligned}
\min \quad & \sum_{i \in I} w_i \left(\sum_{j \in J_i} d_j a_{ij} x_{ij} \right) + \sum_{i \in I} \sum_{j \in J_i} d_j r_{ij} x_{ij} && && (5.3) \\
\text{s.t.} \quad & \sum_{i \in I_j} x_{ij} = 1 && \forall j \in J && (5.4) \\
& \sum_{j \in J_i} d_j a_{ij} x_{ij} \leq C_i y_i && \forall i \in I && (5.5) \\
& x_{ij} \geq 0 && \forall i \in I, j \in J_i && (5.6)
\end{aligned}$$

In the objective function, the production costs (first term) and the transfer costs (second term) are minimized. Equation (5.4) ensures that the demand is satisfied. Constraint (5.5) restricts the production quantity of each plant to its capacity if it is open and to zero if it is closed.

5.3 Solution method

In this section, we first show the transformation of the bilevel model to a MILP as proposed by Cao and Chen (2006) and then apply BD.

5.3.1 Mixed-integer linear program

Cao and Chen (2006) replace the follower problem (5.3) - (5.6) of the previous section with its KKT conditions and assume the partial cooperation assumption (Bard, 1998). For the primal constraints (5.4) and (5.5), the dual variables t_j and u_i are introduced. Then, the LBP can be transformed into the following MIP:

$$\begin{aligned}
\min \quad & \sum_{i \in I} f_i y_i + \sum_{i \in I} p_i \left(C_i y_i - \sum_{j \in J_i} d_j a_{ij} x_{ij} \right) && && (5.7) \\
\text{s.t.} \quad & \sum_{i \in I_j} x_{ij} = 1 && \forall j \in J && (5.8) \\
& \sum_{j \in J_i} d_j a_{ij} x_{ij} \leq C_i y_i && \forall i \in I && (5.9) \\
& \sum_{i \in I} \sum_{j \in J_i} \left(w_i d_j a_{ij} + d_j r_{ij} \right) x_{ij} \leq \sum_{j \in J} t_j + \sum_{i \in I} C_i y_i u_i && && (5.10) \\
& t_j + d_j a_{ij} u_i \leq w_i d_j a_{ij} + d_j r_{ij} && \forall i \in I, j \in J_i && (5.11)
\end{aligned}$$

$$t_j \in \mathbb{R} \qquad \forall j \in J \tag{5.12}$$

$$u_i \leq 0 \qquad \forall i \in I \tag{5.13}$$

$$x_{ij} \geq 0 \qquad \forall i \in I, j \in J_i \tag{5.14}$$

$$y_i \in \{0,1\} \qquad \forall i \in I \tag{5.15}$$

Constraint (5.11) is the dual constraint for the primal variables x_{ij} and equation (5.10) is the optimality condition. The non-linear optimality condition can be replaced by the following set of linear constraints by introducing the auxiliary variable μ_i and a Big M(Cao and Chen, 2006):

$$\sum_{i \in I} \sum_{j \in J_i} (w_i d_j a_{ij} + d_j r_{ij}) \leq \sum_{j \in J} t_j + \sum_{i \in I} \mu_i \tag{5.16}$$

$$\mu_i - C_i u_i \leq -M y_i + M \qquad \forall i \in I \tag{5.17}$$

$$\mu_i - C_i u_i \geq 0 \qquad \forall i \in I \tag{5.18}$$

$$\mu_i \geq -M y_i \qquad \forall i \in I \tag{5.19}$$

$$\mu_i \leq 0 \qquad \forall i \in I \tag{5.20}$$

Constraints (5.17) - (5.20) ensure that μ_i takes the value $C_i u_i$ if y_i is 1 and the value 0 for $y_i = 0$.

5.3.2 Benders decomposition

The MILP of the previous section has complicating variables y_i and the easier continuous variables x_{ij}, t_j, u_i and μ_i. Following Benders (1962), we divided this problem into two easier solvable problems and solved it with BD. The outline of the BD is shown in algorithm 4.

Dual slave problem

The slave problem is defined by fixing the binary decision variables to a solution y^*. The part of the objective function not depending on y is minimized subject to all constraints of the MILP for the fixed y:

$$\min \; -\sum_{i \in I} p_i \sum_{j \in J_i} d_j a_{ij} x_{ij} \tag{5.21}$$

$$\text{s.t.}(5.8), (5.9), (5.16) - (5.20), (5.11) - (5.14)$$

Algorithm 4: Benders decomposition: Main structure

1 Set $y_i^* = 1 \quad \forall i \in I$ (or take any other feasible solution of y) and let $z^* = \infty$ be the upper bound.
2 For $y = y^*$, solve the Benders slave problem and let z^S be the optimal objective value of the slave problem. If the slave problem is bounded, add the corresponding optimality cut to the master problem and update the upper bound $z^* = \min\{z^*, z^S\}$. If the slave problem is unbounded, add the feasibility cut to the master problem.
3 Resolve the master problem. Save the solution y^* and let z_M be the new solution of the master problem.
4 If $z^* = z_M$: Stop, otherwise: Go to Step 2.

In BD, the dual of the slave problem is solved. Therefore, the dual variables α_j, β_i, δ_{ij}, γ, λ_i^1, λ_i^2 and λ_i^3, which correspond to the primal constraints (5.8), (5.9), (5.11) and (5.16) - (5.19), are introduced. The dual constraints (5.23) - (5.26) correspond to the primal variables x_{ij}, t_j, u_i and μ_i.

$$\begin{aligned}
\max \ & \alpha_j + \sum_{i\in I} C_i y_i^* \beta_i + \sum_{i\in I}\sum_{j\in J_i} \left(w_i d_j a_{ij} + d_j r_{ij}\right)\delta_{ij} \\
& \quad + \sum_{i\in I}\left(-M y_i^* + M\right)\lambda_i^1 - \sum_{i\in I} M y_i^* \lambda_i^3 && (5.22)\\
\text{s.t.} \ & \alpha_j + d_j a_{ij}\beta_i + \left(w_i d_j a_{ij} + d_j R_{ij}\right)\gamma \le -p_i d_j a_{ij} && \forall i \in I, j \in J_i \quad (5.23)\\
& -\gamma + \sum_{i\in I_j} \delta_{ij} = 0 && \forall j \in J \quad (5.24)\\
& -C_i\lambda_i^1 - C_i\lambda_i^2 + \sum_{j\in J_i} d_j a_{ij}\delta_{ij} \ge 0 && \forall i \in I \quad (5.25)\\
& -\gamma + \lambda_i^1 + \lambda_i^2 + \lambda_i^3 \ge 0 && \forall i \in I \quad (5.26)\\
& \alpha_j \in \mathbb{R} && \forall j \in J \quad (5.27)\\
& \beta_i, \lambda_i^1 \le 0 && \forall i \in I \quad (5.28)\\
& \gamma \le 0 && (5.29)\\
& \delta_{ij} \le 0 && \forall i \in I, j \in J_i \quad (5.30)\\
& \lambda_i^2, \lambda_i^3 \ge 0 && \forall i \in I \quad (5.31)
\end{aligned}$$

Pareto-optimal cuts

Since the follower problem is a transportation problem which is often highly degenerated, the convergence of the BD can be slow. Magnanti and Wong (1981) proposed a method to strengthen the generated optimality cuts by solving an auxiliary problem. These cuts are not dominated by any other cut for that subproblem and are called pareto-optimal cuts.

Let y^c be a core point of the master problem, which is a point in the relative interior of the feasible region of the master variables $\{0,1\}^{|I|}$, and z_S be the objective value of the dual subproblem (5.22) - (5.31). For calculating such a pareto-optimal cut, the following auxiliary problem is solved:

$$\begin{aligned}
\max \ & \alpha_j + \sum_{i \in I} C_i y_i^c \beta_i + \sum_{i \in I} \sum_{j \in J_i} \left(w_i d_j a_{ij} + d_j r_{ij}\right) \delta_{ij} \\
& + \sum_{i \in I} \left(-M y_i^c + M\right) \lambda_i^1 - \sum_{i \in I} M y_i^c \lambda_i^3 && (5.32) \\
\text{s.t. } & \alpha_j + \sum_{i \in I} C_i y_i^* \beta_i + \sum_{i \in I} \sum_{j \in J_i} \left(w_i d_j a_{ij} + d_j r_{ij}\right) \delta_{ij} \\
& + \sum_{i \in I} \left(-M y_i^* + M\right) \lambda_i^1 - \sum_{i \in I} M y_i^* \lambda_i^3 \qquad = z_S && (5.33) \\
& (5.23) - (5.31)
\end{aligned}$$

The dual slave problem is only extended by equation (5.33), ensuring that an optimal slave problem solution is chosen. Magnanti and Wong (1981) proved that the modified objective function calculates a pareto-optimal cut.

Master problem

Let $(\alpha^*, \beta^*, \gamma^*, \delta^*, \lambda^{1*}, \lambda^{2*}, \lambda^{3*})$ be the optimal solution of the dual slave problem. If the dual slave problem is bounded, then the following optimality cut is added the master problem:

$$\begin{aligned}
z \geq & \alpha_j + \sum_{i \in I} C_i y_i \beta_i^* + \sum_{i \in I} \sum_{j \in J_i} \left(w_i d_j a_{ij} + d_j r_{ij}\right) \delta_{ij}^* \\
& + \sum_{i \in I} \left(-M y_i + M\right) \lambda_i^{1*} - \sum_{i \in I} M y_i \lambda_i^{3*} && (5.34)
\end{aligned}$$

If the dual slave problem is unbounded, the following feasibility cut is added to the master problem:

$$\begin{aligned} 0 \geq & \alpha_j + \sum_{i \in I} C_i y_i \beta_i^* + \sum_{i \in I} \sum_{j \in J_i} \left(w_i d_j a_{ij} + d_j r_{ij} \right) \delta_{ij}^* \\ & + \sum_{i \in I} \left(-M y_i + M \right) \lambda_i^{1*} - \sum_{i \in I} M y_i \lambda_i^{3*} \end{aligned} \tag{5.35}$$

If C_O is the set of optimality cuts and C_F the set of feasibility cuts that are added to the master problem, the master problem is defined as follows:

$$\min \quad \sum_{i \in I} (f_i + p_i C_i) y_i + z \tag{5.36}$$

$$\text{s.t.} \quad c_o \qquad \forall c_o \in C_O \tag{5.37}$$

$$c_f \qquad \forall c_f \in C_F \tag{5.38}$$

$$y_i \in \{0, 1\} \qquad \forall i \in I \tag{5.39}$$

$$z \in \mathbb{R} \tag{5.40}$$

In each iteration, a new optimality or feasibility cut is added and the master problem is solved until the convergence criterion is met.

5.4 Numerical study

To show the efficiency of the method, we generated a new set of DCFSP instances. We used two Capacitated Facility Location Problem datasets as introduced by Avella and Boccia (2009), where fixed opening costs and transportation costs are given. The first set consists of 20 instances with 300 facilities and 1,500 customers which are used as the 1,500 products. The second set consists of 20 instances with 500 facilities and 500 products. In both cases, four different demand/supply ratios $r \in \{5, 10, 15, 20\}$ with 5 instances each are used. The additional needed data was generated randomly and in a similar way as in Cao and Chen (2006): The consumption ratio was set to 1 ($a_{ij} = 1 \quad \forall i \in I, j \in J$). The costs p_i and w_i were generated randomly between 0.4 and 1.0 and $p_i = w_i$.

The tests were performed on an Intel(R) Core(TM) i7-3770 CPU, 3.4GHz, 16GB RAM using Xpress 7.3 and the results are shown in Tables 5.1 and 5.2. We compared the run time of the BD with the run time of the MILP using the formulation of Cao and Chen (2006). Moreover, the number of Benders iterations needed and the percent time improvements are shown.

r	BD iter.	MILP (sec)	BD (sec)	time ratio (in %)
	2	13529	34	0.2513
	3	24457	74	0.3026
5	6	36071	68	0.1885
	7	18840	150	0.7962
	7	42690	101	0.2366
avg	5	27117.4	85.4	0.3150
	6	44321	33	0.0745
	5	27620	33	0.1195
10	3	14057	27	0.1921
	2	4528	32	0.7067
	4	4663	35	0.7506
avg	4	19037.8	32	0.1681
	6	26109	30	0.1149
	3	4503	18	0.3997
15	6	83820	40	0.0477
	2	2099	19	0.9052
	14	19351	72	0.3721
avg	6.2	27176.4	35.8	0.1317
	3	1623	19	1.1707
	3	2829	17	0.6009
20	3	6647	20	0.3009
	12	9373.9	64	0.6827
	6	8227.3	39	0.474
avg	5.4	5740.04	31.8	0.5540

Table 5.1: Computational results for 300×1500 instances of Avella and Boccia (2009)

The results show that the BD can solve all instances significantly faster than the MILP formulation. The average run time for the first set with 300 facilities was 46.25 seconds and 5.15 iterations in the BD. Compared to the MILP, this is an improvement of more than 99% on average. For the second set, all instances were solved on average in 8.55 seconds and with 2.95 iterations using BD, which is again an improvement of more than 99%. Compared to the first set, the run time is even less. As the number of iterations

r	BD iter.	MILP (sec)	BD (sec)	time ratio (in %)
	3	432	10	2.3148
	1	12762	6	0.0470
5	2	20179	10	0.0496
	2	4836	11	0.2275
	1	5046	9	0.1784
avg	1.8	8651	9.2	0.106346
	2	320	7	2.1875
	4	547	10	1.8282
10	2	261	7	2.682
	2	526	6	1.1407
	2	4420	7	0.1584
avg	2.4	1214.8	7.4	0.609154
	4	2931	9	0.3071
	1	1368	5	0.3655
15	6	2908	13	0.447
	3	2755	7	0.2541
	4	615	9	1.4634
avg	3.6	2115.4	8.6	0.406542
	1	1437	5	0.3479
	10	2701	17	0.6294
20	5	3622	11	0.3037
	2	389	6	1.5424
	2	511	6	1.1742
avg	4	1732	9	0.51963

Table 5.2: Computational results for 500 $\times$ 500 instances of Avella and Boccia (2009)

needed was smaller and the slave problem depends on the number of products, it is clear that in the first set with 1,500 products, the run time of the slave problem is higher. Moreover, as many as 37 out of 40 instances were solved in less than 10 iterations.

5.5 Conclusion

We proposed a BD algorithm for solving the DCFSP. The bilevel formulation is transformed into a MILP and then solved with BD. The numerical results show that the run time is significantly improved by the algorithm and real-size instances can be solved efficiently. Admittedly, at this stage the model still looks rather stylized and the formulation should be extended to a real word problem by including further aspects like demand uncertainty.

Chapter 6

Benders decomposition for the Hazmat Transport Network Design Problem

We propose a new method for solving the Hazmat Transport Network Design Problem. In this problem, the government wants to reduce the risk of hazardous accidents for the population by restricting the shipment of hazardous goods on roads. When taking that decision, the government has to anticipate the reaction of the carriers who want to minimize the transportation costs by solving a shortest path problem. We use a bilevel formulation that guarantees stable solutions and transform this model into a mixed-integer linear program by applying the Karush-Kuhn-Tucker conditions. This model is solved to optimality with a multi-cut Benders decomposition. The numerical study shows the computation benefits of the method and run time savings of more than 90%. Moreover, we show that the bilevel model reduces the risk by 35% on average compared to a two-step decision process, which does not anticipate the carriers reaction.

6.1 Introduction

The transportation of hazardous materials is essential for most countries. In 2013, 294.75 million tons of hazardous materials were transported in and through Germany. With 7.15%, hazmat has only a small share of the transported volume in Germany. (Statistisches Bundesamt Wiesbaden, 2015) However, the consequences of hazardous accidents are tremendous, as they can cause significant damages to the area around the accident.

One of the worst accidents of this kind in recent history happened in July 2013 in Lac-Mégantic, QC in Canada. A driverless train with 72 tank cars of petroleum crude oil derailed in the city center and caused the death of at least 42 people. Moreover, at least 30 buildings and 115 businesses were destroyed and it took almost two days for the firefighters to get the situation under control.

According to Erkut et al. (2007), the consequences of such accidents are not only injuries and fatalities, property damage and cleanup costs but also evacuation, product loss, traffic incident delay or environmental damages. To better react in the case of an accident, transported hazmat needs to be indicated and is categorized into the following nine classes: explosives, gases, flammable or combustible liquid, flammable solid, spontaneously combustible and dangerous when wet, oxidizer and organic peroxide, poison and poison inhalation hazard, radioactive, corrosive and, lastly, miscellaneous.

Besides technical improvements and regulations on how to transport these dangerous goods, the regulation of the routing is an important aspect of risk reduction. The government can allow only specific parts of a road network for the shipment of hazardous materials to risk of accidents. Since the carriers are interested in minimizing their cost while satisfying their demand, they are looking for a cost minimal path in the network. Since these two objectives are not the same, the government has to anticipate the reaction of the follower and the problem is modelled as an LBP which is known as the HTNDP.

The contribution of this chapter is first to apply BD to a known mixed-integer formulation for the HTNDP. The numerical study shows that our approach offers significant computational benefits. Second, the necessity of the bilevel formulation is shown by comparing the hierarchical solution to a two-step decision process in which the follower can choose his path in a network that was optimized by neglecting the reaction of the follower. Finally, we show that the price of anarchy increases with the number of commodities.

This chapter is structured as follows. In Section 6.2, the bilevel formulation is presented. The solution algorithm is introduced in Section 6.3. The numerical results are shown in Section 6.4 and the chapter ends with a conclusion.

6.2 Bilevel formulation

The transportation network is represented by a graph $G = (N, A)$ with a set of nodes N and a set of arcs A. K defines the set of commodities shipped through the network. Each commodity $k \in K$ is defined by an origin $o_k \in N$, a destination $d_k \in N$ and the transport volume ϕ_k. For each arc $(i, j) \in A$, the transportation costs c_{ij} and for each arc $(i, j) \in A$ and commodity $k \in K$ the transportation risk r_{ij}^k are given. According to Erkut and Verter (1998), this risk is the probability of an accident multiplied with the accumulated effected population in the area of arc (i, j).

The leader wants to minimize the total risk in the network by deciding whether an arc $(i, j) \in A$ can be used for the shipment with the binary decision variable y_{ij}. Therefore, the leader anticipates the reaction of the follower - the carriers who want to ship the commodities through the network in a cost minimal way. The binary follower decision variable x_{ij}^k is 1 if commodity k is shipped over arc $(i, j) \in A$ and 0 if not.

The leader problem is defined as follows:

$$\min \sum_{k \in K} \sum_{(i,j) \in A} r_{ij}^k \phi_k x_{ij}^k \tag{6.1}$$

$$\text{s.t. } y_{ij} \in \{0, 1\} \qquad \forall (i, j) \in A \tag{6.2}$$

The total risk in the network is minimized while the system decides whether an arc is allowed for shipment or not.

The follower problem is a classical shortest path problem (Hillier and Lieberman, 2009), which is decomposable within the commodities:

$$\min \sum_{k \in K} \left(\sum_{(i,j) \in A} c_{ij}^k x_{ij}^k - \frac{1}{R} \sum_{(i,j) \in A} r_{ij}^k x_{ij}^k \right) \tag{6.3}$$

$$\text{s.t. } \sum_{(i,j) \in A} x_{ij}^k - \sum_{(j,l) \in A} x_{jl}^k = \begin{cases} 0, & \text{if } j \neq o_k, d_k \\ -1, & \text{if } j = o_k \\ 1, & \text{if } j = d_k \end{cases} \qquad \forall j \in N, k \in K \tag{6.4}$$

$$x_{ij}^k \leq y_{ij} \qquad \forall (i, j) \in A, k \in K \tag{6.5}$$

$$x_{ij}^k \in \{0, 1\} \qquad \forall (i, j) \in A, k \in K \tag{6.6}$$

Equation (6.4) is a flow conservation constraint and constraint (6.5) ensures that only arcs that are allowed by the follower can be used. The objective function minimizes the transportation costs for each commodity in the first term. The second term was added by Amaldi et al. (2011) to guarantee stable solutions. R is defined as the maximum risk of a commodity in the network. If the transportation costs c_{ij} are integer, then, for a commodity that has more than one minimum-cost-path, the shortest path with the maximum risk will be chosen since the second term will have the highest negative value. However, a longer path cannot be chosen since the value of the second term can take at maximum the value one, which is the difference between two paths. This ensures that the leader assumes the worst case and a pessimistic bilevel problem is defined. Removing the second term would give an optimistic formulation where the government would assume that carriers will always choose the path with the lowest risk among the shortest paths (if more than one exist).

6.3 Solution approach

In this section, the solution approach based on the idea of Fontaine and Minner (2014, 2016a), is shown. First, the bilevel formulation is transformed into a MILP. In a second step, a multi-cut BD (Birge and Louveaux, 1988) is applied to the MILP.

6.3.1 Mixed-integer linear program

Because of the total unimodularity of the follower problem (6.3) - (6.6), the binary condition (6.6) can be relaxed by the following equation (Amaldi et al., 2011):

$$x_{ij}^k \geq 0 \qquad \forall (i,j) \in A, k \in K \tag{6.7}$$

Equation (6.5) ensures that $x_{ij}^k \leq 1$ holds. This linear follower problem can be replaced by the KKT conditions and transforms the bilevel problem into a non-linear MIP (Bard, 1998).

$$\min \sum_{k \in K} \sum_{(i,j) \in A} r_{ij}^k \phi_k x_{ij}^k \tag{6.8}$$

$$\text{s.t.} \quad \sum_{(i,j) \in A} x_{ij}^k - \sum_{(j,l) \in A} x_{jl}^k = \begin{cases} 0, & \text{if } j \neq o_k, d_k \\ -1, & \text{if } j = o_k \\ 1, & \text{if } j = d_k \end{cases} \qquad \forall j \in N, k \in K \tag{6.9}$$

$$x_{ij}^k \leq y_{ij} \qquad \forall (i,j) \in A, k \in K \tag{6.10}$$

$$\sum_{(i,j) \in A} (c_{ij} - r_{ij}^k / R) x_{ij}^k \leq u_{d_k}^k - u_{o_k}^k + \sum_{(i,j) \in A} v_{ij}^k y_{ij} \qquad \forall k \in K \tag{6.11}$$

$$u_j^k - u_i^k + v_{ij}^k \leq c_{ij} - r_{ij}^k / R \qquad \forall (i,j) \in A, k \in K \tag{6.12}$$

$$v_{ij}^k \leq 0 \qquad \forall (i,j) \in A, k \in K \tag{6.13}$$

$$u_j^k \in \mathbb{R} \qquad \forall j \in N, k \in K \tag{6.14}$$

$$x_{ij}^k \geq 0 \qquad \forall (i,j) \in A, k \in K \tag{6.15}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in A \tag{6.16}$$

u_j^k define the dual variables of the follower constraint (6.4) and v_{ij}^k of constraint (6.5). Equation (6.12) is the dual constraint of the primal variable x_{ij}^k. As in (Cao and Chen, 2006), the optimality condition (6.11) can be linearized by introducing the auxiliary variable w_{ij}^k and a Big M and by replacing (6.11) with the following terms:

$$\sum_{(i,j)\in A} (c_{ij} - r_{ij}^k/R)x_{ij}^k \leq u_{d_k}^k - u_{o_k}^k + \sum_{(i,j)\in A} w_{ij}^k \qquad \forall k \in K \tag{6.17}$$

$$w_{ij}^k \leq v_{ij}^k + M(1-y_{ij}) \qquad \forall (i,j) \in A, k \in K \tag{6.18}$$

$$w_{ij}^k \geq v_{ij}^k \qquad \forall (i,j) \in A, k \in K \tag{6.19}$$

$$w_{ij}^k \geq -My_{ij} \qquad \forall (i,j) \in A, k \in K \tag{6.20}$$

$$w_{ij}^k \leq 0 \qquad \forall (i,j) \in A, k \in K \tag{6.21}$$

6.3.2 Multi-cut Benders decomposition

BD (Benders, 1962) divides the problem into an integer master problem and a continuous slave problem. The decision variables are divided into the complicating binary variables y_{ij} and the easier continuous variables x_{ij}, u_j^k, v_{ij}^k and w_{ij}^k. Both problems are solved iteratively. In the multi-cut version by Birge and Louveaux (1988), the slave problem is further decomposed into $|K|$ slave subproblems. These slave subproblems can be solved in parallel and each subproblem generates a cut that can be added to the master problem.

An outline of the structure of the multi-cut BD is given in Algorithm 5.

Slave problem

The slave subproblems are given by fixing y with y^* and solving (6.8) - (6.10), (6.12) - (6.15) and (6.17) - (6.21) for each $k \in K$ in parallel.

Since the dual variables of the slave problem define the optimality and feasibility cuts of the master problem, the dual slave subproblems are needed. For this formulation, the dual variables α_j,β_{ij}, γ, δ_{ij}, λ_{ij}^1, λ_{ij}^2 and λ_{ij}^3 for the primal constraints (6.9), (6.10), (6.12) and (6.17) - (6.20) are introduced. The dual objective function is maximized subject

Algorithm 5: Multi-cut Benders decomposition: Main structure

1. Set y^* to any feasible solution of y (e.g. $y_{ij} = 1 \quad \forall (i,j) \in A$) and the upper bound $z^* = \infty$.
2. For $y = y^*$, solve for each $k \in K$ the subproblems of the slave problem in parallel. Let z_k^S be the optimal solution value of subproblem k. If all slave problems are bounded, add $|K|$ optimality cuts to the master problem and update the upper bound $z^* = \min\{z^*, \sum_{k \in K} z_k^S\}$. Else add a feasibility cut to the master problem.
3. Solve the master problem and save the solution y^*. Let z_M be the solution of the master problem.
4. Stop if $z^* = z_M$, otherwise go to Step 2.

to the dual constraints (6.23) - (6.26), which correspond to primal decision variables x_{ij},u_j,w_{ij} and v_{ij}. Here, $\chi_{[a=b]}$ defines an indicator function that is 1 if $a = b$ and 0 otherwise.

$$
\begin{aligned}
\max \quad & \alpha_{d_k} - \alpha_{o_k} + \sum_{(i,j)\in A} \beta_{ij} y^*_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r^k_{ij}/R\right) \delta_{ij} \\
& + \sum_{(i,j)\in A} \left(M(1-y^*_{ij})\lambda^1_{ij} - M y^*_{ij} \lambda^3_{ij}\right) && && (6.22) \\
\text{s.t. } & \alpha_j - \alpha_i + \beta_{ij} + \gamma\left(c_{ij} - r^k_{ij}/R\right) \le r^k_{ij}\phi_k && \forall (i,j) \in A && (6.23) \\
& \sum_{(i,j)\in A} \delta_{ij} - \sum_{(j,l)\in A} \delta_{jl} - \gamma\chi_{[j=d_k]} + \gamma\chi_{[j=o_k]} = 0 && \forall j \in N && (6.24) \\
& -\gamma + \lambda^1_{ij} + \lambda^2_{ij} + \lambda^3_{ij} \ge 0 && \forall (i,j) \in A && (6.25) \\
& \delta_{ij} - \lambda^1_{ij} - \lambda^2_{ij} \ge 0 && \forall (i,j) \in A && (6.26) \\
& \alpha_j \in \mathbb{R} && \forall j \in N && (6.27) \\
& \beta_{ij}, \delta_{ij}, \lambda^1_{ij} \le 0 && \forall (i,j) \in A && (6.28) \\
& \gamma \le 0 && && (6.29) \\
& \lambda^2_{ij}, \lambda^3_{ij} \ge 0 && \forall (i,j) \in A && (6.30)
\end{aligned}
$$

If a dual subproblem is unbounded, it is resolved with the following additional constraint to calculate the needed extreme ray for the feasibility cut:

$$\begin{aligned} \alpha_{d_k} - \alpha_{o_k} + \sum_{(i,j)\in A} \beta_{ij} y^*_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r^k_{ij}/R\right) \delta_{ij} + \\ \sum_{(i,j)\in A} \left(M(1 - y^*_{ij})\lambda^1_{ij} - M y^*_{ij} \lambda^3_{ij}\right) \leq M_2 \end{aligned} \tag{6.31}$$

If M_2 is a large number that does not cut off any extreme point but only restricts the extreme rays, the solution of the bounded problem returns the needed extreme ray.

Pareto-optimal cuts

The convergence of BD is highly dependent on the strength of the generated cuts. If the dual subproblem is degenerated, several cuts exist for the same master solution. Magnanti and Wong (1981) proposed a method to generate pareto-optimal cuts. These cuts are not dominated by any other cut and can improve the convergence of the BD significantly.

To generate a pareto-optimal cut, a core point, a point in the relative interior of the feasible region of the complicating variables $\{0,1\}^{|A|}$, y^c is needed. Moreover, let z_k be the objective value of the problem (6.22) - (6.30). Then the following auxiliary problem is solved:

$$\begin{aligned} \max \quad & \alpha_{d_k} - \alpha_{o_k} + \sum_{(i,j)\in A} \beta_{ij} y^c_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r^k_{ij}/R\right) \delta_{ij} \\ & + \sum_{(i,j)\in A} \left(M(1 - y^c_{ij})\lambda^1_{ij} - M y^c_{ij} \lambda^3_{ij}\right) && (6.32) \\ \text{s.t.} \quad & \alpha_{d_k} - \alpha_{o_k} + \sum_{(i,j)\in A} \beta_{ij} y^*_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r^k_{ij}/R\right) \delta_{ij} \\ & + \sum_{(i,j)\in A} \left(M(1 - y^*_{ij})\lambda^1_{ij} - M y_{ij} \lambda^3_{ij}\right) = z_k && (6.33) \\ & (6.23) - (6.30) \end{aligned}$$

Constraint (6.33) forces the model to choose a solution that has the optimal objective value z_k and the objective function generates the dual values of a pareto-optimal cut.

Master problem

Let α_j^{k*},β_{ij}^{k*}, γ^{k*}, δ_{ij}^{k*}, λ_{ij}^{1k*}, λ_{ij}^{2k*} and λ_{ij}^{3k*} be the optimal solution of the solved subproblems. If all subproblems are bounded, the following optimality cuts are added:

$$\begin{aligned}\alpha_{d_k}^{k*} - \alpha_{o_k}^{k*} + \sum_{(i,j)\in A} \beta_{ij}^{k*} y_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r_{ij}^k/R\right) \delta_{ij}^{k*} \\ + \sum_{(i,j)\in A} \left(M(1-y_{ij})\lambda_{ij}^{1k*} - My_{ij}\lambda_{ij}^{3k*}\right) \leq z_k \end{aligned} \tag{6.34}$$

If one or more subproblems of the slave problem are unbounded, the following feasibility cut for all commodities is added:

$$\begin{aligned}\sum_{k\in K} \Bigg(\alpha_{d_k}^{k*} - \alpha_{o_k}^{k*} + \sum_{(i,j)\in A} \beta_{ij}^{k*} y_{ij} + \sum_{(i,j)\in A} \left(c_{ij} - r_{ij}^k/R\right) \delta_{ij}^{k*} \\ + \sum_{(i,j)\in A} \left(M(1-y_{ij})\lambda_{ij}^{1k*} - My_{ij}\lambda_{ij}^{3k*}\right) \Bigg) \leq 0 \end{aligned} \tag{6.35}$$

With C_O being the set of optimality cuts and C_F the set of feasibility cuts, the master problem can be defined as follows:

$$\min \sum_{k\in K} z_k \tag{6.36}$$

$$\text{s.t. } c_o \qquad \forall c_o \in C_O \tag{6.37}$$

$$c_f \qquad \forall c_f \in C_F \tag{6.38}$$

$$z_k \in \mathbb{R} \qquad \forall k \in K \tag{6.39}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in A \tag{6.40}$$

6.4 Numerical study

To show the computational efficiency of our approach, we compared the BD with the MILP formulation. Both were implemented in Xpress 7.6. For all calculations, we used an Intel Core i7 with 8 threads and 32GB RAM and the maximum calculation time was set to 7,200 seconds. The graph of the well-known Sioux-Fall network (LeBlanc, 1975), which consists of 24 nodes and 76 arcs, was used. The cost and risk parameters for each arc were generated randomly between 1 and 100. We used different scenarios defined by the number of commodities $|K| \in \{60, 80, 100, 120, 140\}$. For each of these scenarios, 10 instances were generated. The origin and destination were chosen randomly among all nodes and the demand was also choosen randomly between 1 and 20.

Tables 6.1 - 6.5 shows the detailed numerical results. For each instance, the number of iterations in the BD, the run time for both methods in seconds and the time saving of the BD compared to the MILP formulation are reported. Moreover, three objective values are shown and compared: The objective value of the bilevel formulation (*bilevel*), the over-regulated objective value if the leader ignores the reaction of the follower (*decision L*) and the risk if the follower problem is solved with the leader decision of the previous model (*reaction F*). Δ_1 is the difference between the hierarchical and the over-regulated decision. Δ_2 is the difference between the hierarchical decision and the risk of the follower reaction to the over-regulated decision. Therefore, Δ_2 shows the benefit of the bilevel formulation.

	BD	time (sec)		time-Δ	objective values			Δ_1	Δ_2
ID	iter	BD	MILP	(in %)	bilevel	decision L	reaction F	(in %)	(in %)
1	72	322	2469	86.96	100638	94817	136492	5.78	35.63
2	25	89	142	36.86	86688	83219	106889	4.00	23.30
3	73	270	1041	74.09	80562	77043	108873	4.37	35.14
4	38	137	290	52.97	95020	90869	130758	4.37	37.61
5	49	215	179	-20.12	84602	82254	123210	2.78	45.63
6	86	339	1136	70.16	82664	78414	107325	5.14	29.83
7	67	253	5925	95.72	97409	91463	139850	6.10	43.57
8	47	168	471	64.41	95721	91609	155238	4.30	62.18
9	69	260	497	47.66	96960	91256	123013	5.88	26.87
10	48	163	24	-581.08	98399	95659	130633	2.78	32.76
avg	57.4	222	1217	81.80				4.55	37.25

Table 6.1: Computational results for $|K| = 60$

Already with 60 commodities, BD improves the run time by 81.8% on average. Only in 2 instances, the MILP was still faster. However, these instances are much easier, as the MILP run time is also notably smaller than in the other instances. Moreover, the price of anarchy Δ_1 is only 4.55%, but using that network design and letting the carriers decide would increase the risk by 37.25% on average. This shows that the anticipation of the carriers' reaction is very important and the bilevel formulation should be used.

The average values of all five scenarios are shown in Table 6.6. Here, the scared brackets behind the MILP run time indicate how many instances could not be solved to optimality within the time limit.

ID	BD iter	time (sec) BD	MILP	time-Δ (in %)	objective values bilevel	decision L	reaction F	Δ_1 (in %)	Δ_2 (in %)
1	61	343	2805	87.77	122489	114663	159020	6.39	29.82
2	73	502	7200	93.03	142861	132014	180859	6.81	26.60
3	55	275	2646	89.60	130761	121269	168277	7.04	28.69
4	41	198	99	-100.82	128027	123130	167822	3.82	31.08
5	54	265	557	52.46	136840	130047	184285	4.96	34.67
6	65	343	5730	94.01	144302	135068	213513	6.23	47.96
7	52	264	1583	83.34	125852	118988	170898	5.45	35.79
8	54	266	876	69.62	115830	107383	152710	7.29	31.84
9	58	300	6142	95.11	131236	122525	169774	6.64	29.37
10	52	268	2115	87.33	109420	99982	152974	8.44	39.80
avg	56.5	302	2975	89.84				6.31	33.56

Table 6.2: Computational results for $|K| = 80$

ID	BD iter	time (sec) BD	MILP	time-Δ (in %)	objective values bilevel	decision L	reaction F	Δ_1 (in %)	Δ_2 (in %)
1	48	295	1116	73.56	146890	138272	188854	5.87	28.57
2	43	255	1398	81.75	147985	140961	212845	4.75	43.83
3	60	411	3789	89.15	145249	133565	206689	8.04	42.30
4	77	573	7200	92.04	148284	136453	194195	7.76	30.96
5	130	3054	7200	57.59	155183	142429	207375	7.96	33.63
6	79	569	7200	92.09	161911	148191	220157	8.47	35.97
7	55	369	6055	93.90	158819	150242	213160	5.40	34.22
8	60	419	5894	92.89	168848	159336	236511	5.63	40.07
9	76	536	7200	92.55	161603	149980	234453	7.19	45.08
10	49	320	4907	93.48	187248	172794	263801	7.72	40.88
avg	67.7	680	5196	86.91				6.88	37.55

Table 6.3: Computational results for $|K| = 100$

ID	BD iter	time (sec) BD	MILP	time-Δ (in %)	objective values bilevel	decision L	reaction F	Δ_1 (in %)	Δ_2 (in %)
1	53	391	7200	94.57	177330	165693	254897	6.56	43.74
2	57	410	7200	94.31	187687	169318	249294	10.86	32.82
3	68	584	6781	91.39	195757	181294	244269	7.45	24.78
4	50	425	1503	71.70	193484	182625	263738	5.26	36.31
5	64	498	5477	90.91	180419	169263	261322	6.18	44.84
6	54	377	2580	85.38	213904	201979	285084	5.57	33.28
7	49	353	5529	93.62	187417	174648	244022	6.81	30.20
8	102	988	7200	86.28	187545	174518	261787	6.61	39.59
9	75	594	7200	91.75	192739	179740	271156	6.62	40.69
10	82	1045	7200	85.48	195350	176453	258597	9.60	32.38
avg	65.4	566	5787	90.21				7.15	35.86

Table 6.4: Computational results for $|K| = 120$

ID	BD iter	time (sec) BD	MILP	time-Δ (in %)	objective values bilevel	decision L	reaction F	Δ_1 (in %)	Δ_2 (in %)
1	80	1006	7200	86.02	226529	207033	311925	8.11	37.70
2	56	586	7200	91.86	202219	184667	273781	8.65	35.39
3	99	1713	7200	76.21	220495	201193	284490	8.33	29.02
4	67	738	7200	89.75	236130	215162	306186	8.46	29.67
5	66	672	7200	90.66	209856	191348	288013	8.82	37.24
6	48	441	7200	93.87	232214	212891	297931	8.32	28.30
7	63	576	7200	92.00	218428	201932	290763	7.55	33.12
8	61	582	7200	91.92	219693	204017	282384	7.14	28.54
9	73	675	7200	90.62	237787	217502	319097	8.53	34.19
10	57	554	7200	92.31	214507	198215	292341	7.60	36.29
avg	67	754	7200	89.52				8.15	32.95

Table 6.5: Computational results for $|K| = 140$

K	BD iter	time (sec) BD	time (sec) MILP	time-Δ (in %)	Δ_1 (in %)	Δ_2 (in %)
60	57.4	222	1217 [0]	81.80	4.55	37.25
80	56.5	302	2975 [1]	89.84	6.31	33.56
100	67.7	680	5196 [4]	86.91	6.88	37.55
120	65.4	566	5787 [5]	90.21	7.15	35.86
140	67	754	7200 [10]	89.52	8.15	32.95

Table 6.6: Average computational results

The results show that the average number of iterations of the BD did not increase significantly when the number of commodities was increased. The increased run time of the BD is caused by the higher number of slave subproblems that need to be solved. However, the run time of the MILP model increases significantly. This also leads to the increased number of instances that cannot be solved to optimality within the time limit. In the case of 140 commodities, no instance could be solved to optimality anymore. Even with a maximum run time of 14,400 seconds (4 hours), we could not find the optimal solution for any of these instances and the run time benefit of the BD would just increase. The BD already improves the run time in the smallest scenario by 82% on average. For the larger instances, the improvement increases further up to an average improvement of 90%, even though many MILP models were stopped after the predefined time limit.

The comparison of the objective values shows that the average difference between the bilevel and the over-regulated formulation Δ_1 increases almost linearly in the number of commodities. The increase is clear, as the over-regulated model can directly influence the flow of each commodity. However, the bilevel model can only decide for all commodities and the follower can still react. For a government, this means that different dangerous goods classes help to reduce the risk without directly regulating the carriers' transportation path.

The comparison with the follower reaction Δ_2 shows the importance of the bilevel formulation as the reaction leads to an increased risk of 35% on average.

6.5 Conclusion

We proposed a multi-cut BD method for solving the HTNDP. The bilevel problem is transformed into a MILP. Because of the multi-follower structure, the slave problem can be decomposed into independent subproblems and solved efficiently. The numerical results show that, especially for a large number of shipments, the run time improvements are significant and optimal solutions can be found fast. Moreover, the importance of the bilevel formulation is shown and that the classification of dangerous goods can help reduce the risk without regulating each shipment directly.

Chapter 7

Population-based Risk Equilibration for the Capacitated Multi-Mode Hazmat Transport Network Design Problem

The shipment of hazardous materials is necessary for most countries and many of these products are flammable, explosive or even radioactive. Despite high security standards, accidents still happen and the transportation of hazmat causes fear among the population who faces the risk of those accidents. Therefore, the society requests a fair distribution of risk by the authorities. To fairly distribute the risk, we propose a population-based risk definition that evaluates the risk in each population center. Moreover, we propose different objective functions for equilibrating the risk and extend the bilevel Hazmat Transport Network Design Problem by considering several transportation modes. In this problem, the government wants to equilibrate the risk among the population centers by restricting links to the shipment of hazardous goods. When taking that decision, the government has to anticipate the carriers' reaction who want to minimize the transportation costs for their shipments. This bilevel problem is transformed into single-level mixed-integer linear programs and solved with Xpress. In the numerical results, we show that just equilibrating risk can double the total risk in the network. Both objectives have a convex correlation and therefore an increase of 10% in total risk can already distribute the risk significantly better. Moreover, compared to classical approaches in the literature, we can distribute the risk similarly without increasing the total risk.

7.1 Introduction

Hazardous material accidents can have tremendous consequences for the population. One of the worst accidents of this kind in recent history happened in July 2013 in Lac-Mégantic, QC in Canada. A driverless train with 72 tank cars of petroleum crude oil derailed in the city center and caused the death of at least 42 persons. Moreover, at least 30 buildings and 115 businesses were destroyed and it took almost 2 days to control the fire. However, the transport of hazardous materials is essential not only for industrial countries like Canada, Germany and the United States but also for developing countries. The four most frequently shipped hazardous materials are - with 80% of the transported volume in Canada - crude petroleum, gasoline, fuel oils and non-metallic minerals (Searag et al., 2015). According to Bureau of Transportation Statistics and U.S. Census Bureau (2015), 2,580 million tons of hazardous materials where shipped throughout the United States in 2012. 59.4% of them where transported by truck, 4.3% by rail, 11% by water and 24.3% by pipeline in single mode transportation. Only 1% was shipped via intermodal transportation. In Canada, railways have a much higher relevance. In 2012, 26.1 million tons were transported by rail and 107.4 million tons by truck. A different structure of the network in Germany, which, compared to North America, is very dense, is reflected in the share of used transportation modes: In 2010, 56 million tons were transported by maritime transport, 48 million tons on inland waterways, 63 million tons by rail and 140 million tons by trucks (Statistisches Bundesamt Wiesbaden, 2012).

Thus, the consideration of different transportation modes is essential for the risk calculation whenever you wish to regulate the transport of hazardous materials. The different streams of research investigate the transportation of hazardous material either on roads (e.g. Kara and Verter, 2004) or on rail (e.g. Verma et al., 2011). We want to fill this gap by considering different transportation modes in the HTNDP.

Moreover, the society requests a fair distribution of risk over the population and the government or authority wants to achieve that by deciding if a link of the network is allowed for the transportation of hazardous materials or not. In the literature, the risk is associated with arcs (e.g. Kara and Verter, 2004). This definition neglects the fact that the risk in a population center is influenced by all the links in the area of the population center. This is certainly true if the network consists of different modes, but also if for example, several roads enter or pass by a city. In these definitions either the total risk of the network is minimized or the maximum arc risk is minimized for equilibration. A fair distribution of risk, however, will then strongly depend on the number of arcs in a population center. Moreover, if one arc has to transport a high amount of hazardous material, the maximum risk in the network will be defined by that arc and the distribution of all others arcs gets unimportant with respect to the maximum risk function. Consequently, we introduce a new population-based risk definition to evaluate the risk in population centers. For the fair distribution of risk among the population, different equilibration functions are introduced and compared. The proposed multi-mode multi-commodity bilevel formulation is transformed into a MILP and evaluated in a numerical study to show the benefits of the concept over classical risk definitions. We show that simply eqilibrating risk will also lead to a significant increase of the total risk in the network. Besides, all population centers may end up worse than before. We investigate the trade-off between risk equilibration and risk minimization and show a convex correlation between these two objectives. Therefore, with a small increase of total risk in the network, the distribution can be much better.

The contributions of this chapter are: (1) A new population-based definition of risk and equity risk measures for hazardous material shipments, (2) an extension of the HTNDP to multi-mode shipments and risk equilibration, (3) insights on the trade-off between risk equilibration and risk minimization, (4) a comparison to existing models from the literature (single-mode and maximum arc risk equilibration).

This chapter is structured as follows: In Section 7.2, the problem and its notation are introduced. The population-based risk definition and possible risk measures are shown in Section 7.3. Section 7.4 defines the multi-mode hazmat network design problem and the transformation to a MILP. A numerical study is presented in Section 7.5 before ending with the conclusion and an outlook.

7.2 Problem definition

The transportation network is represented by a graph $G = (N, A)$ with a set of nodes N and a set of arcs A. In a countrywide network, the nodes can be cities, facilities or important points in the network. In a city network, the level of detail needs to be much higher and the nodes represent junctions and entry and exit points of the city. Moreover, we consider different transportation modes $m \in M$. Depending on the detail of the model, these modes are the classical modes train, road, rail, air, water and pipeline; if the risk is equilibrated in a city, we also define different vehicle types as a transportation mode. Each arc $(i, j) \in A$ can have a capacity limit of a_{ij}^m for each mode

$m \in M$. Especially pipeline, rail and water transport systems have limitations on the possible amount of shipments. The road is usually the only mode without any capacity limits.

K is the set of commodities shipped through the network. Each commodity $k \in K$ is defined by an origin $o_k \in N$, a destination $d_k \in N$ and the transport volume ϕ_k. The transportation costs for shipping one unit of commodity $k \in K$ on arc $(i, j) \in A$ with transportation mode $m \in M$ are c_{ij}^{km}. Each commodity can be shipped partly via different transportation modes. However, we do not allow inter-modal transportation, as this is also not often the case in practice. The probability of an incident on arc $(i, j) \in A$ with mode $m \in M$ is given by σ_{ij}^{km}. Similar to the literature, it is assumed that there is no correlation between probabilities and therefore the probabilities are independent.

To equilibrate the risk among the population, we define a set of population centers C with a population P_c. In a global optimization setting, a population center represents a city; if the risk is equilibrated inside a city, these population centers need to represent districts or parts of the city.

Finally, l_{ij}^{mkc} defines the influence of an accident on arc (i, j) of commodity k on the population c using mode m. This influence factor depends on the distance between the population center and the arc, as well as the hazardous material type: The shorter the distance and the more dangerous the material is, the higher is the influence factor. The literature introduces different methods for calculating the influence of an accident on an arc: Batta and Chiu (1988) use a fixed bandwidth around the route segment, Erkut and Verter (1998) define a danger circle and Patel and Horowitz (1994) use a Gaussian plume model to define the impact of airborne hazmat accidents. We assume that these influence factors are given.

The problem is modeled as an LBP, where the leader is represented by the government or an authority. They can decide if the mode of an arc of the network is allowed for the transportation of hazardous materials or not and the decision is modeled by the binary decision variable y_{ij}^m. For simplification, the differentiation between specific hazardous material types is ignored. But the model could easily be extended to include this more realistic setting. The leader decision is subject to the follower optimization problem: The carriers minimize their transportation costs subject to demand satisfaction and capacity restrictions by deciding over the transportation percentage x_{ij}^{km} of commodity $k \in K$ shipped over arc $(i, j) \in A$ on transportation mode $m \in M$.

7.3 Population-based risk definition and evaluation

In the network design literature (e.g. Alp, 1995; Bianco et al., 2009; Kara and Verter, 2004), the risk calculation is associated with arcs. With P_{ij} being the accumulated effected population in the area of arc (i,j), Erkut and Verter (1998) define the risk of an arc by $\sum_{k \in K} \sigma_{ij}^{km} P_{ij} \phi_k x_{ij}^{km}$. In this section, we first introduce the population-based risk definition. Then we define different possible risk equilibration measures and give an example for the differences between the classical risk definition and ours.

7.3.1 Risk definition

In contrast to the classical network design definition of risk on arcs, we define the risk for each population center $c \in C$. Following this definition, we assume that only one accident can happen at the same time in a population center. Therefore, the accidents on the different arcs through a center are independent and the expected risk is defined as follows:

$$R_c(x) := P_c \sum_{m \in M} \sum_{k \in K} \sum_{(i,j) \in A} l_{ij}^{kc} \sigma_{ij}^{km} \phi_k x_{ij}^{km} \tag{7.1}$$

Thus, the risk of a population center is the sum of the transported volume in the influence area of the center weighted with the accident risk and the potential influence factor.

As long as the overall risk in the network is minimized, this risk definition is fully equivalent to the traditional risk definition, only the order of summation is changed. However, the differences can be huge for equilibrating the risk by minimizing the maximum risk.

The network of Figure 7.1 gives an example how the risk measures differ when we minimize the maximum risk. We assume two OD-pairs: 10 units from 1 to 4 via road and 10 units from 1 to 4 via train. In this example, an optimal solution will always ship the 10 train units via 2, as no other solution exists. However, the road commodity has two options, shipping either via 2 or via 3. If the maximum risk on each arc is minimized, both paths are the same from a risk perspective. Both will cause a total risk of 180 and no arc will be forbidden. The carrier will choose the cheapest path via 2 and population A will face a total risk of 280 and population B one of zero. If the maximum population risk is minimized, the government will close the road between 1 and 2 and the carrier will have to ship via 3. The risk for population A would be 100 and for population B 180. In both solutions, the maximum risk on an arc is 90 and the

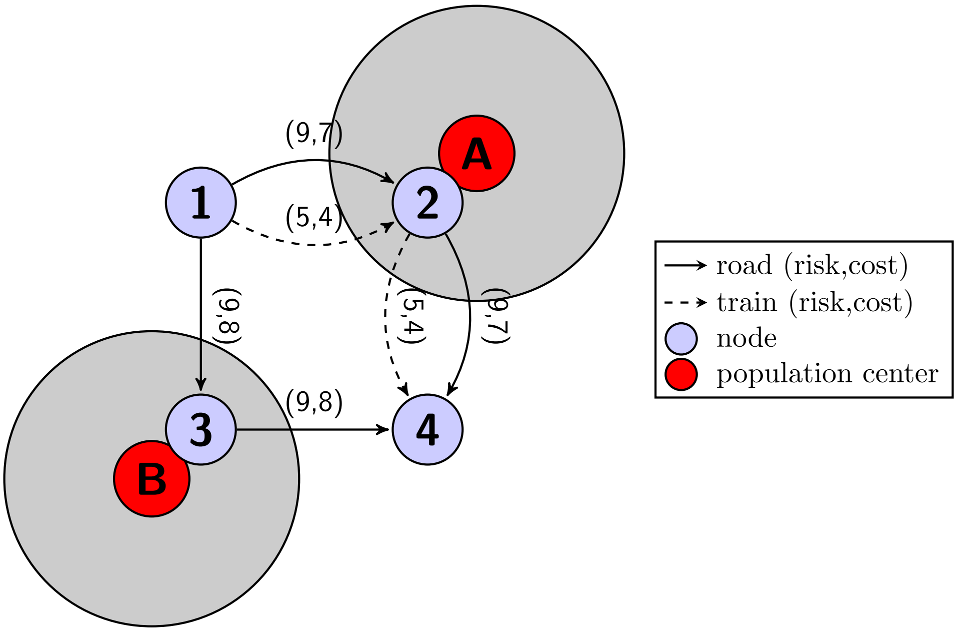


Figure 7.1: Example of the population-based risk definition

classical risk in the network is 280. The only difference is a fair distribution of the risk among the population.

7.3.2 Risk equilibration measures

Erkut and Ingolfsson (2005) summarized different measures the for evaluation of risk on a path and Bianco et al. (2009) equilibrated the risk by minimizing the maximum risk on an arc. Following these risk evaluation ideas, we introduce several possible risk measures for the population-based risk definition of the whole network.

$$\sum_{c \in C} R_c(x) \quad \text{Traditional/Overall risk (Trad)} \tag{7.2}$$

$$\max_{c \in C} R_c(x) \quad \text{Maximum risk (Max)} \tag{7.3}$$

$$\frac{1}{|C|} \sum_{c \in C} \left| R_c(x) - \frac{1}{|C|} \sum_{c' \in C} R_{c'}(x) \right| \quad \text{Average deviation to mean (AdM)} \tag{7.4}$$

$$\max_{c \in C} \left| R_c(x) - \frac{1}{|C|} \sum_{c' \in C} R_{c'}(x) \right| \quad \text{Maximum deviation to mean (MdM)} \tag{7.5}$$

$$\frac{1}{|C|(|C|-1)} \sum_{c,c' \in C | c <> c'} |R_c(x) - R_{c'}(x)| \quad \text{Average deviation among all (AdA)} \tag{7.6}$$

$$\max_{c,c' \in C | c <> c'} |R_c(x) - R_{c'}(x)| \quad \text{Maximum deviation among all (MdA)} \tag{7.7}$$

The traditional risk measure (7.2) sums the risk of all population centers and is equivalent to the arc definition of the risk. The maximum risk (7.3) minimizes the maximal risk in a population center. If each population center is defined by one arc, this definition is equivalent to the maximum arc risk definition by Bianco et al. (2009). The risk measures (7.5) - (7.7) are different deviation measures which are all zero if the risk is perfectly equilibrated and the risk in every population center is the same. While the first two calculate the average and maximum deviation to the mean, the last two give the average difference between all population centers and the maximum difference between two population centers.

As many of these risk measures can still lead to very high risks in some population areas, we introduce social bounds for the risk. Let L be the set of social bounds. Then each population center c can have the bounds b_c^l for each $l \in L$. If the risk is higher than a bound, a penalty p_c^l for each additional shipment is added to the objective value. The new risk is calculated by adding the following term to the objective function.

$$+\sum_{l \in L} p_c^l P_c \max \left\{ 0, \sum_{m \in M} \sum_{k \in K} \sum_{(i,j) \in A} l_{ij}^{kc} \sigma_{ij}^{km} x_{ij}^{km} - b_c^l \right\} \tag{7.8}$$

This gives a piecewise linear increasing objective function. Such functions are also often used to equilibrate the user's travel time in traffic assignment problems (Sheffi, 1985). Moreover, this idea is similar to the idea of conditional value at risk and perceived risk with a risk-averse population. For example, Abkowitz et al. (1992) and Erkut and Ingolfsson (2000) used a non-linear function $f(x) = x^\alpha$ with $\alpha > 1$ to take into account that accidents with high probability and low consequences are less undesirable than low probability-high consequence accidents.

7.4 Capacitated Multi-Mode Hazmat Transport Network Design Problem

In the following section, we first introduce the bilevel formulation for the capacitated multi-mode Hazmat Transport Network Design Problem (cmHTNDP) and explain how the general definition can be adapted to specific network types. Then we transform the model into a MILP.

7.4.1 Bilevel formulation

Besides the already introduced decision variables x_{ij}^{km} and y_{ij}^{m}, z^{km} is the percentage of commodity $k \in K$ shipped with mode m.

All introduced objective functions can be used in this leader problem and all shown transformations of this section can be applied as well. Considering the maximum risk objective function, the leader problem can be defined as follows:

$$\min r_{max} \tag{7.9}$$

$$\text{s.t. } P_c \sum_{m \in M} \sum_{k \in K} \sum_{(i,j) \in A} l_{ij}^{kc} \sigma_{ij}^{km} \phi_k x_{ij}^{km} \leq r_{max} \qquad \forall c \in C \tag{7.10}$$

$$r_{max} \geq 0 \tag{7.11}$$

$$y_{ij}^{m} \in \{0,1\} \qquad \forall (i,j) \in A, m \in M \tag{7.12}$$

The follower problem is a capacitated multi-mode transportation problem. The carriers decide how many percent z^{km} of commodity k are shipped via transportation mode m. Equation (7.14) is the flow conservation constraint and constraint (7.15) ensures that the full demand is divided into the different transportation modes. Constraint (7.16) is the capacity limit for each arc and transportation mode and equation (7.17) ensures that only arcs which are allowed by the leader can be used. In the objective function, the carriers' system-optimum - the overall transportation costs - are minimized.

$$\min \sum_{k \in K} \sum_{m \in M} \sum_{(i,j) \in A} c_{ij}^{km} x_{ij}^{km} \tag{7.13}$$

$$\text{s.t. } \sum_{(i,j) \in A} x_{ij}^{km} - \sum_{(j,l) \in A} x_{jl}^{km} = \begin{cases} 0, & \text{if } j \neq o_k, d_k \\ -z^{km}, & \text{if } j = o_k \\ z^{km}, & \text{if } j = d_k \end{cases} \qquad \forall j \in N, k \in K, m \in M \tag{7.14}$$

$$\sum_{m \in M} z^{km} = 1 \qquad \forall k \in K \tag{7.15}$$

$$\sum_{k \in K} x_{ij}^{km} \phi_k \leq a_{ij}^{m} y_{ij}^{m} \qquad \forall (i,j) \in A, m \in M \tag{7.16}$$

$$x_{ij}^{km} \leq y_{ij}^{m} \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.17}$$

$$x_{ij}^{km} \geq 0 \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.18}$$

$$z^{km} \geq 0 \qquad \forall k \in K, m \in M \tag{7.19}$$

This model is a generalization of the classical HTNDP. Depending on the network and setting, several special cases are possible: By using only one transportation mode and relaxing the capacity constraint, the follower problem is equivalent to the classical shortest path problem by Kara and Verter (2004). In an urban area setting, the only used transportation mode is the road with several vehicle types. Other transportation modes like rail exist, but the decision might be taken in a global network design problem and the caused risk of these modes can be included as constants into the leader objective function. Therefore, the capacity restriction can be neglected and the follower problem is a multi-mode shortest path problem.

By optimizing the risk distribution in a global setting like a province or a country, the modes describe not only different vehicle types, but also rail, pipeline and water transport. Because of the transport via rail, pipeline or water, the capacity restriction is necessary and the follower problem becomes a multi-commodity transportation problem.

7.4.2 Transformation to a mixed-integer linear program

To transform the LBP into a non-linear mixed-integer program, we assume the partial cooperation assumption and the follower problem can be replaced by the KKT conditions (Bard, 1998).

$$\min \; r_{max} \tag{7.20}$$

$$\text{s.t.} \tag{7.21}$$

$$P_c \sum_{m\in M}\sum_{k\in K}\sum_{(i,j)\in A} l_{ij}^{kc}\sigma_{ij}^{km}\phi_k x_{ij}^{km} \le r_{max} \qquad \forall c \in C \tag{7.22}$$

$$\sum_{(i,j)\in A} x_{ij}^{km} - \sum_{(j,l)\in A} x_{jl}^{km} = \begin{cases} 0, & \text{if } j \neq o_k, d_k \\ -z^{km}, & \text{if } j = o_k \\ z^{km}, & \text{if } j = d_k \end{cases} \qquad \forall j \in N, k \in K, m \in M \tag{7.23}$$

$$\sum_{m\in M} z^{km} = 1 \qquad \forall k \in K \tag{7.24}$$

$$\sum_{k\in K} x_{ij}^{km}\phi_k \le a_{ij}^m y_{ij}^m \qquad \forall (i,j) \in A, m \in M \tag{7.25}$$

$$x_{ij}^{km} \le y_{ij}^m \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.26}$$

$$\sum_{k\in K}\sum_{m\in M}\sum_{(i,j)\in A} c_{ij}^{km}x_{ij}^{km} \le \sum_{k\in K} v^k + \sum_{m\in M}\sum_{(i,j)\in A} a_{ij}^m \; s_{ij}^m y_{ij}^m + \sum_{m\in M}\sum_{(i,j)\in A}\sum_{k\in K} t_{ij}^{km} y_{ij}^m \tag{7.27}$$

$$u_j^{km} - u_i^{km} + t_{ij}^{km} + \phi_k s_{ij}^m \leq c_{ij}^{km} \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.28}$$
$$u_{o_k}^{km} - u_{d_k}^{km} + v^k \leq 0 \qquad \forall k \in K, m \in M \tag{7.29}$$
$$v^k, u_j^{km} \in \mathbb{R} \qquad \forall j \in N, k \in K \tag{7.30}$$
$$s_{ij}^m, t_{ij}^{km} \leq 0 \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.31}$$
$$x_{ij}^{km} \geq 0 \qquad \forall (i,j) \in A, k \in K, m \in M \tag{7.32}$$
$$z^{km} \geq 0 \qquad \forall k \in K, m \in M \tag{7.33}$$
$$r_{max} \geq 0 \tag{7.34}$$
$$y_{ij}^m \in \{0,1\} \qquad \forall (i,j) \in A, m \in M \tag{7.35}$$

u_j^{km}, v^k, s_{ij}^m and t_{ij}^{km} define the dual variables of the follower constraints (7.14) - (7.17). Equation (7.28) is the dual constraint of the primal variable x_{ij}^k and equation (7.29) is the dual constraint of the primal variable z^{km}.

As in (Cao and Chen, 2006), the optimality condition in (7.27) can be linearized by introducing the auxiliary variables w_{ij}^m and $\hat{w}_{ij}^{km}$ and a Big $\hat{M}$ and by replacing (7.27) with the following terms:

$$\sum_{k\in K}\sum_{m\in M}\sum_{(i,j)\in A} c_{ij}^{km}x_{ij}^{km} \leq \sum_{k\in K} v^k + \sum_{m\in M}\sum_{(i,j)\in A} a_{ij}^m w_{ij}^m + \sum_{m\in M}\sum_{(i,j)\in A}\sum_{k\in K} \hat{w}_{ij}^{km} \tag{7.36}$$
$$w_{ij}^m \leq s_{ij}^m + \hat{M}(1 - y_{ij}^m) \qquad \forall (i,j) \in A, m \in M \tag{7.37}$$
$$w_{ij}^m \geq s_{ij}^m \qquad \forall (i,j) \in A, m \in M \tag{7.38}$$
$$w_{ij}^m \geq -\hat{M} y_{ij}^m \qquad \forall (i,j) \in A, m \in M \tag{7.39}$$
$$w_{ij}^m \leq 0 \qquad \forall (i,j) \in A, m \in M \tag{7.40}$$
$$\hat{w}_{ij}^{km} \leq t_{ij}^{km} + \hat{M}(1 - y_{ij}^m) \qquad \forall (i,j) \in A, m \in M, k \in K \tag{7.41}$$
$$\hat{w}_{ij}^{km} \geq t_{ij}^{km} \qquad \forall (i,j) \in A, m \in M, k \in K \tag{7.42}$$
$$\hat{w}_{ij}^{km} \geq -\hat{M} y_{ij}^m \qquad \forall (i,j) \in A, m \in M, k \in K \tag{7.43}$$
$$\hat{w}_{ij}^{km} \leq 0 \qquad \forall (i,j) \in A, m \in M, k \in K \tag{7.44}$$

7.5 Numerical study

We use the Sioux Fall network from the literature (Bar-Gera, 2013) to show the benefits of the proposed model. The Sioux Fall network consists of 24 nodes and 76 arcs. Each arc is defined by a length (in km), however other necessary data was generated as follows:

The network is divided into six population centers. The population density of Sioux Fall is 814.4 inhabitants per square kilometer. The population distribution is shown with the definition of the population centers in Figure 7.2. The influence of an arc on a population center was set to 1 if the arc is inside the center, 0 if not. If an arc is contained in more than one center, the influence was proportionally split into two parts (e.g. arc (11,14)).

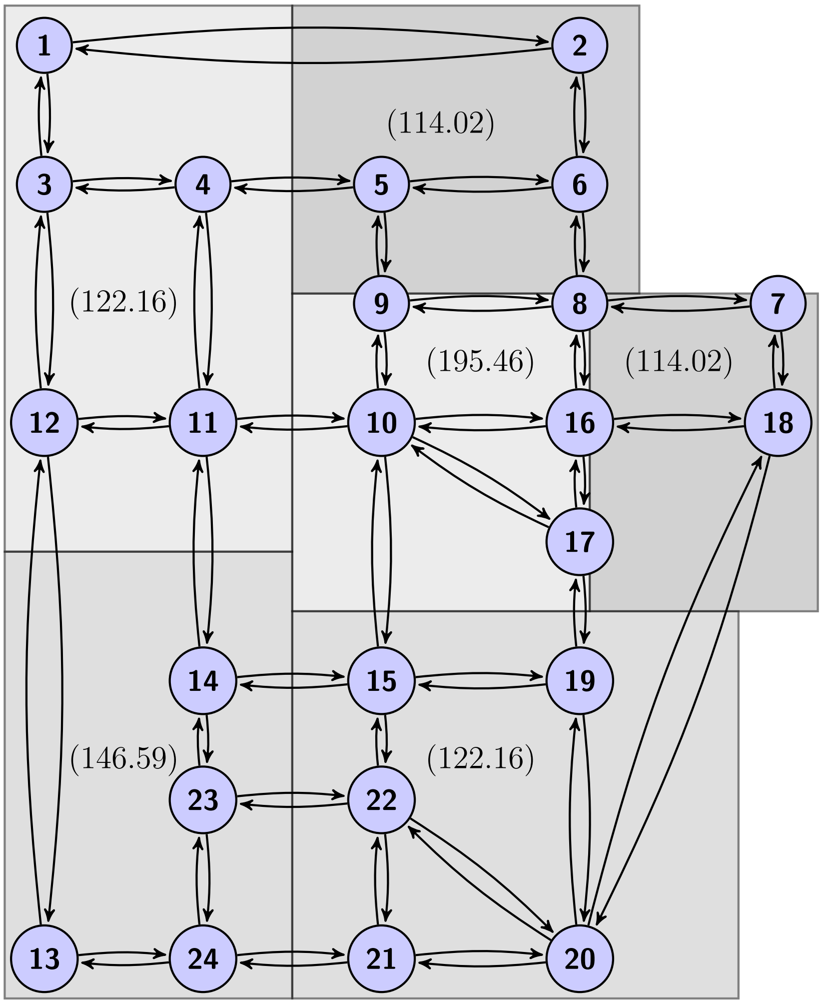


Figure 7.2: Sioux Fall network with population centers and population density

Two vehicle types are used with transportation costs of 1.1 per transported unit per km for the smaller vehicle and 0.9 per km for the larger vehicle. The risk of an accident of the larger vehicle was set 3% than the risk of the smaller one. The accident rate on an arc was generated randomly between 9.56×10^{-9} and 1.08×10^{-7} (Erkut and Gzara, 2008) and σ_{ij}^{km} is the product of the accident rate, the length of the arc and the factor for the vehicle type. We assume only one commodity type. Consequently, the risk for all shipped commodities are the same.

Four different demand scenarios were generated randomly. Nodes $1, 2, 13, 20$ were defined as the enter and exit nodes of the network. 43 commodities were shipped through the network (13 out-flows and 30 in-flows). The demand into the city (in-flows) was generated randomly between 100 and 1,000 and out of the city (out-flows) between 50 and 150.

The model of the previous section was solved in Xpress on an Intel Core i7 with 8 threads and 32GB RAM. In the results, we used the risk measure abbreviations of Section 7.3.2. For the non-linear function with l social bounds (NLl), we used a simple penalty function, which is defined as follows: The risk interval was divided into $l-1$ equidistant segments and from each point p, a further penalty of p^2 was added to the objective function.

In a first test, we evaluated the convergence of the model. Figure 7.3 shows that the lower bound stays zero and therefore the GAP cannot be calculated. This is the case for all measures that calculate the deviation: MdM, AdM, MdA, AdA. As mentioned in the definition of the measures, all population centers have the same risk in an optimal equilibration: zero. Most of the improvements happen within the first 20 minutes

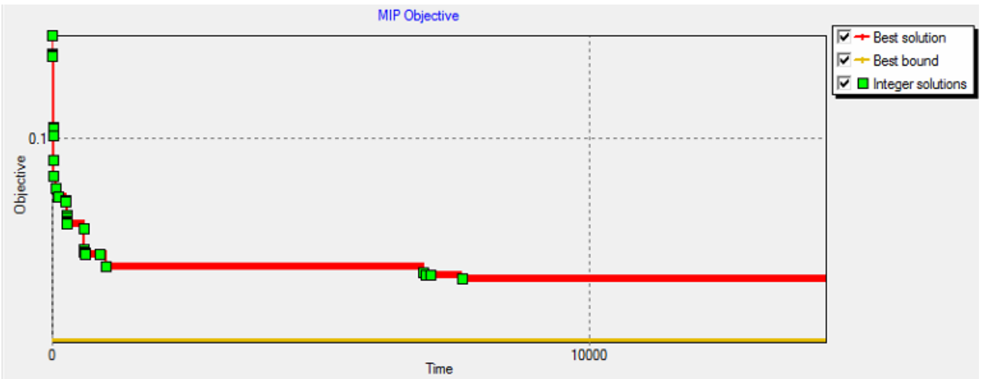


Figure 7.3: Run time analysis for AdA

and there are still some improvements within two hours. Therefore, the time limit of the numerical study was set to 7,200 seconds. For the other objective functions, the convergence was better.

We first show the effect of risk equilibration and the differences between risk equilibration and total risk minimization and discuss possible approaches to combine them. Then, we compare the effect of using different transportation modes in one model before discussing the difference between our model and the equilibration idea from the literature.

7.5.1 Risk equilibration

Table 7.1 shows the results for the different objective functions for demand scenario 1 and the risk of all population centers is reported. The optimized objective function is shown in the first column. Moreover, the optimized risk measure is highlighted in bold.

Objective function	Risk measures Trad	Max	AdM	MdM	AdA	MdA
Trad	**1.574**	0.525	0.155	0.262	0.207	0.419
Max	2.144	**0.370**	0.018	0.055	0.023	0.067
AdM	2.680	0.488	**0.014**	0.041	0.024	0.063
MdM	2.251	0.441	0.042	**0.066**	0.065	0.128
AdA	3.194	0.541	0.004	0.012	**0.007**	0.020
MdA	2.482	0.444	0.028	0.042	0.040	**0.072**
NL7	1.696	0.413	0.070	0.136	0.119	0.266

(a) Risk measures

Objective function	Risk of population centers Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6
Trad	0.525	0.179	0.464	0.105	0.169	0.132
Max	0.370	0.367	0.366	0.303	0.370	0.370
AdM	0.441	0.425	0.488	0.447	0.447	0.433
MdM	0.314	0.371	0.441	0.314	0.421	0.390
AdA	0.520	0.534	0.541	0.532	0.533	0.532
MdA	0.414	0.372	0.444	0.443	0.436	0.372
NL7	0.357	0.286	0.413	0.147	0.284	0.209

(b) Risk of population centers

Table 7.1: Risk evalutation for demand scenario 1

The results show that just minimizing the deviation or the maximum leads to an extreme increase in the overall risk in the network. The results are quite obvious as an equilibrium is only possible on a high level. It shows that, for an equal distribution of risk, almost every population center comes out worse and the total risk increases by more than 100%. Only the non-linear function, which does not try to equalize all populations, distributes the risk better without a dramatic risk increase. As the equilibration measures AdM, MdM, AdA, MdM perform very similar and the maximum risk is only effective if there is no population center with a very high risk, we will use the AdA measure for the following analysis.

7.5.2 Trade-off between risk equilibration and risk minimization

Since the pure risk equilibration increases the total risk significantly, we analyze the trade-off between minimizing the total risk in the network and equilibrating the risk by applying two approaches: First, we use a classical biobjective function that combines the AdA equilibration measure with the overall risk function Trad, which is weighted with $\alpha \in \{0.1, 0.2\}$. In a second approach, we minimize AdA and allow a maximum increase of $\beta \in \{5, 10, 15, 20, 25, 30, 35\}$ percent of the traditional risk by adding an auxiliary constraint.

Table 7.2 shows the results of the first approach for the 4 randomly generated demand scenarios (DSs). The minimal risk solutions are 1.574 (DS 1), 1.388 (DS 2), 1.399 (DS 3) and 1.335 (DS 4). One can see that the risk is better distributed among the population centers without increasing the risk as much as in the pure equilibration measures. Also, a higher weight on the risk minimization (Trad) leads, as expected, to a lower risk with worse equilibration. However, the results also show that it is difficult to find a good α for the objective function. For example, in DS 3 , the traditional risk increases with $\alpha = 0.2$ to 1.762 and in DS 4 to 1.768. The increase in DS 3 is higher than in DS 4. However, for $\alpha = 0.1$ this fact is again reverted.

DS	Obj func	Risk measure		Risk of population centers					
	AdA +	Trad	AdA	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6
1	0.1 Trad	2.179	0.015	0.370	0.362	0.378	0.357	0.370	0.341
	0.2 Trad	2.024	0.031	0.329	0.328	0.406	0.321	0.320	0.320
2	0.1 Trad	2.043	0.003	0.341	0.340	0.346	0.340	0.337	0.339
	0.2 Trad	1.983	0.014	0.340	0.335	0.347	0.320	0.322	0.320
3	0.1 Trad	1.875	0.015	0.322	0.308	0.318	0.307	0.309	0.310
	0.2 Trad	1.762	0.052	0.314	0.304	0.325	0.273	0.273	0.274
4	0.1 Trad	1.809	0.025	0.306	0.295	0.318	0.295	0.301	0.293
	0.2 Trad	1.768	0.030	0.306	0.304	0.308	0.289	0.284	0.278

Table 7.2: Results for 4 demand scenarios using the biobjective risk function

Table 7.3 shows the results for the second approach with the same demand scenarios. The distribution is already much better for an increase of 5 - 15% of the total risk. However, to distribute the risk as fairly as possible, an increase of more than 35%

DS	β	Risk measure Trad	AdA	Risk of population centers Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6
1	0	1.574	0.207	0.525	0.179	0.464	0.105	0.169	0.132
	5	1.653	0.130	0.407	0.264	0.423	0.183	0.195	0.181
	10	1.732	0.103	0.348	0.279	0.423	0.202	0.274	0.205
	15	1.809	0.082	0.346	0.277	0.423	0.242	0.280	0.240
	20	1.885	0.063	0.341	0.286	0.423	0.280	0.281	0.273
	25	1.968	0.047	0.366	0.303	0.397	0.294	0.301	0.305
	30	2.044	0.029	0.331	0.328	0.406	0.322	0.328	0.329
	35	2.125	0.021	0.353	0.364	0.381	0.330	0.351	0.346
2	0	1.388	0.177	0.393	0.175	0.448	0.164	0.134	0.074
	5	1.457	0.138	0.377	0.223	0.401	0.147	0.173	0.136
	10	1.527	0.113	0.375	0.247	0.373	0.171	0.194	0.167
	15	1.596	0.095	0.365	0.266	0.370	0.197	0.199	0.200
	20	1.666	0.084	0.358	0.294	0.363	0.208	0.221	0.222
	25	1.728	0.069	0.357	0.290	0.362	0.243	0.243	0.233
	30	1.804	0.052	0.343	0.340	0.345	0.263	0.260	0.254
	35	1.871	0.041	0.354	0.290	0.366	0.282	0.290	0.290
3	0	1.399	0.159	0.377	0.199	0.412	0.185	0.154	0.072
	5	1.469	0.118	0.356	0.253	0.366	0.170	0.189	0.136
	10	1.539	0.086	0.327	0.308	0.319	0.168	0.237	0.180
	15	1.609	0.064	0.317	0.305	0.326	0.211	0.238	0.211
	20	1.679	0.046	0.321	0.314	0.317	0.243	0.242	0.243
	25	1.741	0.032	0.318	0.308	0.318	0.263	0.265	0.269
	30	1.814	0.018	0.322	0.308	0.314	0.282	0.294	0.293
	35	1.874	0.008	0.321	0.311	0.316	0.313	0.313	0.301
4	0	1.335	0.165	0.347	0.217	0.419	0.099	0.171	0.081
	5	1.401	0.116	0.310	0.286	0.339	0.129	0.225	0.112
	10	1.468	0.088	0.302	0.306	0.309	0.148	0.254	0.148
	15	1.535	0.071	0.301	0.304	0.309	0.180	0.260	0.181
	20	1.600	0.054	0.302	0.306	0.309	0.210	0.254	0.219
	25	1.665	0.035	0.301	0.304	0.310	0.248	0.254	0.248
	30	1.730	0.025	0.293	0.305	0.318	0.274	0.275	0.265
	35	1.802	0.010	0.297	0.303	0.318	0.296	0.293	0.294

Table 7.3: Trade-off between Trad and AdA for 4 demand scenarios using the constrained approach

is necessary. Moreover, the equilibration of the first β steps is mainly achieved by a significant reduction of the risk in the population centers with the highest risk and a shift to low risk population centers. But especially in the last steps, the equilibration is achieved by increasing the risk in low risk population centers without reducing the risk in high risk population centers. For example in demand scenario 1, the risk of population center 1 drops from 0.525 to 0.407 in the first step ($\beta = 5\%$) and to 0.348 in the second step ($\beta = 10\%$). Population center 3 improves in the first step from 0.464 to 0.423. Even though the equilibration improves further for $\beta \geq 15$, this is mostly due to a risk increase in population centers 2, 4, 5 and 6.

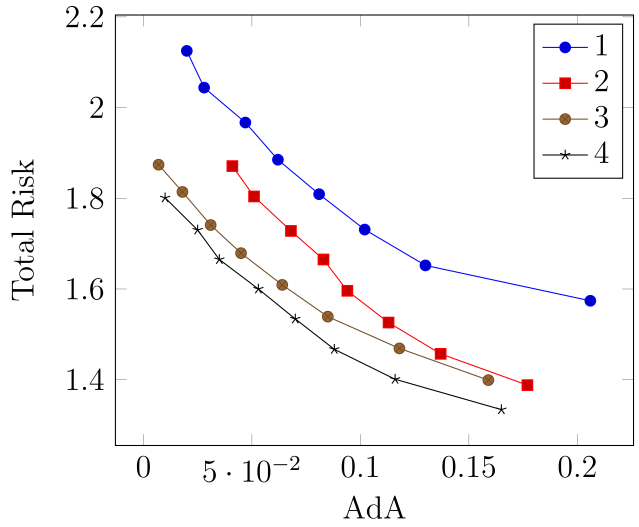


Figure 7.4: Trade-off between Trad and AdA for 4 demand scenarios using the constrained approach

Figure 7.4 shows the pareto-optimal curves of the trade-off between an equilibrated network and a network with a low total risk. The results indicate a convex dependency between both objective functions, which is consistent with our previous findings: With a small increase of total risk, the risk can be much better equilibrated. However, there comes a point from which on the price of total risk in the network for further equilibration is very high.

Besides these two approaches, the non-linear objective function of the previous section showed similar effects and can be a good alternative, especially when risk is percieved differently in different population centers. The biobjective function approach can achieve the same pareto curves as the constrained approach. However, using the second approach, it is easier to control the increase of risk and decide between pareto-optimal decisions and besides, one does not have to find the weights of the objective function.

7.5.3 Comparison to one mode decision model

In this subsection, we investigate the impact of the multi-mode decision model compared to classical single mode models. As in the multi-mode decision model, the mode is part of the decision process. We take the mode decision of the multi-mode model and solve the hazmat network design problem for mode 1 and 2 separately and add the results up (sum). The multi-mode result is compared with the sum of the single mode models. Moreover, the network design of the two single mode models is used in the multi-mode model to see the reaction of the followers on the single mode decisions (reaction). For all models, we used the non-linear objective function with 7 approximation points (NL7) as the non-linear function combined risk equilibration and risk minimization in one function. The detailed results are shown in Table 7.4.

DS		Risk measure		Risk of population centers					
	Model	Trad	AdA	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6
	multi mode	1.696	0.119	0.357	0.286	0.413	0.147	0.284	0.209
1	sum	1.687	0.127	0.381	0.298	0.402	0.130	0.267	0.209
	reaction	1.681	0.130	0.378	0.294	0.410	0.128	0.264	0.207
	multi mode	1.524	0.118	0.356	0.288	0.365	0.147	0.211	0.157
2	sum	1.522	0.118	0.356	0.289	0.365	0.147	0.208	0.157
	reaction	1.522	0.118	0.356	0.289	0.365	0.147	0.208	0.157
	multi mode	1.538	0.089	0.326	0.293	0.329	0.208	0.239	0.142
3	sum	1.547	0.099	0.362	0.267	0.329	0.215	0.233	0.140
	reaction	1.460	0.142	0.386	0.210	0.396	0.188	0.183	0.097
	multi mode	1.419	0.110	0.303	0.286	0.336	0.143	0.240	0.112
4	sum	1.423	0.118	0.308	0.329	0.303	0.090	0.262	0.131
	reaction	1.424	0.118	0.308	0.329	0.303	0.090	0.262	0.131

Table 7.4: Comparison of the multi-mode model with single-level decisions

In scenario 2, there is no difference between the three models. In the other three scenarios, the distribution got worse. In scenario 1, the total risk is reduced by 5%, but the equilibration is worse by 60%. Population centers 1 and 3, which are the ones with the highest risk, increase their risk. This shift towards risk minimization is mostly caused by the reaction of the followers if they are again allowed to change their mode. In scenario 4, the single mode models result even in a higher total risk and a worse equilibration. This worse equilibration is due to the fact that two equilibrated modes do not need to be equilibrated in the same way as when considering several modes.

7.5.4 Comparison to maximum arc risk equilibration

To equilibrate risk, the literature so far proposed to minimize the maximum arc risk (Bianco et al., 2009). In Table 7.5, we compare the solution of a maximum arc risk model to solutions of the pareto curve of the previous section for the four different demand scenarios.

DS	Objective function	Risk measure Trad	AdA	Risk of population centers Pop 1	Pop 2	Pop 3	Pop 4	Pop 5	Pop 6
1	max arc	1.890	0.217	0.591	0.211	0.534	0.210	0.209	0.135
	$\beta = 20$	1.885	0.063	0.341	0.286	0.423	0.280	0.281	0.273
	$\beta = 0$	1.574	0.207	0.525	0.179	0.464	0.105	0.169	0.132
2	max arc	1.789	0.173	0.458	0.264	0.503	0.209	0.211	0.144
	$\beta = 30$	1.804	0.052	0.343	0.340	0.345	0.263	0.260	0.254
	$\beta = 0$	1.388	0.177	0.393	0.175	0.448	0.164	0.134	0.074
3	max arc	1.975	0.170	0.447	0.254	0.508	0.378	0.247	0.142
	$\beta = 35$	1.874	0.008	0.321	0.311	0.316	0.313	0.313	0.301
	$\beta = 0$	1.399	0.159	0.377	0.199	0.412	0.185	0.154	0.072
4	max arc	1.760	0.176	0.450	0.350	0.426	0.178	0.271	0.086
	$\beta = 30$	1.730	0.025	0.293	0.305	0.318	0.274	0.275	0.265
	$\beta = 0$	1.335	0.165	0.347	0.217	0.419	0.099	0.171	0.081

Table 7.5: Comparison to maximum arc risk model

The results show for all demand scenarios that there exists a solution with a similar total risk in the network but a better distribution within the population centers and a solution with a similar distribution within the population centers but significantly smaller total risk. Using maximum arc risk increased the total risk by more than 35% without distributing the risk better. In all scenarios, every population center has a higher risk, than the risk minimal solution. However, compared to the equilibration measures for population centers, this does not lead to a better distribution of risk. The risk distribution remains more or less the same as in the minimal risk solution. Therefore, a similar risk distribution is always possible with the minimal overall risk solution.

7.6 Conclusions

We introduced a new population-based risk definition and extended the HTNDP to a multi-mode problem in oder to address the problem of risk equity. In the numerical study, we showed the superiority of the new definition over the arc risk definition and the necessity to consider multiple modes in the model. We also showed that the pure equilibration of risk increases the total risk significantly and that one has to find a trade-off between equilibration and risk minimization. But because of the convex correlation between these two measures, a small increase in the total risk can lead to a much better equilibration. As the problem is still very difficult to solve, enhancements for solving this problem should be considered in further research.

Chapter 8

Conclusion

8.1 Summary

We introduced a BD algorithm for solving DCLBP. The bilevel formulation is transformed into a MILP using KKT conditions. In this MILP, the binary variables were used as the complicating variables to decompose the problem into a master and a slave problem for applying BD. The efficiency of the method was shown on four different problems: the DNDP (Chapter 3), the DDNDP (Chapter 4), the DCFSP (Chapter 5) and the HTNDP (Chapter 6). Depending on the problem structure, different acceleration methods were used in the BD.

Since the DNDP has non-linear travel time functions in the objective function, we further proposed a linear approximation of these convex functions without introducing binary auxiliary variables. Moreover, the slave problem was decomposed into two subproblems for a fast calculation of the dual variables for the BD. The numerical study showed run time improvements of more than 60% over the MILP.

For computing maintenance schedules in traffic networks, we extended the DNDP to a multi-period model. Because of the problem structure, the slave problem of the DDNDP was decomposed within the periods and a multi-cut version of the BD was used. Even though the BD did not reach convergence, the found solutions were better than a genetic algorithm and simple priority rules, which currently might be used in practice. Further, we showed that, especially with tight budgets, the BD is the only method among all tested approaches that finds feasible solutions.

For the DCFSP and the HTNDP pareto-optimal cuts were generated to further improve the convergence. The results showed run time improvements of more than 90% compared to the MILP formulation for both problems. In the HTNDP, the multi-follower structure was used to decompose the slave problem and apply again the multi-cut version of BD. Moreover, we pointed out that the bilevel formulation in the HTNDP is necessary and single-level approaches can cause a higher risk for the population. The results further underlined that, a good classification of dangerous goods, too, can reduce the risk.

In the last chapter, the single mode risk minimization of the HTNDP was questioned. On the one side, the model was extended to a multi-mode network design problem. On the other side, a new population based risk definition was introduced to distribute the risk fairly among the population. The numerical results showed that pure risk equilibration leads to a very high total risk in the network and every population center in the network can end up with a higher risk. Therefore, a trade-off between risk minimization and risk equilibration is necessary and the authorities need to decide between pareto-optimal solutions. A convex correlation between these two objectives showed that the risk can be significantly better distributed by very slightly incrasing the total risk. Moreover, multiple modes need to be considered and risk equilibration over arcs, as it is done in the literature so far, does not distribute the risk among the population fairly.

8.2 Limitations and future research

The proposed BD algorithm was only tested for network design problems and a bilevel facility selection problem. But bilevel programming has a wide range of applications. Testing the algorithm on other bilevel problems with different structures might provide further insights into limitations and advantages of the BD. The used problems are special cases of bilevel problems with binary leader decision variables. Adapting the algorithm to integer or even continuous leader decision variables, could further extend the applicability of the method.

So far, we only used constraints which depend on the leader decision variables in the leader problem. If leader constraints depend on leader and follower decision variables, the KKT transformation is no longer valid. In this case, the leader also has to anticipate whether or not the follower decision is valid with respect to these so-called coupling constraints. Therefore, extending the approach in this direction would be interesting for future research, as well.

Especially the cmHTNDP showed that further enhancements of the solution method are necessary since realistic networks can be very large. This can be done, for example, by applying the BD to this problem or developing efficient heuristics.

Bibliography

ALK Associates (1994). *ALK's PC*HazRoute (Version 2.0).* ALK Associates, Inc., 1000 Herrontown Road, Princeton, NJ, USA.

Abdulaal, Mustafa and Larry J. LeBlanc (1979). "Continuous equilibrium network design models". In: *Transportation Research Part B: Methodological* 13.1, pp. 19–32.

Abkowitz, Mark D, Mark Lepofsky, and Paul Cheng (1992). "Selecting criteria for designating hazardous materials highway routes". In: *Transportation Research Record* 1333, pp. 30–35.

Abkowitz, Mark, Joseph DeLorenzo, Ron Duych, Art Greenberg, and Tom McSweeney (2001). "Assessing the economic effect of incidents involving truck transport of hazardous materials". In: *Transportation Research Record: Journal of the Transportation Research Board* 1763, pp. 125–129.

Alp, Ertugrul (1995). "Risk-based transportation planning practice: Overall methodology and a case example". In: *Infor* 33.1, p. 4.

Amaldi, Edoardo, Maurizio Bruglieri, and Bernard Fortz (2011). "On the hazmat transport network design problem". In: *Network Optimization Lecture Notes in Computer Science*. Ed. by Julia Pahl, Torsten Reiners, and Stefan Voß. Vol. 6701. Springer Berlin Heidelberg, pp. 327–338.

Avella, Pasquale and Maurizio Boccia (2009). "A Cutting Plane Algorithm for the Capacitated Facility Location Problem". English. In: *Computational Optimization and Applications* 43 (1), pp. 39–65. ISSN: 0926-6003.

Bar-Gera, H. (2013). *Transportation Network Test Problems.* URL: `http://www.bgu.ac.il/~bargera/tntp/` (visited on 07/02/2013).

Bard, J. and J. Moore (1990). "A Branch and Bound Algorithm for the Bilevel Programming Problem". In: *SIAM Journal on Scientific and Statistical Computing* 11.2, pp. 281–292.

Bard, Jonathan F (1998). *Practical Bilevel Optimization: Algorithms and Applications.* Springer.

Bard, Jonathan F and James E Falk (1982). "An explicit solution to the multi-level programming problem". In: *Computers & Operations Research* 9.1, pp. 77–100.

Batta, Rajan and Samuel S Chiu (1988). "Optimal obnoxious paths on a network: transportation of hazardous materials". In: *Operations Research* 36.1, pp. 84–92.

Ben-Ayed, Omar and Charles E Blair (1990). "Computational Difficulties of Bilevel Linear Programming". In: *Operations Research* 38.3, pp. 556–560.

Benders, J.F. (1962). "Partitioning Procedures for Solving Mixed-Variables Programming Problems". In: *Numerische Mathematik* 4.1, pp. 238–252.

Bialas, Wayne F. and Mark H. Karwan (1984). "Two-Level Linear Programming". In: *Management Science* 30.8, pp. 1004–1020.

Bianco, Lucio, Massimiliano Caramia, and Stefano Giordani (2009). "A bilevel flow model for hazmat transportation network design". In: *Transportation Research Part C: Emerging Technologies* 17.2, pp. 175–196.

Bianco, Lucio, Massimiliano Caramia, Stefano Giordani, and Veronica Piccialli (2015). "A Game-Theoretic Approach for Regulating Hazmat Transportation". In: *Transportation Science*.

Birge, John R and Francois V Louveaux (1988). "A multicut algorithm for two-stage stochastic linear programs". In: *European Journal of Operational Research* 34.3, pp. 384–392.

Bracken, Jerome and James T McGill (1973). "Mathematical programs with optimization problems in the constraints". In: *Operations Research* 21.1, pp. 37–44.

Braess, D. (1968). "Über ein Paradoxon aus der Verkehrsplanung". In: *Unternehmensforschung* 12.1, pp. 258–268.

Braess, Dietrich, Anna Nagurney, and Tina Wakolbinger (2005). "On a Paradox of Traffic Planning". In: *Transportation Science* 39.4, pp. 446–450.

Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) (2014a). *Sonderprogramm Brückenmodernisierung.* URL: `http://www.bmvi.de/DE/VerkehrUndMobilitaet/Verkehrstraeger/Strasse/SicherheitVonBruecken/systematische-brueckenertuechtigung_node.html` (visited on 01/20/2015).

— (2014b). *Verkehrsverflechtungsprognose 2030.*

— (2016). *Bundesverkehrswegeplan 2030 (Entwurf März 2016).*

Bureau of Public Roads (1964). *Traffic assignment manual.* U.S. Department of Commerce, Urban Planning Devision. Washington, DC.

Bureau of Transportation Statistics and U.S. Census Bureau (2015). *Transportation-Commodity Flow Survey.* Tech. rep. (Visited on 02/03/2015).

Calvete, Herminia I, Carmen Galé, and María-José Oliveros (2011). "Bilevel model for production–distribution planning solved by using ant colony optimization". In: *Computers & operations research* 38.1, pp. 320–327.

Calvete, Herminia I, Carmen Galé, and José A Iranzo (2014). "Planning of a decentralized distribution network using bilevel optimization". In: *Omega* 49, pp. 30–41.

Candler, Wilfred and Robert Townsley (1982). "A linear two-level programming problem". In: *Computers & Operations Research* 9.1, pp. 59–76.

Cao, Dong and Mingyuan Chen (2006). "Capacitated Plant Selection in a Decentralized Manufacturing Environment: A Bilevel Optimization Approach". In: *European Journal of Operational Research* 169.1, pp. 97–110.

Carotenuto, Pasquale, Stefano Giordani, and Salvatore Ricciardelli (2007). "Finding minimum and equitable risk routes for hazmat shipments". In: *Computers & Operations Research* 34.5, pp. 1304–1327.

Codato, Gianni and Matteo Fischetti (2004). "Combinatorial benders cuts". In: *Integer Programming and Combinatorial Optimization*. Springer, pp. 178–195.

Colson, Benoît, Patrice Marcotte, and Gilles Savard (2005). "Bilevel programming: A survey". In: *4OR* 3.2, pp. 87–107.

— (2007). "An overview of bilevel optimization". In: *Annals of Operations Research* 153.1, pp. 235–256.

Cote, Gilles and Michael A Laughton (1984). "Large-scale mixed integer programming: Benders-type heuristics". In: *European Journal of Operational Research* 16.3, pp. 327–333.

DB Mobility Networks Logistics (DB) (2014). *Bahn startet größtes Modernisierungsprogramm der Infrastruktur*. URL: `http://www.deutschebahn.com/de/presse/presseinformationen/pi_it/8577046/ubd20141208.html?start=0&itemsPerPage=20` (visited on 01/20/2015).

Dempe, Stephan (2002). *Foundations of bilevel programming*. Springer Science & Business Media.

Dempe, Stephan, Vyacheslav Kalashnikov, Gerardo A Pérez-Valdés, and Nataliya Kalashnykova (2015). *Bilevel programming problems: theory, algorithms and applications to energy networks*. Springer.

Eiselt, Horst A and Gilbert Laporte (1997). "Sequential location problems". In: *European Journal of Operational Research* 96.2, pp. 217–231.

Erkut, Erhan and Osman Alp (2007). "Designing a road network for hazardous materials shipments". In: *Computers & Operations Research* 34.5, pp. 1389–1405.

Erkut, Erhan and Fatma Gzara (2008). "Solving the hazmat transport network design problem". In: *Computers & Operations Research* 35.7, pp. 2234–2247.

Erkut, Erhan and Armann Ingolfsson (2000). "Catastrophe avoidance models for hazardous materials route planning". In: *Transportation Science* 34.2, pp. 165–179.

— (2005). "Transport risk models for hazardous materials: revisited". In: *Operations Research Letters* 33.1, pp. 81–89.

Erkut, Erhan and Vedat Verter (1998). "Modeling of transport risk for hazardous materials". In: *Operations research* 46.5, pp. 625–642.

Erkut, Erhan, Stevanus A. Tjandra, and Vedat Verter (2007). "Chapter 9 Hazardous Materials Transportation". In: *Transportation*. Ed. by Cynthia Barnhart and Gilbert Laporte. Vol. 14. Handbooks in Operations Research and Management Science. Elsevier, pp. 539 –621.

Ertogral, Kadir and S David Wu (2000). "Auction-theoretic coordination of production planning in the supply chain". In: *IIE transactions* 32.10, pp. 931–940.

Farahani, Reza Zanjirani, Elnaz Miandoabchi, W.Y. Szeto, and Hannaneh Rashidi (2013). "A review of urban transportation network design problems". In: *European Journal of Operational Research* 229.2, pp. 281–302.

Farvaresh, Hamid and Mohammad Mehdi Sepehri (2011). "A single-level mixed integer linear formulation for a bi-level discrete network design problem". In: *Transportation Research Part E: Logistics and Transportation Review* 47.5, pp. 623–640.

— (2013). "A branch and bound algorithm for bi-level discrete network design problem". In: *Networks and Spatial Economics* 13.1, pp. 67–106.

Fontaine, Pirmin and Stefan Minner (2014). "Benders Decomposition for discrete–continuous linear bilevel problems with application to traffic network design". In: *Transportation Research Part B: Methodological* 70, pp. 163–172.

— (2016a). "A dynamic discrete network design problem for maintenance planning in traffic networks". In: *Annals of Operations Research*, pp. 1–16.

— (2016b). "Benders Decomposition for the Decentralized Facility Selection Problem". Working paper.

— (2016c). "Benders Decomposition for the Hazmat Transport Network Design Problem". Working paper.

Fontaine, Pirmin, Teodor Gabriel Crainic, Michel Gendreau, and Stefan Minner (2016). "Population based Risk Equilibration for the Hazmat Transport Network Design Problem". Working paper.

Friesz, Terry L, Hsun-Jung Cho, Nihal J Mehta, Roger L Tobin, and G Anandalingam (1992). “A simulated annealing approach to the network design problem with variational inequality constraints”. In: *Transportation Science* 26.1, pp. 18–26.

Friesz, Terry L et al. (1993). “The multiobjective equilibrium network design problem revisited: A simulated annealing approach”. In: *European Journal of Operational Research* 65.1, pp. 44–57.

Gao, Ziyou, Jianjun Wu, and Huijun Sun (2005). “Solution algorithm for the bi-level discrete network design problem”. In: *Transportation Research Part B: Methodological* 39.6, pp. 479–495.

Geoffrion, Arthur M (1972). “Generalized Benders Decomposition”. In: *Journal of Optimization Theory and Applications* 10.4, pp. 237–260.

Geoffrion, Arthur M and Glenn W Graves (1974). “Multicommodity distribution system design by Benders decomposition”. In: *Management science* 20.5, pp. 822–844.

Gopalan, Ram, Krishna S Kolluri, Rajan Batta, and Mark H Karwan (1990). “Modeling equity of risk in the transportation of hazardous materials”. In: *Operations Research* 38.6, pp. 961–973.

Grossmann, Ignacio E and Christodoulos A Floudas (1987). “Active constraint strategy for flexibility analysis in chemical processes”. In: *Computers & Chemical Engineering* 11.6, pp. 675–693.

Hansen, Pierre, Brigitte Jaumard, and Gilles Savard (1992). “New branch-and-bound rules for linear bilevel programming”. In: *SIAM Journal on Scientific and Statistical Computing* 13.5, pp. 1194–1217.

Hillier, Frederick S and Gerald J Lieberman (2009). *Introduction to operations research.* 9th ed. Tata McGraw-Hill Education.

Huth, Peter (2015). *Willkommen im Stau: Diese Baustellen halten Berlin 2015 in Atem.* URL: `http://www.bz-berlin.de/berlin/diese-baustellen-halten-berlin-2015-in-atem` (visited on 03/23/2016).

Imamoto, A and B Tang (2008). “Optimal piecewise linear approximation of convex functions”. In: *Proceedings of the World Congress on Engineering and Computer Science*. Citeseer, pp. 1191–1194.

Kara, Bahar Y and Vedat Verter (2004). “Designing a road network for hazardous materials transportation”. In: *Transportation Science* 38.2, pp. 188–196.

Kara, Bahar Yetis, Erhan Erkut, and Vedat Verter (2003). “Accurate calculation of hazardous materials transport risks”. In: *Operations Research Letters* 31.4, pp. 285–292.

Labbé, Martine, Patrice Marcotte, and Gilles Savard (1998). "A bilevel model of taxation and its application to optimal highway pricing". In: *Management science* 44.12-part-1, pp. 1608–1622.

LeBlanc, Larry J (1975). "An algorithm for the discrete network design problem". In: *Transportation Science* 9.3, pp. 183–199.

Lindner-Dutton, Laurel, Rajan Batta, and Mark H Karwan (1991). "Equitable sequencing of a given set of hazardous materials shipments". In: *Transportation Science* 25.2, pp. 124–137.

Luathep, Paramet, Agachai Sumalee, William HK Lam, Zhi-Chun Li, and Hong K Lo (2011). "Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach". In: *Transportation Research Part B: Methodological* 45.5, pp. 808–827.

Magnanti, Thomas L and Richard T Wong (1981). "Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria". In: *Operations Research* 29.3, pp. 464–484.

Marcotte, Patrice, Anne Mercier, Gilles Savard, and Vedat Verter (2009). "Toll policies for mitigating hazardous materials transport risk". In: *Transportation science* 43.2, pp. 228–243.

Martin, Richard Kipp (1999). *Large Scale Linear and Integer Optimization: A Unified Approach*. Kluwer Academic Publisher.

Mathew, Tom V and Sushant Sharma (2009). "Capacity expansion problem for large urban transportation networks". In: *Journal of Transportation Engineering* 135.7, pp. 406–415.

Meng, Qiang and Hai Yang (2002). "Benefit distribution and equity in road network design". In: *Transportation Research Part B: Methodological* 36.1, pp. 19–35.

Migdalas, Athanasios, Panos M Pardalos, and Peter Värbrand (2013). *Multilevel optimization: algorithms and applications*. Vol. 20. Springer Science & Business Media.

Moore, James T and Jonathan F Bard (1990). "The mixed integer linear bilevel programming problem". In: *Operations research* 38.5, pp. 911–921.

Nemhauser, George L and Laurence A Wolsey (1988). *Integer and Combinatorial Optimization*. Wiley New York.

Neumann, Peter (2016). *Verkehrsprobleme und Staus: 2016 gibt es in Berlin noch mehr Baustellen*. URL: `http://www.berliner-zeitung.de/berlin/verkehrsprobleme-und-staus-2016-gibt-es-in-berlin-noch-mehr-baustellen-23355300` (visited on 03/23/2016).

Ng, ManWo, Dung Ying Lin, and S Travis Waller (2009). "Optimal long-term infrastructure maintenance planning accounting for traffic dynamics". In: *Computer-Aided Civil and Infrastructure Engineering* 24.7, pp. 459–469.

Nguyen, Sang (1974). "An Algorithm for the Traffic Assignment Problem." In: *Transportation Science* 8.3, pp. 203–216.

Papadakos, Nikolaos (2008). "Practical enhancements to the Magnanti–Wong method". In: *Operations Research Letters* 36.4, pp. 444–449.

Patel, Minnie H and Alan J Horowitz (1994). "Optimal routing of hazardous materials considering risk of spill". In: *Transportation Research Part A: Policy and Practice* 28.2, pp. 119–132.

Poorzahedy, Hossain and Farhad Abulghasemi (2005). "Application of ant system to network design problem". In: *Transportation* 32.3, pp. 251–273.

Poorzahedy, Hossain and Omid M Rouhani (2007). "Hybrid meta-heuristic algorithms for solving network design problem". In: *European Journal of Operational Research* 182.2, pp. 578–596.

Poorzahedy, Hossain and Mark A. Turnquist (1982). "Approximate algorithms for the discrete network design problem". In: *Transportation Research Part B: Methodological* 16.1, pp. 45–55.

ReVelle, Charles, Jared Cohon, and Donald Shobrys (1991). "Simultaneous siting and routing in the disposal of hazardous wastes". In: *Transportation Science* 25.2, pp. 138–145.

Rei, Walter, Jean-François Cordeau, Michel Gendreau, and Patrick Soriano (2009). "Accelerating Benders decomposition by local branching". In: *INFORMS Journal on Computing* 21.2, pp. 333–345.

Saccomanno, F Frank and AY-W Chan (1985). "Economic evaluation of routing strategies for hazardous road shipments". In: *Transportation Research Record* 1020.

Saharidis, GK and MG Ierapetritou (2009). "Resolution method for mixed integer bilevel linear problems based on decomposition technique". In: *Journal of Global Optimization* 44.1, pp. 29–51.

Sahinidis, NV and Ignacio E Grossmann (1991). "Convergence Properties of Generalized Benders Decomposition". In: *Computers & Chemical Engineering* 15.7, pp. 481–491.

Schmidt, Peter T. and Sascha Karowski (2016). *Hier droht Stau! Der Bauplan fr 2016.* URL: `http://www.merkur.de/lokales/muenchen/stadt-muenchen/muenchen-bauplan-moegliche-stauquellen-jahr-2016-meta-6021787.html` (visited on 03/23/2016).

Schneeweiß, Christoph (2003). *Distributed Decision Making.* 2nd ed. Springer.

Searag, Sagal, Greog Maloney, and Larry McKeown (2015). *Trucking dangerous goods in Canada, 2004 to 2012.* Tech. rep. Statistics Canada. (Visited on 02/03/2015).

Sheffi, Yosef (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Prentice-Hall, Inc.

Sherali, Hanif D and Brian J Lunday (2013). "On generating maximal nondominated Benders cuts". In: *Annals of Operations Research* 210.1, pp. 57–72.

Stackelberg, Heinrich von (1934). *Marktform und Gleichgewicht.* J. springer.

Statistisches Bundesamt Wiesbaden (2012). "Verkehr Gefahrguttransporte". In: *Fachserie 8 Reihe 1.4.*

— (2015). "Verkehr: Gefahrguttransporte 2013". In: *Fachserie 8 Reihe 1.4.*

Suh, Sunduck and TschanghoJohn Kim (1992). "Solving nonlinear bilevel programming models of the equilibrium network design problem: A comparative review". In: *Annals of Operations Research* 34.1, pp. 203–218.

Sun, Longsheng, Mark H. Karwan, and Changhyun Kwon (2015). "Robust Hazmat Network Design Problems Considering Risk Uncertainty". In: *Transportation Science.*

Talbi, El-Ghazali (2013). *Metaheuristics for bi-level optimization.* Vol. 482. Springer.

TomTom International B.V. (2014). *TomTom European Traffic Index 2013.* URL: `http://www.tomtom.com/en_gb/trafficindex/` (visited on 01/20/2015).

— (2016). *TomTom European Traffic Index 2015.* URL: `http://www.tomtom.com/en_gb/trafficindex/` (visited on 03/23/2016).

Van Slyke, Richard M and Roger Wets (1969). "L-shaped linear programs with applications to optimal control and stochastic programming". In: *SIAM Journal on Applied Mathematics* 17.4, pp. 638–663.

Verma, Manish, Vedat Verter, and Michel Gendreau (2011). "A tactical planning model for railroad transportation of dangerous goods". In: *Transportation Science* 45.2, pp. 163–174.

Verter, Vedat and Bahar Y Kara (2008). "A path-based approach for hazmat transport network design". In: *Management Science* 54.1, pp. 29–40.

Völklein, Marco (2015). *Baustellen in München.* URL: `http://www.sueddeutsche.de/muenchen/baustellen-in-muenchen-hier-stehen-sie-im-stau-1.2287759` (visited on 01/20/2015).

Wang, David ZW and Hong K Lo (2010). "Global optimum of the linearized network design problem with equilibrium flows". In: *Transportation Research Part B: Methodological* 44.4, pp. 482–492.

Wang, Shuaian, Qiang Meng, and Hai Yang (2013). "Global optimization methods for the discrete network design problem". In: *Transportation Research Part B: Methodological* 50, pp. 42–60.

Wardrop, John Glen (1952). "Some Theoretical Aspects of Road Traffic Research". In: *ICE Proceedings: Engineering Divisions*. Vol. 1. 3. Thomas Telford, pp. 325–362.

White, Douglas J and G Anandalingam (1993). "A penalty function approach for solving bi-level linear programs". In: *Journal of Global Optimization* 3.4, pp. 397–419.

Xu, Tianze, Heng Wei, and Guanghua Hu (2009). "Study on continuous network design problem using simulated annealing and genetic algorithm". In: *Expert Systems with Applications* 36.2, pp. 1322–1328.