

Methoden der mathematischen Optimierung zur Planung taktischer Wagenrouten im Einzelwagenverkehr

Zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)

von der Fakultät Maschinenbau
der Technischen Universität Dortmund
genehmigte Dissertation

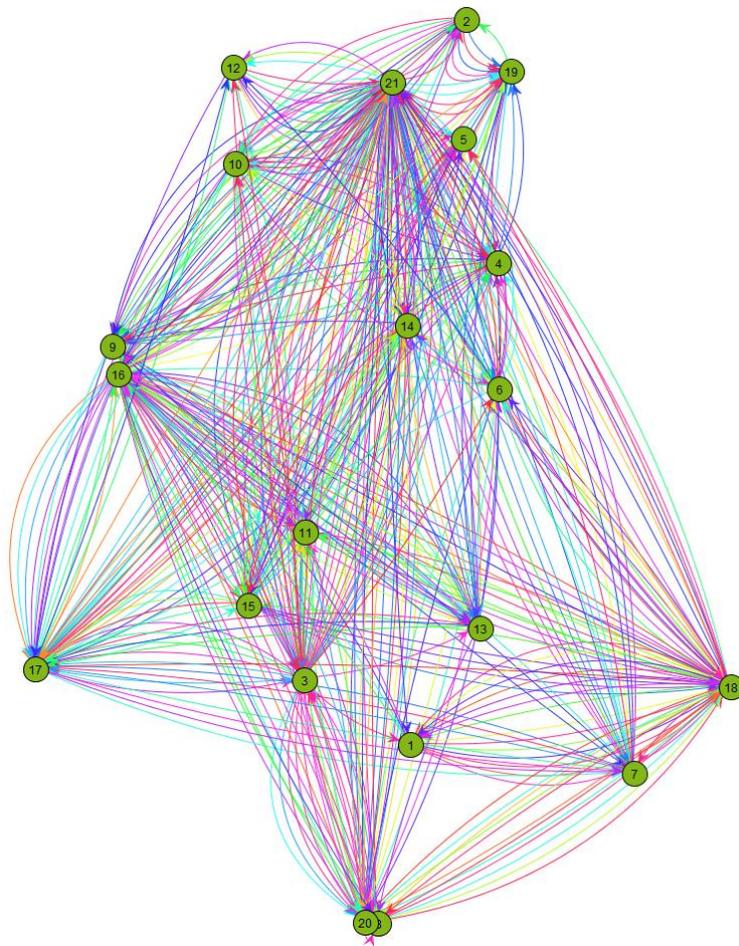
vorgelegt von
Dipl.-Math. Robert Voll
aus Essen

Tag der mündlichen Prüfung: 20.03.2014
1. Gutachter: Prof. Dr.-Ing. Uwe Clausen
2. Gutachter: Prof. Dr. Marco Lübbecke

Dortmund, 2014

*Einer verwirrenden Vielfalt
von Möglichkeiten liegt sehr
häufig eine einzige, simple
Struktur zu Grunde –*

aus Maarten t'Harts
In Unnütz toller Wut



Inhaltsverzeichnis

1	Einleitung	1
1.1	Einführung in die Problemstellung	1
1.2	Ziel und Aufbau der Arbeit	1
2	Logistik des deutschen Einzelwagenverkehrs	5
2.1	Verkehrsträger Schiene	5
2.1.1	Akteure des Schienengüterverkehrs	6
2.1.2	Produktionsformen des Schienengüterverkehrs	7
2.1.3	Deutsche Schieneninfrastruktur	8
2.2	Einzelwagenverkehr in Deutschland	10
2.2.1	Zukunftsansichten des Einzelwagenverkehrs	10
2.2.2	Prozessschritte der Auftragsdurchführung	12
2.2.3	Planungsebenen	12
2.3	Umstellpläne und Wagenrouten	14
2.3.1	Kostenstruktur des Einzelwagenverkehrs	15
2.3.2	Präzisierung der Problemstellung	16
3	Stand von Wissenschaft und Technik	19
3.1	Graphen und Flüsse	19
3.2	Planungsprobleme im Schienenverkehr	20
3.2.1	Abgrenzung von verwandten Planungsproblemen	20
3.2.2	Optimierung von Umstellplänen und taktischen Wagenrouten im Schienengüterverkehr	21
3.3	Grundlagen der Komplexitätstheorie	29
3.4	Grundlagen der mathematischen Optimierung	33
3.4.1	Lösungsverfahren für lineare Optimierungsprobleme	33
3.4.2	Exakte Verfahren der diskreten Optimierung	35
3.4.3	Metaheuristische Optimierungsverfahren	37
4	Entwicklung und Analyse eines Modells für das Wagenroutenproblem	41
4.1	Entwicklung von Modellen zum Wagenroutenproblem	41
4.1.1	Eingabeparameter der Modelle	41
4.1.2	Ein kantenbasiertes Modell	43
4.1.3	Ein pfadbasiertes Modell	46
4.2	Analyse der Modelle (Opt_{arc}) und (Opt_{path})	48
4.2.1	Komplexitätstheoretische Analyse der Modelle	48
4.2.2	Untere Schranken	51

4.2.3	Annahmen zur Modellvereinfachung	53
5	Entwicklung von Lösungsverfahren für das Wagenflussproblem	59
5.1	Ein Column-Generation-Verfahren	59
5.1.1	Struktur der Nebenbedingungen von (Opt_{arc}) und (Opt_{path})	59
5.1.2	Einführung in Column-Generation-Verfahren	60
5.1.3	Anwendung von Column Generation auf (Opt_{path})	61
5.2	Branch-and-Price	67
5.2.1	Vorüberlegungen zum Branch-and-Price	68
5.2.2	Entwicklung von Schnittebenen und Erweiterung zu Branch-and-Price-and-Cut	70
5.2.3	Der Tailing-off-Effekt im Branch-and-Price-and-Cut	71
5.3	Ein metaheuristisches Lösungsverfahren	72
5.3.1	Entwicklung von metaheuristischen Lösungsverfahren	73
5.3.2	Anwendung des Analogieprinzips	73
5.3.3	Eine Matheuristik zur Lösung von (Opt_{arc})	75
6	Numerische Resultate der vorgestellten Verfahren	83
6.1	Herkunft der betrachteten Instanzen	83
6.2	Ergebnisse des kommerziellen Solvers CPLEX	84
6.3	Auswirkungen der Modellvereinfachungen	85
6.4	Resultate von Column Generation und Branch-and-Price-and-Cut	87
6.5	Tuning der Matheuristik	89
6.5.1	Wahl der Nachbarschaftsstruktur	89
6.5.2	Parametertuning	90
6.6	Finale Resultate der Matheuristik	92
7	Validierung und Schlussfolgerungen	97
7.1	Bündelung und Transportkosteneffizienz	97
7.2	Robustheit	100
7.2.1	Analyse der Auswirkung multipler Szenarien in der Optimierung	100
7.2.2	Preis der Robustheit	102
7.3	Entlastung der zentralen Infrastruktur	103
7.3.1	Auslastung von Zügen und Streckenabschnitten	104
7.3.2	Auslastung der Umstellanlagen	105
8	Zusammenfassung und Ausblick	109

1 Einleitung

Die vorliegende Dissertation entstand aus dem Kooperationsprojekt *Ein benutzergeführtes Planungsinstrument zur Bündelung von SGV-Produkten unter Nutzung dezentraler Übergabepunkte* zwischen dem Institut für Transportlogistik der Technischen Universität Dortmund und der Abteilung Verkehrsnetzentwicklung und Verkehrsmodelle (GSV) der Deutschen Bahn AG im Rahmen des *Effizienzclusters LogistikRuhr*¹.

1.1 Einführung in die Problemstellung

Für die kommenden Jahre wird ein stetiges Wachstum des Güterverkehrs erwartet. Gründe dafür liegen neben dem allgemeinen Wirtschaftswachstum in einer zunehmenden Arbeitsteilung und der Globalisierung der Märkte. Trotz aller Fortschritte bei der Liberalisierung des Schienengüterverkehrs in Europa, konnte dieser in der Vergangenheit nur wenig am Wachstum partizipieren. Speziell im Einzelwagenverkehr, einer Produktionsform des Schienengüterverkehrs, bei der kleine Gruppen von Güterwagen von Versendern zu Empfängern transportiert werden, entsteht durch starke Kapitalbindung und geringe Margen ein hoher wirtschaftlicher Druck. Eine konsequente Weiterentwicklung der Produktionskonzepte ist unumgänglich, um den für viele Branchen wichtigen Einzelwagenverkehr zukunftssicher zu machen.

An dieser Stelle kann die logistische Wissenschaft mit Ihren systematischen Planungsmethoden ihre Stärken ausspielen. Eine Vielzahl voneinander abhängender Entscheidungen und der Wunsch nach einem anpassungsfähigen Planungskonzept motivieren den Einsatz mathematischer Optimierungsmethoden im Einzelwagenverkehr. An der interdisziplinären Schnittstelle zwischen Logistik und Mathematik werden in dieser Dissertation mathematische Optimierungsmethoden auf die Routenplanung im Einzelwagenverkehr angewendet, um die Vorteile neuer Planungsmechanismen quantifizierbar zu machen und damit einen praktischen Beitrag zur Weiterentwicklung des Schienengüterverkehrs zu leisten.

1.2 Ziel und Aufbau der Arbeit

Das Ziel dieser Dissertation ist die Entwicklung von Methoden der mathematischen Optimierung zur Planung taktischer Wagenrouten im deutschen Einzelwagenverkehr. Dessen Durchführung findet in einem Netzwerk von Umstellbahnhöfen statt, in denen Züge zerlegt und mit neuer Wagenreihung gebildet werden können, um Konsolidierungseffekte

¹Förderkennzeichen 01IC10L26A

beim Transport zu erzielen. Diese Arbeit steht unter der Leitfrage nach der optimalen Reihenfolge, in der die Wagen des Einzelwagenverkehrs die Umstellbahnhöfe durchlaufen. Die Strukturierung der Arbeit erfolgt ausgehend vom Vorgehensmodell der mathematischen Optimierung (vgl. Abbildung 1.1), welches aus diesem Grunde an dieser Stelle eingeführt wird [CG13]:

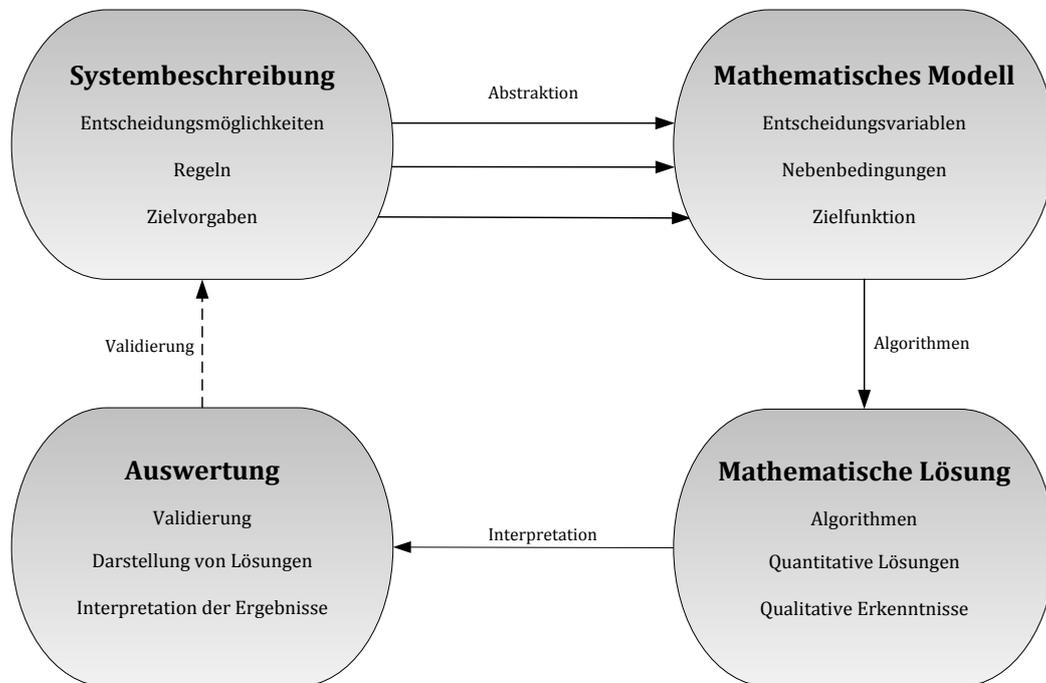


Abbildung 1.1: Schaubild zum Vorgehensmodell der mathematischen Optimierung

Die mathematische Optimierung beschäftigt sich mit der Suche nach der bestmöglichen Lösung von Entscheidungsproblemen. Das zu betrachtende Entscheidungsproblem muss sich durch drei logische Objekte charakterisieren lassen:

1. eine Menge von Entscheidungsmöglichkeiten
2. eine Zielgröße, welche die Wirkung der Entscheidungen quantifizierbar macht
3. ein Regelwerk, das eindeutige Aussagen über die Zulässigkeit einer Entscheidung erlaubt

Die auf diese Weise charakterisierten Entscheidungssysteme umfassen technische, wirtschaftliche und insbesondere logistische Probleme. Das Entscheidungsproblem wird in einem ersten Arbeitsschritt in Form einer Systembeschreibung formuliert.

Die Übertragung des Entscheidungsproblems in mathematische Begriffe wird als Modellierung bezeichnet. Dabei erfolgt eine Darstellung der Entscheidungsmöglichkeiten durch

eine geeignete Menge mathematischer Entscheidungsvariablen. Die Regeln des Problems werden durch Gleichungen, Ungleichungen oder abstrakte Mengenzugehörigkeiten dargestellt. Sie werden als Nebenbedingungen bezeichnet. Die Zielgröße wird in Form der sogenannten Zielfunktion modelliert, um die Qualität von Entscheidungen quantifizierbar zu machen. Die Aufgabe der Optimierung besteht üblicherweise in der Minimierung oder Maximierung einer Zielfunktion.

Der Modellierung folgt die Lösung des mathematischen Optimierungsproblems. Es gilt, eine den Nebenbedingungen entsprechende Variablenbelegung mit dem optimalen Zielfunktionswert zu finden. Neben der quantitativen Ermittlung der Optimallösung konkreter Instanzen besteht eine wichtige Zielsetzung während der Lösungsphase im Finden qualitativer Resultate, d.h. struktureller Erkenntnisse über das Problem.

Den Abschluss des mathematischen Optimierungsprozesses bildet die Auswertung der qualitativen sowie quantitativen Ergebnisse. Dabei sollte in einer Validierungsphase ermittelt werden, ob die modellhaft abgebildeten Charakteristika der Resultate der Systembeschreibung entsprechen. Ist dies nur in unzureichendem Maße der Fall, muss eine Remodellierung stattfinden und der Optimierungsprozess nach der neuen Modellbildung fortgesetzt werden.

Die weitere Arbeit gliedert sich dem Vorgehensmodell entsprechend wie folgt: In Kapitel 2 werden die logistischen Grundlagen des Schienengüterverkehrs mit Fokus auf den Einzelwagenverkehr zusammengestellt. Eine Beschreibung der Planungsprozesse im Einzelwagenverkehr ordnet das Problem der taktischen Routenplanung in den Kontext weiterer Planungsprobleme des Schienengüterverkehrs ein. Eine Analyse gegenwärtiger und künftiger Herausforderungen im Einzelwagenverkehr wird zur Identifikation und Konkretisierung des Problems der taktischen Routenplanung genutzt. Darauf aufbauend wird das zentrale Problem der Arbeit mit allen Anforderungen in Worten formuliert.

In Kapitel 3 wird der Stand von Wissenschaft und Technik zur taktischen Routenplanung sowie zur mathematischen Optimierung dargelegt. Verschiedene wissenschaftliche Ansätze zur taktischen Routenplanung fokussieren unterschiedliche Zielsetzungen, die sich in der Struktur der resultierenden Lösungen bemerkbar machen. Drei dieser strukturellen Eigenschaften (Konsolidierung, Robustheit und Entlastung der zentralen Infrastruktur) werden als nützlich für den deutschen Einzelwagenverkehr identifiziert und den Anforderungen aus Kapitel 2 hinzugefügt.

In Kapitel 4 findet die mathematische Modellierung statt. Die zuvor erstellte Systembeschreibung wird in ein lineares ganzzahliges Optimierungsproblem überführt. Zwei äquivalente Modellformulierungen werden vorgestellt. Es folgt eine Komplexitätsanalyse der Probleme und eine Analyse zur Qualität unterer Schranken bei Anwendung linearer Relaxierungsmethoden. Zur Vereinfachung des Lösungsprozesses werden verschiedene Modellannahmen diskutiert, welche die Anzahl der Entscheidungsvariablen des Modells stark verringern können.

Die Entwicklung von Lösungsverfahren ist das Thema des 5. Kapitels, da das Problem mit generischen Lösungsverfahren nicht in praktikabler Zeit behandelt werden kann. Darüber hinaus ist es nicht möglich, das Modell zu vereinfachen oder die Instanzen zu verkleinern, ohne den praktischen Nutzen der Optimierung einzuschränken. Eine

Analyse der Nebenbedingungsstruktur zeigt auf, dass das Problem bei vorübergehender Vernachlässigung gewisser Nebenbedingungen in simple Teilprobleme zerfällt. Diese Eigenschaft ermöglicht den Einsatz einer speziellen Klasse von Dekompositionsverfahren, den Column-Generation-Verfahren. Deren Einbindung in ein Branch-and-Bound-Schema führt zu strukturellen Erkenntnissen und einem Branch-and-Price-Algorithmus, der bessere Resultate als kommerzielle Lösungsverfahren liefert. Die Lösung von Realinstanzen, wie sie vom Forschungspartner DB Mobility Logistics zur Verfügung gestellt wurden und die vollständige Planung des deutschen Einzelwagenverkehr erlauben würden, gelingt jedoch auch mit Branch-and-Price nicht. Daher wird mit Hilfe wissenschaftlicher Designprinzipien ein metaheuristischer Ansatz entwickelt, der die Methode der lokalen Suche mit exakten Lösungsverfahren zu einer sogenannten Matheuristik verbindet. Diese Matheuristik ermittelt sukzessive neue Routen für Teilmengen der Sendungen und verbessert so die globale Lösung iterativ.

Kapitel 6 enthält die numerischen Resultate der zuvor genannten Column-Generation- und Branch-and-Price-Verfahren für eine Reihe von Instanzen, die auf Realdaten basieren. Des Weiteren wird die im vorangegangenen Kapitel entwickelte Matheuristik ausgiebig getestet, um die am besten geeigneten Nachbarschaftsstrukturen und Parameterwahlen zu ermitteln. Abschließende Optimierungsläufe zeigen auf, dass die entwickelte Matheuristik zur Lösung realer Instanzen geeignet ist und in praxistauglicher Zeit befriedigende Resultate liefert.

Eine Validierung der in Kapitel 3 identifizierten Modelleigenschaften findet in Kapitel 7 statt. Eine Instanz, die ein Teilnetz des deutschen Einzelwagenverkehrsnetzes abbildet und für die ein Wagenrouting auf Basis aktueller Planungsmaßstäbe gegeben ist, wird mit dem neuen Lösungsverfahren gelöst. Anhand von Vergleichen verschiedener Eigenschaften und Kennzahlen beider Lösungen kann die Überlegenheit der in dieser Arbeit entwickelten Modellierung und der passenden Lösungsmechanismen gegenüber derzeitigen Planungsmethoden aufgezeigt werden.

2 Logistik des deutschen Einzelwagenverkehrs

In diesem Kapitel wird der Schienengüterverkehr (SGV) sowie ausgewählte Eigenschaften des Verkehrsträgers Schiene in Deutschland beschrieben. Es werden die Produktionsformen des Schienengüterverkehrs erläutert. Der Einzelwagenverkehr (EWV) zählt zu den Produktionsformen des Schienengüterverkehrs und wird als Kernthema dieser Arbeit genauer dargestellt. Dabei werden typische Planungsprobleme des EWV identifiziert. Eine weitere Fokussierung auf die Wagenroutenplanung durch Umstellpläne und eine formale Systembeschreibung leiten zu den nachfolgenden Kapiteln über.

2.1 Verkehrsträger Schiene

Definition 1 (Schienengüterverkehr)

Im Schienengüterverkehr wird die Verkehrsleistung durch den Transport von Güterwagen mittels Güterzügen auf Schienen erbracht [CG13].

Die Eigenschaften des Verkehrsträgers Schiene sind maßgeblich für die wirtschaftliche Leistungsfähigkeit und die Organisation des SGV. Der niedrige Rollwiderstand zwischen Schiene und Rad sorgt bei Eisenbahnen für einen sehr geringen spezifischen Energiebedarf. Lange Beschleunigungs- und Bremswege sind aus dem gleichen Grund charakteristisch für Eisenbahnen. Die Eisenbahn kann daher ihre energetischen Vorteile in erster Linie bei möglichst langen, ununterbrochenen Fahrten ausnutzen. Dieser Vorteil verstärkt sich bei Transporten großer Massen. Die Spurführung der Schiene sorgt überdies für eine hohe Sicherheit beim Transport von Gefahrgütern. Organisatorische Vorteile ergeben sich aus dem Fehlen von Sonntagsfahrverboten und einer hohen Planbarkeit der Transporte [Fra12b].

Gleichzeitig bildet die Notwendigkeit von Schienen eine starke infrastrukturelle Einschränkung: Während nahezu jeder Versender von Gütern über einen Zugang zum öffentlichen Straßennetz verfügt, ist die Dichte von Zugangsstellen zum öffentlichen Schienennetz eher gering [Ber01]. Dies erschwert die Bedienung in der Fläche. Der Netzzugang wird im Abschnitt 2.1.3 über die deutsche Schieneninfrastruktur vertieft. Die Vorhaltung von Infrastruktur und Rollmaterial führt zu einer hohen Kapitalbindung bei den Eisenbahngesellschaften. Die Schiene war im Jahre 2012 in Deutschland gemessen an der Verkehrsleistung, dem Produkt aus Transportentfernung und Gewicht, der zweitbedeutendste Verkehrsträger. Mit einer Verkehrsleistung von beinahe 110 Mrd. Tonnenkilometern (tkm) hatte der Schienengüterverkehr laut statistischem Bundesamt [Sta13] einen Anteil von 17,2 % am deutschen Modal Split, der die relativen Anteile aller Ver-

kehrsträger widerspiegelt. Obwohl sich dieser Anteil in der jüngeren Vergangenheit nicht signifikant verändert hat, ist – mit Ausnahme der Jahre der globalen Wirtschaftskrise 2008 und 2009 sowie der europäischen Staatsschuldenkrise 2012 – eine Steigerung der Verkehrsleistung des deutschen SGV festzustellen (vgl. Abbildung 2.1). Prognosen deuten auf eine Fortsetzung dieses Trends hin [MD10].

Die Bundesregierung kommuniziert seit geraumer Zeit den Wunsch zur Stärkung des SGV [Bia13], um mehr Verkehrsleistung auf die Schiene zu verlagern. Neben politischen Interventionen haben jedoch auch diverse ökonomische und gesellschaftliche Vorgänge Einfluss auf den SGV. Eine Reihe dieser exogenen Einflussfaktoren, wie Demographie und Urbanisierung, wurde von der Unternehmensberatung McKinsey in [MD10] zusammengetragen und auf ihre Auswirkungen für den SGV hin analysiert. Die Prognose der Studie lautet, dass bis zum Jahr 2025 eine Erhöhung des Anteils des SGV am deutschen Modal Split auf 21 % und damit eine Stärkung der Position im intermodalen Wettbewerb möglich sei. In Erwartung einer wachsenden Verkehrsleistung bezeichnen die Berater von McKinsey den SGV als unverzichtbar für die deutsche Wirtschaft. Weitere Analysen zur Zukunft des Schienengüterverkehrs bietet Abschnitt 2.2.1.

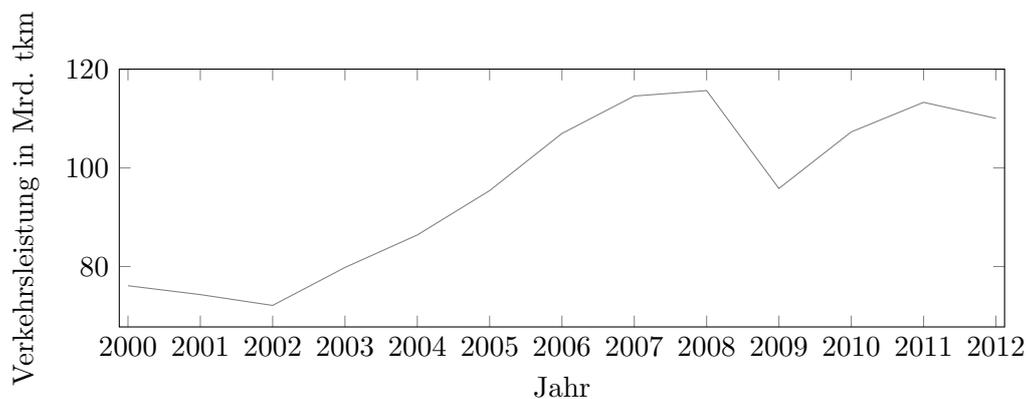


Abbildung 2.1: Entwicklung der Verkehrsleistung im SGV seit 2000 [Sta13]

2.1.1 Akteure des Schienengüterverkehrs

Unternehmen des SGV werden nach dem Allgemeinen Eisenbahngesetz (AEG) in öffentliche und nichtöffentliche Eisenbahnen eingeteilt. Öffentliche Eisenbahnunternehmen stellen im Gegensatz zu nichtöffentlichen Eisenbahnen ihre Leistungen jeder juristischen Person bzw. jedem Unternehmen zur Verfügung. Des Weiteren werden *Eisenbahninfrastrukturunternehmen* (EIU) und *Eisenverkehrsunternehmen* (EVU) unterschieden [CG13]:

Eisenbahninfrastrukturunternehmen

EIU sind öffentliche oder privatrechtlich organisierte Unternehmen, welche eine Eisenbahninfrastruktur betreiben. Zur Eisenbahninfrastruktur zählen Schienenwege, Strom-

fernleitungen, Anlagen des SGV, wie Bahnhöfe und Zugbildungsanlagen (vgl. Abschnitt 2.1.3) sowie Abstellgleise. Die Aufgaben der EIU bestehen in Bau, Wartung und Bereitstellung von Schienenwegen und Dienstleistungen in Anlagen des SGV. Die Bereitstellung von Fahrwegkapazitäten zu einer bestimmten Zeit auf einem bestimmten Streckenabschnitt wird dabei als *Trasse* bezeichnet. Das größte öffentliche EIU in Deutschland ist die DB Netze AG. Dem AEG entsprechend muss jedes öffentliche EIU einen unabhängigen und diskriminierungsfreien Zugang zu seinen Dienstleistungen anbieten.

Eisenverkehrsunternehmen

EVU sind öffentliche oder privatrechtlich organisierte Unternehmen, die Eisenbahnverkehrsleistungen erbringen. Darunter sind in erster Linie Beförderungsleistungen für Passagiere oder Güter zu verstehen. Die EVU sind in der Regel im Besitz von Rollmaterial, d.h. Triebfahrzeugen und Wagons. Zur Durchführung ihrer Aufgaben sind sie auf die Nutzung der Infrastruktur und somit auf die EIU angewiesen. Mit einer Verkehrsleistung von 78,5 Mrd. tkm und einem Marktanteil von 71,4 % im Jahre 2012 ist DB Schenker Rail Marktführer im deutschen SGV [Deu13].

2.1.2 Produktionsformen des Schienengüterverkehrs

Im SGV werden grundsätzlich drei Produktionsformen unterschieden, die in Deutschland weitestgehend unabhängig voneinander durchgeführt werden. Die folgenden Beschreibungen sind an die Ausführungen von Nikutta und Pahl [NP09] angelehnt:

1. Im *Ganzzugverkehr* (GV) werden vollständige Züge ohne Zwischenhalt vom Versender zum Empfänger transportiert. Diese Produktionsform wird üblicherweise für große Sendungsaufkommen verwendet und nutzt bei ausreichender Auslastung der Züge die Vorteile der Schiene voll aus. Massengüter aus Montan- und Chemieindustrie eignen sich besonders gut für den GV, weil sie ausreichende Aufkommen erzeugen und sich überdies ihres Gewichts wegen für den Transport auf der Schiene anbieten. Der GV zeichnet sich durch eine hohe Transportgeschwindigkeit aus. Die ununterbrochene Fahrt ist sehr gut planbar und somit effizient in Produktions- und Logistikketten einzubinden. Obwohl etwa die Hälfte des Transportaufkommens des SGV über den GV abgewickelt wird, verursacht er nur etwa ein Drittel der gesamten Verkehrsleistung.
2. Im *Einzelwagenverkehr* (EWW) werden einzelne Wagen oder kleine Wagengruppen vom Versender zum Empfänger transportiert. Die Größe der Wagengruppen rechtfertigt dabei aus wirtschaftlicher Sicht meist keinen Direkttransport. Ein derart geringes Aufkommen wird auch als *nicht zugwürdig* bezeichnet. Daher werden Wagen verschiedener Versender und/oder verschiedener Empfänger für den Hauptlauf gebündelt und gemeinsam transportiert. EWW und GV werden in Abgrenzung zum kombinierten Verkehr auch als *konventioneller SGV* bezeichnet. Dem EWW wird das Unterkapitel 2.2 gewidmet, da er den Fokus dieser Arbeit bildet.

3. Im *Kombinierten Verkehr* (KV) werden Ladeeinheiten, wie Container oder Wechselbrücken, in Railports oder Umschlagbahnhöfen per Kran auf Güterzüge verladen [CG13]. Auch die *Rollende Landstraße*, d.h. der Transport vollständiger Lkw über die Schiene, ist ein typisches Beispiel für den KV. Vor- und Nachlauf werden zumeist über die Straße abgewickelt. Der Kunde hat dadurch den Vorteil, dass er keinen eigenen Gleisanschluss besitzen muss, um Sendungen in den SGV einzuschleusen. Für den Hauptlauf findet in der Regel eine Einspeisung in den Einzelwagenverkehr statt. Es existieren jedoch auch KV-Gesellschaften, die pendelnde Ganzzüge zwischen Umschlagbahnhöfen verkehren lassen [SS12].

2.1.3 Deutsche Schieneninfrastruktur

Schienenwege

Das deutsche Schienenverkehrsnetz umfasste im Jahre 2012 ca. 37.700 km [Sta13] mit der Normalspurbreite von 1.435 mm. Davon sind ca. 59 % elektrifiziert, d.h. mit elektrischen Oberleitungen ausgestattet. Dabei handelt es sich in Deutschland um ein Wechselstromnetz mit 16,7 Hz und 15 kV Spannung. Auf nicht elektrifizierten Strecken kommen üblicherweise Diesellokomotiven zum Einsatz. Es werden abhängig von der Geschwindigkeit linien- und punktförmige Zugbeeinflussungssysteme eingesetzt, die neben der Geschwindigkeit auch die Mindestzugfolgezeiten zwischen zwei aufeinander folgenden Zügen bestimmen. Diese liegen üblicherweise im Bereich weniger Minuten [SS12]. Der SGV und der Schienenpersonenverkehr nutzen in Deutschland die gleichen Schienenwege, d.h. es gibt Strecken, auf denen sowohl Personenzüge als auch Güterzüge verkehren. Dabei wird der Personenverkehr im Allgemeinen priorisiert. Als Folge dieser Priorisierung verbleiben für die zumeist langsameren Güterzüge zwei Alternativen: Entweder muss der Güterverkehr in der Nacht stattfinden, wenn der Personenverkehr stark reduziert ist, oder die Güterzüge müssen Personenzügen den Weg zum Überholen frei machen. Für diesen Zweck sind entlang der Schienenstrecke Überholgleise installiert. Diese zur Hauptstrecke parallelen Gleisabschnitte können befahren werden, um Züge passieren zu lassen [CG13]. Die Länge der Überholgleise sowie der kürzeste zulässige Abstand zwischen zwei Hauptsignalen beschränken die Maximallänge der Güterzüge [SS12]. Abhängig vom Streckenabschnitt sind maximale Zuglängen von 700 m bis 740 m in Deutschland üblich. Auf ausgewählten Strecken können jedoch Verbesserungen der Infrastruktur auch den Einsatz längerer Züge ermöglichen. Auf diese Weise wurde die zulässige Zuglänge auf der Strecke von Maschen in das dänische Padborg auf 835 m vergrößert [DB 12].

Ebenso wie die Gesamtlänge eines Güterzuges ist auch die Anzahl der Achsen und das Höchstgewicht beschränkt: Das Gesamtgewicht eines Zuges im deutschen SGV liegt üblicherweise bei maximal 1.600 t. Schwerere Züge bedürfen einer amtlichen Sondergenehmigung [Fie05]. Des Weiteren gibt es eine streckenabhängige Höchstlast pro Achse, die bis zu 22,5 t betragen kann. Die Gewichts- und Längenbeschränkung der Züge begrenzt die Gesamtkapazität eines Zuges. Die wirtschaftlichen Folgen dieser Beschränkungen werden von Clausen und Voll [CV13] erläutert und in Abschnitt 2.3.1 erneut aufgegriffen.

Knoten des Schienengüterverkehrs

Die Knoten des konventionellen SGV sind Gleisanschlüsse, Güterverkehrsstellen und Satellitenbahnhöfe sowie verschiedene Typen von Umstellbahnhöfen. Der Zugang zum Schienennetz erfolgt in der Regel über einen der mehr als 3.000 *Gleisanschlüsse* in Deutschland [NP09]. Es handelt sich dabei um Gleisanlagen, die auf den Werksgeländen der Versender liegen und Eigentum des Versenders sind. Gleisanschlüsse können direkt in das öffentliche Schienennetz eingebunden sein oder eigene Schienenwege zu einer Übergabestelle des öffentlichen Netzes beinhalten. An den Gleisanschlüssen erfolgt die Beladung der Eisenbahnwagen mit güterabhängiger Beladetechnik [CG13].

Die Wagen mehrerer Gleisanschlüsse werden an *Güterverkehrsstellen* oder *Satellitenbahnhöfen* (Sat) für den Nahverkehr gebündelt. Einige dieser Satellitenbahnhöfe verfügen über eigene Rangiermittel. Sie werden als *Satellitenbahnhöfe mit Rangiermitteln* (SmR) bezeichnet und können schon zu den Umstellbahnhöfen gezählt werden. Gemeinsam mit sogenannten *Knotenpunktbahnhöfen* (Kbf) bündeln sie den Nahverkehr zu zugwürdigen Mengen, die dann in die *Rangierbahnhöfe* transportiert werden [NP09]. Die Rangierbahnhöfe (Rbf) sind riesige Gleisanlagen, in denen Züge in einzelne Wagen zerlegt und neu angeordnet werden. Sie dienen überdies als Gateways für den internationalen SGV, der in Deutschland einen signifikanten Anteil der Verkehrsleistung erzeugt [SS12]. Sie sind für gewöhnlich in drei Gruppen paralleler Gleise aufgeteilt (vgl. Abbildung 2.2).

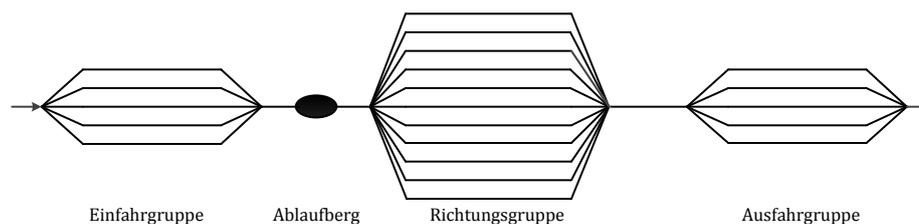


Abbildung 2.2: Schematische Darstellung eines Rangierbahnhofes aus [CG13]

In der Einfahrgruppe werden Züge getrennt. Eine Rangierlok schiebt sie über einen Ablaufberg. Der Ablaufberg ist eine natürliche oder künstliche Erhöhung im Bereich des Rbf. Nach Überqueren der Bergkuppe rollen die Wagen allein von ihrer Gewichtskraft beschleunigt in die Richtungsgruppe. Die Richtungsgruppe ist ebenfalls eine Gruppe paralleler Gleise, in der mit Hilfe komplizierter Weichensysteme eine Sortierung der anrollenden Wagen nach Ausgangsrelationen vorgenommen wird. Die Gleise der Richtungsgruppe werden als Richtungsgleise bezeichnet. Die Anzahl der Richtungsgleise bestimmt die Anzahl der Züge, die parallel gebildet werden können. Sind alle Wagen eines Zuges auf einem Richtungsgleis versammelt, werden die Wagengruppen in die Ausfahrgruppe gezogen. Nach einem abschließenden technischen Check verlassen sie von dort aus den Rbf [DB 11].

2.2 Einzelwagenverkehr in Deutschland

Dieses Unterkapitel beinhaltet detailliertere Darstellungen zum EWW. Zunächst wird die Zukunft des SGV im Allgemeinen und des EWW im Speziellen kritisch bewertet und Verbesserungspotentiale identifiziert. Anschließend werden die Prozesse des Einzelwagenverkehrs beschrieben. Dazu wird der Transportvorgang aus Sicht des Kunden und des Planers dargestellt. Es wird mit der Planung von Wagenrouten im Einzelwagenverkehr das zentrale Thema dieser Arbeit eingeführt. Abschließend wird eine zielführende Formulierung des anschließend zu lösenden Planungsproblems herausgearbeitet.

2.2.1 Zukunftsaussichten des Einzelwagenverkehrs

Im EWW wurden in den vergangenen Jahren mehr als 30 % der Gesamtverkehrsleistung des SGV produziert. DB Schenker Rail kann als Quasimonopolist im EWW angesehen werden [Mon11, SS12]. Die aktuelle Bedeutung dieser Produktionsform allein sichert jedoch nicht ihr zukünftiges Bestehen und ihren wirtschaftlichen Erfolg. Die schrittweise Einstellung des EWW ist in mehreren europäischen Ländern, darunter Frankreich und Italien, bereits besiegelt [Fra12a]. In weiteren europäischen Ländern Verschiedene Aspekte führen zu einer wirtschaftlich schwierigen Situation im EWW. Durch die im Rahmen des Sanierungsprogramms *Marktorientiertes Angebot Cargo* (MORA-C) durchgeführte Verringerung bedienter Gleisanschlüsse wurde die Flächenbedienung durch den EWW eingeschränkt. In der Folge erschwerte sich der Netzzugang für Versender kleiner Mengen. Dies führte gerade bei Kleinkunden zu einer Verlagerung von Verkehren auf die Straße. Siegmann [Sie12] schreibt, dass außerdem die Transportzeit im EWW „schon die Grenze der Marktwürdigkeit gegenüber dem LKW-Transport“ darstelle. Kleine Margen bei hoher Kapitalbindung schaffen hohe Markteintrittsbarrieren und verringern die Attraktivität für die EVU.

Diesen Problemen stehen jedoch auch positive Aussichten gegenüber: Der EWW sei „zwar eine Herausforderung, jedoch auch eine Chance“, wie der DB Schenker Rail Vorsitzende konstatierte [Fra12a]. Steigende Kosten für fossile Brennstoffe und ein gesteigertes Bewusstsein für Ökologie machen den Verkehrsträger Schiene interessanter. Die Nutzung der Schiene führt bei hinreichender Auslastung und bei entsprechendem Energiemix bei der Stromerzeugung zu erheblich weniger CO₂-Emissionen und geringeren externen Kosten als der Straßenverkehr [Fra12b]. Daher besteht der politische Wunsch nach einer Verlagerung zusätzlicher Verkehre auf die Schiene [Bia13]. Der Wunsch, das gesamte, für die Zukunft prognostizierte Wachstum der Verkehrsleistung über die Schiene abzuwickeln, wird dabei jedoch an der wirtschaftlichen Realität scheitern. Während die Wirtschaftlichkeitsschwelle des Schienengüterverkehrs je nach Berechnungsgrundlage bei mindestens 200 km Transportentfernung liegt, übersteigt dies den Radius von rund 90 % aller innerdeutschen Transporte [Fra12b]. Steigende Serviceanforderungen von Privatkunden und der Güterstruktureffekt, d.h. die Tendenz zum Versand hochwertiger, kleinteiliger Produkte, erhöhen überdies die Anzahl zeitkritischer Sendungen. Für sie stellt der Schienentransport auf Grund langer Transportzeiten in der Regel keine Alternative dar. Nichtsdestotrotz ist durchaus mit einem Wachstum des Verkehrsaufkommens

auf der Schiene und damit auch des Einzelwagenverkehrs zu rechnen. Die wirtschaftlichen Faktoren, die den Güterverkehr beeinflussen, sind nur schwierig zu prognostizieren, wie bereits in einer gemeinsamen Studie von McKinsey und der Deutschen Bahn festgestellt wurde [MD10]. Drei wesentliche Entwicklungen, die ein Wachstum des Schienenverkehrsaufkommens in Deutschland in Aussicht stellen, sind zu beobachten:

1. In Deutschland wächst das Güterverkehrsaufkommen langsamer als die Verkehrsleistung. Dies ist nicht zuletzt auf größere Entfernungen zwischen Versendern und Empfängern zurückzuführen. Da der Verkehrsträger Schiene – hinreichend hohes Aufkommen stets vorausgesetzt – besonders auf der Langstrecke ökonomisch und ökologisch effizient ist, steigern längere Transportstrecken seine Attraktivität.
2. Bedeutende deutsche Industrieunternehmen, wie BASF und Porsche, investieren vermehrt in den SGV. Die Allianz pro Schiene [All07] listet 13 Beispiele deutscher Firmen auf, die im SGV einen wichtigen Bestandteil zukünftiger Logistikketten sehen. Einige der gelisteten Unternehmen haben mit der Einrichtung neuer Gleisanschlüsse Zeichen für die Stärkung des innerdeutschen Eisenbahnnetzes gesetzt.
3. Die weiter andauernde Liberalisierung des europäischen SGV führt vor allem zu dessen Internationalisierung [EKJ06]. Die dadurch verbesserte Konkurrenzfähigkeit zum LKW lässt auch eine Stärkung des europäischen SGV erwarten. Mit Blick auf Deutschlands geographische Lage als Transitland innerhalb Europas ist mit einer steigenden Verkehrsleistung auf der Schiene zu rechnen [DM07]. Dies gilt nicht zuletzt für den EWV [Sie12].

Eine wachsende Verkehrsleistung stellt große Herausforderungen an die Infrastruktur und die Logistik dar. Auf vielen Hauptachsen des SGV sind schon heute Engpässe vorhanden [SS12]. Ihnen kann mit Infrastrukturverbesserungen oder durch bessere Ausnutzung der vorhandenen Infrastruktur durch bessere Planungsmechanismen begegnet werden [NA10]. Der Ausbau von Schieneninfrastruktur ist kostspielig und nur langfristig realisierbar [Vog11]. Posner [Pos08] gibt überdies zu bedenken, dass die Liberalisierung den Wettbewerb auf der Schiene immer weiter voran treibt. Auf Grund des entstehenden Preisdrucks werde es zunehmend schwieriger für die EVU Gewinne zu erwirtschaften. Langfristige Investitionen seien daher wenig attraktiv. Die Unternehmensberatung Oliver Wyman [DM07] kommt zu ähnlichen Ergebnissen.

Effektivere Planungsmechanismen allein bieten zwar keinen Ersatz für verbesserte Infrastruktur, sie können jedoch eine Ergänzung darstellen, wenn mit ihrer Hilfe eine Entlastung von Engpässen erfolgt und vorhandene Kapazitäten besser genutzt werden können. Eine effektivere Planung kann Leistungssteigerungen ohne höhere Kapitalbindung bewirken. Eine Weiterentwicklung der Planungsprozesse bildet daher ein angemessenes Mittel, um den SGV der Zukunft wettbewerbsfähig zu gestalten.

2.2.2 Prozessschritte der Auftragsdurchführung



Abbildung 2.3: Verkürzte Darstellung des Angebotsprozesses

Der Durchführung von Wagentransporten im EWV beginnt mit einer Angebotsanfrage des Kunden beim EVU. Diese Anfrage enthält die benötigte Anzahl an Leerwagen sowie Wünsche für das Zeitfenster zwischen der Anlieferung von Wagen und deren Abholung. Das EVU nimmt anschließend für den Versand durch das Schienennetz eine grobe Vorausplanung sowie eine entsprechende Kostenabschätzung vor. Dazu muss die Strecke der Wagen vorbestimmt werden. Der Kunde erhält anschließend ein Angebot zur Transportdurchführung, das bereits Zeitangaben einschließt. Nimmt der Kunde das Angebot in Anspruch, reserviert das EVU beim EIU die notwendigen Trassen. Anschließend passt das EVU seine Triebfahrzeug-, Personal- und Wagendisposition an, so dass der Kunde zur vereinbarten Zeit Leerwagen erhält und diese nach der Beladung auch abgeholt werden können. Eine Mindestanzahl an Wagen zur Einspeisung in das Netz besteht nicht. Die Anmeldung von Transporten erfolgt sehr dynamisch: Transportmengen müssen nur wenige Stunden im Voraus angemeldet werden [NP09]. Eine vereinfachte Darstellung des Prozesses ist in Abbildung 2.3 gegeben.

2.2.3 Planungsebenen

Die genannten Schritte der Angebotserstellung und der anschließenden Ressourcendisposition stellen für das EVU ein sehr komplexes System von Planungsschritten dar. Basierend auf der Arbeit von Assad [Ass80] werden in diesem Abschnitt die Planungsschritte des SGV mit Fokus auf den EWV dargestellt. Assad unterscheidet zunächst drei hierarchische Planungsebenen: strategische, taktische und operative Planungsentscheidungen. Die Planungsebenen unterscheiden sich hinsichtlich ihres zeitlichen Wirkungshorizontes und der notwendigen Kapitalinvestitionen. Die strategischen Entscheidungen betreffen vor allem die Verfügbarkeit von Infrastruktur und Ressourcen. Darin enthalten sind die folgenden Planungsaufgaben:

- Design und Änderung des Schienennetzes
- Eröffnung, Schließung und Erhalt von Bahnanlagen
- langfristige, stark aggregierte Planung zukünftiger Wagenflüsse, um benötigte Infrastrukturkapazitäten abschätzen zu können

Die strategischen Entscheidungen sind sehr kapitalintensiv und daher sowohl in ihrer Umsetzung als auch in ihrer zeitlichen Wirkung sehr langfristig angelegt. Für wirksame Entscheidungen müssen langfristige Prognosen über Verkehrsaufkommen und wirtschaftliche Trends berücksichtigt werden. Der Planungshorizont kann je nach Typ der Entscheidung Jahre bis Jahrzehnte betragen.

Taktische Entscheidungen betreffen hingegen den mittelfristigen Einsatz der vorhandenen Ressourcen. Zu den taktischen Entscheidungen zählen die folgenden:

- Verteilung und Konsolidierung aggregierter Wagenflüsse, d.h. eine Vorausplanung der Wagenrouten auf Basis von Prognosedaten
- Zuordnung von Rangieraufgaben zu den Umstellbahnhöfen
- Planung, auf welchen Wegstrecken, wie viele Züge einzusetzen sind
- Vorabanmeldung von Trassen beim EIU
- mittelfristige Verteilungsmuster für die Triebfahrzeuge
- eine den vorherigen Planungen angepasste vorläufige Personalplanung

Die genannten Entscheidungen basieren auf kurzfristiger und exakter prognostizierten Daten als die strategischen Planungsschritte. Sie lassen eine unterjährige Planung zu. Die taktische Planung sollte in der Lage sein, saisonale Schwankungen zu berücksichtigen. Daher ist bspw. eine quartalsweise Anpassung der Planung sinnvoll. Eine zu kleinschrittige, z.B. tagesgenaue, Planung ist jedoch nicht möglich [NP09], wie in Unterkapitel 4.1 eingehender begründet wird.

Die Vielzahl aufeinander abzustimmender Entscheidungen stellt eine große Herausforderung für ein EVU dar. Eine integrierte Planung aller Details ist beinahe unmöglich. Daher ist ein sequentieller Planungsansatz sinnvoll.

Die taktischen Pläne müssen in der täglichen Betriebsführung verfeinert werden. Die kurzfristige Annahme von Aufträgen, wie in Abschnitt 2.2.2 beschrieben, macht kurzfristige Plananpassungen ständig nötig. Daher müssen die folgenden, operativen Planungsschritte teils in Echtzeit, teils auf Wochenbasis durchgeführt werden:

- exakte Zeitpläne für jedes Fahrzeug (mit Abfahrts- und Ankunftszeiten in allen Knoten des Netzwerkes)
- Zuordnung von Fahrzeugen zu Gleisen und Fahrzeiten
- genaue Distribution der Triebfahrzeuge auf der Basis der dynamischen Transportaufkommen
- lokale Verteilung von Leerwagen zur Kundenbedienung
- globale Verteilung von Leerwagen zur Vermeidung lokaler Engpässe
- Reihenfolge der Bearbeitung von Rangieraufgaben in den Umstellbahnhöfen

- Ausgleich kurzfristiger Ressourcen- und Infrastrukturausfälle durch Wartung und Reparatur

Einige dieser Planungsaspekte spielen auch in der taktischen Planung eine Rolle. Die globale Leerwagenverteilung kann basierend auf den aggregierten Verkehrsströmen, die auf der taktischen Ebene geplant werden, vorbestimmt werden. Die Planung allgemeiner Wartungsintervalle ist vorhersehbar. Die Reaktion auf kurzfristige Ausfälle kann hingegen gravierenden Einfluss auf verschiedene Planungsaspekte haben und muss zeitnah erfolgen.

2.3 Umstellpläne und Wagenrouten

Der erste Schritt der taktischen Planung besteht in der Bestimmung von Wagenrouten. Basierend auf Prognosedaten zu aggregierten Transportaufkommen gilt es, einen Plan zur Routenführung jedes Wagens zu bestimmen. Die Zuordnung von Rangieraufgaben zu Umstellbahnhöfen bietet einen anderen Blickwinkel auf die Bestimmung von Wagenrouten, wie im Folgenden kurz ausgeführt werden soll.

Nach Übergabe einer Sendung von einem Kunden an das EVU werden die Wagen der Sendung zunächst zur nächstgelegenen Güterverkehrsstelle oder zum nächsten Satellitenbahnhof transportiert und dort mit Wagen weiterer lokaler Versender gebündelt. In der Regel gibt es von einem Gleisanschluss nur eine einzige Schienenverbindung zu einem Satellitenbahnhof. Somit ist die Routenführung zunächst alternativlos. Die weitere Streckenführung ab dem Satellitenbahnhof ist nicht zwangsläufig durch die Schienenwege vorherbestimmt. Von den Satellitenbahnhöfen kann es Schienenverbindungen in mehrere Richtungen geben. Handelt es sich um einen SmR, stehen bereits mehrere Richtungsgleise zur Verfügung, um eine Sortierung nach Ausgangszügen vorzunehmen. Die Sortierung wird gemäß einem Umstellplan vorgenommen.

Definition 2 (Umstellplan)

Der Umstellplan bestimmt die Zuordnung jedes Wagen zu einem Richtungsgleis.

Die Zuordnung zum Richtungsgleis ist gleichzusetzen mit der Zuordnung zu einem ausgehenden Zug. Der auf einem Richtungsgleis gebildete Wagenzug wird von einem Triebfahrzeug zum nächsten Umstellbahnhof gebracht. Die Umstellpläne in jedem Umstellbahnhof müssen logischerweise so gestaltet sein, dass jeder Wagen nach einer endlichen Anzahl von Umstellungen sein Ziel erreicht, um dort dem Empfänger übergeben zu werden. Sind die Umstellpläne jedes Bahnhofs bekannt, kann daraus die Streckenführung für jeden Wagen bestimmt werden. Wäre andererseits die Strecke jedes Wagens als Sequenz besuchter Umstellbahnhöfe gegeben, könnten daraus Umstellpläne bestimmt werden. Umstellpläne bilden somit die Wirkung einer netzweiten Wagenroutenplanung knotenintern ab. Dies zeigt die Überschneidungen bei der Entscheidung für einen bestimmten Umstellplan und die Festlegung einer Route für jeden Wagen.

2.3.1 Kostenstruktur des Einzelwagenverkehrs

Zur Bewertung eines Wagenroutenplanes ist eine quantifizierbare Größe nötig, anhand derer seine Qualität bemessen werden kann. Der in den vorangegangenen Abschnitten erläuterte ökonomische Druck auf den EWV motiviert die Betrachtung der Produktionskosten. Die Struktur der Gesamtkosten, die einem EVU entstehen, wenn es alle Transportaufträge einem bestimmten Routenplan entsprechend durchführt, sollen in diesem Abschnitt genauer untersucht werden. Der Vollständigkeit halber sei jedoch erwähnt, dass auch andere Zielgrößen, wie eine Erhöhung des Servicelevels durch kürzere Transportzeiten, betrachtet werden könnten. Siegmann nennt vier Kostenelemente, die durch den Einsatz jedes Zuges in einem Produktionssystem mit Umstellungen entstehen [Sie97]:

- *Traktionskosten* beinhalten die Kosten für das Triebfahrzeug, den Triebfahrzeugführer und für den Energieverbrauch. Die Kosten für das Triebfahrzeug enthalten dabei sowohl Abschreibungen als auch Unterhaltungskosten für das Fahrzeug. Siegmann führt an, dass die beiden erstgenannten Teilkosten ihrem Charakter nach eher zeitabhängig sind. Die Energiekosten sind hingegen abhängig von Beschaffenheit und Länge einer Fahrtstrecke, aber im Wesentlichen unabhängig von der Auslastung des Zuges. Sie können unter Berücksichtigung einer mittleren Reisegeschwindigkeit für eine gegebene Fahrtstrecke bestimmt werden. Wird eine konkrete Zugverbindung etabliert, sind die Traktionskosten als Fixkosten zu betrachten.
- *Trassenkosten* sind Kosten, die das EVU für die Nutzung der Schienenwege an das EIU zu entrichten hat. In Abhängigkeit von verschiedenen technischen Faktoren [DB 13], z.B. zulässigen Achslasten und Höchstgeschwindigkeiten, entstehen entfernungsabhängige Kosten. Ähnlich den Traktionskosten sind die Trassenkosten bei Etablierung einer Zugverbindung als Fixkosten anzusehen. Die Trassenvergabe wird vom zuständigen EIU geregelt. Für regelmäßige Verkehre werden sogenannte Regeltrassen über Rahmenverträge langfristig zu günstigen Konditionen vergeben [DB 13]. Zusätzlich existiert die Möglichkeit Bedarfstrassen anzumelden. Diese werden vorangemeldet und können gegen ein vergleichsweise geringes Entgelt bei Nichtnutzung abbestellt werden. Kurzfristig benötigte Fahrwegkapazitäten werden als sogenannte Adhoc-Trassen zur Verfügung gestellt [NP09]. Daher ist es im Zweifelsfalle ökonomisch sinnvoll, mehr Kapazitäten zu bestellen als unbedingt erforderlich sind.
- *Zugbildungskosten* entstehen bei der Zerlegung und Bildung von Zügen in Umstellbahnhöfen. Die Kosten, die für das Umstellen aller im Zug mitfahrenden Wagen anfallen, können einer Zugfahrt als variable Kosten zugerechnet werden. Sie sind von der Auslastung des Zuges, d.h. der Anzahl der angehängten Wagen, abhängig. Die Kosten für eine Umstellung können vom Umstellbahnhof abhängen. In großen Anlagen mit hohem Automatisierungsgrad sind die auf die einzelnen Wagen umzulegenden Kosten geringer als in weniger technisierten Knoten.
- *Wagenkosten* bilden die Kosten des Wagenparks ab. Dem EVU entstehen jedoch bis auf Reparaturen keine direkten Kosten durch den Wageneinsatz oder dessen

Laufleistung, sondern lediglich durch Abschreibungen. Die Wagenkosten können von der Planung der Wagenflüsse nur mittelbar beeinflusst werden, wenn der Wagenpark auf Grund effektiverer Routen verkleinert werden könnte. Aus diesem Grund werden die Wagenkosten im Weiteren vernachlässigt.

Zusammenfassend ist festzuhalten, dass die Kosten einer Zugfahrt als Summe streckenabhängiger Fixkosten und einer auslastungsabhängigen Kostenkomponente für die Zugbildung modelliert werden können.

2.3.2 Präzisierung der Problemstellung

Die Systemkomplexität des deutschen EWV und insbesondere die Optimierung seiner Prozesse machen eine rechnergestützte Planung in der Praxis sinnvoll. Dies gilt insbesondere für eine effiziente Routenplanung der etwa 20.000 Eisenbahnwagons, die täglich im EWV transportiert werden. Zur späteren Umsetzung in ein Modell wird an dieser Stelle die zentrale Planungsaufgabe dieser Arbeit in Worten beschrieben. Diese Systembeschreibung orientiert sich dabei am Vorgehensmodell der mathematischen Optimierung aus Unterkapitel 1.2:

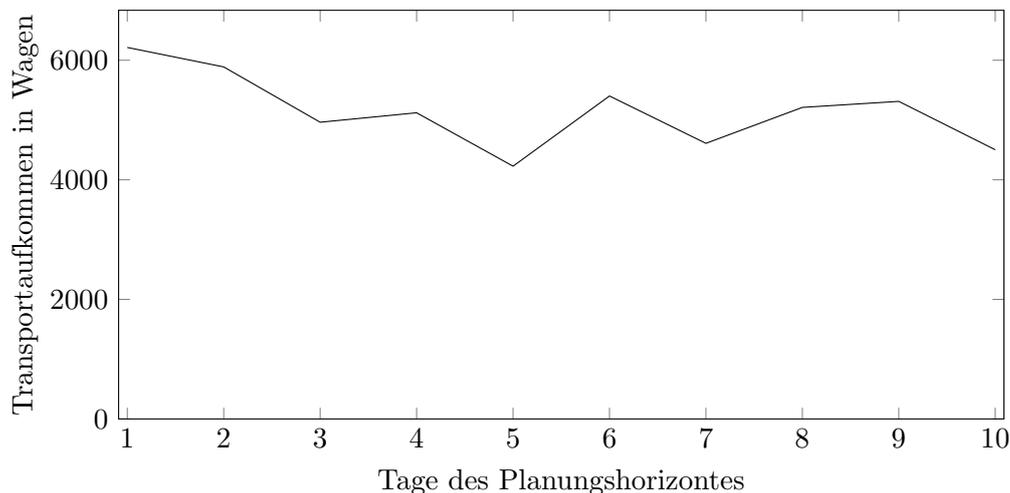


Abbildung 2.4: Darstellung des regionalen Wagenaufkommens für 10 Werktage

Zu bestimmen ist ein Wagenroutenplan für den deutschen EWV, der gesamtkosteneffizient für alle Transportaufträge eines gegebenen Planungszeitraums ist. Die Wagen einer *Relation*, d.h. mit gleichem Start- und Zielbahnhof, werden als Wagengruppe betrachtet, die während ihres Transportes nicht getrennt wird. Da die Wagenaufkommen äußerst volatil sind, d.h. von Tag zu Tag schwanken (vgl. Abbildung 2.4), ist es unzureichend, das Aufkommen eines Mustertages zu berücksichtigen [NP09]. Der zu berechnende Routenplan muss stattdessen für mehrere Szenarien zulässig und in der Summe kostenminimal sein, weil nicht nur die Gesamtzahl der Relationen, sondern vor allem die Aufkommen einzelner Relationen stark schwanken. Eine untertägige Planung ist indes nicht möglich,

da die prognostizierten Aufkommensszenarien keine entsprechende Genauigkeit aufweisen und somit keine solide Datengrundlage für eine Optimierung taktischer Routenpläne verfügbar ist. Die Prognosemethoden schränken überdies die Möglichkeit ein, exakte Parameter für Transportzeiten zu erhalten, weshalb im vorliegenden Fall eine Beschränkung der Umstellvorgänge jedes Wagens als implizite Begrenzung der Transportzeit eingesetzt wird. Die Kosteneffizienz wird anhand der in Abschnitt 2.3.1 definierten Kostenstruktur bemessen. Die betrachteten Wagenströme werden aggregiert auf das Netz der Umstellbahnhöfe betrachtet. Da von Knoten ohne Rangiermittel in der Regel nur eine einzige Schienenverbindung zu einem Umstellbahnhof existiert, sind in diesen Knoten auch keine Entscheidungen zu Umstellplänen zu treffen [Bru06]. Die Planung erfolgt unter der Maßgabe, dass die folgenden Systemregeln eingehalten werden:

- alle gegebenen Transportaufträge müssen korrekt ausgeführt, d.h. alle Sendungen vom Versender zum Empfänger transportiert werden.
- die für den Transport aller Sendungen notwendigen Zugverbindungen sind derart zu gestalten, dass gegebene Längen- und Gesamtgewichtsrestriktionen der Züge auf jedem Streckenabschnitt eingehalten werden.
- die Kapazität jedes Umstellbahnhofs ist zu berücksichtigen. Die Kapazität wird dabei sowohl in Form der Ablaufbergkapazität, d.h. der Anzahl möglicher Wagenumstellungen pro Zeitraum, als auch der Anzahl verfügbarer Richtungsgleise bemessen.
- für jeden Wagen ist eine Höchstzahl erlaubter Umstellvorgänge gegeben.
- alle Wagen, die sowohl Versenderknoten als auch Empfängerknoten gemein haben, d.h. zur gleichen Relation gehören, werden auf gleichem Weg durch das Schienennetz transportiert.
- die Planung soll eine möglichst hohe Robustheit gegenüber Aufkommensschwankungen besitzen. Dazu ist es erforderlich, dass der Wagenroutenplan für alle betrachteten Szenarien zulässig ist.

3 Stand von Wissenschaft und Technik

Ziel dieses Kapitels ist die Darstellung von Konzepten zur Erstellung von Umstellplänen und zur Wagenroutenplanung aus der Wissenschaft sowie der Praxis des SGV. Es werden in Unterkapitel 3.1 zunächst Begriffe der Graphentheorie eingeführt, die zum Verständnis der weiteren Arbeit notwendig sind. Danach werden in Unterkapitel 3.2 wissenschaftliche Veröffentlichungen zu Umstellplänen und entsprechende Softwaresysteme diskutiert. In Unterkapitel 3.3 werden Begriffe der Komplexitätstheorie eingeführt und für das Verständnis wichtige Optimierungsprobleme definiert. Unterkapitel 3.4 enthält eine Zusammenfassung notwendiger Kenntnisse der mathematischen Optimierung.

3.1 Graphen und Flüsse

Ein *Graph* ist ein Tupel $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ aus einer Menge von *Knoten* \mathcal{N} und einer Menge von *Kanten* $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. Ein Graph heißt *gerichtet*, wenn alle Kanten aus \mathcal{E} einen eindeutigen Startknoten $i \in \mathcal{N}$ und einen eindeutigen Endknoten $j \in \mathcal{N}$ besitzen. Eine gerichtete Kante wird dann mit $e = (i, j)$ bezeichnet. Die eingehenden Kanten eines Knotens $i \in \mathcal{N}$ seien mit $\delta_i^+ = \{(j, i) \mid (j, i) \in \mathcal{E}\}$ bezeichnet. Analog werden die ausgehenden Kanten $\delta_i^- = \{(i, j) \mid (i, j) \in \mathcal{E}\}$ bezeichnet. Ein Graph heißt *vollständig*, wenn es zwischen jedem Paar von Knoten $i, j \in \mathcal{N}$ eine Kante $e = (i, j)$ gibt. Eine Folge von Kanten e_1, e_2, \dots, e_k heißt *Kantenzug* oder *Kantenfolge*. Ein Kantenzug mit gleichem Start- und Endknoten heißt *Kreis*. Ein Graph ohne Kreise wird als *kreisfrei* bezeichnet. Ein kreisfreier Kantenzug heißt *Weg*. Die Abbildung

$$\text{dist} : \mathcal{N} \times \mathcal{N} \mapsto \mathbb{R} \quad (3.1)$$

bezeichne den Abstand zweier Knoten i und j im Graph, d.h. die Länge der kürzesten Verbindung zwischen den betreffenden Knoten. Ein Graph erfüllt die Dreiecksungleichung, wenn für drei beliebige Knoten i_1, i_2, i_3 stets die folgende Gleichung gilt:

$$\text{dist}(i_1, i_3) \leq \text{dist}(i_1, i_2) + \text{dist}(i_2, i_3) \quad (3.2)$$

Graphen können verwendet werden, um physische Netzwerke, z.B. Schienennetze, logisch abzubilden. Um Bewegungen physischer Einheiten, wie z.B. von Zügen oder Eisenbahnwagen, abzubilden, wird sich sogenannter Flüsse bedient. Der Begriff eines Flusses zwischen zwei Knoten wird zunächst abstrakt als mathematische Abbildung eingeführt:

Definition 3 (s-t-Fluss)

Es seien $s, t \in \mathcal{N}$, $s \neq t$. Eine Funktion $x : \mathcal{E} \mapsto \mathbb{R}$ heißt (s, t) -Fluss, wenn gilt

$$\sum_{e \in \delta_i^+} x(e) = \sum_{e \in \delta_i^-} x(e) \quad \forall i \notin \{s, t\} \quad (3.3)$$

Stellt man sich einen Fluss als Menge von Objekten vor, die sich von einer Quelle s zu einer Senke t bewegen, so muss der Definition gemäß jedes Objekt, das in einen Knoten eintritt, diesen auch wieder verlassen. Ausnahmen sind die Quelle, aus der sich Objekte heraus-, und die Senke, in die sich Objekt hineinbewegen. Flüsse stellen ein geeignetes Mittel dar, um die Bewegung physischer Einheiten in einem Netzwerk zu modellieren.

3.2 Planungsprobleme im Schienenverkehr

Zunächst werden wissenschaftliche Methoden für Planungsprobleme des SGV kurz dargestellt. Der Fokus wird dabei auf die Planung von Wagenrouten eingeschränkt, welche in der wissenschaftlichen Literatur in der Regel über die Optimierung von Umstellplänen abgebildet wird. Anschließend werden wissenschaftliche Ansätze sowie kommerzielle Softwaresysteme für Umstellpläne und Wagenrouten betrachtet.

3.2.1 Abgrenzung von verwandten Planungsproblemen

Die Optimierung verschiedenster Planungsprobleme des Schienenverkehrs wurde in vielen wissenschaftlichen Arbeiten untersucht. Dabei ist zunächst anzumerken, dass sich die Planungsaufgaben im SGV deutlich von denen im Schienenpersonenverkehr (SPV) unterscheiden. Die verwendeten Methoden sind im Allgemeinen nicht übertragbar, wie bereits Busseck et al. in [BWZ97] zusammenfassten. Aus diesem Grund wird trotz höchst erfolgreicher Projekte im SPV, wie z.B. [KHA⁺09], im Folgenden nur noch Literatur betrachtet, die sich auf den SGV bezieht.

Die Wagenroutenplanung behandelt die Bewegung von Wagen. Die Planung von Wagenbewegungen ist zunächst logisch von der Fahrplangestaltung, d.h. der zeitlichen Planung von Zugbewegungen abzugrenzen, wie von Harrod et al. in ihrem Übersichtsartikel [HGC⁺10] beschreiben. Die Bewegung von Wagen ist jedoch physisch unmittelbar mit Zugbewegungen verbunden. Daher erscheint eine gemeinsame Planung sinnvoll und wurde für den nordamerikanischen SGV bereits von Keaton [Kea89, Kea92], Huntley et al. [HBSM95] sowie von Gorman [Gor98] betrachtet. Die resultierenden Pläne werden als *Operationspläne* bezeichnet. Eine aktuellere Arbeit von Zhu et al. [ZCG11] behandelt unter der Bezeichnung *Scheduled Service Network Design* ebenfalls die integrierte Planung von Wagenbewegungen und Zugfahrplänen. Eine Übertragbarkeit der Ergebnisse auf den deutschen EWV liegt jedoch nicht vor, wie im folgenden Absatz begründet wird. In der Europäischen Union und damit insbesondere in Deutschland sind EIU und EVU per Gesetz getrennte Gesellschaften. Bevor ein EVU einen Transport durchführen kann, muss beim EIU eine Trasse angemietet werden (vgl. Abschnitt 2.2.2). Dieses organisatorische Hindernis liegt in Nordamerika nicht vor, da sich die dortigen Schienennetze in Privatbesitz, d.h. im Besitz der EVU, befinden [Pos08]. Während die Fahrplanerstellung in Nordamerika einen rein firmeninternen Prozess darstellt, ist in Deutschland eine kooperative Absprache zwischen EVU und EIU notwendig. Somit ist es einem EVU nicht möglich, ohne Absprache mit dem zuständigen EIU, einen integrierten Zeitplan für Wagen- und Zugbewegungen zu gestalten. In Folge dessen ist ein sequentieller Ansatz für Wagenroutenplanung und Fahrplanerstellung in Deutschland die Praxis [NP09].

Im ersten Schritt wird bestimmt, in welcher Reihenfolge ein Wagen die Umstellbahnhöfe durchläuft. Dies geschieht durch die Erstellung von Umstellplänen, deren Festlegung laut Unterkapitel 2.3 zum Wagenrouting äquivalent ist. Dabei ist zwar zu berücksichtigen, wie viele Züge voraussichtlich notwendig sein werden, um die Transporte durchzuführen, nicht aber deren genaue Abfahrtszeiten. Erst im Anschluss werden die Wagen konkreten Zügen zugeordnet, wie es Jha et al. [JAG07] beschrieben haben. Für weitere Erläuterungen zur Bestimmung konkreter Zeitpläne sei auf die Übersichtsartikel von Cordeau et al. [CTV98] sowie Harrod et al. [HGC⁺10] verwiesen.

Die genannte Planungsabfolge stellt nicht die einzig denkbare Sequenzierung dar. Lin et al. [LWJ⁺12] berichten, dass der übliche Planungsablauf im chinesischen Schienengüterverkehr mit der Bestimmung von Zugverbindungen und deren Häufigkeit beginnt. Anschließend werden vollständig getaktete Fahrpläne erstellt. Die zur Erfüllung der Transportaufträge notwendigen Wagenbewegungen richten sich dann nach diesem Fahrplan. Die Erstellung von Umstellplänen ist damit nur ein untergeordnetes Problem, das am Ende der Planungshierarchie steht. Crevier et al. [CSC12] untersuchen die kombinierte Optimierung von Fahrplänen und Preissystemen. Jin et al. [JZL13] integrieren in die Planung von Zugfahrten zusätzlich ein Personalplanungsproblem. Im Folgenden werden ausschließlich Arbeiten vorgestellt, die vornehmlich die Erstellung von Umstellplänen, respektive die Routenplanung im SGV, zum Thema haben.

3.2.2 Optimierung von Umstellplänen und taktischen Wagenrouten im Schienengüterverkehr

Die Erstellung von Umstellplänen hängt von geographisch unterschiedlichen Umständen ab, die sich aus technischen, rechtlichen oder organisatorischen Zusammenhängen ergeben. Drei Klassen von Umstellplänen sind zu unterscheiden, die nach absteigender Zahl ihrer Freiheitsgrade vorgestellt werden: Individuelle Wagenrouten, Blockingpläne und Leitwegepläne. Nach deren Vorstellung werden alle drei Konzepte hinsichtlich ihrer Übertragbarkeit auf den deutschen EWW bewertet und verglichen. Anschließend wird ein Überblick über kommerzielle Softwaresysteme gegeben.

Umstellpläne mit individuellem Wagenrouting

Die größtmögliche Anzahl an Freiheitsgraden bei der Routenplanung ist gegeben, wenn es erlaubt ist, jeden Wagen entlang einer *individuellen Route* durch das Netzwerk zu transportieren. Fukusawa et al. [FAPU02] betrachten das Problem der individuellen Wagensteuerung im brasilianischen EWW. Sie beziehen dabei die Verteilung von Leerwagen mit in ihre Planung ein.

Campetella et al. [CLPR06] optimieren Wagenflüsse in Italien auf Basis individueller Wagenrouten, d.h. es wird die Trennung von Sendungen mit gleichem Versender und Empfänger zugelassen. Dieses Vorgehen sei in Italien üblich. Sie integrieren dabei Teile weiterer Planungsschritte in ihr Optimierungsproblem. Das Schweizer Expressnetz für Einzelwagen war Gegenstand der Forschung von Ceselli et al. [CGL⁺08]. Auch sie lassen die individuelle Steuerung jedes Wagens zu. Sie nehmen dabei eine hierarchische Einteil-

lung in Teilprobleme vor, die mit der Optimierung von Umstellplänen beginnt. Dabei wird die besondere Struktur des schweizerischen Einzelwagenverkehrsnetzes ausgenutzt, in dem es nur zwei große Umstellbahnhöfe gibt. Es wird das Ziel verfolgt, Rangierarbeiten gleichmäßig auf die genannten Umstellbahnhöfe zu verteilen.

Blockingpläne

Im Falle der sogenannten *Blockingpläne* werden Wagen nicht mehr individuell geroutet. Wagen, die Versender- und Empfängerknoten gemein haben, d.h. derselben Relation zugeordnet sind, werden in der Regel während ihres Transportes zum Empfänger nicht getrennt. In Ausnahmen wird die Trennung von Wagenflüssen der gleichen Relation zugelassen. Folgende Definition erklärt den Namen Blockingplan:

Definition 4 (Block)

Als Block wird eine Gruppe von Wagen bezeichnet, die auf einer Strecke zwischen zwei Knoten gemeinsam transportiert werden. Die enthaltenen Wagen müssen nicht notwendigerweise einer einzigen Relation anzugehören. Start- und Endknoten des Blocks müssen nicht mit den Start- oder Endknoten der beteiligten Relationen übereinstimmen.

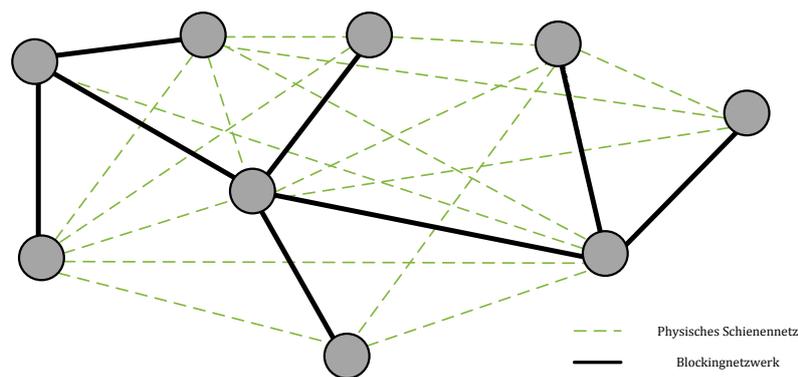


Abbildung 3.1: Beispiel eines Blockingnetzwerkes als virtuelles Teilnetz des Schienennetzes

Jeder Wagen wird somit Bestandteil verschiedener Blöcke, während er von seinem Versenderknoten zu seinem Empfängerknoten transportiert wird. Die Entscheidung, einen Block zwischen zwei Knoten zu bilden, impliziert die Notwendigkeit einer Zugverbindung zu dessen Transport. Ein Zug kann mehrere Blöcke transportieren. Es entsteht eine virtuelle Verbindung zwischen den betreffenden Knoten, die auch als *Blockkante* bezeichnet wird. Alle Blockkanten einer Instanz, d.h. eines konkreten Planungsfalls, bilden das sogenannte *Blockingnetzwerk*. Das Blockingnetzwerk ist ein virtuelles Teilnetz des physischen Schienennetzes (vgl. Abbildung 3.1). Die Zuordnung von Wagen zu Blöcken wird durch den Blockingplan abgebildet.

Definition 5 (Blockingplan)

Der Blockingplan bestimmt die Blöcke, als deren Teil die Wagen jeder Relation transportiert werden.

Der Blockingplan ist damit eine Variante eines Umstellplanes. Aus ihm lässt sich die Gestalt des Blockingnetzwerkes ableiten. Das Konzept der Blockingpläne wird vornehmlich in den USA und Kanada angewendet. Zahlreiche wissenschaftliche Beiträge zu deren Erstellung mit Hilfe mathematischer Methoden wurden veröffentlicht, wie die nachfolgende, chronologisch gegliederte Aufstellung der Modelle verschiedener Autoren aufzeigt: Das erste Optimierungsmodell für Blockingpläne wurde 1980 von Bodin et al. [BGS80] veröffentlicht. Unter Verwendung einer nichtlinearen Zielfunktion wurde das als *Railroad Blocking Problem* (RBP) bezeichnete Problem modelliert. Die Nichtlinearität des Modells entsteht bei der Modellierung von verspätungsabhängigen Kosten bei der Umstellung von Wagengruppen. Neben Kosten für jede Wagenumstellung werden die Transportkosten wagenweise in der Zielfunktion berücksichtigt. Dies bedeutet, dass die Etablierung eines Blocks zwischen zwei Knoten keine direkten Kosten impliziert. Hingegen verursacht der Transport jedes einzelnen Wagens entfernungsabhängige Kosten. Es werden somit keine Traktions- oder Trassenkosten betrachtet. Die verwendete Kostenfunktion ist folglich nicht geeignet, um die Kosten des deutschen EWV abzubilden. Von den in Abschnitt 2.3.2 geforderten Nebenbedingungen werden lediglich die Rangierkapazität der Knoten und die Längenrestriktion für die Wagenzüge berücksichtigt. Eine Lösung des Problems war auf Grund der beschränkten technischen sowie algorithmischen Möglichkeiten zu jener Zeit nur mit manuellen Eingriffen möglich.

Ebenfalls im Jahre 1980 veröffentlichte Assad [Ass80] ein Modell zum Aufbau von Blockingplänen. Die betrachtete Kostenfunktion setzt sich aus Umstell- und Transportkosten zusammen. Dabei werden die Transportkosten auf Basis von Zugkilometern berechnet. Diese Modellierung deckt sich mit den Kostenstrukturen aus Abschnitt 2.3.1. Kapazitäten in den Knoten werden ebenso betrachtet wie Kapazitäten der Wagenzüge. Ihre Kapazität wird dabei in der Anzahl transportierter Wagen bemessen. Eine implizite Beschränkung der Transportzeit besteht durch die Einführung von Strafkosten für Verspätungen in der Zielfunktion. Die Modellierung ist nur mit Einschränkungen auf das Problem dieser Arbeit übertragbar, weil sich der Autor auf Netzwerke mit einer Linienstruktur konzentriert, d.h. die Knoten sind geographisch in einer Linie angeordnet. Einen Lösungsalgorithmus schlägt Assad nicht vor.

Ein pfadbasierter Ansatz aus dem Jahre 1983 stammt von Crainic et al. [CFR83]. Betrachtet wird ein pfadbasiertes, nichtlineares Modell, dessen Kostenfunktion sich aus zugkilometerabhängigen Transportkosten und Strafkosten für Umstellungen zusammensetzt. Die Flussvariablen sind kontinuierlich gewählt. Daher ist es erlaubt, die Wagen einer Relation über verschiedene Strecken zu ihrem Ziel zu transportieren. Dies ist auch nötig, da das Modell die Priorisierung bestimmter Sendungen zulässt. Die Nichtlinearität des Modells macht den Einsatz von Heuristiken zur Problemlösung notwendig. Die Trennbarkeit einer Relation verhindert die Übertragung auf den deutschen EWV.

Newton [New97] entwickelt ein pfadbasiertes Modell für das Blockingproblem. Sein Hauptaugenmerk liegt auf dem Design des Blockingnetzwerkes. Dazu wird innerhalb des

Modells bestimmt, ob eine Zugverbindung etabliert wird oder nicht. Relationen dürfen auf mehrere Pfade aufgeteilt werden. Die Flussstärke auf jeder Kante ist beschränkt, d.h. es werden Zugkapazitäten berücksichtigt. Die Kapazität der Knoten wird in Wagen und in der Anzahl ausgehender Blocks bemessen. Die Menge der potentiellen Pfade ist vor Beginn der Optimierung explizit gegeben und kann spezielle implizite Bedingungen enthalten. Newton nennt bspw. eine Beschränkung der Umstellungen pro Wagen. Die Einschränkung auf eine vorab definierte Menge potentieller Pfade vereinfacht dabei den Lösungsprozess. Es ist dadurch jedoch nicht auszuschließen, dass potentielle Optimallösungen, die bei Betrachtung aller Pfade gefunden werden könnten, nicht in der Formulierung enthalten sind. Die Kostenfunktion setzt sich aus Transportkosten, die pro Wagenkilometer berechnet werden, und Umstellkosten für jeden Wagen zusammen. Dies steht im Widerspruch zu den Anforderungen aus Abschnitt 2.3.2 und beschränkt die Übertragbarkeit der Methoden. Zur Lösung des Modells entwickeln Newton et al. [NBV98] ein Column-Generation-Verfahren, bei dem sie nur eine eingeschränkte Anzahl erlaubter Pfade für jede Relation zulassen. Damit ist es möglich, alle getesteten Instanzen mit einstelligen Optimalitätslücken zu lösen. Barnhart et al. [BHV00] entwickeln hierzu eine Methode auf Basis von Lagrange-Relaxierungen. Die Lagrange-Multiplikatoren werden dabei mit Hilfe einer Subgradientenmethode ermittelt. Auch in diesem Fall macht die Einschränkung der potentiellen Pfade eine beinahe optimale Lösung der betrachteten Instanzen möglich.

Jin erweitert in seiner Dissertation [Jin98] Newtons Modell [New97] um Aspekte der Robustheit. Er entwickelt verschiedene Modelle, welche die dynamische Nachfrage im EWW abbilden. Ein Modell, das als *statischer Worst-Case-Fall* bezeichnet wird, erlaubt die Einbindung multipler Szenarien, wie sie in der vorliegenden Arbeit untersucht wird. Es werden jedoch, wie auch in Newtons Modell, keine Zugkosten betrachtet und die Wagenflüsse einer Relation dürfen getrennt werden. Weitere Modelle, die den Erwartungswert unter multiplen Szenarien minimieren oder dynamische Änderungen während des Planungshorizontes berücksichtigen, werden ebenfalls vorgestellt. Zur Lösung der Optimierungsprobleme werden Column-Generation-Verfahren und Lagrange-Relaxierungen verwendet.

Ahuja et al. [AJL07] präsentieren ein Modell, das ebenfalls die Bildung des Blockingnetzwerkes fokussiert. Berücksichtigt werden Rangiergleiskapazitäten und Ablaufbergkapazitäten in den Knoten. Des Weiteren wird die Anzahl der Wagen beschränkt, die über eine Kante transportiert werden dürfen. Die Kosten für den Transport von Wagen wird proportional zu den gefahrenen Wagenkilometern berechnet. Eine Trennung der Wagen einer Relation ist nicht erlaubt. Eine auf einer Nachbarschaftssuche basierende Verbesserungsheuristik wurde in ein Softwaresystem implementiert, auf das in Abschnitt 3.2.2 näher eingegangen wird. Die Heuristik ist in der Lage, zulässige Lösungen für extrem große Realinstanzen mit bis zu 50.000 Sendungen zu erzeugen. Ein Optimalitätsnachweis oder eine Güteabschätzung werden nicht erbracht.

Yue et al. [YZYF11] entwickeln ein Modell für das Blockingproblem in China. Dies steht nur im scheinbaren Widerspruch zur Aussage von Lin et al. [LWJ⁺12], in China werde kein Blocking betrieben, da ein spezielles nordchinesisches Teilnetz betrachtet wird, das für den Transport von Kohle genutzt wird. Es wird ein nichtlineares Modell vor-

geschlagen, das eine Vielzahl gekoppelter Kostenfaktoren enthält. Zur Lösung wird ein Ameisenalgorithmus (vgl. Abschnitt 3.4.3) eingesetzt. Die Ergebnisse sind auf Grund der stochastischen Einflüsse des Algorithmus nur bedingt aussagekräftig.

Blockingpläne für den iranischen SGV werden von Yaghini et al. [YFN11] optimiert. Das betrachtete Optimierungsmodell weist starke Ähnlichkeiten zu dem Modell von Ahuja et al. [AJL07] auf. Es werden jedoch keine Umstellkosten in der Zielfunktion berücksichtigt. Zur Lösung wird ein Ameisenalgorithmus eingesetzt, der in der Lage ist, innerhalb weniger Sekunden zulässige Lösungen für die betrachteten Testinstanzen zu generieren und zu verbessern. Güteabschätzungen zeigen eine hohe Genauigkeit des Verfahrens. In einem weiteren Beitrag wird von Yaghini et al. [YABS13] eine Tabusuche für das Blockingproblem entwickelt.

Tabelle 3.1: Überblick über die Nebenbedingungen der Blockingmodelle

	Bodin [BGS80]	Assad [Ass80]	Crainic [CFR83]	Newton [New97]	Ahuja [AJL07]	Yue [YZYF11]	Yaghini [YFN11]
Flüsse trennbar	x	x		x			
Zeitlimit	*	*	*			x	
Länge max.	x					x	x
Gewicht max.						x	x
Wagenzahl max.		x	x	x	x	x	x
Zugkosten		x	x			x	
Wagenkosten	x		x	x	x	x	x
Umstellkosten	x	x	x	x	x	x	x

*durch zusätzliche Kosten in der Zielfunktion abgebildet

In Tabelle 3.1 sind die Eigenschaften der Modelle in komprimierter Form zusammengefasst. Dabei ist auffällig, dass bei der Mehrzahl der Modelle die Transportkosten nicht in Abhängigkeit von Zugkilometern, sondern linear in Wagenkilometern berechnet werden. Diese Modellierung wird den Kostenstrukturen im deutschen EWV, wie sie in Abschnitt 2.3.1 vorgestellt wurden, nicht gerecht. Die Ausrichtung der Kostenstruktur in der Modellierung des nordamerikanischen SGV hat jedoch praktische Gründe. Zwischen dem nordamerikanischen und dem deutschen SGV bestehen erhebliche Unterschiede hinsichtlich Infrastruktur und Besitzverhältnissen: Während sich die Eisenbahninfrastruktur in Europa im Besitz der von den EVU unabhängigen EIU befindet, gehören die nordamerikanischen Schienenwege überwiegend privat finanzierten Eisenbahnunternehmen, die selbst Transportleistungen anbieten [Pos08]. Es gibt eine Trennung von Güterverkehr und Personenverkehr auf der Schiene, während in Deutschland eine gemischte Nutzung unter Priorisierung des Personenverkehrs üblich ist. Die in Abschnitt 2.1.3 beschriebenen Überholgleise, auf denen Güterzüge warten, während schnellere Personenzüge vorbeifahren, sind somit in dieser Form nicht notwendig. Eine Längenbeschränkung für Güterzüge wird daher in Nordamerika nicht durch überholenden Verkehr erzwungen. In Folge dessen sind Zuglängen von mehreren Kilometern möglich. Des Weiteren ist die nordamerikanische Infrastruktur für den Einsatz von doppelstöckigen Frachtzügen geeignet, der die Zugkapazitäten weiter erhöhen und somit eine Verschiebung der Kostenstruktur bewirkt. Der Fixkostenanteil an den Transportgrenzkosten eines Wagens

ist bei hinreichender Auslastung wesentlich geringer als im Falle kurzer Züge. Eine zu den Wagenkilometern proportionale Kostenstruktur bildet die realen Kosten sinnvoll ab. Diese Erkenntnisse wurden im Rahmen der Promotion bereits in [CV13] veröffentlicht. Die Vernachlässigung von Bündelungseffekten in Form von zugabhängigen Kosten hat einen großen algorithmischen Vorteil. Werden die Kosten linear in den gefahren Wagenkilometern berechnet, besteht ein linearer Zusammenhang zwischen optimalem Routing aller Wagen und dem korrespondierenden optimalen Blockingnetzwerk. Das bedeutet, dass die Berechnung eines optimalen Blockingnetzwerkes oder Routings vom algorithmischen Standpunkt einfach ist, wenn das jeweils andere gegeben ist. Dieser strukturelle Zusammenhang lässt sich algorithmisch ausnutzen und versetzt die darauf basierenden Algorithmen in die Lage, Ergebnisse sehr effizient zu berechnen.

Leitwegepläne

In Deutschland wird derzeit noch eine Form von Umstellplänen verwendet, die als *Leitwegepläne* bezeichnet werden. Ein Leitwegeplan ist ein spezieller Blockingplan, der zwei bestimmte Nebenbedingungen berücksichtigt. Dazu werden zunächst die Umstellbahnhöfe des EWV in Hierarchielevel unterteilt. Auf dem höchsten Hierarchielevel stehen die *Rangierbahnhöfe* (Rbf). Die mittlere Hierarchiestufe umfasst die sogenannten *Knotenpunktbahnhöfe* (Kbf). Die unterste Hierarchiestufe bilden die Satellitenbahnhöfe mit Rangiermitteln (SmR). Die Umstellkapazitäten der jeweiligen Bahnhofstypen stehen zumeist in Relation zu ihrem Hierarchielevel, d.h. in der Regel sind Rbf größer als Kbf, welche wiederum größer als SmR sind. Die Anzahl der jeweiligen Umstellbahnhöfe im deutschen Schienennetz ist hingegen umgekehrt proportional zu ihrem Hierarchielevel: In Deutschland gibt es derzeit nach gängiger Einteilung 9 Rbf, 27 Kbf und 129 SmR. Mit Hilfe der Hierarchielevel lässt sich ein Leitwegeplan folgendermaßen definieren:

Definition 6 (Leitwegeplan)

Ein Leitwegeplan ist ein Blockingplan, der zusätzlich folgende Nebenbedingungen erfüllt:

1. *Hierarchiebedingung*: Jeder Wagen darf entlang der Hierarchie der Umstellbahnhöfe des EWV streng monoton aufwärts bis zum Rbf gelangen. Auf dem Hierarchielevel Rbf darf er für zwei Umstellvorgänge verbleiben, auf allen anderen Hierarchieleveln nur für einen. Anschließend wird er der Hierarchie folgend streng monoton abwärts geleitet.
2. *Leitwegebedingung*: Für je zwei Umstellbahnhöfe gibt es nur eine einzige Route, d.h. Umstellreihenfolge, über die Wagen vom erstgenannten zum zweitgenannten Umstellbahnhof geleitet werden dürfen.

Zur Verdeutlichung der Definition werden die Hierarchiebedingung und die Leitwegebedingung kurz beispielhaft dargestellt: Die Hierarchiebedingung sagt aus, in welcher Reihenfolge jeder Wagen das Netzwerk der Umstellbahnhöfe durchlaufen darf (vgl. Abbildung 3.2). Der übliche Weg eines Wagens von einem SmR zu einem anderen führt ihn zunächst zu einem Kbf, von wo aus er zu einem Rbf weiter transportiert wird. Der Wagen wird von dort aus zu einem weiteren Rbf transportiert. Ein Transport zu einem dritten

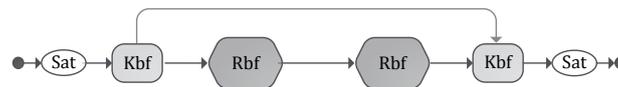


Abbildung 3.2: Schematische Darstellung von Leitwegen im EWW [CG13]

Rbf ist erlaubt, falls dies sinnvoll ist. Ansonsten findet ein Transport abwärts entlang der Hierarchien über einen Kbf zum Ziel-SmR statt. Je nach Aufkommen zwischen zwei Knoten ist es erlaubt, ein Hierarchielevel zu überspringen. So kommt es in der Realität häufig vor, dass zwischen zwei Kbf eine Direktverbindung ohne zwischenzeitliche Umstellungen in einem Rbf etabliert wird. Das beschriebene System wird auch als flexibles Knotenpunktsystem bezeichnet [Ber01].

Die Leitwegebedingung ist ein Regelsystem, das in dieser Form nur im deutschen EWW Anwendung findet: Zwei Wagen mit gemeinsamem Zielbahnhof, deren Routen sich in einem Umstellbahnhof kreuzen, werden für den Rest ihrer Reise über die gleichen Routen geführt. Dieses System wurde aus historischen Gründen eingeführt. Einem Güterwagen, dessen Zielbahnhof bekannt ist, kann mit Hilfe der Leitwegebedingung von jedem Bahnhof des Netzwerkes eine eindeutige Strecke und damit ein eindeutiger Nachfolgebahnhof zugeordnet werden. Dieses Vorgehen hat den Vorteil, dass ein mit seinem Ziel markierter Wagen in jedem Bahnhof schnell seinem Ausgangszug und damit seinem Richtungsgleis zugeordnet werden kann. Auf diese Weise sinkt die Fehleranfälligkeit des Systems, wenn Wagen manuell zu Richtungsgleisen zugeordnet werden.

Die Optimierung von Leitwegeplänen ist Thema der Diplomarbeit von Schülldorf [Sch08] und der Dissertation von Homfeld [Hom12]. Teilergebnisse dieser Arbeiten wurden zuvor bereits von Fügenschuh et al. [FHS09] veröffentlicht. Homfeld entwickelt ein Modell, das Transportkosten sowohl in Zugkilometern als auch in Wagenkilometern bemisst. Auch Umstellkosten werden in der Zielfunktion berücksichtigt. Neben Gewichts- und Längenrestriktionen sind Kapazitätsrestriktionen in den Umstellbahnhöfen sowie Fahrzeitrestriktionen im Modell enthalten. Das Modell von Homfeld erfüllt beinahe alle Kriterien, die in Abschnitt 2.3.2 definiert wurden. Nicht berücksichtigt ist der für die vorliegende Arbeit zentrale Punkt der Robustheit, d.h. die Möglichkeit, stabile Lösungen für mehrere Szenarien gleichzeitig zu berechnen. Die Leitwege- und Hierarchiebedingungen, die Homfeld in sein Modell integriert, verringern die Freiheitsgrade der Lösungen. Es ist ihm daher möglich, auch bei Instanzen mit Realgröße gute untere Schranken für die Zielfunktionswerte zu ermitteln. Es werden mehrere Verfahren zum Auffinden zulässiger Lösungen entwickelt und vorgestellt. Für Realinstanzen erweist sich ein sogenannter *Rip-up-and-Reroute-Ansatz* als besonders vielversprechend. Es handelt sich dabei um eine Spezialform einer lokalen Suche. Aus numerischen Tests an ausgewählten Instanzen schlussfolgert er, dass eine Vernachlässigung der Leitwegebedingung keinen signifikanten Vorteil brächte. Eine Untersuchung, welchen Einfluss die wesentlich restriktivere Hierarchiebedingung hat, wird nicht durchgeführt.

Vergleich der drei Klassen von Umstellplänen

Individuelle Wagenroutings sind auf Grund ihrer Freiheitsgrade sehr flexibel. Die große Anzahl von Gestaltungsmöglichkeiten macht die Behandlung des Problems mit mathematischen Methoden nicht zwangsläufig einfacher als eine stärker restringierte Problemvariante. Im Falle realer Probleminstanzen des deutschen EWW würde die große Anzahl notwendiger Variablen die Optimierung erschweren. Dies ist kein Grund, der für sich genommen als Nachteil der individuellen Wagenrouten auszulegen ist. Hingegen sprechen praktische Gründe gegen die Umsetzung, die von Fügenschuh et al. [FHS09] angeführt wurden: Der Umstellprozess wird in deutschen Umstellbahnhöfen nicht voll automatisiert durchgeführt. Wenn individuelle Routen für jeden Wagen erlaubt sind, kann es bspw. passieren, dass mehrere Wagen, die vom selben Versender stammen und für den selben Empfänger bestimmt sind, auf unterschiedlichen Zügen den Bahnhof verlassen. Auf Grund der großen Zahl von Möglichkeiten ergibt sich ein hohes Fehlerpotential bei der Zuweisung von Richtungsgleisen. Die Autoren prophezeien daher, dass das individuelle Wagenrouting in Deutschland nicht eingeführt werde, solange Menschen den Umstellprozess aktiv begleiten.

Die Modellierung der Transportkosten auf Basis der gefahrenen Wagenkilometer führt zu einer besonderen strukturellen Eigenschaft optimaler Blockingpläne, die hinsichtlich dieser Kostenfunktion entstehen:

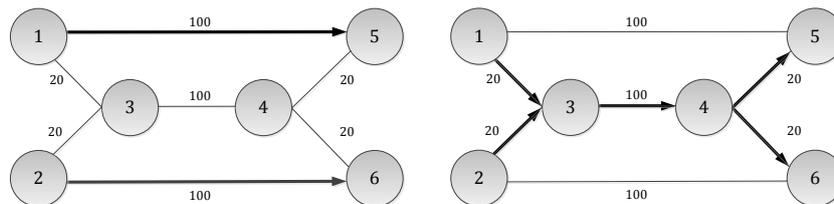


Abbildung 3.3: Optimallösungen für 2 Relationen $1 \rightarrow 5$ und $2 \rightarrow 6$:
links minimierte Wagenkilometer, rechts minimierte Zugkilometer

Bei Vernachlässigung von Kosten auf Basis von Zugkilometern werden in der Optimallösung Umwege einzelner Wagen nur dann in Kauf genommen, wenn die Kapazitäten gewisser Knoten dies erzwingen, aber nicht weil Kostenvorteile durch die Bündelung von Wagenströmen entstehen [CV13]. Der Unterschied zwischen Lösungen, die auf Wagen- bzw. Zugkosten basieren, wird in Abbildung 3.3 illustriert.

Zur Vereinfachung der Beispiele wurden nur Transportkosten berücksichtigt, die proportional zu den an den Kanten angegebenen Streckenlängen seien. Während bei der Minimierung von Wagenkosten jeder Wagen entlang seines kürzesten Weges durch das Netzwerk transportiert wird, führt die Modellierung der Zugkosten zu einer Bündelung der Verkehre auf der in der Abbildung zentralen Kante (3, 4). Die Folge der Vernachlässigung von Zugkosten bei Blockingplänen ist eine starke Verteilung der Transporte und der Umstellarbeiten. Diese Verteilung ist ein wichtiger struktureller Effekt, der bei entspre-

chender Modellierung in den Optimallösungen zu sehen ist und im Vergleich zu Leitwegeplänen zu einer Entlastung der zentralen Infrastruktur führt.

Eine Eigenschaft der Leitwegepläne ist hingegen die Konzentration der Umstellarbeiten auf wenige ausgewählte Knoten, nämlich die Rangierbahnhöfe. Die Zentralisierung der Umstellarbeiten auf wenige Bahnhöfe war ein Teilziel der Konsolidierungsbemühungen im EWV im Rahmen des MORA-C-Programms [NP09] (vgl. Abschnitt 2.2.1). Eine Abkehr vom wenig flexiblen Leitwegesystem kann sich jedoch speziell im Hinblick auf die stark belasteten Rangierbahnhöfe und Hauptkorridore des Schienennetzes in Zukunft als notwendig erweisen und motiviert die Anwendung der Blockinglogik [HH13].

Eine Synthese der jeweils positiven Eigenschaften der Blocking- und Leitwegepläne, d.h. eine gleichmäßige Verteilung der Umstellarbeiten und die gleichzeitige Bündelung von Zugverkehren, wäre wünschenswert. Die gleichmäßige Nutzung der Infrastruktur und die Transportbündelung bilden gemeinsam mit der Robustheit gegenüber Aufkommenschwankungen die zentralen strukturellen Anforderungen an die Modellierung der Wagenflusssteuerung, die in Kapitel 4 durchgeführt wird.

Softwaresysteme zur Erstellung von Umstellplänen

Die hohe wirtschaftliche Relevanz der Erstellung von Umstellplänen lässt vermuten, dass kommerzielle Softwarelösungen für dieses Einsatzfeld existieren. Für die Bedürfnisse nordamerikanischer EVU existieren verschiedene Softwarelösungen zur Erstellung von Blockingplänen. So bietet die Firma Innovative Scheduling das Produkt *Innovative Railroad Blocking Optimizer* (IRBO) an. Es basiert auf der Arbeit von Ahuja et al. [AJL07]. Die Unternehmensberatung Oliver Wyman [OD12] bietet die Software MultiRail an, die auf der Arbeit von Ireland et al. [ICF⁺04] beruht. Vergleichbare Produkte, die in Deutschland Anwendung finden, sind nicht bekannt. Zu nennen sind interne Implementierungen bei DB Mobility Logistics, welche Algorithmen aus der Arbeit von Homfeld [Hom12] aufgreifen. Eine kommerziell verfügbare Softwarelösung wird seitens der Deutschen Bahn nicht verwendet. Insbesondere ist bei der Deutschen Bahn kein System im Einsatz, das die genannten Robustheitsaspekte berücksichtigen kann. Eine Marktstudie des Fraunhofer Instituts für Materialfluss und Logistik [SK07] zum Thema Softwarelösungen für EVU nennt ebenfalls keine Softwareprodukte für Umstellpläne.

3.3 Grundlagen der Komplexitätstheorie

Die Komplexitätstheorie ist ein Teilgebiet der Informatik, das sich mit der Bestimmung der Mindestressourcen befasst, die zur algorithmischen Lösung eines Entscheidungsproblems notwendig sind. Diese Definition und die Ausführungen dieses Abschnitts stützen sich auf Wegener [Weg03]. Von größtem praktischen Interesse sind im Allgemeinen die Ressourcen Laufzeit und benötigter Speicher. Die Komplexitätstheorie abstrahiert jedoch die Komplexität auf die Anzahl notwendiger arithmetischer Operationen, um hardwareunabhängige Resultate zeigen zu können. Typische Erkenntnisse der Komplexitätstheorie machen Aussagen darüber, ob ein Problem mit einer gewissen Anzahl

arithmetischer Standardoperationen lösbar ist. Das Vorgehen bei der Lösung ist oft zweitrangig. Trotz dieser sehr theoretischen Sichtweise liefert die Komplexitätstheorie sehr wichtige Erkenntnisse darüber, welche Form von Lösungsansätzen überhaupt zielführend sein können, um ein spezifisches Problem zu lösen, und welche auszuschließen sind. Es folgt daher eine kurze Einführung von Begriffen und Schreibweisen aus der Komplexitätstheorie.

Definition 7 (Polynomielle Laufzeit)

Es sei \mathcal{Q} ein Problem, das algorithmisch zu lösen ist, und $t(\mathcal{Q})$ die Laufzeit eines Algorithmus, der \mathcal{Q} korrekt löst. Dazu sei n ein Maß für die Größe einer Instanz von \mathcal{Q} . Der Algorithmus \mathcal{A} hat eine polynomielle Laufzeit, wenn ein Polynom p existiert, so dass für jede Instanzgröße n von \mathcal{Q} gilt:

$$t(\mathcal{Q}) \leq p(n).$$

Man sagt, ein Algorithmus hat eine Laufzeit von $\mathcal{O}(n^k)$, wenn eine Konstante c existiert, so dass gilt:

$$t(\mathcal{Q}) \leq cn^k.$$

Das Problem \mathcal{Q} gehört zur Klasse \mathcal{P} der polynomiell lösbaren Probleme, wenn ein Algorithmus mit polynomieller Laufzeit existiert, der \mathcal{Q} löst. Die Darstellung $\mathcal{O}(n^k)$ wird als Landausymbolik bezeichnet.

Die Problemklasse \mathcal{P} umfasst effizient, d.h. polynomiell, lösbare Probleme, obwohl hohe Koeffizienten k möglich und sehr lange Laufzeiten somit nicht ausgeschlossen sind. Eine bisher unbewiesene Vermutung der Komplexitätstheorie besagt, dass es eine mit \mathcal{P} nicht identische Menge von Problemen \mathcal{NP} gibt, für die lediglich die Überprüfung einer Lösung in Polynomialzeit möglich ist. Diese Vermutung, die meist kurz mit $\mathcal{P} \neq \mathcal{NP}$ bezeichnet wird, gilt als eines der zentralen Probleme der Informatik. Viele Aussagen der Komplexitätstheorie sind von der Form "Falls $\mathcal{P} \neq \mathcal{NP}$ gilt, so gilt auch...". In dieser Arbeit wird die allgemein übliche Hypothese unterstellt, dass $\mathcal{P} \neq \mathcal{NP}$ gilt [Weg03]. Des Weiteren existieren die sogenannten \mathcal{NP} -schweren Probleme, deren Komplexität derart ist, dass die Existenz eines polynomiellen Lösungsalgorithmus eines \mathcal{NP} -schweren Problems die polynomielle Lösbarkeit aller Probleme der Klasse \mathcal{NP} impliziert.

Die Schnittmenge der \mathcal{NP} -schweren Probleme und der Menge \mathcal{NP} wird als Menge \mathcal{NPC} der \mathcal{NP} -vollständigen Probleme bezeichnet. Die Probleme in \mathcal{NPC} haben drei Gemeinsamkeiten: Sie sind erstens alle auf einfache Weise, d.h. in polynomieller Zeit, aufeinander zurückzuführen und bilden somit eine Äquivalenzklasse innerhalb der \mathcal{NP} -schweren Probleme. Zweitens ist eine Lösung eines \mathcal{NP} -vollständigen Problems auf Grund der Zugehörigkeit zur Menge \mathcal{NP} in Polynomialzeit auf Zulässigkeit prüfbar. Drittens ist für kein Problem in \mathcal{NPC} ein polynomieller Algorithmus bekannt, da die Probleme \mathcal{NP} -schwer sind. Da viele bekannte und wohl erforschte Probleme nachweislich in \mathcal{NPC} liegen, gelten sie als besonders schwierig, weil die Lösung eines der vielen Probleme zu einer Lösung für alle anderen geführt hätte. Zu den prominentesten Beispielen \mathcal{NP} -vollständiger Probleme zählen das *Travelling-Salesman-Problem* und das *Rucksackproblem* [Sch00].

Der Beweis, dass ein Optimierungsproblem \mathcal{NP} -vollständig ist, zeigt, dass die exakte Lösung und zeitliche Effizienz nicht gleichzeitig für alle Instanzen erreichbar sind. Dabei bedeutet Exaktheit, dass die bestmögliche Lösung gefunden und darüber hinaus bewiesen wird, dass es sich um die bestmögliche Lösung handelt. Es existieren zwei übliche Methoden, dieser Schwierigkeit zu begegnen. In beiden Fällen wird partiell auf Effizienz oder Exaktheit verzichtet:

1. *Exakte, nicht polynomielle Verfahren:* Obwohl es für \mathcal{NP} -schwere Probleme nicht möglich ist, Verfahren zu entwickeln, die exakt und effizient sind, ist es möglich, Verfahren zu entwickeln, die nicht für alle Instanzen, wohl aber für Instanzen von praktischer Relevanz in angemessener Zeit optimale Lösungen ermitteln. So kann bspw. für gewisse Probleme eine exponentielle Laufzeit für Instanzen kleiner und mittlerer Größe akzeptabel sein. Zu den exakten Verfahren gehören *Branch-and-Bound* sowie *Branch-and-Price*, die im Folgenden erläutert werden.
2. *Heuristische Verfahren:* Bei der Anwendung heuristischer Verfahren wird auf eine beweisbar beste Lösung verzichtet. Heuristische Verfahren haben sich in der Vergangenheit bei der Lösung vieler schwerer Optimierungsprobleme als hilfreich erwiesen. Weitere Erläuterungen zu (Meta-)Heuristiken folgen in Abschnitt 3.4.3.

Komplexität ausgewählter Optimierungsprobleme

Im Folgenden werden drei wichtige Optimierungsprobleme definiert und in ihre Komplexitätsklassen eingeordnet.

Definition 8 (Lineares Optimierungsproblem)

Es seien $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ und $A \in \mathbb{R}^{m \times n}$

$$\begin{aligned} (LP) \quad & \min_x cx \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

Lineare Optimierungsprobleme, auch *Lineare Programme* (LP) genannt, sind dadurch charakterisiert, dass sowohl ihre Zielfunktion als auch ihre Nebenbedingungen durch lineare Funktionen gegeben sind. Die Nebenbedingungen werden in der obigen Darstellung durch die Matrix A ausgedrückt. Die linearen Strukturen erlauben die schnelle und exakte Lösung des Problems, wie der folgende Satz zeigt. Der Beweis für diesen Satz wurde 1984 von Karmarkar [Kar84] durch Einführung der Innere-Punkte-Verfahren erbracht.

Satz 1 (Komplexität von LP)

Lineare Programme sind in Polynomialzeit lösbar. Somit gilt $(LP) \in \mathcal{P}$.

Eine weitere Klasse von Optimierungsproblemen stellen die ganzzahligen linearen Optimierungsprobleme (engl.: *integer linear program* (ILP)) dar. Wie auch beim LP sind Nebenbedingungen und Zielfunktion von linearer Gestalt. Das ILP zeichnet sich dadurch aus, dass ein zulässiger Lösungsvektor ganzzahlig sein muss. Ein Vektor wird ganzzahlig genannt, wenn jede seiner Komponenten ganzzahlig ist.

Definition 9 (Ganzzahliges Lineares Optimierungsproblem)

Es seien $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und $A \in \mathbb{R}^{m \times n}$.

$$\begin{aligned} (ILP) \quad & \min_x cx \\ \text{s.t.} \quad & Ax \leq b \\ & x \in \mathbb{Z}^n \end{aligned}$$

Die Ganzzahligkeitsbedingungen machen aus dem in Polynomialzeit lösbaren (LP) ein \mathcal{NP} -vollständiges Problem [GJ97]:

Satz 2 (Komplexität von ILP)

Ganzzahlige lineare Optimierungsprobleme sind NP-vollständig. Es gilt daher $(ILP) \in \mathcal{NPC}$.

Ein Spezialfall des allgemeinen ganzzahligen Optimierungsproblems ist das ganzzahlige Mehrgüterflussproblem (engl.: *integer multicommodity flow problem* (IMCFP)):

Definition 10 (Ganzzahliges Mehrgüterflussproblem)

Es sei N eine Menge von Knoten, die durch Kanten $(i, j) \in \mathcal{N} \times \mathcal{N}$ verbunden sind, \mathcal{K} eine Menge von Gütern, die jeweils eine Quelle und eine Senke haben. Zu jeder Kante sei eine maximale Kapazität u_{ij} gegeben.

$$\begin{aligned} (IMCFP) \quad & \min \sum_{(i,j) \in \mathcal{A}, k \in \mathcal{K}} x_{ij}^k \\ \text{s.t.} \quad & \sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} \quad \forall (i, j) \in \mathcal{A} \\ & \sum_{i \in \mathcal{N}} x_{ji}^k - \sum_{i \in \mathcal{N}} x_{ij}^k = B(k, j) \quad \forall k \in \mathcal{K} \quad \forall (i, j) \in \mathcal{A} \\ & x_{ij}^k \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall (i, j) \in \mathcal{A} \end{aligned}$$

$$B(k, j) = \begin{cases} 1, & \text{falls } i \text{ Quelle von Gut } k \text{ ist} \\ -1, & \text{falls } i \text{ Senke von Gut } k \text{ ist} \\ 0, & \text{sonst} \end{cases}$$

Das ganzzahlige Mehrgüterflussproblem ist hilfreich, um eine Reihe verschiedener Transportprobleme zu modellieren. Die Ganzzahligkeit der Variablen korrespondiert im Falle von Transportproblemen mit der Untrennbarkeit der Güter. So stellt z.B. ein Container eine Einheit dar, die nicht in kleineren Untereinheiten transportiert werden kann. Für Transportnetzwerke, in denen die Trennung möglich ist, wie z.B. in Wasser- oder Datennetzen, ist das Mehrgüterflussproblem effizient als LP lösbar. Das ganzzahlige Mehrgüterflussproblem hingegen gehört laut Ahuja et al. [AMO93] zu den schwersten gängigen Optimierungsproblemen. Der folgende Satz untermauert diese Aussage [GJ97]:

Satz 3 (Komplexität von IMCFP)

Das Entscheidungsproblem, ob das ganzzahlige Mehrgüterflussproblem (IMCFP) eine zulässige Lösung besitzt, ist bereits für zwei Güter \mathcal{NP} -vollständig.

Ein ebenfalls häufig im Zusammenhang mit Transportproblemen vorkommendes Problem ist das *Kürzeste-Wege-Problem*:

Definition 11 (Kürzeste-Wege-Problem)

Gegeben sei ein Graph $\mathcal{G} = (\mathcal{N}, \mathcal{V})$ mit nicht-negativen Kantenlängen (Gewichten) c_{ij} sowie zwei Knoten $s \neq t$. Gesucht ist ein Kantenzug minimaler Länge, der s und t verbindet.

Das kürzeste Wege Problem ist in Polynomialzeit lösbar. Wird das Problem um eine Nebenbedingung erweitert, welche die Anzahl der Kanten q beschränkt, die der kürzeste Weg beinhalten darf, bleibt es effizient lösbar [CLRS10]:

Satz 4 (Komplexität des Kürzeste-Wege-Problems)

Das Kürzeste-Wege-Problem wird vom Algorithmus von Dijkstra mit Laufzeit $\mathcal{O}(|\mathcal{N}|^2)$ exakt gelöst. Der Algorithmus von Bellman-Ford löst das Kürzeste Wege Problem mit einer Kantenzahlbegrenzung von q in $\mathcal{O}(|\mathcal{N}|^3)$. Er bestimmt nicht nur den kürzesten Weg mit höchstens q Kanten, sondern auch die kürzesten Wege, die jeweils höchstens $1, 2, \dots, q - 1$ Kanten benutzen.

3.4 Grundlagen der mathematischen Optimierung

3.4.1 Lösungsverfahren für lineare Optimierungsprobleme

In diesem Abschnitt werden wichtige Eigenschaften linearer Programme zusammengetragen, die die Basis der später betrachteten Column-Generation-Verfahren bilden. Es wird zunächst der Begriff des Polyeders eingeführt.

Definition 12 (Polyeder)

Eine Menge $\mathcal{P} \subset \mathbb{R}^n$ heißt Polyeder, falls eine Matrix $A \in \mathbb{R}^{m \times n}$ und ein Vektor $b \in \mathbb{R}^m$ existieren, für die gilt

$$\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}. \quad (3.4)$$

Ein beschränktes Polyeder wird Polytop genannt.

Anschaulich ist ein Polyeder eine Geometrie, die von mehreren Ebenen begrenzt wird, wie in Abbildung 3.4 beispielhaft dargestellt. Der Bereich der zulässigen Lösungen eines LP ist ein Polyeder [KV12]:

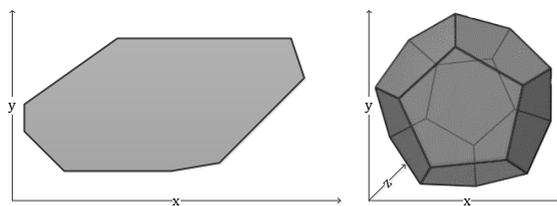


Abbildung 3.4: Beispiele beschränkter Polyeder in 2 und 3 Dimensionen

Satz 5 (Charakterisierung von Optimallösungen linearer Programme)

Die Lösungsmenge eines linearen Programmes ist ein Polyeder. Wenn ein LP Optimallösungen besitzt, dann befindet sich mindestens eine von ihnen in einer Ecke des Lösungspolyeders. Die Lösungsmenge eines beschränkten LP ist ein Polytop.

Der von Dantzig entwickelte Simplex-Algorithmus macht sich die Eigenschaft, dass jedes lösbare LP eine Ecklösung hat, zu Nutze. Der Algorithmus wird häufig mit der Allegorie "Springe von Ecke zu Ecke, bis keine Verbesserung mehr möglich ist!" beschrieben. Die Ecken eines Polyeders können durch sogenannte Basislösungen charakterisiert werden. Neue Basislösungen können durch einfache Transformationen bekannter Basislösungen effizient berechnet werden. Zur Prüfung der Optimalität einer Basislösung wird auf Dualitätstechniken zurückgegriffen, die im Folgenden eingeführt werden.

Definition 13 (Duale Programme)

Zu jedem beschränkten linearen Programm der zuvor definierten Form (LP) – in diesem Zusammenhang als primales Programm bezeichnet – existiert ein duales lineares Programm:

$$\begin{aligned} (LP_{dual}) \quad & \max_{\pi} \pi b \\ \text{s.t.} \quad & \pi A \leq c \\ & \pi \geq 0 \end{aligned}$$

Jeder primalen Ungleichung entspricht eine duale Variable und jeder primalen Variable eine duale Ungleichung. Die beiden folgenden Sätze zeigen, dass Dualität kein rein theoretisches Konstrukt ist, sondern eine enge Verbindung zwischen primalem und dualem Programm besteht [KV12]:

Theorem 1 (Schwacher Dualitätssatz)

Falls das primale und das duale Programm zulässige Lösungen x und π besitzen, gilt:

$$cx \leq \pi b \tag{3.5}$$

Theorem 2 (Starker Dualitätssatz)

Hat (LP) eine endliche Optimallösung x^* , so existiert eine Optimallösung π^* von (LP_{dual}) , und die optimalen Zielfunktionswerte stimmen überein. Es gilt also:

$$cx^* = \pi^* b \tag{3.6}$$

Das Prinzip der Dualität kann genutzt werden, um die Optimalität einer Basislösung eines LPs zu überprüfen. Der *revidierte Simplexalgorithmus* betrachtet nur eine Teilmenge der Variablen des Gesamtproblems. Dieses Vorgehen ist sinnvoll, weil in der Regel nur ein kleiner Anteil der Variablen in der Optimallösung von Bedeutung sind, d.h. nicht den Wert Null haben. Die Identifikation relevanter Variablen ist auf Grund der *reduzierten Kosten* (vgl. Definition 14) möglich, wie Satz 6 und Korollar 1 zeigen.

Definition 14 (Reduzierte Kosten)

Es sei \underline{x} eine zulässige Lösung von (LP) und $\underline{\pi}$ eine zulässige Lösung des zugehörigen dualen Programms (LP_{dual}) . Die reduzierten Kosten r_i der Variable \underline{x}_i werden wie folgt definiert:

$$r_i = c_i - \underline{\pi}A_i. \quad (3.7)$$

Die reduzierten Kosten können mit Hilfe der *Komplementärschlupfbedingung* zur Charakterisierung optimaler Lösungen eines LP eingesetzt werden [KV12]:

Satz 6 (Komplementärschlupfbedingung)

Es sei x^* eine zulässige Lösung von (LP) und π^* eine zulässige Lösung von (LP_{dual}) . Folgende Aussagen sind äquivalent:

1. x^* und π^* sind Optimallösungen von (LP) bzw. (LP_{dual}) .
2. Die Zielfunktionswerte cx^* der primalen und π^*b der dualen Lösung stimmen überein.
3. $(c - \pi^*A)x^* = 0$ und $\pi^*(b - Ax^*) = 0$

Aus der Komplementärschlupfbedingung kann folgende Aussage abgeleitet werden:

Korollar 1 (Reduzierte Kosten zur Identifikation im Optimum relevanter Variablen)

Es sei x^* Optimallösung von (LP) und π^* Optimallösung von (LP_{dual}) . Alle Komponenten der primalen Optimallösung x^* , die reduzierte Kosten $r_i > 0$ haben, haben den Wert 0 und somit keinen Einfluss auf den Zielfunktionswert.

Beweis: Da x^*, π^* Optimallösungen der jeweiligen Probleme sind, gilt laut Satz 6

$$(c - \pi^*A)x^* = 0. \quad (3.8)$$

Aufgrund von Definition 14 ist dies äquivalent zu $r_i x_i^* = 0$. Somit muss mindestens einer der beiden Faktoren 0 sein. Daraus folgt, dass $x_i^* = 0$, falls $r_i = c - \pi^*A \neq 0$. Damit ist das Korollar bereits bewiesen. Der Fall $r_i < 0$ kann im Optimum nicht gelten, da π^* optimal und somit zulässig ist in (LP_{dual}) . Damit gilt die im Modell enthaltene Nebenbedingung:

$$\pi A \leq c \Leftrightarrow r_i \geq 0 \quad (3.9)$$

□

Die zuvor genannten Sätze bilden die theoretische Grundlage für die Column-Generation-Verfahren, die in Kapitel 5 eingeführt werden.

3.4.2 Exakte Verfahren der diskreten Optimierung

Branch-and-Bound-Verfahren

Die Lösung allgemeiner ganzzahliger Optimierungsprobleme ist \mathcal{NP} -schwer [Sch00]. Eine Standardmethode zur Lösung von ILP ist das *Branch-and-Bound-Verfahren*. Es basiert auf einer Zerteilung des Lösungsraumes. Die Teilmengen des Lösungsraumes

werden mit Hilfe sogenannter Relaxierungstechniken sukzessive daraufhin untersucht, ob die gesuchte Optimallösung in der betrachteten Teilmenge liegen kann. *Relaxierung* bezeichnet dabei die Vereinfachung eines Problems. Die Vernachlässigung der Ganzzahligkeitsbedingungen eines ILP führt zu einem LP, welches mit den Methoden aus Abschnitt 3.4.1 gelöst werden kann. Man spricht in diesem Fall auch von *linearer Relaxierung*. Die Lösung des relaxierten Problems liefert eine sogenannte *untere Schranke* (engl.: *lower bound* (LB)) für den Zielfunktionswert des ILP, d.h. der optimale Zielfunktionswert des ILP ist größer als die ermittelte Schranke LB. Diese Information kann auf zweierlei Weise benutzt werden: Erstens kann die untere Schranke zur Abschätzung der Qualität eines zulässigen Punktes \bar{x} des ILP verwendet werden. Die Differenz $Z(\bar{x}) - LB$ wird als *Optimalitätslücke* (engl.: *Gap*) bezeichnet. Sie ist ein Maß für die theoretisch mögliche Verbesserung der bereits bekannten Lösung \bar{x} . Die untere Schranke wird zweitens im Branch-and-Bound-Verfahren algorithmisch genutzt. Für das betrachtete ILP sei eine zulässige Lösung \bar{x} gegeben. Wird nun eine Teilmenge der ILP-Lösungsmenge gefunden, über der die LP-Relaxierung einen höheren Wert hat als der Zielfunktionswert $Z(\bar{x})$ der bekannten Lösung, kann ausgeschlossen werden, dass die Optimallösung in dieser Teilmenge liegt. Der Branch-and-Bound-Algorithmus teilt das Grundproblem sukzessive in kleinere Teilprobleme auf und identifiziert dabei Teilmengen, in denen die Optimallösung nicht liegen kann. Auf diese Weise ist der Algorithmus in der Lage, Teilmengen der Lösungsmenge auszuschließen, ohne die darin enthaltenen Punkte explizit zu enumerieren. Der Branch-and-Bound-Prozess kann in Form eines Baumes (vgl. Abbildung 3.5) visualisiert werden. In jedem Knoten des Baumes ist ein LP zu lösen, um eine untere Schranke für die Zielfunktionswerte in jenem Teilbaum zu bestimmen, dessen Wurzel der betrachtete Knoten darstellt.

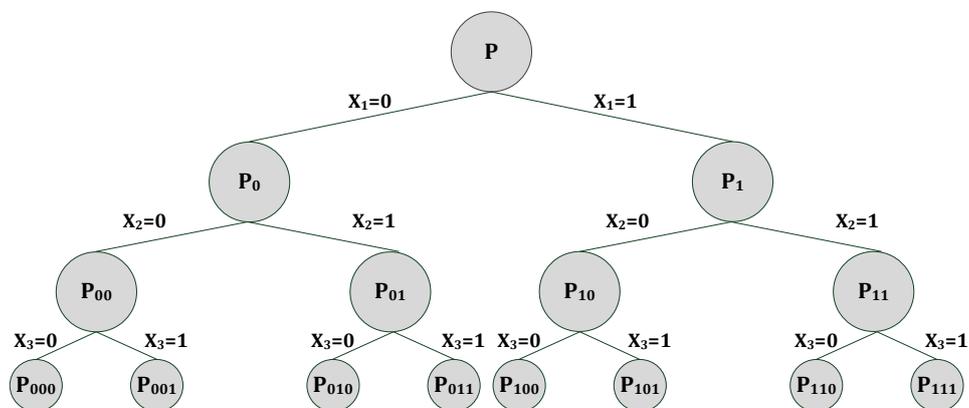


Abbildung 3.5: Branch-and-Bound-Baum

Aus der Theorie der ganzzahligen Optimierung ist bekannt, dass der Branch-and-Bound-Algorithmus endlich ist, wenn der Lösungsraum beschränkt ist [Sch00]. Die Optimierungsprobleme, die in dieser Arbeit behandelt werden, sind beschränkt. Daher stellt

Branch-and-Bound ein mögliches Lösungsverfahren für die betrachteten Planungsaufgaben dar.

Schnittebenenverfahren

Das *Schnittebenenverfahren* ist eine Erweiterung des zuvor beschriebenen Branch-and-Bound-Verfahrens. Die Konvergenzgeschwindigkeit des Branch-and-Bound-Algorithmus hängt sehr stark von der Qualität der unteren Schranken ab. Für viele reale Probleme sind die LP-Schranken sehr schwach, d.h. ihr Wert unterscheidet sich extrem vom Zielfunktionswert der besten erreichbaren ganzzahligen Lösung. Dies hat zur Folge, dass der Branch-and-Bound-Algorithmus viele Variablen fixieren muss, bevor ein Ast verworfen werden kann. Zur Stärkung der LP-Relaxierung wird beim Schnittebenenverfahren daher versucht, zusätzliche Nebenbedingungen einzufügen, die für das ILP, jedoch nicht für das LP redundant sind. Dadurch wird der Lösungsraum der LP-Relaxierung verkleinert. Der Lösungsraum des LP approximiert den Lösungsraum des ILP somit besser. Dieser Vorgang wird in Abbildung 3.6 visualisiert. Das Einfügen der Schnittebenen wird zu Beginn der Bearbeitung eines Teilproblems durchgeführt. Das Auffinden guter Schnittebenen ist jedoch ein algorithmisch hochkomplexes Problem. Neben problemspezifischen Schnittebenen gibt es eine Theorie allgemeiner Schnittebenen, die bereits in den 1950er Jahren von Gomory vorangetrieben wurde (vgl. z.B. [NW99]). Das Schnittebenenverfahren wird mit dem Branch-and-Bound-Verfahren kombiniert. Das resultierende Verfahren wird *Branch-and-Cut* genannt und bspw. im kommerziellen Solver CPLEX von IBM verwendet.

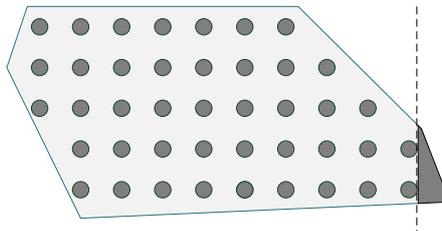


Abbildung 3.6: Visualisierung des Schnittebenenverfahrens

3.4.3 Metaheuristische Optimierungsverfahren

Heuristische Optimierungsverfahren garantieren im Gegensatz zu den zuvor beschriebenen Verfahren der exakten Optimierung nicht, dass sie die Optimallösung für jede Instanz eines Optimierungsproblems finden. Die Komplexitätstheorie zeigt für viele praktische Optimierungsprobleme, dass sie nicht in praktikabler Zeit exakt gelöst werden können, da sie \mathcal{NP} -vollständig sind. Daher ist der Verzicht auf die bewiesene Optimalität, d.h. der Einsatz heuristischer Verfahren, ein sinnvoller Weg zur Bearbeitung komplexer Optimierungsprobleme. Ungeachtet dessen werden (meta-)heuristische Verfahren von der wissenschaftlichen Welt häufig mit Argwohn betrachtet. Diese und die weiteren Ausführungen dieses Abschnitts basieren auf den Arbeiten von Sörensen [SG10, Sör13].

Bezeichnung 1 (Metaheuristik)

Der Begriff *Metaheuristik* wird auf zweierlei Weise verwendet:

1. Eine Metaheuristik ist ein hochgradig problemunabhängiger Rahmen aus Vorschriften und Strategien zur Entwicklung heuristischer Optimierungsalgorithmen.
2. Der Begriff bezeichnet auch die problemspezifische Implementierung einer Heuristik entsprechend diesen Vorschriften und Strategien verwendet.

Eine Metaheuristik im Sinne der Definition als Handlungsrahmen ist somit kein Algorithmus, d.h. keine Abfolge konkreter Schritte, sondern eher eine Sammlung von Konzepten. Metaheuristiken lassen sich sinnvoll in drei Kategorien einteilen:

- *Konstruktive Metaheuristiken* bauen, ihrer Nomenklatur entsprechend, zulässige Lösungen auf. Dies passiert häufig durch stückweises Zusammensetzen von Teillösungen, wie es einfache Greedyheuristiken tun. Durch Hinzunahme stochastischer Effekte entstehen sogenannte GRASP-Heuristiken [FR95], die eine bedeutende Entwicklung in der Disziplin der Metaheuristiken darstellen. Auch die auf der Simulation von Ameisenkolonien basierenden Ameisenalgorithmen von Dorigo [SD10] sind ein Beispiel für konstruktive Metaheuristiken.
- *Populationsbasierte Metaheuristiken* basieren auf der Auswahl und Kombination von Lösungen aus einem aktuell betrachteten Lösungspool, der auch *Population* bezeichnet wird. Sprachliche Analogien zu antropologischen und biologischen Begriffen sind bei der Gattung der *evolutionären Algorithmen* besonders häufig zu finden. Biologische Prinzipien und Metaphern dienten in der Vergangenheit oft als Inspiration für neue Verfahren. Dieses Vorgehen führte jedoch auch dazu, dass Metaphern nicht nur als Inspiration, sondern auch als Legitimation für teils absurde Verfahren verwendet wurden, wie Sörensen anmerkt [Sör13].
- *Lokale-Suche-Verfahren* sind Metaheuristiken, die eine bekannte Lösung verbessern sollen. Dazu wird in sogenannten Nachbarschaften einer gegebenen zulässigen Momentanlösung gesucht. Nachbarschaften bestehen in der Regel aus einer Menge potentieller Lösungen, die durch Änderung einzelner Eigenschaften aus der Momentanlösung hervorgehen. Bekannte Spezialfälle der lokalen Suche sind das *Simulated Annealing* [KGV83], das von der Bewegungsfreiheit von Molekülen beim Abkühlen einer Metallstruktur inspiriert ist, und die *Tabusuche* [SG10], die wiederholte Überprüfungen einzelner Lösungen zu umgehen versucht, indem sie Teile des zurückliegenden Suchverlaufs in die weitere Suche miteinbezieht. Beide Ansätze verfolgen das Ziel, die Konvergenz des Verfahrens gegen ein lokales, aber nicht globales Optimum zu verhindern. Abhängig von der gewählten Nachbarschaft bleibt die Momentanlösung auch nach einer Iteration die beste bekannte Lösung. Dies kann zur Stagnation des Verfahrens führen. Das Ziel ist es daher stets, lokale Optima wieder verlassen zu können.

Dies geschieht im Fall der *Variable Neighborhood Search* (VNS) durch systematische Änderungen der Nachbarschaftsstruktur [PR10]. Im Gegensatz zur VNS wird

bei der *Variable Depth Neighborhood Search* (VDNS) die Struktur der Nachbarschaft beibehalten und nur die Größe der Nachbarschaft vergrößert. In Abbildung 3.7 werden die Konzepte der VNS und VDNS konzeptionell gegenüber gestellt.

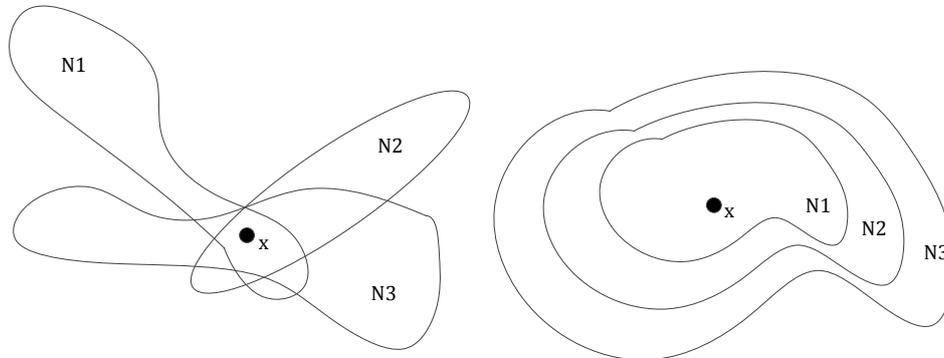


Abbildung 3.7: links VNS, rechts VDNS mit je drei Nachbarschaften N_1, N_2, N_3 um eine Momentanlösung x (angelehnt an [PR10])

Die *Very Large Scale Neighborhood Search* (VLSN) wählt extrem große Nachbarschaften und durchsucht diese mit speziellen Techniken, die den jeweiligen Nachbarschaftsstrukturen angepasst sind. Für die Anwendung ist eine spezielle Modellstruktur erforderlich [AEOP02]. Alle lokalen Suchverfahren sind mit dem Problem lokaler Optima konfrontiert. Das bedeutet, dass eine Lösung, die bzgl. der gewählten Nachbarschaften nicht mehr verbessert werden kann, nicht das globale Optimum sein muss. Daher spricht man davon, dass das Verfahren lokalen Optima entkommen muss.

Eine von den klassischen Metaheuristiken abweichende Klasse von Algorithmen, die sogenannten *Matheuristiken* (engl.: *matheuristics*), hat in den vergangenen Jahren verstärkte Aufmerksamkeit in der wissenschaftlichen Literatur gefunden [SG10].

Definition 15 (Matheuristiken)

Matheuristiken sind hybride Optimierungsverfahren, die Aspekte metaheuristischer Ansätze mit Methoden der exakten Optimierung kombinieren.

Diese sehr allgemein gehaltene Definition trifft auf eine große Menge von Optimierungsalgorithmen zu. Matheuristiken sollen die Vorteile von exakten und metaheuristischen Ansätzen verbinden. Eine Reihe erfolgreicher Beispiele findet sich bei Maniezzo et al. [MSV09]. Konkrete Verfahren werden an späterer Stelle eingeführt, wenn sie zur Darstellung des Vorgehens benötigt werden.

4 Entwicklung und Analyse eines Modells für das Wagenroutenproblem

Das in Abschnitt 2.3.2 beschriebene Planungsproblem wird in diesem Kapitel mathematisch modelliert, um eine Bearbeitung der Probleme mittels mathematischer Methoden zu ermöglichen. Es folgen eine Komplexitätstheoretische Analyse der Modelle sowie Betrachtungen zu unteren Schranken und vereinfachenden Annahmen im Unterkapitel 4.2.

4.1 Entwicklung von Modellen zum Wagenroutenproblem

In diesem Unterkapitel findet die mathematische Modellierung auf Basis der Systembeschreibung in Abschnitt 2.3.2 statt. Dazu werden zunächst die Ziele und relevanten Parameter bestimmt. Anschließend werden in den Abschnitten 4.1.2 und 4.1.3 zwei äquivalente mathematische Formulierungen hergeleitet.

4.1.1 Eingabeparameter der Modelle

Die notwendigen Parameter, die Einfluss auf die Wagenrouten haben, können in zwei Gruppen unterteilt werden. Zum einen müssen Informationen über das zugrunde liegende physische Netzwerk bekannt sein. Zum anderen müssen Daten über das zu transportierende Wagenaufkommen vorliegen. Die Parameter, die das Netzwerk beschreiben, können ihrerseits in Knoten- und Kantenparameter unterteilt werden.

Knotenparameter

Das physikalische Netzwerk besteht aus der Menge der Umstellbahnhöfe \mathcal{N} , d.h. aus den SmR, Kbf und Rbf. Die Bezeichnungen entstammen der Hierarchie des heutigen Leitwegesystems (vgl. Abschnitt 3.2.2), welche im hier entwickelten Modell jedoch keine Bedeutung mehr hat. Für jeden Knoten müssen Informationen über die Rangierkapazität vorliegen. Dazu sei die Zugbildungskapazität, parametrisiert als maximale Anzahl von Zügen, die dort pro Tag gebildet werden können, gegeben. Die Zugbildungskapazität R_i eines Knotens $i \in N$ wird in erster Linie von der Anzahl verfügbarer Richtungsgleise in der Ausfahrgruppe und der Rangiertechnik bestimmt. Die Rangiertechnik definiert dabei für gewöhnlich die durchschnittliche Bildungszeit eines Zuges und somit die durchschnittliche Anzahl der Züge, die an einem Tag auf einem Richtungsgleis gebildet werden können. Des Weiteren sei die *Ablaufbergkapazität* B_i gegeben, d.h. die maximale Anzahl an Wagen, die an einem Tag in einem Knoten umgestellt werden können. Obwohl kleinere Umstellbahnhöfe keinen Ablaufberg besitzen, sondern mit Rangierlokomotiven arbeiten, ist der Begriff Ablaufbergkapazität gängig. Zusätzlich seien die Kosten U_i für die

Durchführung eines Umstellvorgangs für einen einzelnen Wagen in jeder Anlage $i \in N$ gegeben.

Kantenparameter

Die Kanten \mathcal{A} des Netzwerkes bilden die physischen Schienenverbindungen zwischen den Bahnhöfen ab. Das Netzwerk ist im Sinne der Graphentheorie vollständig, d.h. es gibt für jedes Paar von Knoten i, j eine Verbindung $e = (i, j)$. Die Verbindungen können in dem Sinne indirekt sein, dass das Durchfahren (ohne Halt) oder Vorbeifahren an einem anderen Knoten nötig sein kann, um ein Ziel zu erreichen. Basierend auf den Betriebskosten eines Zuges sei für jedes Knotenpaar $(i, j) \in \mathcal{A}$ der Kostensatz c_{ij} für die Durchführung einer Direktzugverbindung bekannt. Ferner müssen die infrastrukturell und gesetzlich bedingten Obergrenzen für die Länge L_{ij} und das Gewicht W_{ij} eines Wagenzuges auf jeder Kante $(i, j) \in \mathcal{A}$ bekannt sein. Besteht eine Kante $(i, j) \in \mathcal{A}$ aus mehreren physikalischen Schienenwegen, wird das Minimum der jeweiligen Obergrenzen gewählt.

Aufkommensdaten

Zur Bestimmung effizienter Wagenrouten ist Wissen über das Aufkommen zu transportierender Wagen auf allen Relationen von Nöten. Im hier vorliegenden Problem halten drei Parameter Einzug in die Modellierung. Zu berücksichtigen ist die Anzahl aller Wagen v_d^k der Relation $k \in \mathcal{K}$, die an einem Tag $d \in \mathcal{D}$ auf einer Relation zu transportieren sind, und die aggregierten Längen l_d^k und Gewichte w_d^k dieser Wagen. Das Tupel dieser Informationen aller Relationen $Q_d = \{(v_d^k, l_d^k, w_d^k) \mid \text{Relation } k \text{ hat positives Aufkommen}\}$ wird als *Aufkommensmatrix* bezeichnet. Historische Aufkommensmatrizen sind für zurückliegende Zeiträume bekannt, für künftige taktische Planungszeiträume jedoch nicht exakt prognostizierbar. Der Grund dafür liegt in der dynamischen Auftragsannahme im EWW, die in Abschnitt 2.2.2 thematisiert wurde. Neben unvollständigen Informationen über das Wagenaufkommen konkreter Tage, sprechen mehrere praktische Argumente gegen eine tägliche Änderung der Umstellpläne:

1. Eine tägliche Änderung der Umstellpläne würde die Frage aufwerfen, was mit Wagen zu tun ist, die nicht innerhalb einer Planungsperiode von ihrer Quelle zum Ziel gebracht werden. Da die Mehrheit der Wagen nicht innerhalb von 24 Stunden ihr Ziel erreicht, ist ein nicht zyklischer Plan somit kaum praktikabel.
2. Trotz der fortschreitenden Automatisierung der Umstellbahnhöfe werden große Teile der Umstellungen manuell koordiniert. Die Fehleranfälligkeit stiege durch stetig wechselnde Routen und daraus resultierende Zugzusammenstellungen.
3. Der Planung der Umstellpläne folgen, wie in Abschnitt 2.2.3 dargestellt wurde, eine Reihe weiterer komplexer Planungsschritte, die einer robusten, d.h. insbesondere nicht täglich wechselnden, Planungsgrundlage bedürfen.
4. Zur Nutzung konkreter Trassen ist die Absprache mit dem EIU notwendig. Ständig wechselnde Routen erhöhen nicht nur den Verwaltungsaufwand, sondern auch die Kosten, da mehr Adhoc-Trassen benötigt werden (vgl. Abschnitt 2.3.1).

Die angeführten Argumente rechtfertigen die bisherige Praxis, die Umstellpläne saisonal zu aktualisieren [NP09]. Aus planerischer Sicht stellt sich jedoch die Frage, wie das Wagenaufkommen für einen längeren Zeitraum sinnvoll abgebildet werden kann. Mit Hilfe statistischer Verfahren ist eine tagesgenaue Mittelfristprognose, d.h. für einige Wochen, mit hinreichender Genauigkeit möglich. Die hohe Volatilität des täglichen Wagenaufkommens macht jedoch eine repräsentative Abbildung des Aufkommens in einer einzelnen Aufkommensmatrix schwer möglich. So kann es vorkommen, dass auf einer Relation an einem Tag der Woche Dutzende von Wagen zu transportieren sind, während an den übrigen Tagen der Woche überhaupt kein Aufkommen vorliegt. Die Abbildung dieses Umstandes in einer Matrix durch Mittelung des Aufkommens stellt keine sinnvolle Planungsgrundlage dar [CV11].

Aus den zuvor genannten Gründen wird kein Umstellplan bestimmt, der optimal für eine einzelne Matrix ist. Das Ziel ist es, einen Umstellplan zu erstellen, der optimal ist für eine Menge von Aufkommensmatrizen $Q_{\mathcal{D}}$. Die Menge \mathcal{D} kann entweder als eine Menge von Szenarien interpretiert werden oder als Zeitscheiben des Planungszeitraumes, z.B. Tage einer Woche. Die Matrizen seien statisch, d.h. vor und während der Optimierung vollständig bekannt und keinen Änderungen unterworfen.

Für jede Relation wird überdies eine Beschränkung der Umstellvorgänge vorgegeben. Dies ist äquivalent dazu, dass jeder Wagen einer Relation k über maximal S_k Kanten des Netzwerkes transportiert wird. Diese Restriktion wird als Ersatz für eine explizite Fahrzeitrestriktion aufgenommen, da keine empirischen Prozesszeiten für das hier zu entwickelnde Produktionsverfahren vorliegen.

4.1.2 Ein kantenbasiertes Modell

Es wird unter Zuhilfenahme der zuvor definierten Parameter ein Optimierungsproblem definiert, dessen Lösung einen optimalen Umstellplan im Sinne des in Abschnitt 2.3.2 beschriebenen Problems darstellt.

Entscheidungsvariablen

Das Modell soll der Bestimmung optimaler Umstellpläne dienen. Äquivalent dazu ist laut Unterkapitel 2.3 die Bestimmung der optimalen Wagenrouten. Es wird definiert, dass mit dem Begriff Relation im Folgenden die Menge der Wagen auf der betreffenden Relation gemeint ist. Es wird für jede Kombination einer Relation $k \in \mathcal{K}$ und einer Kante $(i, j) \in \mathcal{A}$ eine Variable eingeführt:

$$x_{ij}^k = \begin{cases} 1, & \text{wenn Relation } k \text{ über Kante } (i, j) \text{ transportiert wird} \\ 0, & \text{sonst} \end{cases} \quad (4.1)$$

Die Variablen x_{ij}^k werden im Folgenden als Flussvariablen bezeichnet. Um die systemweiten Transportkosten darstellen zu können, werden Variablen eingeführt, welche die Anzahl täglich zu transportierender Wagenzüge auf einer Kante $(i, j) \in \mathcal{A}$ bestimmen:

$$n_{ij} = \text{Anzahl Wagenzüge, die täglich über die Kante } (i, j) \text{ transportiert werden} \quad (4.2)$$

Die Variablen n_{ij} definieren das Design eines Netzwerkes, das dem zuvor definierten Blockingnetzwerk entspricht. Die Variablen n_{ij} werden daher Designvariablen genannt. Die Anzahl verschiedener Szenarien, d.h. Aufkommensmatrizen, aus Abschnitt 4.1.1 kann auch als Anzahl der Tage des Planungszeitraumes interpretiert werden. Es sei darauf hingewiesen, dass sowohl Fluss- als auch Designvariablen unter diesen Szenarien invariant sind, also über den gesamten Planungshorizont hinweg nicht verändert werden. Gesucht ist ein einziger Plan, der für alle Szenarien, d.h. für alle Aufkommensmatrizen Q_d , zulässig ist und die Gesamtkosten über alle Szenarien minimiert. Diese Modellierung ist äquivalent zum Begriff der absoluten Robustheit aus der robusten Optimierung, wie er von Kouvelis und Yu [KY97] definiert und bereits von Jin [Jin98] für Blockingpläne formuliert wurde.

Zielfunktion

Ziel der Optimierung ist die Minimierung der Gesamtkosten des Verkehrs zwischen den betrachteten Knoten des EWW. Diese Gesamtkosten setzen sich aus Transport- und Umstellkosten zusammen. Die Transportkosten des Einzelwagenverkehrs lassen sich am besten als Kosten je Zugkilometer abbilden, wie in Abschnitt 2.3.1 besprochen wurde. Daher wird der Zielfunktionsterm für die Transportkosten als Summe der Transportkosten für die Blocks aller Kanten des Netzwerkes gebildet und über alle Tage des Planungszeitraums aufsummiert.

$$\text{Transportkosten} = \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} c_{ij} n_{ij} \quad (4.3)$$

Die Umstellkosten werden wagenweise modelliert. Dazu wird jede Wagenumstellung eingepreist. Es wird angenommen, dass jeder Zug, der in einem Umstellbahnhof hält, vollständig in Einzelwagen zerlegt wird. Die Kosten einer Umstellung U_i werden abhängig vom Knoten $i \in \mathcal{N}$ berechnet, in dem die Umstellung stattfindet.

$$\text{Umstellkosten} = \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} \sum_{k \in \mathcal{K}} U_i v_d^k x_{ij}^k \quad (4.4)$$

Die Zielfunktion wird als Summe der zuvor genannten Kostenbestandteile definiert:

$$Z(x, n) = \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} [c_{ij} n_{ij} + \sum_{k \in \mathcal{K}} U_i v_d^k x_{ij}^k] \quad (4.5)$$

Nebenbedingungen

Der in Worten formulierten Problembeschreibung in Abschnitt 2.3.2 entsprechend werden die folgenden Nebenbedingungen in das Modell aufgenommen. Die Erfüllung aller Transportaufträge korrespondiert mathematisch mit der *Flusserhaltungsbedingung* (4.6) für jede Relation $k \in \mathcal{K}$ und jeden Knoten $j \in \mathcal{N}$.

$$\sum_{i \in \mathcal{N}} x_{ji}^k - \sum_{i \in \mathcal{N}} x_{ij}^k = B(k, j) \quad (4.6)$$

$$\text{mit } B(k, j) = \begin{cases} 1, & \text{wenn } j \text{ Quelle von Relation } k \text{ ist} \\ -1, & \text{wenn } j \text{ Senke von Relation } k \text{ ist} \\ 0, & \text{sonst} \end{cases}$$

Die linke Seite der Gleichung bestimmt die Differenz des ausgehenden Flusses und des eingehenden Flusses der jeweils betrachteten Relation in einem Knoten $j \in \mathcal{N}$. Diese Differenz muss in der Quelle 1 sein, weil der Fluss die Quelle verlassen muss. Analog muss die Differenz in der Senke der Relation -1 sein, damit der Fluss die Senke erreicht. In allen übrigen Knoten muss ein Flussgleichgewicht herrschen.

Die nächsten beiden Restriktionen betreffen die Kapazität eines Wagenzuges auf der Kante (i, j) . Dort müssen hinreichend viele Züge verkehren, damit alle Wagen, deren Route über diese Kante verläuft, transportiert werden können. Die Anzahl der Züge, die eine Kante $(i, j) \in \mathcal{A}$ befahren, muss immer so groß gewählt werden, dass ihr Produkt mit der maximalen Wagenzuglänge größer als die Summe der Längen aller Wagen ist, welche über die Kante zu transportieren sind. Die *Längenrestriktion* (4.7) muss an jedem Tag des Planungszeitraumes erfüllt sein, d.h. für jede Aufkommenskonfiguration aus $Q_{\mathcal{D}}$.

$$\sum_{k \in \mathcal{K}} l_d^k x_{ij}^k \leq L_{ij} n_{ij} \quad \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \quad (4.7)$$

Eine analoge Nebenbedingung wird auch für die *Gewichtsrestriktion* eines Wagenzuges eingeführt:

$$\sum_{k \in \mathcal{K}} w_d^k x_{ij}^k \leq W_{ij} n_{ij} \quad \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \quad (4.8)$$

Die Nebenbedingungen (4.7) und (4.8) stellen den Zusammenhang zwischen Fluss- und Designvariablen her. Wenn die zu transportierenden Wagen die jeweilige Maximallänge oder das Maximalgewicht eines Wagenzuges auf einer Kante $(i, j) \in \mathcal{A}$ überschreiten, muss ein weiterer Zug auf der Kante eingesetzt werden. Fluss- und Designvariablen werden gleichzeitig optimiert und ihre optimalen Variablenbelegungen bedingen einander. In einem Knoten können nicht unbegrenzt viele Züge umgestellt werden. Eine Größe, in der die Kapazität einer Anlage bemessen werden kann, ist die Anzahl ihrer Richtungsgleise (vgl. Unterkapitel 2.1.3). Die maximale Anzahl an Zügen, die an einem Tag in einer Anlage gebildet werden können, ist als Knotenparameter gegeben. Daraus wird die folgende *Rangiergleisrestriktion* abgeleitet:

$$\sum_{j \in \mathcal{N}} n_{ij} \leq R_i \quad \forall i \in \mathcal{N} \quad (4.9)$$

Ferner wird die Anzahl an Wagen beschränkt, die pro Tag umgestellt werden können. Die Restriktion wird als *Ablaufbergrestriktion* bezeichnet:

$$\sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k x_{ij}^k \leq B_i \quad \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \quad (4.10)$$

Die Anzahl der Umstellvorgänge jeder Relation ist beschränkt. Äquivalent ist auch eine Beschränkung der Anzahl der Netzwerkkanten S_k , die von einer Relation auf einem

zulässigen Weg benutzt werden dürfen:

$$\sum_{(i,j) \in \mathcal{A}} x_{ij}^k \leq S_k \quad \forall k \in \mathcal{K} \quad (4.11)$$

Die Definition der Entscheidungsvariablen beinhaltet bereits deren Ganzzahligkeit. Die genannten Nebenbedingungen ergeben zusammen mit der Zielfunktion $Z(x, n)$ das kantenbasierte Modell zur Optimierung robuster Umstellpläne. Es wird im Folgenden als (Opt_{arc}) bezeichnet.

$$\begin{aligned} (Opt_{arc}) \quad & \min_{x,n} \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} [c_{ij} n_{ij} + \sum_{k \in \mathcal{K}} U_i v_d^k x_{ij}^k] \\ \text{s.t.} \quad & \sum_{i \in \mathcal{N}} x_{ji}^k - \sum_{i \in \mathcal{N}} x_{ij}^k = B(k, j) && \forall k \in \mathcal{K} \quad \forall j \in \mathcal{N} \\ & \sum_{k \in \mathcal{K}} l_d^k x_{ij}^k \leq L_{ij} n_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\ & \sum_{k \in \mathcal{K}} w_d^k x_{ij}^k \leq W_{ij} n_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\ & \sum_{j \in \mathcal{N}} n_{ij} \leq R_i && \forall i \in \mathcal{N} \\ & \sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k x_{ij}^k \leq B_i && \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \\ & \sum_{(i,j) \in \mathcal{A}} x_{ij}^k \leq S_k && \forall k \in \mathcal{K} \\ & n_{ij} \in \mathbb{N} && \forall (i, j) \in \mathcal{A} \\ & x_{ij}^k \in \{0, 1\} && \forall k \in \mathcal{K} \quad \forall (i, j) \in \mathcal{A} \end{aligned}$$

Das Modell (Opt_{arc}) kann der Klasse der *kapazitierten Mehrgüternetzwerk Design Probleme mit Fixkosten* (engl.: *fixed charge multicommodity capacitated network design problem* (FMCNDP)) zugeordnet werden.

4.1.3 Ein pfadbasiertes Modell

Das Optimierungsmodell (Opt_{arc}) ist kantenbasiert. Es wird für jede Relation und Kante entschieden, ob die Kante im Pfad der Relation enthalten ist. Diese Formulierung wird oft als kompakte Formulierung bezeichnet. Demgegenüber stehen sogenannte extensive Formulierungen, die im Falle von Flussproblemen in *Quelle-Senke-Pfaden* (vgl. Unterkapitel 3.1) formuliert sind. Es ist möglich, alle zulässigen Quelle-Senke-Verbindungen einer Relation aufzuzählen. Für jede Relation $k \in \mathcal{K}$ ist genau ein Pfad aus der Menge der für

k zulässigen Pfade $\mathcal{P}(k)$ zu wählen, um eine Lösung zu generieren. Die zahlreichen Flussvariablen und Flusserhaltungsbedingungen (4.6) aus dem kantenbasierten müssen nicht mehr betrachtet werden, weil sie durch eine geeignete Definition der zulässigen Quelle-Senke-Pfaden stets erfüllt sind. Es wird daher zunächst der Begriff *Quelle-Senke-Pfad* formal definiert. Die Forderung aus dem kantenbasierten Modell, dass ein zulässiger Pfad nur eine beschränkte Menge Kante verwenden darf, wird in die Definition eines Pfades integriert.

Definition 16 (Quelle-Senke-Pfad)

Eine Kantenfolge $p=e_1, e_2, \dots, e_l$ heißt Quelle-Senke-Pfad oder einfach Pfad der Länge l für Relation $k \in \mathcal{K}$, wenn p ein kreisfreier Kantenzug ist, der Quelle und Senke von k verbindet. Es gelte immer $l \leq S_k$, d.h. jeder Pfad erfülle die Restriktion (4.11). Die Menge aller zulässigen Pfade der Relation k werde als $\mathcal{P}(k)$ bezeichnet.

Es besteht ein mathematischer Zusammenhang zwischen Flüssen über Kanten und Flüssen, die mittels Pfaden dargestellt sind. Zunächst werden Koeffizienten x_{ij}^{kp} definiert, die einen von einem Pfad induzierten Fluss in einen Kantenfluss transformiert.

$$x_{ij}^{kp} = \begin{cases} 1, & \text{wenn Pfad } p \text{ der Relation } k \text{ über Kante } (i, j) \text{ führt} \\ 0, & \text{sonst} \end{cases} \quad (4.12)$$

Zusätzlich werden Binärvariablen $\lambda_p^k \in \{0, 1\}$ eingeführt, welche die Auswahl eines Pfades für eine Relation ermöglichen.

$$\lambda_p^k = \begin{cases} 1, & \text{wenn Relation } k \text{ über Pfad } p \text{ transportiert wird} \\ 0, & \text{sonst} \end{cases} \quad (4.13)$$

Außerdem wird eine Nebenbedingung in das Problem eingefügt, welche die Wahl genau eines Pfades für jede Relation erzwingt.

$$\sum_{p \in \mathcal{P}(k)} \lambda_p^k = 1 \quad \forall k \in \mathcal{K} \quad (4.14)$$

Mit Hilfe dieser Definitionen wird ein Zusammenhang zwischen den kantenbasierten und den pfadbasierten Flüssen hergestellt.

$$x_{ij}^k = \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \quad (4.15)$$

Das Modell (Opt_{arc}) wird durch die Substitution (4.15) als pfadbasiertes Modell (Opt_{path}) reformuliert:

$$\begin{aligned} (Opt_{path}) \quad & \min_{\lambda, n} \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} [c_{ij} n_{ij} + \sum_{k \in \mathcal{K}} U_i v_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp}] \\ \text{s.t.} \quad & \sum_{k \in \mathcal{K}} l_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq L_{ij} n_{ij} \quad \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \end{aligned}$$

$$\begin{aligned}
\sum_{k \in \mathcal{K}} w_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} &\leq W_{ij} n_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\
\sum_{j \in \mathcal{N}} n_{ij} &\leq R_i && \forall i \in \mathcal{N} \\
\sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} &\leq B_i && \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \\
\sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} &= 1 && \forall k \in \mathcal{K} \quad \forall (i, j) \in \mathcal{A} \\
n_{ij} &\in \mathbb{N} && \forall (i, j) \in \mathcal{A} \\
\lambda_p^k &\in \{0, 1\} && \forall k \in \mathcal{K} \quad \forall p \in \mathcal{P}(k)
\end{aligned}$$

Es sei noch einmal darauf hingewiesen, dass die Flusserhaltungsbedingungen (4.6) und die Beschränkung der Kantenzahl (4.11) in der Definition eines Pfades (vgl. Definition 16) codiert sind. Das pfadbasierte Modell (Opt_{path}) hat dadurch erheblich weniger Nebenbedingungen. Dafür ist die Anzahl der zu betrachtenden Variablen nun exponentiell in der Instanzgröße. Bei Verwendung üblicher Lösungsmethoden wäre (Opt_{path}) nicht mehr im Computer darstellbar, weil die Menge zulässiger Pfade eine gigantische Größe annähme. Der Nutzen dieser Formulierung wird sich jedoch im Rahmen der Column-Generation-Verfahren erschließen. Zum Abschluss dieses Abschnittes wird ein Zusammenhang zwischen (Opt_{arc}) und (Opt_{path}) als Lemma festgehalten.

Lemma 1 (Zielfunktionswerte)

(Opt_{arc}) und (Opt_{path}) haben den gleichen optimalen Zielfunktionswert.

Beweis: Substitution (4.15) zeigt die Äquivalenz der beiden Formulierungen. Somit unterscheiden sich auch die Zielfunktionswerte nicht. \square

4.2 Analyse der Modelle (Opt_{arc}) und (Opt_{path})

In diesem Unterkapitel findet die komplexitätstheoretische Einordnung der zuvor definierten Modelle (Opt_{arc}) und (Opt_{path}) sowie eine Analyse zur Ermittlung unterer Schranken durch lineare Relaxierung statt. In Abschnitt 4.2.3 werden vereinfachende Modellannahmen diskutiert.

4.2.1 Komplexitätstheoretische Analyse der Modelle

In diesem Abschnitt wird gezeigt, dass das Finden einer zulässigen Lösung für (Opt_{arc}) und (Opt_{path}) ein \mathcal{NP} -vollständiges Entscheidungsproblem ist. Für den Beweis wird eine häufig verwendete Graphentransformation benötigt, welche die Äquivalenz von Kanten- und Knotenkapazitäten aufzeigt [AMO93]:

Lemma 2 (Äquivalenz von Kanten- und Knotenkapazitäten)

Jeder gerichtete Graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ mit Knotenkapazitäten $c(i)$ für jeden Knoten $i \in \mathcal{E}$ kann in Polynomialzeit in einen gerichteten Graphen $\mathcal{G}' = (\mathcal{N}', \mathcal{E}')$ ohne Knotenkapazitäten überführt werden. Die Transformation eines Graphen mit Kantenkapazitäten auf einen Graphen, der nur knotenkapazitiert ist, ist ebenfalls in Polynomialzeit möglich.

Das folgende Theorem beweist, dass kein Polynomialzeitalgorithmus existiert, der eine zulässige Lösung für (Opt_{arc}) findet:

Theorem 3 (Komplexität von (Opt_{arc}))

Das zu (Opt_{arc}) assoziierte Entscheidungsproblem ist \mathcal{NP} -vollständig, d.h. das Finden einer zulässigen Lösung ist \mathcal{NP} -vollständig.

Beweis: Es muss gezeigt werden, dass die Zulässigkeit einer Lösung von (Opt_{arc}) in Polynomialzeit bzgl. der Instanzgröße verifiziert werden kann und dass eine Transformation von (Opt_{arc}) auf ein beliebiges \mathcal{NP} -vollständiges Problem ebenfalls in Polynomialzeit möglich ist. Die Zulässigkeit einer gegebenen Lösung kann durch Überprüfung aller Nebenbedingungen geschehen. Ihre Anzahl ist polynomiell beschränkt und die Überprüfung in Linearzeit möglich. Daher ist nur noch die Reduzierbarkeit von (Opt_{arc}) auf ein \mathcal{NP} -vollständiges Problem zu zeigen. Dies geschieht per Widerspruchsbeweis.

Aus Theorem 3 ist bereits bekannt, dass das Entscheidungsproblem zum ganzzahligen Mehrgüterflussproblem ($IMCFP$) \mathcal{NP} -vollständig ist. Um zu zeigen, dass (Opt_{arc}) ebenfalls in \mathcal{NPC} liegt, wird angenommen, (Opt_{arc}) liege nicht in \mathcal{NPC} . Es wird gezeigt, dass jede Instanz des ($IMCFP$) eine Instanz von (Opt_{arc}) ist. Aus der Annahme $(Opt_{arc}) \notin \mathcal{NPC}$ würde $(IMCFP) \notin \mathcal{NPC}$ folgen. Dies stände im Widerspruch zu Theorem 3. Somit kann die Annahme $(Opt_{arc}) \notin \mathcal{NPC}$ nicht wahr sein.

Es sei also eine beliebige Instanz des ganzzahligen Mehrgüterflussproblems ($IMCFP$) gegeben:

$$\begin{aligned} \min \quad & \sum_{(i,j) \in \mathcal{A}, k \in \mathcal{K}} x_{ij}^k \\ \text{s.t.} \quad & \sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} \quad \forall (i,j) \in \mathcal{A} \\ & \sum_{j \in \mathcal{N}} x_{ij}^k - x_{ji}^k = B(k,j) \quad \forall k \in \mathcal{K} \quad \forall (i,j) \in \mathcal{A} \\ & x_{ij}^k \in \{0,1\} \quad \forall k \in \mathcal{K} \quad \forall (i,j) \in \mathcal{A} \end{aligned}$$

Nach Definition 10 sind im ($IMCFP$) Kantenkapazitäten gegeben. Diese können Lemma 2 entsprechend ohne Beschränkung der Allgemeinheit als Knotenkapazitäten $\Gamma(i)$ aufgefasst werden. Das resultierende Probleme werde als ($IMCFP'$) bezeichnet. Folgende Parameterwahlen für (Opt_{arc}) zeigen, dass (Opt_{arc}) eine Verallgemeinerung von ($IMCFP'$) ist:

$$\begin{aligned}
|\mathcal{D}| &:= 1 \\
c_{ij} &:= 0 && \forall (i, j) \in \mathcal{A} \\
U_i &:= 1 && \forall i \in \mathcal{N} \\
v_1^k &:= 1 && \forall k \in \mathcal{K} \\
l_1^k = w_1^k = L_1^k = W_1^k &:= 0 && \forall k \in \mathcal{K} \\
R_i &:= 0 && \forall i \in \mathcal{N} \\
B_i &:= \Gamma(i) && \forall i \in \mathcal{N} \\
S_k &:= \infty && \forall k \in \mathcal{K}
\end{aligned}$$

Das Problem (Opt_{arc}) hat mit diesen Spezifikationen folgende Gestalt:

$$\begin{aligned}
&\min_{x, n} \sum_1 \sum_{(i, j) \in \mathcal{A}} [0n_{ij} + \sum_{k \in \mathcal{K}} 1x_{ij}^k] \\
\text{s.t.} \quad &\sum_{i \in \mathcal{N}} x_{ji}^k - \sum_{i \in \mathcal{N}} x_{ij}^k = B(k, j) \\
&\sum_{k \in \mathcal{K}} 0x_{ij}^k \leq 0n_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\
&\sum_{k \in \mathcal{K}} 0x_{ij}^k \leq 0n_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\
&\sum_{j \in \mathcal{N}} n_{ij} \leq 0 && \forall i \in \mathcal{N} \\
&\sum_{j \in \mathcal{N}, k \in \mathcal{K}} 1x_{ij}^k \leq \Gamma(i) && \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \\
&\sum_{(i, j) \in \mathcal{A}} x_{ij}^k \leq \infty && \forall k \in \mathcal{K} \quad \forall (i, j) \in \mathcal{A} \\
&n_{ij} \in \mathbb{N} && \forall (i, j) \in \mathcal{A} \\
&x_{ij}^k \in \{0, 1\} && \forall k \in \mathcal{K} \quad \forall p \in \mathcal{P}(k)
\end{aligned}$$

Die Nebenbedingungen $\sum_{j \in \mathcal{N}} n_{ij} \leq 0$ implizieren $n_{ij} = 0$ für alle Kanten $(i, j) \in \mathcal{A}$. Nach Entfernen der n_{ij} und aller nun redundanten Nebenbedingungen aus dem Problem, bleibt die Instanz von ($IMCFP'$) übrig. Da laut Annahme (Opt_{arc}) nicht \mathcal{NP} -vollständig ist, ist auch ($IMCFP'$) nicht \mathcal{NP} -vollständig. ($IMCFP'$) kann mit Lemma 2

in Polynomialzeit auf (*IMCFP*) zurück transformiert werden. Somit wäre das Entscheidungsproblem zu (*IMCFP*) nicht \mathcal{NP} -vollständig. Daraus ergibt sich der Widerspruch und das Theorem ist bewiesen. \square

Aus der Äquivalenz der Probleme (Opt_{arc}) und (Opt_{path}) folgt, dass auch das zu (Opt_{path}) assoziierte Entscheidungsproblem \mathcal{NP} -vollständig ist. Es ist damit gezeigt, dass für (Opt_{arc}) und (Opt_{path}) bereits das Finden einer zulässigen Lösung und folglich insbesondere das Finden der Optimallösung ein komplexes Problem darstellt.

Korollar 2

Das zu (Opt_{path}) assoziierte Entscheidungsproblem ist \mathcal{NP} -vollständig.

Beweis: Gleichung (4.15) zeigt, dass Lösungen von (Opt_{arc}) und (Opt_{path}) in Polynomialzeit in Lösungen des jeweils anderen Problems transformiert werden können. \square

Aus Theorem 3 kann geschlussfolgert werden, dass es nicht möglich ist, einen Algorithmus zu entwerfen, der (Opt_{arc}) effizient, d.h. in Polynomialzeit, und exakt für alle Instanzen löst. Da jedoch nicht angestrebt wird, das Modell für beliebige Instanzen, sondern für Instanzen aus der Praxis des Schienengüterverkehrs zu lösen, muss die Suche nach einem effizienten, exakten Algorithmus nicht aussichtslos sein.

4.2.2 Untere Schranken

Dieser Abschnitt befasst sich mit der Bestimmung unterer Schranken für das zuvor definierte Modell (Opt_{arc}). Bei der Zusammenfassung der Vor- und Nachteile nordamerikanischer Blockingmodelle in Unterkapitel 3.2.2 wurde der algorithmische Vorteil wagenbasierter Transportkostenstrukturen genannt. In (Opt_{arc}) werden die Transportkosten auf Basis von Zugkilometern modelliert. Die Notwendigkeit von Zugkapazitäten auf einer Kante wird durch Wagenflüsse auf der jeweiligen Kante induziert, d.h. ein Zug verkehrt nur auf einer Kante, wenn dort Wagen zu transportieren sind. Umgekehrt gilt, dass Wagen nur bei Vorliegen entsprechender Zugkapazitäten über eine Kante transportiert werden können. Im Modell wird dieser Zusammenhang durch die Längen- und Gewichtsbeschränkungen dargestellt:

$$\sum_k l_d^k x_{ij}^k \leq L_{ij} n_{ij} \quad \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \quad (4.16)$$

$$\sum_k w_d^k x_{ij}^k \leq W_{ij} n_{ij} \quad \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \quad (4.17)$$

Angenommen, gewisse Wagenflüsse auf einer Kante (i', j') seien bekannt, d.h. die Flussvariablen $x_{i'j'}^k$ für alle Relationen $k \in \mathcal{K}$ gegeben. Dann kann die Mindestmenge notwendiger Züge $\underline{n}_{i'j'}$ zum Transport jener Wagen mit Hilfe von Formel (4.18) bestimmt werden.

$$\underline{n}_{i'j'} = \max_{d \in \mathcal{D}} \left[\left\{ \frac{\sum_{\mathcal{K}} x_{i'j'}^k l_d^k}{L_{i'j'}}, \frac{\sum_{\mathcal{K}} x_{i'j'}^k w_d^k}{W_{i'j'}} \right\} \right] \quad (4.18)$$

Der Maximumterm und die aufrundenden Gaußklammern in der Formel machen den Zusammenhang zwischen den Flussvariablen und der optimalen Belegung der Designvariablen n_{ij} nichtlinear. Die Designvariablen modellieren die Entscheidung, eine Zugverbindung auf einer Strecke zu etablieren, d.h. durch die Wahl der Designvariablen werden Transportkapazitäten auf einer jeweiligen Kante geschaffen. Diese Transportkapazitäten müssen nicht notwendigerweise vollständig ausgenutzt werden. Die verbleibenden freien Kapazitäten können zum Transport weiterer Wagen genutzt werden, deren Grenzkosten in Folge der freien Kapazität null sind. Deshalb ist der kürzeste Weg einer Relation nicht gleichzeitig der kostengünstigste Weg. In der Optimallösung tritt hierdurch ein sichtbarer Bündelungseffekt auf. Dieser Umstand ändert sich, wenn die Designvariablen nicht mehr ganzzahlig, sondern kontinuierlich gewählt werden. Dies geschieht bei der linearen Relaxierung der Designvariablen, die von vielen exakten Lösungsansätzen (vgl. Unterkapitel 3.4.2) benutzt wird. Da die Leistungsfähigkeit jener Ansätze stark von der Genauigkeit abhängt, mit der die Relaxierungen das Originalproblem approximieren, ist eine genauere Betrachtung sinnvoll.

Da linear relaxierte Modelle Zugkosten nicht sinnvoll abbilden können, fällt der im nicht relaxierten Modell sichtbare Bündelungseffekt weg. Die Struktur der Optimallösungen beider Probleme unterscheidet sich stark. Aus algorithmischer Sicht ist noch gravierender, dass die optimalen Zielfunktionswerte der Modelle stark voneinander abweichen und das relaxierte Modell somit keine engen unteren Schranken für das nicht-relaxierte Modell liefert. Dieser Umstand spiegelt sich auch in der polyedrischen Struktur der Probleme wider:

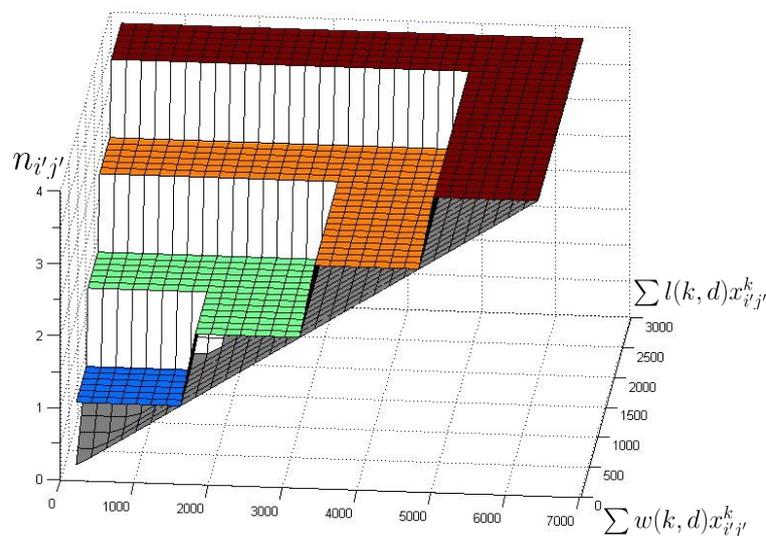


Abbildung 4.1: Vergleich zwischen relaxierter und nicht-relaxierter Kostenstruktur

In Abbildung 4.1 sind die Graphen zweier Funktionen dargestellt. Auf den Achsen sind aggregierte Längen und Gewichte der als bekannt angenommenen Wagenströme abgetragen. Die darüber liegende Funktion ordnet einem Wagenaufkommen mit Länge und Gewicht die Anzahl notwendiger Züge zu dessen Transport zu. In diesem Beispiel werden die in Deutschland gängigen Kapazitätsgrenzen für Wagenzüge von 700m und 1600t verwendet. Die stufenförmige Funktion, kurz Treppenfunktion (farbig dargestellt), stellt dabei den Zusammenhang zwischen Wagenfluss und Zugzahl her, wie er in (Opt_{arc}) modelliert ist. Die darunter verlaufende lineare Funktion (grau) bildet den gleichen Zusammenhang im Falle der linearen Relaxierung der Designvariablen ab. Der Abstand zwischen Treppenfunktion und linearer Funktion ist ein Maß dafür, wie sehr die relaxierte Kostenfunktion die Transportkosten für ein gegebenes Aufkommen unterschätzt. Gerade im Falle geringfügiger Aufkommen auf einer Kante, wie sie durch die kleinen Sendungsgrößen im EWV naturgemäß vorkommen, werden die Transportkosten massiv unterschätzt. Eine schwache Approximation der realen Transportkosten ist die Folge. Da die Transportkosten bei Realinstanzen einen gewichtigen Anteil an den Gesamtkosten haben, liefert die lineare Relaxierung sehr schwache untere Schranken für das Problem (Opt_{arc}) . Einen ähnlichen Zusammenhang hat bereits Homfeld für den Fall eindimensionaler Kapazitätsbeschränkungen, d.h. bei Beschränkung der Zuglänge, festgestellt [Hom12]. Schwache Schranken durch lineare Relaxierung sind eine allgemeine Eigenschaft von Netzwerkproblemen mit Fixkosten, wie Hewitt et al. anmerken [HNS10]. Gendron und Crainic haben diese Problematik ebenfalls untersucht [GC94]. Sie betrachten Modelle mit kontinuierlichen Flussvariablen und binären Designvariablen. Eine Übertragung ihrer Ergebnisse zur Verbesserung der Schranken für (Opt_{arc}) ist daher nur bedingt möglich. Die geringe Qualität unterer Schranken ist entscheidend für die Entwicklung angemessener Optimierungsalgorithmen und die Bewertung von Güteabschätzungen [MW83]. In Kapitel 6 wird das Modell (Opt_{arc}) für verschiedene Instanzen mit der kommerziellen Optimierungssoftware CPLEX untersucht. Die Ergebnisse untermauern die Thesen zum Einfluss der schwachen Schranken auf die Lösbarkeit des Modells.

4.2.3 Annahmen zur Modellvereinfachung

Die zuvor definierten Modelle (Opt_{arc}) und (Opt_{path}) stellen ein Abbild der Entscheidungsprobleme bei der Routenplanung im EWV dar. Die Optimierung dieser Modelle kann unter Umständen zu Lösungen führen, die Strukturen aufweisen, die bei der Umsetzung in der Realität unerwünscht wären. Ein Beispiel hierfür könnte ein enormer Umweg für eine Relation sein, der zwar rechnerisch minimale Kosten verursacht, bei dem die Kostenersparnis gegenüber einer direkteren Route jedoch so geringfügig ist, dass in der Praxis stets die kürzere Route gewählt würde. Um praxisferne Lösungsstrukturen zu verhindern, ist es üblich, Vorannahmen an Modelle zu formulieren und jene Strukturen in sogenannten Preprocessing-Verfahren zu verbieten. Neben einer erhöhten Praxisnähe der Lösungen bieten Preprocessing-Verfahren auch mathematische Vorteile bei der Ermittlung von Lösungen: Die Anzahl der Variablen im Modell (Opt_{arc}) wächst quadratisch mit der Anzahl der Knoten und linear in der Anzahl der Relationen. Die Anzahl möglicher Lösungen wächst jedoch exponentiell in der Anzahl der Variablen. Dieses Wachstum des

Lösungsraumes wird umgangssprachlich als *kombinatorische Explosion* bezeichnet. Um die Anzahl der im Problem relevanten Variablen zu verringern, ist es hilfreich, die Werte gewisser Variablen zu fixieren.

Zwei Klassen von Preprocessingverfahren können unterschieden werden. Es existieren exakte Preprocessings, welche die Eigenschaft haben dass sie zwar den Lösungsraum verkleinern, dabei aber garantieren, keine Optimallösungen auszuschließen. Diese Garantie wird von der zweiten Klasse, den heuristischen Preprocessings, nicht gegeben. Sie basieren auf den oben erwähnten Annahmen an die Gestalt der Optimallösung. Werden zu scharfe vereinfachende Annahmen getroffen, ist es möglich den Lösungsraum so sehr zu verkleinern, dass dabei Optimallösungen ausgeschlossen werden. Im Folgenden werden Techniken aus beiden Klassen für das Modell (Opt_{arc}) vorgestellt. Tests zu deren Wirkungsweise finden sich im Ergebniskapitel 6.

Exakte Preprocessings

Zu betrachten sind Variablen, für die a priori zu bestimmen ist, dass sie in der Optimallösung von (Opt_{arc}) fixe Werte annehmen. Es handelt sich dabei um Fixierungen, die in der Netzwerkoptimierung üblich sind.

1. *Pseudokreise*: Der Graph enthält Kanten, die von einem Knoten zu sich selbst führen. In der Optimallösung liegt auf ihnen kein Fluss an. Daher wird festgelegt:

$$x_{ii}^k := 0 \quad \forall k \in \mathcal{K} \quad \forall i \in \mathcal{N} \quad (4.19)$$

$$n_{ii} := 0 \quad \forall i \in \mathcal{N} \quad (4.20)$$

2. Zu Quelle/Senke adjazente Knoten: Haben die Wagen einer Relation k ihre Senke $d(k)$ erreicht, werden sie von dort nicht weiter transportiert. Ebenso wenig werden sie nach Verlassen der Quelle $o(k)$ noch einmal dorthin transportiert. Diese Erkenntnis lässt sich auf die Flussvariablen übertragen.

$$x_{ij}^k := 0, \text{ falls } j = o(k) \vee i = d(k) \quad (4.21)$$

3. *Knoten mit geringer Umstellkapazität*: Das Aufkommen eines Knotens verringert dessen freie Umstellkapazität. Relationen von zu großem Umfang können in jenen Knoten nicht umgestellt werden. Die Grundlast eines Knoten $i' \in \mathcal{N}$ im Szenario $d \in \mathcal{D}$ beträgt

$$Grundlast(i', d) = \max \left\{ \frac{\sum_{k \in \mathcal{K}} x_{i'j}^k l_d^k}{L_{i'j}}, \frac{\sum_{k \in \mathcal{K}} x_{i'j}^k w_d^k}{W_{i'j}} \right\}. \quad (4.22)$$

Dabei sind $L_{i'j}$ und $W_{i'j}$ die maximal zulässigen Längen und Gewichte auf den aus Knoten i' herausführenden Kanten. Wenn die Differenz aus Knotenkapazität des Knotens i' und seiner Grundlast an mindestens einem Tag d kleiner ist als die Anzahl der Züge, die eine Relation k' an jenem Tag induziert, kann sie in i' nicht umgestellt werden. Somit gilt in diesem Fall:

$$x_{ij}^{k'} := 0 \quad \text{und} \quad x_{j i'}^{k'} := 0 \quad \forall j \in \mathcal{N}. \quad (4.23)$$

Heuristische Preprocessings durch vereinfachende Annahmen

Die hier aufgeführten Preprocessingschritte sind heuristischer Natur. Ihre Formulierung geht zum Teil aus Fachgesprächen mit Experten bei DB Mobility Logistics hervor oder wurden von Schülldorf im Rahmen seiner Diplomarbeit [Sch08] vorgeschlagen:

1. *Relative Umwege*: Die Transportkostenstruktur im vorliegenden Problem macht es theoretisch möglich, dass für die Wagen einer Relation – relativ zur Länge der Direktverbindung zwischen Quelle und Senke – enorme Umwege in Kauf genommen werden. Es ist nämlich möglich, dass ihre Grenzkosten bei Mitfahrt auf anderen Zugverbindungen null betragen. Es kann dennoch sinnvoll sein, relative Umwege auszuschließen. Verschiedene praktische und theoretische Vorteile lassen sich daraus ableiten, die an einem fiktiven Beispiel diskutiert werden sollen (vgl. Abbildung 4.2 links):

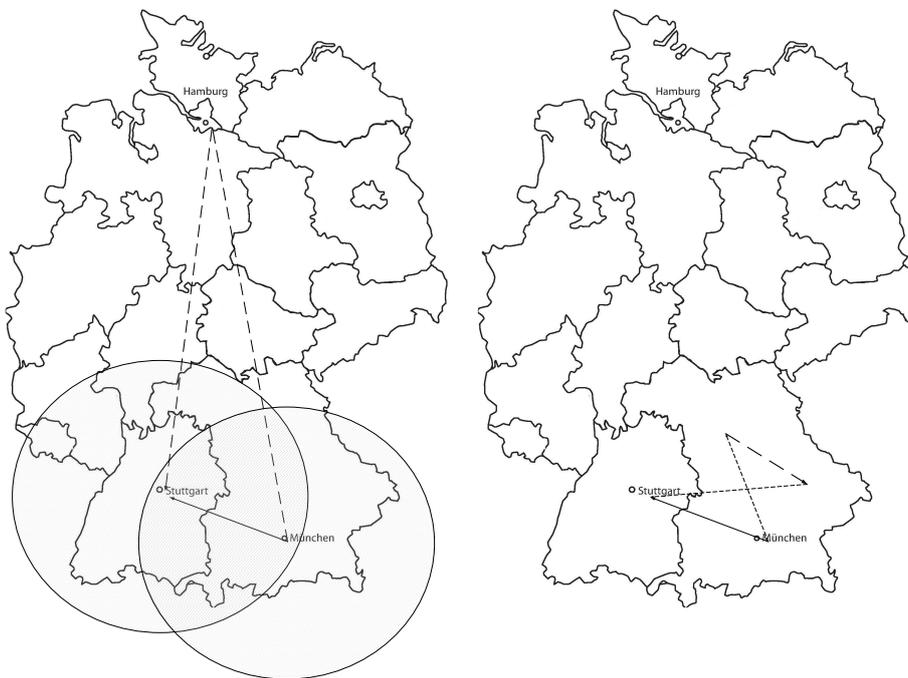


Abbildung 4.2: Illustration heuristischer Preprocessings (links: Relative Umwege, rechts: Entgegengesetzte Kanten)

Es sei in München ein Aufkommen von zwei Wagen gegeben, die nach Stuttgart zu transportieren sind. Die Länge der Direktverbindung beträgt ca. 220 km. Es fahre täglich ein Zug mit freier Kapazität von München nach Maschen und von dort fahre wiederum ein Zug mit freier Kapazität nach Stuttgart. Die zwei Wagen könnten über diesen Umweg transportiert werden, ohne Transportkosten zu verursachen. Zunächst einmal besteht der praktische Einwand, dass ein derartiger

Umweg vermutlich nur ungern in Kauf genommen würde. Dies trüge nicht zur Akzeptanz des Verfahrens in der Praxis bei. Zum anderen ist zu erwarten, dass die prognostizierten Wagenaufkommen, auf deren Basis die Wagenrouten optimiert werden, nicht vollständig zutreffend sind. In Folge dessen wäre es denkbar, dass die freien Kapazitäten, auf deren Grundlage der lange Umweg geplant wird, nicht vorhanden sind, wenn gewisse Aufkommen in der Prognose unterschätzt wurden. In der Realität könnte der Umweg dann zu positiven Grenzkosten führen, weil ein weiterer Zug notwendig wird. Das Verbot großer Umwege stabilisiert daher die Lösung gegenüber Schwankungen der Aufkommensdaten. Es kann jedoch dazu führen, dass kostengünstige Routen verboten werden. Es wird daher folgende Regel vorgeschlagen: Es sei $dist(k)$ die Länge der Direktverbindung zwischen Quelle $o(k)$ und Senke $d(k)$ einer Relation $k \in \mathcal{K}$ und Δ ein skalarer Faktor. Die Relation k wird über keinen Knoten $j \in \mathcal{N}$ transportiert, der außerhalb der Umkreise mit Radius $\Delta \times dist(k)$ um Quelle $o(k)$ und die Senke $d(k)$ liegt. Je kleiner Δ gewählt wird, desto weniger Freiheitsgrade bestehen für die Route der Relation.

2. *Gegenläufige Kanten:* Aus den zuvor geschilderten Gründen kann es theoretisch kostengünstig sein, dass Wagen einer Relation für eine Teilstrecke in entgegengesetzter Richtung zur Direktverbindung transportiert werden. Diese *gegenläufigen Kanten* (i, j) haben die Eigenschaft, dass ihr Anfangspunkt $i \in \mathcal{N}$ näher am Ziel $d(k)$ einer Relation k als an deren Quelle $o(k)$ liegt und gleichzeitig ihr Endpunkt $j \in \mathcal{N}$ näher an der Quelle der Relation als an deren Ziel liegt (vgl. Abbildung 4.2 rechts). Das Verbot gegenläufiger Kanten wird folgendermaßen formalisiert:

$$dist(o(k), j) \leq dist(o(k), i) \wedge dist(d(k), i) \leq dist(d(k), j) \Rightarrow x_{ij}^k := 0 \quad (4.24)$$

3. *Günstiger Direkttransport:* Die Grenzkosten für den Transport eines Wagens sind nicht a priori zu bestimmen, weil die Grenzkosten in Abhängigkeit von der Auslastung eines Zuges Anteile an dessen Fixkosten enthalten. Es ist jedoch möglich, die notwendigen Kosten für den Transport eines Wagens auf dem kürzesten Weg zwischen Quelle und Senke nach oben abzuschätzen. Im teuersten Fall wird für den Transport einer Relation eine Direktzugverbindung etabliert. Diese zunächst hohen Kosten stehen jedoch der Ersparnis von Umstellkosten gegenüber, die entstünden, wenn die Relation über weitere Knoten zu ihrer Senke transportiert würde. Es sei k eine Relation und (i, j) eine Kante und es gelte $i \neq o(k)$ und $j \neq d(k)$, d.h. die Kante beginnt nicht in der Quelle und endet nicht in der Senke von k . Wird Relation k über Kante (i, j) transportiert, entstehen Umstellkosten in i und j . Ist die Summe dieser Umstellkosten größer als die Direkttransportkosten, kann die Kante verboten werden.

Es mag auf den ersten Blick nicht ersichtlich sein, weshalb es sich hierbei nicht um ein exaktes Preprocessing handelt. Ist ein Knoten i Quelle vieler Relationen mit geringem Aufkommen, für die auf Grund der Regel außer den Direktverbindungen

alle Kanten verboten sind, müssten viele Direktzüge den Knoten verlassen. Dies kann durch einen Mangel an Richtungsgleisen, der eine Bündelung der ausgehenden Transporte erfordern würde, unmöglich sein. Im schlimmsten Fall kann dies zu einer so starken Einschränkung des Lösungsraums führen, dass keine zulässige Lösung mehr enthalten ist. Dieses Preprocessing ist daher mit Bedacht einzusetzen und im Zweifelsfall zu verwerfen.

In Kapitel 6 werden die vorgestellten Preprocessings empirisch getestet und somit ihre Wirksamkeit bewertet.

5 Entwicklung von Lösungsverfahren für das Wagenflussproblem

Numerische Tests mit CPLEX 12.4 lassen den Schluss zu, dass die optimale Lösung von Instanzen von (Opt_{arc}) mit generischen Lösungsverfahren für ganzzahlige Programme aktuell nicht praktikabel ist (vgl. Kapitel 6). Dieser Umstand motiviert die Entwicklung problemspezifischer Optimierungsmethoden. In den Unterkapiteln 5.1 und 5.2 werden Column-Generation-Verfahren und Branch-and-Price-Algorithmen eingeführt. Anschließend wird in Unterkapitel 5.3 ein Ansatz entwickelt, der in den Bereich der Matheuristiken fällt.

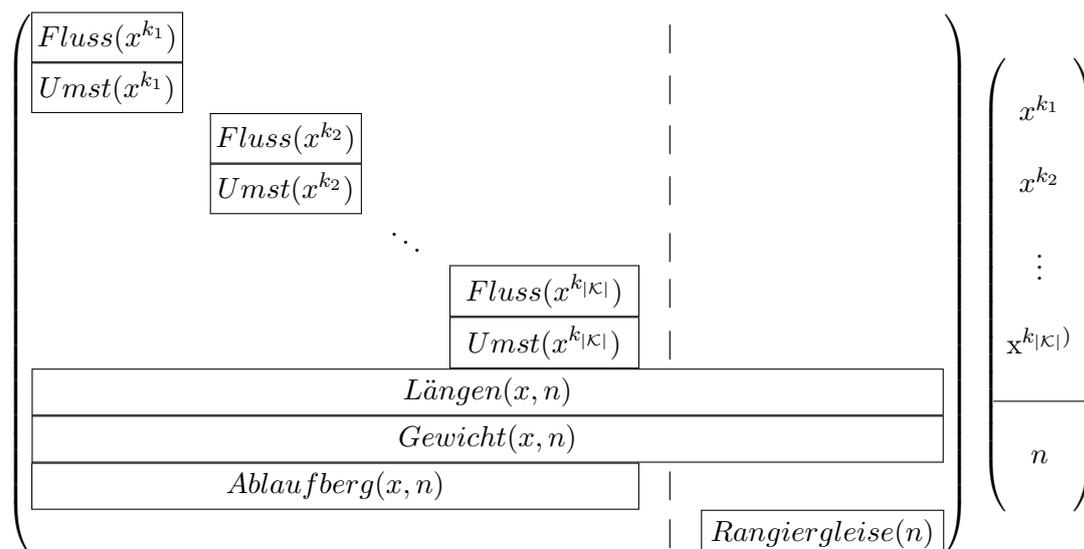
5.1 Ein Column-Generation-Verfahren

Die Struktur der Nebenbedingungen von (Opt_{arc}) führt zu schwachen unteren Schranken durch lineare Relaxierung, wie in Abschnitt 4.2.2 begründet wurde. Diese Schwäche der Schranken verhindert eine effiziente Lösung des nicht relaxierten Problems (Opt_{arc}) mit Hilfe der in CPLEX implementierten Branch-and-Cut-Algorithmen. Auch für den Fall, dass eine gute ganzzahlige Lösung bekannt ist, müssen viele Knoten des Suchbaumes durchlaufen werden, um Äste des Baumes zu verwerfen. Darüber hinaus ist für Instanzen mit Realgröße nicht einmal die lineare Relaxierung in angemessener Zeit lösbar und somit keine Schranke bekannt. Beide Probleme motivieren den Versuch, zunächst die Lösung der linearen Relaxierung effizienter zu gestalten. Dazu wird Column Generation (CG) auf die Relaxierung des zu (Opt_{arc}) äquivalenten Modells (Opt_{path}) angewendet. Im nachfolgenden Unterkapitel 5.2 wird CG in ein Branch-and-Bound-Schema integriert werden, um auch ganzzahlige Lösungen des Modells (Opt_{path}) zu finden.

5.1.1 Struktur der Nebenbedingungen von (Opt_{arc}) und (Opt_{path})

Die Nebenbedingungen des Optimierungsproblems definieren einen polyedrischen Zulässigkeitsbereich. Es ist sinnvoll, die Struktur der Nebenbedingungen zu untersuchen, wenn eine Verbesserung der Lösungsgeschwindigkeit angestrebt wird. Die Nebenbedingungen lassen sich in Form der Matrix in Abbildung 5.1 darstellen, bei deren Betrachtung eine spezielle Struktur auffällt.

Der linke Teil der Matrix enthält die Spalten, die sich auf die Flussvariablen der Relationen $x^{k_1}, \dots, x^{k_{|\mathcal{K}|}}$ beziehen, während die rechten Spalten auf die Designvariablen n Bezug nehmen. Die Nebenbedingungen zur Flusserhaltung (4.6) (kurz *Fluss*) und die Begrenzung der Umstellvorgänge (4.11) (kurz *Umst*) betreffen jeweils nur die Flussvariablen

Abbildung 5.1: Struktur der Nebenbedingungsmatrix von (Opt_{arc})

einer Relation. Die übrigen Nebenbedingungen enthalten Flussvariablen mehrerer Relationen und zum Teil Designvariablen. Bereiche außerhalb der eingezeichneten Kästchen enthalten nur Nullen. Es ergibt sich eine sogenannte Block(diagonal)struktur, für die spezielle Lösungsverfahren existieren: die sogenannten Column-Generation-Verfahren. Zu jeder Relation existiert ein eigener Block. Die einzelnen Blöcke hängen durch weitere Nebenbedingungen voneinander ab. Bei Vernachlässigung der *koppelnden Nebenbedingungen* (engl: *forcing oder coupling constraints*) bleibt für jede Relation ein Subproblem mit Flusserhaltungsbedingung und Umstellbeschränkung übrig. Dies entspricht exakt der Definition eines Pfades im pfadbasierten Optimierungsmodell (Opt_{path}) . Auch in der kantenbasierten Version ist für jede Relation nur ein Pfad zu bestimmen. Wie die Blockstruktur ausgenutzt werden kann, um effizient Lösungen zu berechnen, wird im folgenden Abschnitt dargestellt.

5.1.2 Einführung in Column-Generation-Verfahren

Die Blockdiagonalstruktur der Nebenbedingungsmatrix wird von sogenannten *Column-Generation-Verfahren* ausgenutzt [BJN⁺98]. Es handelt sich dabei um ein Dekompositionsverfahren für lineare Programme, das auf der in Korollar 1 gezeigten Identifikation im Optimum relevanter Variablen beruht. Die reduzierten Kosten von Variablen aus Definition 14 geben Auskunft darüber, ob die jeweilige Variable im Optimum aktiv, d.h. ungleich null, sein kann. Die sukzessive Identifikation dieser Variablen kann auf die lineare Relaxierung des pfadbasierten Modells (Opt_{path}) angewendet werden. Für jeden möglichen Pfad p einer Relation k enthält das Modell eine Entscheidungsvariable λ_p^k . Ist die Variable im Optimum nicht aktiv, d.h. nimmt sie den Wert null an, kann der korrespondierende Pfad nicht Teil der Optimallösung sein. Damit ist es möglich, jenen Pfad zu vernachlässigen und die Betrachtung auf eine Teilmenge möglicher Pfade zu be-

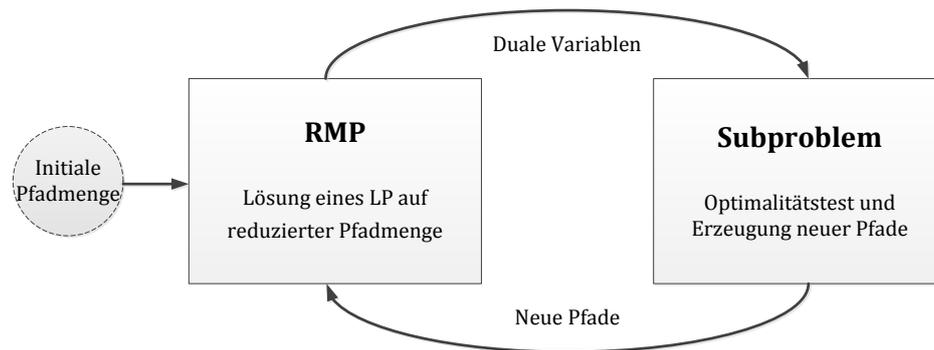


Abbildung 5.2: Schaubild zu Column-Generation-Verfahren

schränken, wodurch auch der zur Darstellung nötige Speicheraufwand extrem reduziert wird. Column-Generation-Verfahren bestehen aus zwei sich sequentiell wiederholenden Prozessen (vgl. Abbildung 5.2):

1. Das zu lösende LP wird in diesem Zusammenhang als *Master Problem* (MP) bezeichnet. Das Master Problem wird mit einer Startmenge von Pfaden initialisiert, die eine zulässige Lösung des betrachteten LP ermöglichen. Das MP auf einer reduzierten Pfadmenge wird als *Restricted Master Problem* (RMP) bezeichnet. Das RMP kann auf Grund seiner beschränkten Größe deutlich schneller gelöst werden als das MP. Die bei Verwendung eines Simplexverfahrens ebenfalls berechneten dualen Variablen im Optimum des RMP werden an das sogenannte Subproblem übergeben.
2. Das *Subproblem* bestimmt mit Hilfe von Korollar 1, ob die bisher betrachtete Pfadmenge alle für das MP relevanten Pfade enthält. Falls dies nicht der Fall ist, werden die dualen Variablen genutzt, um neue Pfade zu erzeugen, die für die nächste Iteration an das RMP übergeben werden [Des06]. Die koppelnden Nebenbedingungen sind Teil des RMP, werden im Subproblem jedoch nur in Form dualer Informationen berücksichtigt. Der Vorgang der Pfaderzeugung wird im kommenden Abschnitt 5.1.3 am Beispiel von (Opt_{path}) erläutert.

5.1.3 Anwendung von Column Generation auf (Opt_{path})

Drei Komponenten machen ein Column-Generation-Verfahren aus (vgl. Abbildung 5.2): die initiale Pfadmenge, das RMP und das Subproblem. Das MP und somit das RMP sind durch die lineare Relaxierung von (Opt_{path}) gegeben:

$$\begin{aligned}
(\text{MP}) \quad & \min_{x,n} \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} [c_{ij} n_{ij} + \sum_{k \in \mathcal{K}} U_i \sum_{p \in \mathcal{P}(k)} v_d^k \lambda_p^k x_{ij}^{kp}] \\
\text{s.t.} \quad & \sum_{k \in \mathcal{K}} l_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq L_{ij} n_{ij} && \forall d \in \mathcal{D} \quad \forall (i,j) \in \mathcal{A} \\
& \sum_{k \in \mathcal{K}} w_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq W_{ij} n_{ij} && \forall d \in \mathcal{D} \quad \forall (i,j) \in \mathcal{A} \\
& \sum_{j \in \mathcal{N}} n_{ij} \leq R_i && \forall i \in \mathcal{N} \\
& \sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq B_i && \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \\
& \sum_{p \in \mathcal{P}(k)} \lambda_p^k = 1 && \forall k \in \mathcal{K} \quad \forall p \in \mathcal{P}(k) \\
& 0 \leq \lambda_p^k \leq 1 && \forall k \in \mathcal{K} \quad \forall p \in \mathcal{P}(k) \\
& n_{ij} \geq 0 && \forall (i,j) \in \mathcal{A}
\end{aligned}$$

Die initiale Pfadmenge

Die initiale Pfadmenge muss für jede Relation mindestens einen Pfad enthalten. Des Weiteren muss aus den betrachteten Pfaden eine zulässige Lösung des MP konstruierbar sein. Es wird die Menge der direkten Pfade von der Quelle $o(k)$ zur Senke $d(k)$ für jede Relation k definiert:

$$\mathcal{P}_{\text{direkt}}(k) = \{p \in \mathcal{P}(k) \mid x_{ij}^{kp} = 1, \text{ falls } i = o(k) \wedge j = d(k), x_{ij}^{kp} = 0 \text{ sonst}\} \quad (5.1)$$

Testrechnungen auf verschiedenen Instanzen zeigen, dass die oben definierte Menge der Direktpfade in der Regel eine zulässige Startmenge für das MP erzeugt (vgl. Kapitel 6). Es kann gezeigt werden, dass theoretisch Fälle existieren, in denen die Menge $\mathcal{P}_{\text{direkt}}(k)$ keine zulässige initiale Pfadmenge liefert. Auf die Konstruktion eines solchen Falles wird an dieser Stelle jedoch verzichtet.

Ist der Ansatz mit $\mathcal{P}_{\text{direkt}}(k)$ nicht erfolgreich, ist es möglich, mit Hilfe sogenannter *Big-M-Methoden* eine Startlösung zu erzeugen [Des06]. Bei diesen in der linearen Optimierung üblichen Verfahren werden zunächst virtuelle Pfade eingeführt, die sehr hohe Kosten haben, dabei jedoch die Zulässigkeit der Lösung erzwingen. Eine vergleichbare Methodik wird in Unterkapitel 5.3 zum Finden einer zulässigen Startlösung für (Opt_{arc}) verwendet.

RMP

Das RMP hat die gleiche Struktur wie das MP. Es sind lediglich weniger Variablen bzw. Pfade enthalten. Das RMP ist ein lineares Programm, das mit Standardmethoden,

wie dem Simplexverfahren oder durch Innere-Punkte-Methoden, gelöst werden kann. Da professionell implementierte Codes für diese Verfahren existieren, kann auf Softwareprodukte, wie CPLEX oder GuRoBi, zurückgegriffen werden. Diese Verfahren bestimmen neben der primalen Lösung des RMP auch eine Optimallösung des dualen Problems. Die dualen Variablen werden im Subproblem benötigt und bieten die Möglichkeit, in jeder Iteration des Column-Generation-Verfahrens eine untere Schranke für das MP zu berechnen. Es werden daher Bezeichnungen für die dualen Variablen derjenigen Nebenbedingungen eingeführt, in denen Pfadvariablen vorkommen.

$$\begin{aligned}\mu &\Rightarrow \sum_{k \in \mathcal{K}} l_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq L_{ij} n_{ij} \\ \eta &\Rightarrow \sum_{k \in \mathcal{K}} w_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq W_{ij} n_{ij} \\ \sigma &\Rightarrow \sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k \sum_{p \in \mathcal{P}(k)} \lambda_p^k x_{ij}^{kp} \leq B_i \\ \tau &\Rightarrow \sum_{p \in \mathcal{P}(k)} \lambda_p^k = 1\end{aligned}$$

Der Zielfunktionswert Z_{RMP}^* der Lösung des RMP liefert keine untere Schranke für den optimalen Zielfunktionswert Z_{MP}^* des MP, weil die Pfadmenge im RMP kleiner ist als im MP. Es gilt

$$Z_{MP}^* \leq Z_{RMP}^*. \quad (5.2)$$

Mit Hilfe der Duallösungen können dennoch untere Schranken für den Zielfunktionswert Z_{MP}^* des MP bestimmt werden. Damit ist auch eine untere Schranke für (Opt_{path}) bekannt, falls eine optimale Lösung des MP nicht möglich ist. Dazu sind zwei Methoden bekannt: Die Methode von Farley [Far90] hat sich als wenig effizient für das vorliegende Problem erwiesen und wird daher nicht eingehender betrachtet. Die von Lasdon [Las70] eingeführte und bewiesene Schranke benutzt die aus den dualen Variablen bestimmten reduzierten Kosten. Im Falle des hier betrachteten MP muss für jede Relation $k \in \mathcal{K}$ das Minimum der reduzierten Kosten r_{kp} aller Variablen λ_p^k bestimmt werden, um die *Lasdonschranke* LB_{Lasdon} zu bestimmen:

$$LB_{Lasdon} = Z_{RMP}^* + \sum_k \min_{p \in \mathcal{P}(k)} r_{kp} \leq Z_{MP}^* \leq Z_{RMP}^* \quad (5.3)$$

Die relevanten r_{kp} werden in jeder Iteration für das Subproblem bestimmt. Der Aufwand zur Bestimmung der Lasdonschranke LB_{Lasdon} ist daher minimal. Die Lasdonschranke wächst nicht zwangsläufig monoton. Grund dafür sind sogenannte *Oszillationen* der Werte der dualen Variablen [LD05], die zur Berechnung der Schranke verwendet werden. Die Lasdonschranke, die in der letzten Iteration errechnet wird, muss daher nicht zwangsläufig besser sein als die zuvor berechneten. In diesem Fall liefert das Maximum aller berechneten Lasdonschranken eine sinnvolle untere Schranke.

Subproblem

Im Subproblem wird geprüft, ob die aktuelle Menge von Pfaden im RMP hinreichend groß ist, um das Optimum des MP zu enthalten. Für diese Prüfung wird Korollar 1 verwendet. Das Korollar wird auch zur Auswahl neuer Pfade benutzt, falls die Pfadmenge noch zu klein ist. Zunächst werden die reduzierten Kosten r_{kp} einer Pfadvariablen λ_p^k bestimmt. Zur Berechnung sind zwei Werte erforderlich, deren Differenz anschließend gebildet wird: vom Kostenkoeffizienten der Variable in der Zielfunktion wird das Produkt aus Nebenbedingungskoeffizienten und den Werten der zur jeweiligen Nebenbedingung dualen Variable subtrahiert.

$$\begin{aligned}
 r_{kp} &= \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} U_i x_{ij}^{kp} \\
 &\quad - \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} \mu_{ij}^d w_d^k x_{ij}^{kp} - \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} \eta_{ij}^d l_d^k x_{ij}^{kp} \\
 &\quad - \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} \sigma_i^d v_d^k x_{ij}^{kp} - \tau^k
 \end{aligned} \tag{5.4}$$

Um den unübersichtlichen Ausdruck handhabbar zu machen, wird der folgende nur von den Dualvariablen und Parametern abhängige Ausdruck Ω_{ij}^k definiert.

$$\Omega_{ij}^k = \sum_{d \in \mathcal{D}} [U_i - \mu_{ij}^d l_d^k - \eta_{ij}^d w_d^k - \sigma_i^d v_d^k] \tag{5.5}$$

Mit Hilfe von Ω_{ij}^k aus Formel (5.5) lassen sich die reduzierten Kosten (5.4) kompakt darstellen:

$$r_{kp} = \sum_{(i,j) \in \mathcal{A}} \Omega_{ij}^k x_{ij}^{kp} - \tau^k \tag{5.6}$$

Es sollen diejenigen Pfadvariablen mit minimalen reduzierten Kosten¹ bestimmt werden, welche die lokal gesehen größeren Verbesserungen des aktuellen Zielfunktionswertes des RMP erwarten lassen. Dies ist äquivalent zur Minimierung der reduzierten Kosten:

$$\min_{k,p} r_{kp} = \min_{k,p} \sum_{(i,j) \in \mathcal{A}} [\Omega_{ij}^k x_{ij}^{kp}] - \tau^k \tag{5.7}$$

Pfade sind durch ihre Kantenflüsse x_{ij}^{kp} definiert (vgl. Formel (4.12)). Um der Definition eines Pfades zu genügen, muss ein Kantenfluss die Flusserhaltungsbedingung (4.6) und die Umstellbeschränkung (4.11) erfüllen. Ein für die Relation k zulässiger Pfad p , dessen assoziierte Pfadvariable λ_p^k minimale reduzierte Kosten hat, ist daher Lösung des folgenden Minimierungsproblems (Sub):

¹Der Begriff der reduzierten Kosten wurde ursprünglich für Maximierungsprobleme und ist daher möglicherweise irreführend: Bei Minimierungsproblemen sind die minimalen reduzierten Kosten zu bestimmen

$$\begin{aligned}
(\text{Sub}) \quad & \min_x \sum_{(i,j) \in \mathcal{A}} [\Omega_{ij}^k x_{ij}^{kp}] - \tau^k \\
\text{s.t.} \quad & \sum_{i \in \mathcal{N}} x_{ji}^{kp} - \sum_{i \in \mathcal{N}} x_{ij}^{kp} = B(k, j) \quad \forall (i, j) \in \mathcal{A} \\
& \sum_{(i,j) \in \mathcal{A}} x_{ij}^{kp} \leq S_k \\
& B(k, j) = \begin{cases} 1, & \text{wenn } j \text{ Quelle von Relation } k \text{ ist} \\ -1, & \text{wenn } j \text{ Senke von Relation } k \text{ ist} \\ 0, & \text{sonst} \end{cases}
\end{aligned}$$

Die Lösung des Subproblems wird auch als *Pricing* bezeichnet. Das Problem (Sub) ist ein *Kürzeste-Wege-Problem mit Kantenbeschränkung*, das laut Satz 4 in kubischer Zeit vom Bellman-Ford-Algorithmus gelöst werden kann. Dabei wird nicht nur der kürzeste Pfad mit S_k Kanten, sondern gleichzeitig auch die kürzesten Pfade mit $2, \dots, S_k - 1$ Kanten berechnet. Auf diese Weise können innerhalb jeder einzelnen Iteration des Column-Generation-Verfahrens mehrere Pfade für jede Relation erzeugt werden. Diese Erhöhung der Anzahl generierter Pfade pro Iteration kann zu einer verringerten Anzahl nötiger Iterationen zum Auffinden des Optimums führen [LD05]. Dabei bleibt die Lösungszeit der Subprobleme polynomiell beschränkt und könnte zur weiteren Beschleunigung parallelisiert werden, um die heute üblichen Mehrkernprozessoren auszunutzen. Teilaspekte der hier eingeführten Column-Generation-Verfahren wurden bereits im Rahmen der Promotion in [VC11] veröffentlicht.

Ausgewählte Ergebnisse des Column-Generation-Verfahrens und der Tailing-off-Effekt

In Tabelle 5.1 sind Ergebnisse exemplarisch ausgewählter Testinstanzen mit einem Zeitlimit von zwölf Stunden aufgelistet. Die vollständigen Ergebnislisten und Beschreibungen zur Implementierung und zum Aufbau der Instanzen finden sich in Kapitel 6. Die Spalten T_{CPLEX} und T_{CG} enthalten die jeweilige Lösungszeit bis zur Optimalität, welche jedoch auf zwölf Stunden begrenzt wurde. Im Falle des CG-Verfahrens wurde der relative Abstand ω zwischen der besten Lasdonschranke LB_{Lasdon} und der oberen Schranke (UB) für den MP-Zielfunktionswert Z_{RMP}^* in der letzten Iteration berechnet:

$$\omega = \frac{Z_{RMP}^* - LB_{Lasdon}}{Z_{RMP}^*}$$

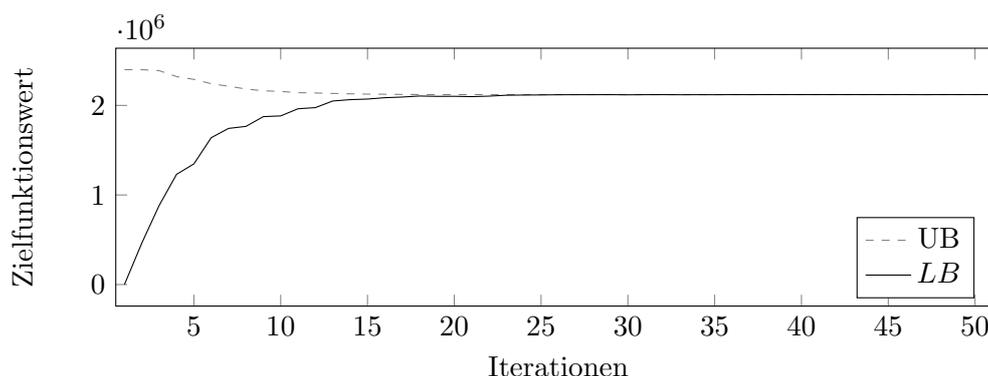
Die Spalten *Pfade* und *Iter* enthalten die Gesamtzahl der im CG erzeugten Pfade und die Anzahl der Iterationen des Verfahrens.

Tabelle 5.1: Ergebnisse der Tests mit Column Generation

	T_{CPLEX}	T_{CG}	ω	Pfade	Iter
DB-21-254-10a	681s	100s	0,00%	2.471	51
DB-21-254-10b	425s	1.059s	0,00%	4.336	139
gen-50-224-10	–	43200s	0,03%	101.118	1162
gen-100-547-7	–	43200s	26,43%	41.719	79

Das CG-Verfahren ist in der Lage, alle Instanzen zur Optimalität zu lösen, die auch mit Hilfe von CPLEX gelöst werden können. Dabei ist CG in einem Fall sogar schneller. Größere Instanzen, wie die in Tabelle 5.1 zuletzt aufgeführten *gen-50-224-10* und *gen-100-547-7*, können auch von CG nicht zur Optimalität gelöst werden. Bei der zweitgenannten Instanz ist zu beobachten, dass die Anzahl der durchgeführten Iterationen gemessen an der Laufzeit sehr gering ist. Dies deutet darauf hin, dass die Lösung der RMP viel Zeit in Anspruch nimmt. Es werden jedoch nicht-triviale untere Schranken berechnet, die zuvor nicht bekannt waren.

Es fällt auf, dass die Instanz *gen-50-224-10* im Gegensatz zu *DB-21-254-10a* und *DB-21-254-10b* nicht effizient gelöst wird, obwohl die jeweilige Anzahl der Relationen nur geringfügig abweicht. Die größere Anzahl der Knoten in *gen-50-224-10* ist eine mögliche Erklärung. Ein Vergleich der Lösungsverläufe zeigt jedoch vor allem ein algorithmisches Problem auf:

Abbildung 5.3: Verlauf von LB_{Lasdon} und UB für *DB-21-254-10a*

In den Abbildungen 5.3 und 5.4 sind jeweils die Verläufe der oberen (UB) und unteren Schranken (LB) über den durchgeführten Iterationen abgetragen. In beiden Diagrammen fällt auf, dass sich die obere Schranke nach vergleichsweise wenigen Iterationen nicht mehr verändert. Im Falle von *DB-21-254-10a* wird das Optimum innerhalb von 33 Iterationen erreicht. In den restlichen Iterationen werden immer mehr Pfade erzeugt, die jedoch in der Optimallösung keinen Beitrag leisten. Der Beweis, dass Optimalität vorliegt, dauert erheblich länger als das Finden des Optimums. Der Verlauf der unteren Schranke gibt einen Hinweis auf den Konvergenzverlauf des Verfahrens. Obwohl Abbildung 5.3 suggeriert, dass die Schranken gleich sind, ist die LB bis zur letzten Iteration

geringfügig kleiner als die UB.

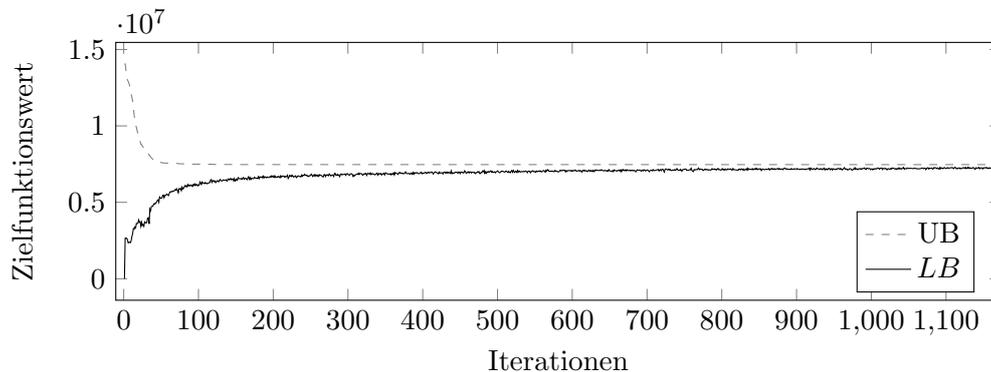


Abbildung 5.4: Verlauf von LB_{Lasdon} und UB für *gen-50-224-10*

Auch im Falle von *gen-50-224-10* stagniert die obere Schranke nach wenigen Iterationen (vgl. Abbildung 5.4). Die untere Schranke bleibt deutlich unterhalb der oberen Schranke. Die Vermutung, das Optimum sei bereits gefunden, ist naheliegend und nicht unbegründet. Das Verfahren benötigt jedoch eine hohe Zahl weiterer Iterationen, bis der Optimalitätsnachweis erbracht werden kann. Dieser Effekt wird als *Tailing-off-Effekt* bezeichnet und ist bei Column-Generation-Verfahren nicht unüblich. Wie Lübbecke und Desrosiers schreiben, ist der Effekt bisher nur teilweise verstanden [LD05]. Der Effekt wird im vorliegenden Fall vornehmlich durch die Degeneriertheit des Master Problems hervorgerufen, welche bei der hier vorliegenden Set-Covering-Struktur üblich ist. Eine Möglichkeit des Umgangs mit dem Tailing-off-Effekt, bei der die Lasdonschranke benutzt wird, wird im Rahmen der Branch-and-Price-Verfahren in Unterkapitel 5.2 umgesetzt.

Zusammenfassung Column Generation

Das Column-Generation-Verfahren ist ein Lösungsverfahren, das auf die lineare Relaxierung des pfadbasierten Modells (Opt_{path}) angewendet werden kann. Sukzessive werden neue Pfade in Polynomialzeit erzeugt, die dem RMP hinzugefügt werden. Für große Probleminstanzen liefert das Verfahren untere Schranken, die ansonsten nicht berechnet werden können, obwohl die Optimallösung der LP nicht gefunden werden.

5.2 Branch-and-Price

Das vorgestellte Column-Generation-Verfahren liefert untere Schranken für (Opt_{path}). Erst die Einbindung in ein klassisches Branch-and-Bound-Schema (vgl. Abschnitt 3.4.2) erlaubt die Lösung des ganzzahligen Modells. Branch-and-Bound besteht aus einem strukturierten Vergleich zwischen der jeweils besten verfügbaren Lösung und lokalen unteren Schranken. Die unteren Schranken kann das Column-Generation-Verfahren liefern, welches dazu in jedem Knoten des Branch-and-Bound-Baumes durchzuführen ist.

Es werden also in jedem Knoten neue Pfade generiert und damit das mit dem Knoten assoziierte LP gelöst. Das resultierende Verfahren ist exakt und als Branch-and-Price (B&P) bekannt [BJN⁺98].

Im Folgenden werden zunächst Details des Branch-and-Price diskutiert. Anschließend werden problemspezifische Schnittebenen entwickelt, welche die Lösung beschleunigen, und der Umgang mit dem Tailing-off-Effekt im Branch-and-Price erklärt.

5.2.1 Vorüberlegungen zum Branch-and-Price

Im Gegensatz zu klassischen Branch-and-Bound-Verfahren existieren für Branch-and-Price nur in bedingtem Maße generische Softwarelösungen, die die jeweilige Problemstruktur optimal ausnutzen. Definition und Implementierung des sogenannten Branchingschemas, d.h. der Aufbau des Entscheidungsbaumes und die Reihenfolge der Traversierung, sind in der Regel individuell zu gestalten [DL10], obwohl auch Arbeiten zur Entwicklung generischer Regeln existieren [Van11]. Daher ist zu diskutieren, wie das Branchingschema im vorliegenden Fall auszusehen hat. Dabei sind zwei teils konkurrierende Ziele zu verfolgen:

- (a) Einerseits sollten Branchingentscheidungen getroffen werden, die starken Einfluss auf den Zielfunktionswert der Teilprobleme haben. Ansonsten muss der Baum bis in große Tiefen durchsucht werden, bevor der Zielfunktionswert des LP im betrachteten Knoten größer ist als der Zielfunktionswert der besten bekannten Lösung. In Folge dessen könnten Äste des Baumes erst spät im Verlauf der Suche verworfen werden. Dadurch stiege die Anzahl zu durchsuchender Knoten.
- (b) Andererseits sollte das Branchingschema die Suche in Teile des Baumes führen, in denen vielversprechende, ganzzahlige Lösungen zu finden sind. Dies ist notwendig, weil ganzzahlige Lösungen aus praktischer Sicht wichtig sind und weil gute ganzzahlige Lösungen das frühere Verwerfen von Ästen erlauben. Daher ist es sinnvoll, Entscheidungen in jeder Iteration derart zu treffen, dass die Anzahl fraktionaler, d.h. nicht-ganzzahliger, Variablen verringert wird.

Das zu lösende Modell (Opt_{path}) hat zwei Typen von Entscheidungsvariablen: die Designvariablen n_{ij} und die Pfadvariablen λ_p^k . Die Kostenstruktur des EWV, die in der Zielfunktion von (Opt_{path}) modelliert wurde, setzt sich aus Transport- und Umstellkosten zusammen. Im Falle praxisnaher Instanzen entfällt der größere Anteil der Gesamtkosten auf die Transportkosten, welche durch die Designvariablen bestimmt sind. In Anbetracht der Argumente aus (a) ist es naheliegend, zunächst über die n_{ij} zu branchen. Numerische Tests haben jedoch gezeigt, dass durch diese Entscheidung mehr Pfadvariablen fraktional werden. Dieser Effekt läuft dem in (b) genannten Ziel guter ganzzahliger Lösungen zuwider.

Eine Alternative ist das Branching über die Pfadvariablen. Die Variablen λ_p^k bestimmen, ob die Wagen der Relation k über Pfad p transportiert werden. Ein Branching über diese Variablen erzeugt demnach zwei Äste des Baumes. Im Falle $\lambda_p^k = 1$ wird der Pfad p für Relation k erzwungen, im Falle $\lambda_p^k = 0$ wird der Pfad verboten. Während

das Erzwingen eines Pfades zu einem sehr unsymmetrischen Entscheidungsbaum führt, entstehen beim Verbieten eines Pfades algorithmische Probleme im weiteren Verlauf des Verfahrens: In jedem Knoten des Baumes werden mit Hilfe des ins CG-Verfahren eingebetteten Kürzeste-Wege-Algorithmus neue Pfade generiert. Dabei kann jedoch nicht vermieden werden, dass der zuvor verbotene Pfad p immer wieder als kürzester Pfad erzeugt wird. Es müsste daher zunächst der zweitkürzeste Pfad erzeugt werden. Werden weitere Pfade für Relation k verboten, verstärkt sich dieser Effekt. Die effiziente Lösung der Subprobleme in tieferen Regionen des Baumes wird stark beeinträchtigt, da bei κ verbotenen Pfaden ein κ -kürzester Weg zu berechnen wäre. Das Branching über die Pfadvariablen λ_p^k ist daher zu vermeiden.

An dieser Stelle ist die Äquivalenz der extensiven Formulierung (Opt_{path}) und der kompakten Formulierung (Opt_{arc}) hilfreich, die durch Gleichung (4.12) gezeigt wurde. Die pfadbasierten Flüsse in (Opt_{path}) setzen sich aus Kantenflüssen, wie sie (Opt_{arc}) charakterisieren, zusammen. Es besteht daher die Möglichkeit, über die Kantenflussvariablen x_{ij}^k zu branchen. Die Entscheidung $x_{ij}^k = 0$ bedeutet in diesem Fall, dass alle Pfade der Relation k verboten werden, in denen die Kante (i, j) enthalten ist. Dies kann im Subproblem des CG-Verfahrens auf einfache Weise umgesetzt werden, indem die Kante (i, j) aus dem Graph entfernt wird und dadurch nicht mehr in neuen Pfaden auftreten kann. Komplizierter ist hingegen die Entscheidung $x_{ij}^k = 1$, welche die Wahl von Pfaden erzwingt, die die Kante (i, j) enthalten. Das Erzwingen bestimmter Kanten beeinflusst die Generierung neuer Pfade im Subproblem des CG-Verfahrens. Das dort zu lösende Kürzeste-Wege-Problem ist unter der Bedingung, dass bestimmte Kanten gezwungenermaßen zu verwenden sind, nicht mehr in Polynomialzeit zu lösen. Im Gegensatz zum allgemeinen Fall ist im hier betrachteten Modell (Opt_{path}) jedoch die Anzahl erlaubter Kanten pro Pfad beschränkt. Dies ist hilfreich, wenn Entscheidungen der Form $x_{ij}^k = 1$ getroffen werden, wie im Folgenden sukzessive erläutert wird (vgl. auch Abbildung 5.5). Dazu werde die Anzahl der erzwungenen Kanten einer Relation mit q bezeichnet:

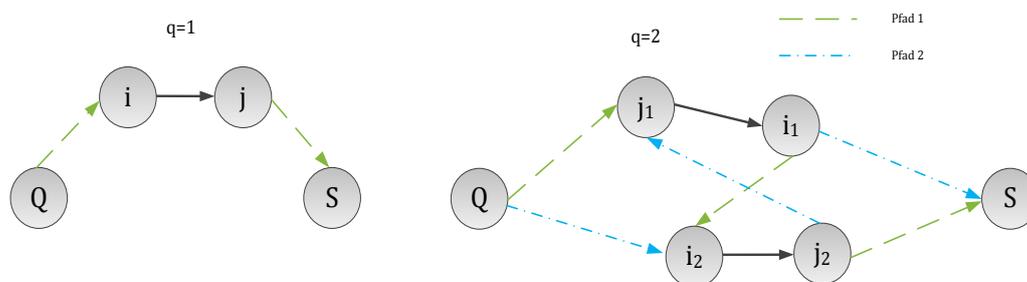


Abbildung 5.5: Visualisierung des Subproblems für $q = 1$ und $q = 2$

- $q = 1$: Einen Pfad zu erzeugen, der genau eine gegebene Kante (i, j) enthalten muss, ist simpel. Er setzt sich aus einem kürzesten Weg von der Quelle Q zu Knoten i , der Kante (i, j) selbst und einem kürzesten Weg von j zur Senke S der Relation zusammen. Dabei ist auf die Beschränkung der Anzahl benutzter Kanten zu achten, die in der Pfaddefinition 16 enthalten ist. Dieses Problem kann in $\mathcal{O}(|N|^2)$ gelöst werden.
- $q = 2$: Bei zwei erzwungenen Kanten $e_1 = (i_1, j_1)$ und $e_2 = (i_2, j_2)$ wird das Problem komplexer, wenn unbekannt ist, in welcher Reihenfolge die beiden Kanten zu benutzen sind. Ist eine Reihenfolge, z.B. e_1 vor e_2 , gegeben, ist das Problem auf den Fall $q=1$ zurückzuführen: Ein kürzester Pfad über e_1 und e_2 setzt sich aus einem kürzesten Weg von der Quelle der Relation über e_1 nach i_2 (Fall $q=1$), der Kante e_2 selbst und einem kürzesten Weg von i_2 zur Senke zusammen. Da im Allgemeinen die Kantenreihenfolge nicht bekannt ist, müssen für $q=2$ zwei Reihenfolgen betrachtet werden.
- $q > 2$: Das Prinzip, das für $q=2$ beschrieben wurde, kann rekursiv auch für größere Werte von q fortgesetzt werden. In diesem Fall wächst allerdings die Anzahl zu betrachtender Reihenfolgen schnell an. Es sind stets $q!$ Reihenfolgen zu betrachten. Die Laufzeit ist mit $\mathcal{O}(q!|\mathcal{N}|^2)$ abzuschätzen.

Der beschriebene Ansatz für das Subproblem nutzt die allgemeine Idee der Bellman-Optimalität [AMO93]: Ein kürzester Pfad besteht aus Teilpfaden, die ebenfalls minimale Länge haben. Der rekursive Ansatz mit $q!$ Reihenfolgen ist nicht mehr in Polynomialzeit lösbar. Da jedoch q für jede Relation durch die maximale Anzahl S_k erlaubter Kanten für einen Pfad beschränkt ist, ist der Aufwand für die Berechnung vertretbar, solange S_k klein genug ist. S_k wird als maximale Anzahl der Umstellungen jedes Wagens im Praxisfall höchstens als fünf gewählt, weil ansonsten Transportzeitversprechen nicht einzuhalten sind. Damit ist der Parameter hinreichend klein. Um die Rechenzeit zu verkürzen, ist es möglich, die Berechnung abubrechen, sobald der erste zulässige Pfad mit positiven reduzierten Kosten gefunden wurde. Alternativ kann das Subproblem auch mit Hilfe eines entsprechenden Lösungsverfahrens als ganzzahliges Programm gelöst werden. Die rekursive Berechnung neuer Pfade im beschriebenen Branchingschema macht eine Anwendung des Verfahrens praktikabel. Zur Beschleunigung des Verfahrens werden dem Branch-and-Price im Folgenden weitere Elemente hinzugefügt.

5.2.2 Entwicklung von Schnittebenen und Erweiterung zu Branch-and-Price-and-Cut

Das Branch-and-Price-Verfahren berechnet unter Berücksichtigung der bisherigen Entscheidungen in jedem Knoten des Entscheidungsbaumes eine untere Schranke für den Zielfunktionswert. Bessere Schranken erlauben es dabei, Äste des Baumes, in denen die Optimallösung nicht liegt, früher zu verwerfen. In Abschnitt 4.2.2 wurde die Problematik schwacher Schranken anhand der Struktur der Polyeder diskutiert, die den zulässigen Lösungsbereich der linearen Relaxierungen darstellen. Das Kernproblem bestand darin,

dass die Zugkosten nach der Relaxierung nicht mehr sinnvoll abgebildet wurden. In Folge dessen ging auch der angestrebte Bündelungseffekt (vgl. Abschnitt 3.2.2) verloren. Im Verlauf des Branch-and-Price-Verfahrens kann diese Problematik zumindest in Teilen mit Hilfe angepasster Schnittebenen vermindert werden, die für bessere untere Schranken sorgen:

Das beschriebene Branching erzwingt oder verbietet für bestimmte Relationen die Benutzung bestimmter Kanten. Wird die Entscheidung $x_{ij}^k = 1$ getroffen, muss Relation k über die Kante (i, j) transportiert werden. Aus der Anschauung und den Längen- und Gewichtsrestriktionen (4.7) und (4.8) wird offensichtlich, dass hierzu Transportkapazitäten erforderlich sind. Es sei $\mathcal{K}_{fix}(i, j)$ die Teilmenge aller Relationen \mathcal{K} , für welche die Entscheidung $x_{ij}^k = 1$ getroffen wurde. Dann lässt sich die folgende Schnittebene für die Kante (i, j) ableiten:

$$n_{ij} \geq \left[\max_{d \in \mathcal{D}} \left\{ \frac{\sum_{\mathcal{K}_{fix}(i,j)} l_d^k}{L_{ij}}, \frac{\sum_{\mathcal{K}_{fix}(i,j)} w_d^k}{W_{ij}} \right\} \right] \quad (5.8)$$

Die aggregierten Aufkommen der Relationen in $\mathcal{K}_{fix}(i, j)$ induzieren eine Mindestzahl an Zügen, die durch den Term im Klammerinneren zu berechnen ist. Maßgeblich ist dabei der aufkommenstärkste Tag, der durch den Maximumterm ermittelt wird. Das Aufrunden spiegelt die Ganzzahligkeit der Designvariablen n_{ij} wider. Die Schnittebenen (5.8) enthalten als Spezialfall die folgende Menge klassischer Schnittebenen aus der Designtheorie für Mehrgüternetzwerke, die bspw. bei Hewitt et al. [HNS10] Erwähnung finden:

$$n_{ij} \geq x_{ij}^k \quad (5.9)$$

Die beschriebenen Schnittebenen können in jedem Knoten effizient berechnet und dem Problem hinzugefügt werden. Die Schnittebenen sind nur lokal gültig, da sie von den bisher betrachteten Entscheidungen abhängen. Sie können somit hilfreich sein, um bessere Schranken in den Knoten des Entscheidungsbaumes zu erhalten und werden daher zur Erweiterung des B&P zu einem Branch-and-Price-and-Cut-Verfahren (B&P&C) verwendet.

5.2.3 Der Tailing-off-Effekt im Branch-and-Price-and-Cut

Im Abschnitt 5.1.3 wurde der Tailing-off-Effekt beschrieben, der bei einigen, speziell größeren Instanzen dazu führte, dass die lineare Relaxierung nicht zur Optimalität gelöst werden konnte. Die Lösung der Relaxierungen bis zur Optimalität ist jedoch beim Branch-and-Bound bzw. Branch-and-Price nicht zwingend erforderlich, sofern es möglich ist, in jedem Knoten des Entscheidungsbaumes eine untere Schranke zu bestimmen. Die im Rahmen der Column-Generation-Verfahren beschriebene Lasdonschranke bietet genau diese Möglichkeit. Die Lasdonschranke ist nicht die theoretisch beste, aber dennoch eine gültige untere Schranke. Sie erlaubt die Anwendung eines Verfahrens, das von Lübbecke und Desrosiers [LD05] als verfrühter Abbruch (*engl.: Early Termination (ET)*) bezeichnet wird. Die Erzeugung neuer Pfade im Rahmen des CG-Verfahrens wird bei

der ET abgebrochen, sobald z.B. der relative Abstand zwischen der aktuellen Lösung des RMP und der Lasdonschranke (5.1.3) ein vorgegebenes Maß ω unterschreitet. Alternativ wäre auch eine Zeitvorgabe ein denkbare Abbruchkriterium. Das Branch-and-Price bleibt trotz dieser Anpassung ein exaktes Verfahren. Die Schranken in den Knoten sind zwar schwächer als bei vollständiger Berechnung aller Pfade, können dafür aber in einem Bruchteil der Zeit berechnet werden.

Zusammenfassung Branch-and-Price-and-Cut

Branch-and-Price ist eine spezielle Variante des Branch-and-Bound, die eingesetzt werden kann, um das pfadbasierte Modell (Opt_{path}) zu lösen. Die Besonderheit liegt in der Lösung der linearen Relaxierungen in den Knoten des Entscheidungsbaumes, welche mit Hilfe von CG durchgeführt werden. Im CG wird die Subproblemstruktur ausgenutzt, indem nur Kürzeste-Wege-Probleme gelöst werden. Da diese Struktur abhängig von den Branchingentscheidungen verloren geht, gestaltete sich das Design eines Branchingschemas als vergleichsweise komplex. Die Diskussion verschiedener Varianten führte zu der Erkenntnis, dass das Branching über die Variablen der kompakten Formulierung (Opt_{arc}) vielversprechend ist, wodurch jedoch die Pfaderzeugung erschwert wird. Diesem nicht unüblichen Problem kann im vorliegenden Problem mit einem rekursiven Ansatz begegnet werden. Das Branching bietet überdies den Vorteil, dass aus den Entscheidungen problemspezifische Schnittebenen abgeleitet werden können, um die in den Knoten bestimmten unteren Schranken zu verbessern. Die dem CG-Verfahren inhärente Problematik des Tailing-Off-Effekts kann umgangen werden, indem die Pfaderzeugung an geeigneter Stelle abgebrochen wird. Numerische Ergebnisse, welche die Überlegenheit des Branch-and-Price-and-Cut gegenüber kommerziellen Solvern zeigen, finden sich in Kapitel 6. Die Lösung von Realinstanzen in praxistauglicher Zeit und Lösungsgüte ist jedoch auch mit dem entwickelten Branch-and-Price-and-Cut derzeit nicht möglich. Verschiedene Strategien zur Verbesserung des Verfahrens aus [DDS02] wurden erfolglos getestet. Daher werden im Folgenden metaheuristische Ansätze zur Lösung von (Opt_{path}) bzw. (Opt_{arc}) betrachtet. Die hier vorgestellten Ergebnisse zu CG und B&P&C wurden bereits im Rahmen früherer Arbeiten veröffentlicht [VC11, VC13a].

5.3 Ein metaheuristisches Lösungsverfahren

Trotz der guten Ergebnisse, die das Branch-and-Price-Verfahren aus Abschnitt 5.2 für einige kleine Instanzen liefert, ist die Lösung von Instanzen realer Größe mit Hilfe exakter Verfahren derzeit nicht praktikabel. Angesichts der Tatsache, dass auch moderne Verfahren schon bei der Lösung der linearen Relaxierung großer Instanzen an ihre Grenzen stoßen, ist die exakte Lösung des Problems höchst anspruchsvoll. Die Betrachtung (meta-)heuristischer Lösungsansätze zum Auffinden praxistauglicher Lösungen erscheint daher naheliegend. Es wird zunächst ein Vorgehen festgelegt und anschließend ein metaheuristischer Ansatz mit Komponenten der exakten Optimierung, d.h. eine Matheuristik, entwickelt.

5.3.1 Entwicklung von metaheuristischen Lösungsverfahren

Wie bereits in Abschnitt 3.4.3 angedeutet, existiert eine Vielzahl metaheuristischer Ansätze zur Lösung ganzzahliger Optimierungsprobleme. Sörensen [Sör13] gibt Empfehlungen, wie eine systematische Entwicklung und Erforschung von Metaheuristiken zu wissenschaftlichem Mehrwert führen kann. Aus seinen Empfehlungen lässt sich ein mehrstufiges Vorgehen ableiten, das in dieser Arbeit den Rahmen für die Entwicklung einer Metaheuristik vorgeben wird:

1. Der erste Schritt zur Lösung eines gegebenen Problems besteht in der Identifikation und Auswahl verwandter Optimierungsprobleme, für die wirksame Metaheuristiken bekannt sind. Die Anwendung eines *Analogieprinzips* soll die Übertragung erfolgversprechender Strukturen ermöglichen.
2. Der nächste Schritt besteht in der *Zerlegung der Metaheuristiken in ihre Komponenten*, um den Einfluss der einzelnen Komponenten auf das Gesamtverfahren zu analysieren. Dies ist insbesondere erforderlich, wenn mehrere metaheuristische Ansätze miteinander kombiniert werden. Die komponentenweise Betrachtung einer Metaheuristik erlaubt es, Gemeinsamkeiten guter Verfahren für ein Problem zu identifizieren. Darüber hinaus wird es möglich, unwirksame Strukturen, wie z.B. überflüssige stochastische Komponenten, zu entfernen und damit *schlanke* Ansätze zu entwickeln.
3. Die Entwicklung einer adaptiven Metaheuristik, bei der der Benutzer keine Parameter einstellen muss, ist zwar wünschenswert, aber in der Regel schwer zu realisieren. Die Anzahl der Parameter sollte demnach möglichst klein gehalten werden. Dadurch wird es möglich, das *Parametertuning* systematisch und transparent durchzuführen. Dazu schlagen Silberholz und Golden [SG10] eine Teilung der Menge der Testinstanzen vor. Eine repräsentative Teilmenge aller Instanzen sollte gewählt werden, um das Parametertuning durchzuführen. Die übrigen Instanzen sollten mit den ermittelten Parametern gelöst werden. Dadurch wird verhindert, dass die Testresultate verfälscht werden, weil die Parameter auf alle Testinstanzen abgestimmt sind.
4. Der letzte Schritt besteht im *systematischen Testen* der Metaheuristik unter Verwendung statistischer Analysemethoden. Die Herkunft und Zusammenstellung aller Instanzen sollte transparent sein.

Die Anwendung des beschriebenen Vorgehens auf das Modell (Opt_{arc}) wird im Folgenden beschrieben.

5.3.2 Anwendung des Analogieprinzips

Dem oben beschriebenen Vorgehen folgend gilt es, metaheuristisch gelöste Optimierungsprobleme zu identifizieren, die mit dem vorliegenden Problem (Opt_{arc}) verwandt sind. Dazu wurden drei wissenschaftliche Arbeiten aus der jüngeren Vergangenheit ausgewählt.

1. Ahuja et al. [AJL07] betrachten in ihrer Arbeit das Blockingproblem. Wie in Abschnitt 3.2.2 dargestellt wurde, ist das Blockingproblem verwandt mit dem im Modell (Opt_{arc}) modellierten Wagenflussproblem. Der zentrale Unterschied besteht in den nahezu unbeschränkten Zugkapazitäten. Der vorgeschlagene Algorithmus betrachtet eine zulässige Lösung und durchläuft sukzessive die Knoten des Netzwerkes. Es wird lokal überprüft, ob die Etablierung neuer Zugverbindungen Einsparungen bringen kann. Anschließend werden Relationen den neuen Verbindungen entsprechend geroutet. Mit den neuen Routen wird die gegebene Lösung aktualisiert. Der Vorgang wird wiederholt, bis ein lokales Optimum erreicht ist. Die Suche nach einem globalen Optimum wird nicht beschrieben. Nichtsdestotrotz ist das Verfahren den Autoren zufolge auch bei Realinstanzen so erfolgreich, dass es bei mehreren nordamerikanischen Eisenbahnunternehmen eingesetzt wird.
2. Homfeld präsentiert in seiner Dissertation [Hom12] ein Verfahren zur Optimierung von Leitwegeplänen, das er als *Rip-up-and-Reroute* bezeichnet. Mit Hilfe eines komplizierten Bewertungsschemas werden Relationen gewählt, die neu geroutet werden. Um lokalen Optima zu entkommen, wird mit Hilfe eines Simulated-Annealing-Ansatzes eine zulässige Lösung aus einem Lösungspool gewählt. Das Verfahren enthält also eine stochastische Komponente.
3. Das Modell (Opt_{arc}) gehört zur Klasse der *kapazitierten Mehrgüternetzwerk Design Probleme mit Fixkosten (FMCNDP)*, wie in Unterkapitel 4.1 bereits festgestellt wurde. Hewitt et al. [HNS10] stellen eine Metaheuristik für derartige Netzwerke vor. Sie betrachten ein verallgemeinertes Modell für das Netzwerkdesign. Ihre Metaheuristik, die sie *IP Search* nennen, erzeugt mit verschiedenen Prozeduren aus einer gegebenen Lösung sehr große Nachbarschaften, die mit Hilfe eines Solvers für ganzzahlige Programme durchsucht wird. Die Ergebnisse von Optimierungsläufen für verschiedene Instanzen zeigen die Leistungsfähigkeit des Verfahrens.

Die Gemeinsamkeit der drei genannten Verfahren ist offensichtlich: Eine gegebene Lösung wird mit Hilfe einer lokalen Suche iterativ verfeinert. Obwohl die Nachbarschaften auf unterschiedlichen Wegen erzeugt werden, findet bei allen Verfahren ein Umrouten einer gewissen Anzahl von Sendungen bzw. Relationen statt. Die Hauptunterschiede liegen neben den zu Grunde liegenden Modellen in der Strategie zur Vermeidung lokaler Optima. Während sich Ahuja et al. mit einer lokal optimalen Lösung begnügen, lässt Homfeld in Form des Simulated Annealing stochastische Einflüsse zu. Auch Hewitt et al. versuchen, lokale Optima durch stochastische Komponenten zu meiden.

Das Durchsuchen der Nachbarschaften findet bei Ahuja et al. mit Hilfe einer Heuristik und bei Hewitt et al. durch exakte Verfahren der ganzzahligen Optimierung statt. Homfeld betrachtet das sequentielle Routen von Relationen entlang angepasster kürzester Wege sowie das Routen ganzer Gruppen von Relationen mittels ganzzahliger Optimierung. Der Vorteil der ganzzahligen Optimierung besteht vor allem darin, dass mit ihrer Hilfe auch sehr große Nachbarschaften effizient durchsucht werden können.

Auf Grund der Analyse der beschriebenen Ansätze wird im Folgenden eine Metaheuristik mit folgenden Eigenschaften konzipiert:

- Eine gegebene zulässige Lösung wird mit Hilfe einer lokalen Suche sukzessiv verbessert.
- Die Nachbarschaften werden durch Wahl umzuroutender Relationen aus der aktuellen Lösung erzeugt.
- Zur Vermeidung lokaler Optima werden im Sinne einer variablen Nachbarschaftssuche (vgl. Abschnitt 3.4.3) wechselnde Nachbarschaften erzeugt, aber es wird auf stochastische Elemente verzichtet, um verlässliche, deterministische Ergebnisse zu erhalten.
- Das Durchsuchen der Nachbarschaften wird mit Hilfe exakter Verfahren der ganzzahligen Optimierung durchgeführt.

Da die Suche nach neuen Lösungen in den Nachbarschaften mit Hilfe exakter Verfahren geschieht, kann das zu entwickelnde Verfahren als Matheuristik (vgl. Definition 15) bezeichnet werden. Puchinger und Raidl [PR08] bezeichnen Matheuristiken, die eine übergeordnete Metaheuristik mit exakten Verfahren zur Lösung von Subproblemen benutzen, als *integrativ*. Die Tabelle 5.2 ordnet das neue Verfahren in den Kontext der oben genannten, relevanten Ansätze ein.

	stochastische Elemente	deterministisch
heuristische Suche	Homfeld [Hom12]	Ahuja [AJL07]
exakte Suche	Homfeld [Hom12], Hewitt [HNS10]	Matheuristik

Tabelle 5.2: Einordnung der Matheuristik in frühere Ansätze

5.3.3 Eine Matheuristik zur Lösung von (Opt_{arc})

Es wird der folgende Rahmenalgorithmus 1 für die Lösung von (Opt_{arc}) konzipiert, dessen Elemente nachfolgend näher zu definieren sind:

Die dynamische Vergrößerung des Parameters γ , der die Nachbarschaftsgröße steuert, ist notwendig, wenn bei zu langer Stagnation der aktuellen Lösung (\hat{x}, \hat{n}) keine Nachbarschaften mehr erzeugt werden, die noch nicht durchsucht wurden. Dies entspricht dem Konzept der VDNS aus Abschnitt 3.4.3. Bei hinreichend langer Laufzeit wachsen die betrachteten Nachbarschaften kontinuierlich bis die Nachbarschaft des Subproblems der gesamten zulässigen Menge entspricht. Vom theoretischen Standpunkt aus betrachtet ist Algorithmus 1 somit exakt. In Wirklichkeit ist dadurch jedoch nichts gewonnen, weil das zu lösende Subproblem in diesem Fall dem vollständigen Problem (Opt_{arc}) entspricht. Nichtsdestotrotz erhöht der VDNS-Ansatz die Wahrscheinlichkeit, einmal erreichte, lokale Optima wieder verlassen zu können. Ideen zu diesem Rahmenalgorithmus wurden bereits vorab in [VC13b] veröffentlicht.

Algorithmus 1 Rahmenalgorithmus Matheuristik**Input:** Anzahl umgerouteter Relationen pro Iteration $\gamma \in \mathbb{N}$, Zeitlimit T_{max} *PHASE I:*Konstruiere eine zulässige Startlösung (\hat{x}, \hat{n}) *PHASE II:***while** $t < T_{max}$ **do** Wähle eine Menge Γ mit γ freien Relationen Finde optimale Routen für die Relationen in Γ Aktualisiere (\hat{x}, \hat{n}) **if** alle Umgebungen der Größe γ durchsucht **then** Vergrößere γ **end if****end while****Output:** (\hat{x}, \hat{n})

Zur Implementierung sind vor allem drei Elemente des Rahmenalgorithmus 1 zu konkretisieren: die Konstruktion einer Startlösung, die Wahl der Menge Γ der freien Relationen und die Suche nach optimalen Routen für die Relationen in Γ . Die Schritte sind leichter zu verstehen, wenn sie in umgekehrter Reihenfolge ihres Auftretens im Rahmenalgorithmus erklärt werden:

Finden optimaler Routen für die Relationen in Γ

Gegeben seien eine zulässige Lösung (\hat{x}, \hat{n}) und eine Teilmenge aller Relationen $\Gamma \subseteq \mathcal{K}$. Die Relationen $k \in \Gamma$ werden als *frei* bezeichnet. Gesucht ist eine zulässige Lösung der betrachteten Instanz von (Opt_{arc}) , die einen besseren Zielfunktionswert als (\hat{x}, \hat{n}) besitzt. Da nur die Relationen $k \in \Gamma$ neu zu routen sind, können die Flussvariablen aller nicht freien Variablen auf den Wert der gegebenen Lösung fixiert werden:

$$x_{ij}^k = \hat{x}_{ij}^k \quad \forall k \notin \Gamma \quad (5.10)$$

Über den Flussvariablen der freien Relationen muss optimiert werden. Des Weiteren müssen alle Designvariablen Teil der Optimierung sein. Durch die Festlegung der Routen der Flüsse der Relationen $k \notin \Gamma$ können jedoch die unteren Schranken (5.11) für die Designvariablen bestimmt werden. Die untere Schranke ist analog zu den Schnittebenen (5.8) aus dem B&P&C-Verfahren:

$$n_{ij} \geq \left[\max_{d \in \mathcal{D}} \left\{ \frac{\sum_{k \in \mathcal{K} \setminus \Gamma} l_d^k}{L_{ij}}, \frac{\sum_{k \in \mathcal{K} \setminus \Gamma} w_d^k}{W_{ij}} \right\} \right] \quad (5.11)$$

Das zu lösende Modell enthält bei einer Instanzgröße von $|\mathcal{N}|$ Knoten maximal $\frac{|\mathcal{N}|^2}{2}$ Kanten. Das Modell besteht aus maximal $\frac{|\mathcal{N}|^2}{2}$ Design- und $\gamma \frac{|\mathcal{N}|^2}{2}$ Flussvariablen, d.h.

insgesamt höchstens $(\gamma + 1) \frac{|\mathcal{N}|^2}{2}$ Variablen. Die Größe des Subproblems wächst somit linear mit dem Parameter γ . Wird γ größer gewählt, wächst auch die durchsuchte Nachbarschaft und damit das Verbesserungspotential der Optimierung.

Ein Subproblem mit γ freien und vielen fixierten Relationen ist leichter zu lösen als ein Problem mit gleich vielen freien, aber keinen fixierten Relationen. Der Grund dafür liegt darin, dass die fixierten Relationen bereits ein fast vollständiges Transportnetz aufspannen, das nur geringfügig angepasst werden muss, um die freien Relationen ebenfalls zu transportieren. Modellseitig zeigt sich dies in sehr engen unteren Schranken während der Optimierung durch den Solver. Je kleiner der Anteil der freien Relationen an der Gesamtzahl der Relationen $\frac{\gamma}{|\mathcal{K}|}$ ist, desto enger sind die unteren Schranken und desto kürzer ist die für das Subproblem benötigte Lösungszeit.

Der Parameter γ steuert somit das Verhältnis zwischen der potentiellen Verbesserung und der Laufzeit des Optimierungsschrittes. Eine statistische Analyse, wie γ vorteilhaft zu wählen ist, findet sich in Abschnitt 6.5.2.

Definition der Nachbarschaft und Wahl von Γ

Die im Rahmenalgorithmus 1 durchzuführende Wahl der freien Relationen Γ entspricht der Wahl der Nachbarschaft in der lokalen Suche, wie sie in Abschnitt 3.4.3 beschrieben wurde. Die Wahl einer Menge freier Relationen Γ induziert die folgende Nachbarschaft:

$$\mathcal{N}(\hat{x}, \hat{n}) = \{(x, n) \mid x_{ij}^k = \hat{x}_{ij}^k \quad \forall k \notin \Gamma\} \quad (5.12)$$

Die Struktur der Nachbarschaft ist entscheidend für die Performanz einer lokalen Suche und enthält, umgangssprachlich ausgedrückt, die Intelligenz des Verfahrens. Die Nachbarschaft muss so gewählt sein, dass sie in jedem Schritt in sinnvoller Zeit durchsucht werden kann. Dies hängt, wie oben gezeigt, in erster Linie von der Wahl des Parameters γ ab. Besonders entscheidend für den Erfolg einer lokalen Suche ist jedoch auch die Fähigkeit des Verfahrens, ein einmal erreichtes lokales Optimum in den folgenden Iterationen zu verlassen. Auch diese Eigenschaft des Verfahrens ist in erster Linie eine Folge der Nachbarschaftswahl. Daher ist der Wahl der γ freien Relationen, von denen die Nachbarschaft induziert wird, eine hohe Bedeutung beizumessen. Im Folgenden werden drei unterschiedliche Nachbarschaftsstrukturen definiert und diskutiert.

1. Nachbarschaft auf Basis freier Zugkapazitäten

Die Zielfunktion von (Opt_{arc}) wird dominiert von den Transportkosten, die von den Designvariablen n_{ij} induziert werden. Deshalb ist die gezielte Bündelung von Verkehrsströmen ein wichtiger Effekt der Optimierung. Bündelung bedeutet, dass die Wagen verschiedener Relationen mit einem Zug transportiert werden. Sie werden also im Modell über eine gemeinsame Kante transportiert. Kanten, auf denen wenig gebündelt wird, d.h. auf denen nur wenige Relationen verkehren, müssen jedoch nicht weniger effizient genutzt sein als solche, auf denen viel Bündelung stattfindet. Die Auslastung der Züge auf einer Kante gibt bessere Hinweise auf die Effizienz der Routingentscheidungen für die Relationen, die über diese Kante transportiert werden. Diese Erkenntnis wird im

Folgenden zur Wahl der freien Relationen Γ eingesetzt.

Gegeben sei eine zulässige Lösung $(\widehat{x}, \widehat{n})$. Die Auslastung einer Kante bzgl. dieser Lösung hängt vom Aufkommen im betrachteten Szenario $d \in \mathcal{D}$ ab. Es wird die freie Kapazität (5.13) auf einer Kante $(i, j) \in A$ in Szenario $d \in \mathcal{D}$ berechnet.

$$u(i, j, d) = \widehat{n}_{ij} - \max_{d \in \mathcal{D}} \left\{ \frac{\sum_{k \in \{k | \widehat{x}_{ij}^k = 1\}} l_d^k}{L_{ij}}, \frac{\sum_{k \in \{k | \widehat{x}_{ij}^k = 1\}} w_d^k}{W_{ij}} \right\} \quad (5.13)$$

Je größer die ungenutzte Zugkapazität auf einer Kante ist, desto größer sind die Grenzkosten für den Transport jedes Wagens auf dieser Kante. Da im Modell jedoch in der Regel nicht nur ein Szenario betrachtet wird, sollte die Freikapazität über die Anzahl der Szenarien $|\mathcal{D}|$ gemittelt werden:

$$\overline{u(i, j)} = \sum_{d \in \mathcal{D}} \frac{u(i, j, d)}{|\mathcal{D}|} \quad (5.14)$$

Algorithmus 2 zur Wahl der Menge der freien Relationen Γ nutzt diese Auslastungskennzahl $\overline{u(i, j)}$:

Algorithmus 2 Nachbarschaft auf Basis freier Zugkapazitäten

Input: zulässige Lösung $(\widehat{x}, \widehat{n})$, Parameter γ

for all $(i, j) \in A$ **do**

Berechne $\overline{u(i, j)}$

end for

Initialisiere $\Gamma = \emptyset$

while $|\Gamma| < \gamma \wedge \exists (i, j) : \overline{u(i, j)} > 0$ **do**

Wähle Kante (i', j') mit maximaler durchschnittlicher Freikapazität:

$$(i', j') = \arg \max_{(i, j) \in A} \overline{u(i, j)}$$

Füge Relationen, die auf (i', j') verkehren, den freien Relationen hinzu:

$$\Gamma := \Gamma \cup \{k \in \mathcal{K} \mid \widehat{x}_{i', j'}^k = 1\}$$

Setze $\overline{u(i', j')} := -1$

end while

Output: Menge freier Relationen Γ

Aus Gründen der Übersichtlichkeit ist in Algorithmus 2 Folgendes nicht beschrieben: Damit bei Stagnation des Zielfunktionswertes nicht mehrmals in Folge die gleiche Nachbarschaft erzeugt wird, werden in einer Tabuliste diejenigen Kanten vermerkt, die bei der Erzeugung der Nachbarschaft bereits untersucht wurden. Wird keine verbesserte Lösung gefunden, werden in der nächsten Iteration andere Kanten untersucht und somit eine andere Nachbarschaft definiert. Bei Stagnation des Zielfunktionswertes wird

somit die Nachbarschaftsstruktur angepasst. Dieses Vorgehen ist charakteristisch für VNS-Ansätze (vgl. Abschnitt 3.4.3). Die wechselnde Struktur der Nachbarschaft verringert die Wahrscheinlichkeit, dass der Algorithmus gegen ein lokales, jedoch nicht global Optimum konvergiert. Testergebnisse zur Leistungsfähigkeit des Verfahrens werden in Unterkapitel 6.6 vorgestellt.

2. Nachbarschaft auf Basis von Knotenauslastungen

Der von Ahuja et al. [AJL07] mit praktischem Erfolg eingesetzte Algorithmus wählt Nachbarschaften, indem er iterativ die Knoten durchläuft und Relationen wählt, die dort umgestellt werden. Eine leichte Variation dieser Methodik wird in Algorithmus 3 eingesetzt. Eines der Entwicklungsziele bei der Modellierung von (Opt_{arc}) war die Entlastung zentraler Knoten, wie es in den nordamerikanischen Blockingmodellen üblich ist. Um diesem Ziel gerecht zu werden, werden die Knoten in Algorithmus 3 in absteigender Reihenfolge ihrer Auslastung \bar{U}_i durchlaufen. Die Auslastung wird bzgl. der verwendeten Richtungsleise R_i eines Knotens $i \in \mathcal{N}$ berechnet.

Algorithmus 3 Nachbarschaft auf Basis von Knotenauslastungen

Input: zulässige Lösung (\hat{x}, \hat{n}) , Parameter γ

for all $i \in \mathcal{N}$ **do**

Berechne Auslastung $\bar{U}_i = \sum_j \hat{n}_{ij} / R_i$

end for

Initialisiere $\Gamma = \emptyset$

while $|\Gamma| < \gamma \wedge \exists i : \bar{U}_i > 0$ **do**

Wähle Knoten i' mit maximaler Auslastung \bar{U}_i :

$$i' = \arg \max_{i \in \mathcal{N}} \bar{U}_i$$

Füge Relationen, die in Knoten i' umgestellt werden, den freien Relationen hinzu:

$$\Gamma := \Gamma \cup \{k \in \mathcal{K} \mid \exists j \in \mathcal{N} : \widehat{x}_{i'j}^k = 1\}$$

Setze $\bar{U}_{i'} := -1$

end while

Output: Menge freier Relationen Γ

Ebenso wie in Algorithmus 2 wird auch die Wahl der Nachbarschaft auf Basis von Knotenauslastungen um eine Tabuliste betrachteter Knoten ergänzt, um im Sinne eines VNS-Ansatzes ein wiederholtes Durchsuchen einer Nachbarschaft bei Stagnation des Zielfunktionswertes zu vermeiden.

3. Nachbarschaft auf Basis von Randomisierung

Obwohl eine Festlegung auf deterministische Verfahren, d.h. solche ohne stochastische Effekte, erfolgt ist, wird ein randomisierter Ansatz präsentiert. Dies geschieht, um eine

Beurteilung zu ermöglichen, ob die beiden zuvor entwickelten Nachbarschaften auf Basis freier Zugkapazitäten bzw. Knotenauslastungen besser als eine zufällige Auswahl sind.

Algorithmus 4 Nachbarschaft auf Basis von Randomisierung

Input: zulässige Lösung (\hat{x}, \hat{n}) , Parameter γ

Initialisiere $\Gamma = \emptyset$
while $|\Gamma| < \gamma$ **do**
 Wähle gemäß einer Gleichverteilung zufällig eine Relation $k' \in \mathcal{K}$
 Füge die gewählte Relation den freien Relationen hinzu:

$$\Gamma := \Gamma \cup \{k'\}$$

end while

Output: Menge freier Relationen Γ

Der randomisierte Ansatz in Algorithmus 4 hat den Vorteil, dass er einfach und schnell zu implementieren ist. Des Weiteren kann auf die Anwendung von VNS- und VDNS-Varianten verzichtet werden, weil die unerwünschte Konvergenz gegen lokale Optima bereits durch die Randomisierung unwahrscheinlicher wird. Nichtsdestotrotz wäre es möglich, die Nachbarschaftsgröße auch im randomisierten Fall dynamisch anzupassen. Zur Verwendung randomisierter Algorithmen ist anzumerken, dass der Nachteil nicht konstanter Ergebnisse bei unabhängigen Testläufen in Kauf genommen werden muss. Ein empirischer Vergleich der drei vorgestellten Nachbarschaftsstrukturen wird in Abschnitt 6.5.1 präsentiert. Zur konkreten Umsetzung von Rahmenalgorithmus 1 ist nur noch die Konstruktion einer Startlösung notwendig.

Startlösung für die Matheuristik

Die Konstruktion einer zulässigen Startlösung ist für viele Instanzen ein äußerst komplexes Unterfangen. Wie bereits bei der Komplexitätsanalyse in Abschnitt 4.2.1 festgestellt wurde, ist das Finden einer zulässigen Lösung ein \mathcal{NP} -vollständiges Problem. Im Rahmen der Algorithmenentwicklung wurden verschiedene konstruktive Heuristiken getestet. Keine von ihnen war in der Lage, für sämtliche betrachteten Instanzen zulässige Startlösungen zu generieren. In diesem Abschnitt wird eine Methode vorgestellt, die für alle betrachteten Instanzen (vgl. Kapitel 6) erfolgreich war.

Die Erzeugung einer zulässigen Lösung von (Opt_{arc}) ist mit Hilfe eines sogenannten Big-M-Ansatzes unter Verwendung der *Phase II* des Rahmenalgorithmus 1 möglich. Die Big-M-Methode ist ein Trick zum Finden zulässiger Lösungen aus der linearen Optimierung [Sch00]: Zunächst wird ein Hilfsproblem mit künstlichen Variablen definiert, welche Zulässigkeit im Hilfsproblem erzwingen und gleichzeitig extrem hohe Kosten in der Zielfunktion verursachen. Dadurch ist es möglich, die Lösung des Hilfsproblems sukzessive zu verbessern, bis keine künstlichen Variablen mehr in der Lösung enthalten sind. Zur Übertragung dieses Vorgehens auf das Modell (Opt_{arc}) werden künstliche binäre Variablen y_{ij} mit folgenden Eigenschaften eingeführt:

- Gilt $y_{ij} = 1$, so bestehen auf der Kante (i, j) unendlich hohe Gewichts- und Längenkapazitäten G^∞ und L^∞ . Um einen numerischen Umgang zu ermöglichen, werden G^∞ und L^∞ lediglich extrem groß gewählt.
- Die Variable y_{ij} verursacht in der Zielfunktion unendliche Kosten. Auch hier wird der Umgang mit unendlich großen Zahlen durch einen extrem großen Koeffizienten M umgangen. Der große Koeffizient M ist namensgebend für die Big-M-Methode.

Die praktische Interpretation der Variablen y_{ij} wäre die Einführung einer Zugverbindung unendlicher Kapazität und mit unendlich hohen Kosten. Es entsteht das folgende Hilfsproblem (Opt_{Start}):

$$\begin{aligned}
 (Opt_{Start}) \quad & \min_{x,n,y} \sum_{d \in \mathcal{D}} \sum_{(i,j) \in \mathcal{A}} [c_{ij}n_{ij} + \sum_{k \in \mathcal{K}} U_i x_{ij}^k + M y_{ij}] \\
 \text{s.t.} \quad & \sum_{i \in \mathcal{N}} x_{ji}^k - \sum_{i \in \mathcal{N}} x_{ij}^k = Balance(k, j) && \forall k \in \mathcal{K} \quad \forall j \in \mathcal{N} \\
 & \sum_{k \in \mathcal{K}} l_d^k x_{ij}^k \leq L_{ij}n_{ij} + L^\infty y_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\
 & \sum_{k \in \mathcal{K}} w_d^k x_{ij}^k \leq W_{ij}n_{ij} + W^\infty y_{ij} && \forall d \in \mathcal{D} \quad \forall (i, j) \in \mathcal{A} \\
 & \sum_{j \in \mathcal{N}} n_{ij} \leq R_i && \forall i \in \mathcal{N} \\
 & \sum_{j \in \mathcal{N}, k \in \mathcal{K}} v_d^k x_{ij}^k \leq B_i && \forall i \in \mathcal{N} \quad \forall d \in \mathcal{D} \\
 & \sum_{(i,j) \in \mathcal{A}} x_{ij}^k \leq S_k && \forall k \in \mathcal{K} \\
 & n_{ij} \in \mathbb{N} && \forall (i, j) \in \mathcal{A} \\
 & y_{ij} \in \{0, 1\} && \forall (i, j) \in \mathcal{A} \\
 & x_{ij}^k \in \{0, 1\} && \forall (i, j) \in \mathcal{A} \quad \forall k \in \mathcal{K}
 \end{aligned}$$

Der Solver CPLEX findet für dieses Problem stets in kurzer Zeit eine zulässige, wenn auch beliebig teure Lösung. Auf diese schlechte Startlösung von (Opt_{Start}) kann die Phase II des Rahmenalgorithmus 1 angewendet werden. Mit ihrer Hilfe werden so lange Relationen umgeroutet und Designvariablen n_{ij} angepasst, bis für alle Hilfsvariablen $y_{ij} = 0$ gilt. Phase II generiert stets zulässige Lösungen. (Opt_{Start}) und (Opt_{arc}) sind identisch, wenn alle künstlichen Variablen null sind. Phase II sucht eine solche Lösung, weil die Kosten für die künstlichen Variablen so hoch sind, dass ihre Minimierung das

primäre Ziel der Optimierung ist. Dieser Prozess kann beschleunigt werden, indem bei der Lösung des Subproblems den künstlichen Variablen eine hohe Priorität gegeben wird. Wie eingangs erwähnt, war es mit Hilfe der Big-M-Methode möglich, für alle betrachteten Instanzen eine zulässige Lösung zu generieren.

Zusammenfassung Matheuristiken

In diesem Abschnitt wurde eine Matheuristik, d.h. eine hybride Methode aus Metaheuristik und exakten Verfahren, zur Lösung von (Opt_{arc}) entwickelt. Nach einer Vorstellung von Designmethoden für Metaheuristiken wurden erfolgreiche metaheuristische Ansätze zu vergleichbaren Problemstellungen analysiert. Es zeigte sich, dass alle erfolgreichen Ansätze zu vergleichbaren Themen die Gemeinsamkeit aufwiesen, eine auf lokaler Suche basierende Verbesserungsheuristik zu verwenden, d.h. dass einzelne Relationen aus einer Lösung entfernt und auf Basis der übrigen, fixierten Relationen neu geroutet werden. Im Rahmen des Versuchs der Übertragung auf das vorliegende Problem wurden verschiedene Nachbarschaftsstrukturen diskutiert, die mit Hilfe exakter Verfahren effizient durchsucht werden können. Dabei wurde stets das Ziel verfolgt, die Matheuristik zu befähigen, ein einmal erreichtes lokales Optimum wieder zu verlassen. Dazu wurden mehrere metaheuristische Techniken verwendet. Zum einen sind die Nachbarschaften so konstruiert, dass sie sich von Iteration zu Iteration strukturell unterscheiden (VNS). Es werden immer neue Kombinationen von freien Relationen gewählt. Des Weiteren wird die Nachbarschaftsgröße, gesteuert vom Parameter γ , dynamisch angepasst (VDNS), wenn die Variation der Nachbarschaften allein keine Verbesserung bringt. Die VDNS kommt also erst zum Einsatz, wenn die VNS keine Verbesserungen mehr erzielt. Die Techniken VNS und VDNS verringern gemeinsam die Wahrscheinlichkeit, dass das Verfahren einem lokalen Optimum nicht mehr entkommen kann.

Die Konstruktion einer Startlösung, die anschließend mit Hilfe der Matheuristik verbessert werden kann, gestaltete sich schwierig. Ein Big-M-Ansatz wurde genutzt, um eine zulässige Lösung eines Hilfsproblem mit Hilfe von Phase II der Matheuristik so zu verbessern, dass eine Startlösung für (Opt_{arc}) entsteht. Die Leistungsfähigkeit der Matheuristik wird in zahlreichen Tests zur Nachbarschafts- und Parameterwahl in Unterkapitel 6.5 aufgezeigt.

6 Numerische Resultate der vorgestellten Verfahren

In diesem Kapitel werden Ergebnisse von Testläufen der in Kapitel 5 entwickelten Algorithmen vorgestellt. Dazu wird zunächst die Herkunft der Instanzen erläutert. In Unterkapitel 6.2 werden Testergebnisse des kommerziellen Solvers CPLEX für (Opt_{arc}) dargestellt. Anschließend wird die Wirkung der Preprocessing-Verfahren in Unterkapitel 6.3 analysiert. Es folgen die Resultate der neu entwickelten Verfahren Column Generation und Branch-and-Price-and-Cut in Unterkapitel 6.4. In Unterkapitel 6.5 findet das Parametertuning für die Matheuristik statt, bevor die Ergebnisse der Matheuristik vorgestellt werden. Alle Rechnungen dieses Kapitels wurden auf einem Arbeitsplatzrechner mit 16 GB Arbeitsspeicher und einem Intel-*i5*-Prozessor durchgeführt.

6.1 Herkunft der betrachteten Instanzen

Zehn Datensätze unterschiedlicher Größe werden für die numerischen Tests in diesem Kapitel verwendet. Im Rahmen des Forschungsprojektes, aus dem diese Dissertation entstanden ist, stellte der Projektpartner *DB Mobility Logistics* drei Datensätze (*DB-21-254-10*, *DB-107-1584-10*, *DB-147-3272-5*) zur Verfügung, die als Realdaten betrachtet werden können. Der erstgenannte Datensatz wurde dabei in zwei Varianten hinsichtlich des Aufkommens geliefert, die durch die Namenszusätze *a* und *b* gekennzeichnet sind. Weitere sechs Datensätze wurden auf Basis dieser Instanzen künstlich erzeugt.

Die Bezeichnung der Datensätze ist folgendermaßen aufgebaut: Das Präfix *DB* steht für einen Originaldatensatz, wohingegen *gen* für einen künstlich generierten Datensatz steht. Es folgen die drei Kennzahlen *Anzahl der Knoten*, *Anzahl der Relationen* und *Anzahl der betrachteten Szenarien*. So steht z.B. *gen-50-525-6* für einen künstlich generierten Datensatz mit 50 Knoten und 525 Relationen in 6 zu betrachtenden Szenarien.

Die Erzeugung der künstlichen Datensätze wurde folgendermaßen durchgeführt: Eine realitätsnahe Netzstruktur ist nur schwierig nachzubilden. Daher wird in der Regel auf das Netzwerk, bestehend aus Knoten und ihren Distanzen, aus den Originaldatensätzen zurückgegriffen. Die Netzstruktur wird nur für drei Datensätze (*gen-50-224-10*, *gen-50-525-6*, *gen-100-547-7*) künstlich erzeugt: Ausgehend von einer gegebenen Anzahl von Knoten N , werden Knotenkoordinaten auf einer Fläche von 880x640 km (ungefähre Ausmaße der Bundesrepublik Deutschland) randomisiert gleichverteilt erzeugt. Die Kapazitäten der Knoten werden entsprechend den etwaigen Verhältnissen aus den Originaldatensätzen gewählt: 7 % der Knoten können 150 Züge pro Tag umstellen, 20 % der Knoten können 60 Züge pro Tag umstellen, die übrigen Knoten haben eine Kapazität von 20 Zügen pro Tag. Die Ablaufbergkapazität eines Knotens wurde pauschal mit 25

Wagen je Richtungsgleis bemessen. Die Distanz zwischen zwei Knoten wurde als euklidischer Abstand, multipliziert mit einem pauschalen Umwegfaktor von 1,3, berechnet. Die Erzeugung der Wagenaufkommen auf einer Relation, gemessen in Länge, Gewicht und Anzahl Wagen, wurde auf folgende Weise vorgenommen: Die Anzahl der Relationen sei gegeben. Es wird eine entsprechende Anzahl Quelle-Senke-Paare randomisiert gleichverteilt gewählt. Für die zu den Quelle-Senke-Paaren korrespondierenden Relationen wird anschließend ein Wagenaufkommen generiert. Dazu wird für jedes Szenario, deren Anzahl ebenfalls gegeben ist, eine zufällige Wagenzahl zwischen 1 und 50 bestimmt. Diese Wagenzahl wird anschließend mit einer zufälligen Durchschnittslänge der Wagen zwischen 8 und 30 m sowie mit einem Durchschnittsgewicht zwischen 15 und 35 t multipliziert, um die Längen und Gewichte der Relationen zu erhalten. Auf diese Weise entstanden die Datensätze *gen-21-100-10*, *gen-21-220-6* und *gen-107-5763-10*.

Die sonstigen Parameter für das Modell wurden auf Basis realistischer Werte gewählt, die vom Projektpartner bestätigt wurden: Ein durchschnittlicher Zugkilometer kostet demnach 6 €, eine Wagenumstellung 25 €. Detailliertere Daten sind in der Regel nicht verfügbar und können nur aus globalen Kostendaten ermittelt werden. Die maximale Anzahl erlaubter Umstellungen S_k einer Relation k wurde dem heutigen Leitwegekonzept entsprechend auf fünf begrenzt, da auf diese Weise die heutigen Servicestandards, d.h. insbesondere zeitliche Vorgaben, nicht durch zusätzliche Rangiervorgänge unterschritten werden. Die Maximallängen und -gewichte wurden auf allen Kanten mit 700 m und 1600 t bemessen.

6.2 Ergebnisse des kommerziellen Solvers CPLEX

Die zehn im vorangegangenen Abschnitt beschriebenen Instanzen des Problems (Opt_{arc}) wurden mit Hilfe des kommerziellen Solvers CPLEX in der Version 12.4 getestet. Dazu wurde das Modell in der Modellierungssprache GAMS implementiert und anschließend an den Solver übergeben. Die Ergebnisse der Testläufe nach drei bzw. zwölf Stunden sind in Tabelle 6.1 vermerkt. Dabei enthalten die Spalten *Obj* den Zielfunktionswert der besten gefunden Lösung, *LB* die beste gefundene untere Schranke und *Gap* die daraus resultierende relative Optimalitätslücke nach den jeweiligen Laufzeiten.

Tabelle 6.1: Resultate des kommerziellen Solvers CPLEX 12.4

	3h			12h		
	Obj	LB	Gap	Obj	LB	Gap
gen-21-100-10	559.025	266.594	52,31%	476.975	266.594	44,11%
gen-21-220-6	4.256.184	3.546.451	16,68%	4.256.184	3.546.451	16,68%
DB-21-254-10a	2.625.383	2.251.940	14,22%	2.595.860	2.251.940	13,25%
DB-21-254-10b	3.879.096	2.902.515	25,18%	3.537.360	2.902.515	17,95%
gen-50-224-10	–	0	–	–	7.345.232	–
gen-50-525-6	–	0	–	–	0	–
gen-100-547-7	–	0	–	–	0	–
DB-107-1584-10	–	0	–	–	0	–
gen-107-5763-10	–	0	–	–	0	–
DB-147-3272-5	–	0	–	–	0	–

Die von CPLEX erzielten Ergebnisse können zusammenfassend als praktisch nicht verwertbar bezeichnet werden. Für die Mehrzahl der Instanzen wird innerhalb der gegebenen Laufzeit von zwölf Stunden keine zulässige Lösung gefunden. Für die kleineren Instanzen gelingt dies zwar, zufriedenstellend sind die Lösungen mit Optimalitätslücken im teils deutlich zweistelligen Bereich jedoch nicht. Für die vier größten Instanzen konnte nicht einmal die lineare Relaxierung gelöst werden. Somit ist keine nichttriviale, untere Schranke für den Zielfunktionswert dieser Instanzen bekannt.

Obwohl CPLEX nicht der einzige kommerzielle Solver ist, der zur Lösung in Frage kommt, gilt er neben Gurobi dennoch als einer der leistungsstärksten. Als Grund sind neben der Größe der Instanzen vor allem die schwachen Schranken zu nennen, die bei der linearen Relaxierung von (Opt_{arc}) berechnet werden (vgl. Abschnitt 4.2.2).

6.3 Auswirkungen der Modellvereinfachungen

In Abschnitt 4.2.3 des Modellkapitels wurden verschiedene Modellvereinfachungen vorgestellt, deren Wirkung auf die Modellgröße an dieser Stelle untersucht werden sollen. Es wurden zwei Typen von Vereinfachungen identifiziert: exakte Preprocessings, die den Lösungsraum des Modells nicht verkleinern und heuristische Preprocessings, die auf praktischen Annahmen über die Struktur der Optimallösung basieren. Exemplarisch wurden drei Instanzen ausgewählt, um zu analysieren, wie effektiv die jeweiligen Preprocessing-Verfahren wirken. In Tabelle 6.2 wurde notiert, um welchen Anteil die jeweiligen Preprocessings die Anzahl der Flussvariablen im Modell (Opt_{arc}) reduziert haben.

Tabelle 6.2: Prozentuale Variablenreduktion der Preprocessings aus Abschnitt 4.2.3

	DB-21-254-10a	gen-100-547-7	DB-107-1584-10
Exakte Preprocessings	5,14%	1,18%	7,69%
Günstiger Direkttransport	0,72%	0,00%	12,61%
Gegenläufige Kanten	19,24%	7,59%	17,72%
Relativer Umwegfaktor $\Delta = 2$	33,07%	7,84%	37,54%
Relativer Umwegfaktor $\Delta = 1,5$	50,40%	29,74%	54,65%
Relativer Umwegfaktor $\Delta = 1$	79,25%	82,76%	85,57%
Kumuliert bei Umwegfaktor $\Delta = 1,5$	65,79%	34,64%	69,61%

Die Wirkung der exakten Preprocessings ist begrenzt. Die Anzahl der Variablen reduziert sich bei den betrachteten Instanzen um maximal 7,69 %. Diese Vereinfachungen können dafür durchgeführt werden, ohne den Lösungsraum zu verkleinern. Das Verbot relativer Umwege führt zu sehr starken Reduktionen der Variablenzahl, wenn der erlaubte Umwegfaktor klein genug gewählt ist. Dabei sei daran erinnert, dass ein Umwegfaktor von $\Delta = 1,5$ Knoten von der weiteren Betrachtung ausschließt, die mehr als 1,5 mal so weit von Quelle und Senke entfernt sind, wie die Länge der Direktverbindung beträgt. Ein Umweg mit relativem Umwegfaktor $\Delta = 1,5$ ist somit mindestens dreimal so lang wie die Direktverbindung der jeweiligen Relation. In der untersten Zeile der Tabelle wurde die Wirkung aller Preprocessings kumuliert betrachtet, da die einzelnen

Preprocessings Überschneidungen beim Verbot gewisser Kanten aufweisen können. Für den kombinierten Fall wurde der für praktische Anwendungen plausible Umwegfaktor $\Delta = 1,5$ gewählt.

Untere Schranken nach Modellvereinfachungen

Die Variablenreduktion erlaubt eine effizientere Lösung der linearen Relaxierung des Modells (Opt_{arc}), welche sich als komplexes Problem (vgl. Abschnitt 4.2.2) darstellt. Da für Instanzen von praktisch relevanter Größe derzeit kein Verfahren bekannt ist, das eine befriedigende untere Schranke in praxistauglicher Zeit liefert, werden hier unter Inkaufnahme extremer Laufzeiten untere Schranken für die in dieser Arbeit betrachteten Instanzen von (Opt_{arc}) gesucht. Hierdurch wird es möglich, die Leistungsfähigkeit der Lösungsverfahren für (Opt_{arc}) besser zu beurteilen. Es bieten sich verschiedene Möglichkeiten an, um eine untere Schranke für den optimalen Zielfunktionswert zu bestimmen:

- Es ist möglich, eine untere Schranke mit Hilfe von CPLEX zu finden. CPLEX löst die lineare Relaxierung und versucht, die Schranke durch Hinzunahme von Schnittebenen zu verbessern. Dazu ist es nützlich, eine zulässige ganzzahlige Lösung, welche mit der Matheuristik erzeugt werden kann, zu übergeben. Dieses Verfahren ist jedoch nur für hinreichend kleine Instanzen erfolgreich, da CPLEX für große Instanzen die lineare Relaxierung oft nicht zur Optimalität lösen kann.
- Ist durch die Verwendung der Matheuristik eine zulässige Lösung x^1 des nicht-relaxierten Problems (Opt_{arc}) bekannt, kann auch dann eine untere Schranke bestimmt werden, wenn CPLEX die lineare Relaxierung nicht zur Optimalität löst. CPLEX verwendet üblicherweise ein duales Simplexverfahren, das während der Optimierung zulässige duale Lösungen π erzeugt. Die bekannte ganzzahlige Lösung x^1 ist insbesondere im relaxierten primalen Problem zulässig. Nach dem schwachen Dualitätssatz aus Kapitel 3 ist der duale Zielfunktionswert der Lösung π eine untere Schranke für den Zielfunktionswert des primalen Problems und damit auch des ganzzahligen Problems (Opt_{arc}). Diese Schranke kann jedoch nur so gut sein wie die Optimallösung der linearen Relaxierung.
- Wie bereits in Unterkapitel 6.4 gezeigt werden konnte, ist das CG-Verfahren in einigen Fällen schneller als CPLEX, wenn es um die Lösung der linearen Relaxierung geht. Damit ist auch CG ein geeignetes Verfahren zur Erzeugung unterer Schranken von (Opt_{arc}).
- Falls das CG-Verfahren nicht konvergiert, können die Lasdonschranken, die während der Lösung mitberechnet werden, ebenfalls als untere Schranken dienen.

In Tabelle 6.3 ist die jeweils beste der oben genannten unteren Schranken für die in Unterkapitel 6.1 definierten Instanzen vermerkt. Dabei wird in der Spalte *Methode* das verwendete Verfahren angegeben, wobei *CG/Lasdon* für eine Lösung per Column Generation steht und $CPLEX_{dual}$ für die beschriebene duale Schranke. Vor Berechnung

der Schranken wurde die Größe der Probleme jeweils mit den Preprocessing-Verfahren aus Unterkapitel 4.1 reduziert. Die dabei verwendeten Parameter Δ für die relativen Umwege und T_{max} für die maximale Laufzeit sind ebenfalls in der Tabelle vermerkt. Allen Verfahren wurde eine zulässige ganzzahlige Lösung übergeben, die mit Hilfe der Matheuristik aus Unterkapitel 5.3 erzeugt wurde.

Tabelle 6.3: Untere Schranken für alle Instanzen

	Δ	Beste LB	T_{max} [h]	Methode
gen-21-100-10	1,5	266.594,6	12	<i>CPLEX</i>
gen-21-220-6	1,5	3.625.228,6	12	<i>CPLEX</i>
DB-21-254-10a	1,5	2.283.203,8	12	<i>CPLEX</i>
DB-21-254-10b	1,5	2.942.289,5	12	<i>CPLEX</i>
gen-50-224-10	1,5	7.461.049,0	12	<i>CPLEX</i>
gen-50-525-6	1,5	15.152.882,9	24	<i>CG/Lasdon</i>
gen-100-547-7	1,5	14.691.346,9	24	<i>CPLEX_{dual}</i>
DB-107-1584-10	1	24.598,1*	240	<i>CPLEX_{dual}</i>
gen-107-5763-10	1	44.817,6*	240	<i>CPLEX_{dual}</i>
DB-147-3272-5	1	15.603,5*	240	<i>CPLEX_{dual}</i>

* Ergebnisse aus Datenschutzgründen durch Skalierung anonymisiert

Es ist zu erkennen, dass CPLEX in der Regel die besten unteren Schranken liefert. Die hier genannten Schranken werden verwendet, um die Lösungsqualität des Branch-and-Price-and-Cut-Verfahrens in Unterkapitel 6.4 sowie der Matheuristik in Unterkapitel 6.5 zu bewerten.

6.4 Resultate von Column Generation und Branch-and-Price-and-Cut

In Unterkapitel 5.1 wurden CG-Verfahren zur Lösung der linearen Relaxierung von (Opt_{path}) eingeführt. Darauf aufbauend wurde in Unterkapitel 5.2 ein Branch-and-Price-and-Cut-Verfahren zur Bestimmung ganzzahliger Lösungen von (Opt_{path}) entwickelt. Beide Verfahren wurden in Matlab R2011b implementiert. Die RMP im CG-Verfahren wurde mit Hilfe von CPLEX 12.4 gelöst. Als Medium zwischen Matlab und CPLEX wurde die Programmiersprache GAMS verwendet, da sich diese Schnittstelle als schneller herausstellte als die direkte Schnittstelle zwischen Matlab und CPLEX.

Resultate des Column-Generation-Verfahrens

Die initiale Pfadmenge des CG-Verfahrens wurde wie in Abschnitt 5.1.3 beschrieben gewählt. In Tabelle 6.4 sind die Resultate des Verfahrens notiert. Die maximale Laufzeit wurde auf zwölf Stunden begrenzt. In der Spalte *CPLEX* wurde die Lösungszeit vermerkt, die CPLEX benötigt, um die Optimallösung ohne vorheriges Preprocessing zu ermitteln, was für die letzten sechs Instanzen nicht gelang. Die Spalte T_{CG} enthält die Lösungszeit des CG-Verfahrens. Wenn innerhalb von zwölf Stunden keine Optimallösung gefunden wird, liefert CG mit der Lasdonschranke trotzdem eine untere Schranke, deren

relative Qualität Gap_{CG} (vgl. Tabelle 6.4) notiert wurde. Die letzten beiden Spalten enthalten jeweils die Anzahl generierter Pfade und die Gesamtzahl der CG-Iterationen.

Tabelle 6.4: Resultate des Column-Generation-Verfahrens

	CPLEX[s]	T_{CG} [s]	Gap_{CG}	#Pfade	Iterationen
gen-21-100-10	78	174	0,00%	2.746	103
DB-21-254-10a	681	100	0,00%	2.471	51
DB-21-254-10b	425	1.059	0,00%	4.336	139
gen-21-220-6	100	934	0,00%	6.827	131
gen-50-224-10	–	43.200	0,03%	101.118	1.162
gen-50-525-6	–	43.200	0,03%	26.318	92
gen-100-547-7	–	43.200	26,43%	41.719	79
DB-107-1584-10	–	43.200	19,72%	40.684	32
gen-107-5763-10	–	43.200	25,21%	45.176	9
DB-147-3272-5	–	43.200	52,86%	67.038	24

Das CG-Verfahren findet für alle Instanzen, die CPLEX lösen kann, ebenfalls die Optimallösung. Bei den übrigen Instanzen ist es mit CG wenigstens möglich, eine untere Schranke zu generieren, die im Falle der Instanzen *gen-50-224-10* und *gen-50-525-6* mit einer Genauigkeit von je 0,03 % so nah an der bestmöglichen LP-Schranke liegt, dass der Abstand vernachlässigt werden kann. Bei den vier größten Instanzen kann die Lücke zur Lasdonschranke innerhalb der Maximallaufzeit nicht geschlossen werden. Es werden dennoch untere Schranken ermittelt, die für das im Folgenden getestete Branch-and-Price-Verfahren entscheidend sind.

Resultate des Branch-and-Price-and-Cut-Verfahrens

Das B&P&C-Verfahren nutzt das CG-Verfahren, dessen Tests darauf schließen lassen, dass es geeignet ist, um untere Schranken für (Opt_{path}) zu berechnen. In Kombination mit der Early Termination aus Abschnitt 5.2.3 wurden die Resultate in Tabelle 6.5 erzeugt. Die dafür nötigen Parameterwerte ω wurden in Abhängigkeit von der Instanzgröße und den Erfahrungen aus den CG-Tests gewählt. Die Spalte T_{max} der Tabelle enthält die vorgegebene Maximallaufzeit des B&P&C-Verfahrens. Spalte T_{BF} enthält die Zeit, nach der die Bestlösung für die jeweilige Instanz gefunden wurde. Die Optimalitätslücke zur besten globalen unteren Schranke, die während des B&P&C-Verfahrens bestimmt wird, steht in Spalte $Gap_{B\&P\&C}^{B\&P\&C}$. Des Weiteren wurde die Optimalitätslücke der B&P&C-Lösung zur jeweils besten bekannten Schranke aus Tabelle 6.3 in der Spalte $GAP_{bestLB}^{B\&P\&C}$ vermerkt. Zum Vergleich wurden die Optimalitätslücken der jeweils besten zulässigen Lösung von CPLEX bzgl. der gleichen Schranke in der Spalte GAP_{bestLB}^{CPLEX} aufgelistet, um einen Vergleich der Lösung von CPLEX und B&P&C zu ermöglichen.

Das B&P&C-Verfahren liefert für alle Instanzen, die von CPLEX gelöst werden können, bessere zulässige Lösungen als der kommerzielle Solver. Im Fall der kleinen Realinstanz *DB-21-254-10a* ist die Optimalitätslücke mit 2,56 % sehr zufriedenstellend. Des Weiteren werden für vier Instanzen, zu denen zuvor keine zulässigen Lösungen bekannt waren, Lösungen gefunden. Für die beiden größten Instanzen kann mit B&P&C keine

Tabelle 6.5: Resultate des Branch-and-Price-and-Cut-Verfahrens

	ω	$T_{max}[h]$	$T^{BF}[s]$	$Gap_{B\&P\&C}^{B\&P\&C}$	$Gap_{bestLB}^{B\&P\&C}$	Gap_{bestLB}^{CPLEX}
gen-21-100-10	0,1	4	11702s	75,63%	45,64%	50,02%
gen-21-220-6	0,1	4	11705s	25,54%	10,01%	19,83%
DB-21-254-10a	0,1	4	12401s	17,44%	2,56%	14,22%
DB-21-254-10b	0,1	4	12412s	30,00%	15,26%	22,99%
gen-50-224-10	0,3	6	8493s	42,02%	27,99%	–
gen-50-525-6	0,4	6	2573s	60,33%	39,82%	–
gen-100-547-7	0,4	6	21216s	53,25%	53,11%	–
DB-107-1584-10	0,4	6	4899s	35,52%	16,36%	–
gen-107-5763-10	0,4	12	–	–	–	–
DB-147-3272-5	—	12	–	–	–	–

zulässige Lösung gefunden werden. Die Optimalitätslücken sind nicht sehr befriedigend, was auf eine eher geringe Qualität der ganzzahligen Lösungen hindeutet. Die unteren Schranken, die während des B&P&C ermittelt werden, sind erheblich schwächer als die Schranken aus Tabelle 6.3. Der Grund dafür liegt darin, dass durch die Anwendung von Early Termination, schwache globale Schranken gefunden werden. Außerdem fällt auf, dass die besten Lösungen häufig früh gefunden, danach aber nicht mehr verbessert werden können. Es ist festzuhalten, dass das B&P&C-Verfahren bei jeder Instanz CPLEX übertrifft, aber aus praktischer Sicht keine befriedigenden Lösungen liefert.

6.5 Tuning der Matheuristik

In diesem Unterkapitel werden die notwendigen Einstellungen der in Unterkapitel 5.3 entwickelten Matheuristik ermittelt und dokumentiert. Dazu wird in Abschnitt 6.5.1 zunächst durch numerische Experimente an einer ausgewählten Teilmenge der Abschnitt 6.1 definierten Instanzen ermittelt, welche der drei vorgestellten Nachbarschaftsstrukturen geeignet ist. Anschließend findet in Abschnitt 6.5.2 die Wahl geeigneter Parameter statt. Die finalen Resultate der Matheuristik werden in Unterkapitel 6.6 dargestellt. Bei allen Tests wurden die Teilprobleme der Matheuristik mit Hilfe von CPLEX 12.4 gelöst. In jeder Iteration wurde die bis dahin beste bekannte Lösung als Startlösung an CPLEX übergeben.

6.5.1 Wahl der Nachbarschaftsstruktur

Drei mögliche Nachbarschaftsstrukturen wurden in Unterkapitel 5.3 diskutiert:

1. auf Basis freier Zugkapazitäten
2. auf Basis von Knotenauslastungen
3. auf Basis von Randomisierung

Tabelle 6.6 fasst die Ergebnisse von Tests mit allen drei Nachbarschaften zusammen. Es wurde, wie im Vorgehen zur Entwicklung einer Matheuristik beschrieben, eine repräsentative Menge von vier Instanzen ausgewählt. Die Laufzeit wurde für jeden Test

auf drei Stunden begrenzt. Der Parameter γ wurde so eingestellt, dass in jedem Iterationsschritt der Matheuristik 10 % aller Relationen umgeroutet werden. Die randomisierte Wahl wurde in je drei unabhängigen Testläufen durchgeführt, um die durch die Randomisierung entstehende Abweichung der Ergebnisse besser einordnen zu können. Die Spalten *Fz* und *Knoten* enthalten die Ergebnisse für die auf Zugauslastungen bzw. Knotenauslastungen basierenden Nachbarschaften. Es folgen die durchschnittlichen Ergebnisse mit randomisierter Nachbarschaft in der Spalte *Ø Rand*.

Tabelle 6.6: Optimalitätslücken in Abhängigkeit von Nachbarschaftsstrukturen

	T_{max} [h]	γ	Fz	Knoten	Ø Rand
gen-21-220-6	3	22	93,46%	93,40%	93,71%
DB-21-254-10a	3	25	5,70%	6,05%	14,92%
gen-50-224-10	3	22	17,77%	28,04%	18,11%
DB-107-1584-10	3	75	13,56%	16,51%	11,86%

Die Ergebnisse in Tabelle 6.6 lassen die Schlussfolgerung zu, dass die auf Kantenauslastungen basierende Nachbarschaftsstruktur am effektivsten ist. Obwohl die anderen Verfahren für jeweils eine Instanz besser sind, kann sich das Verfahren auf Basis freier Zugkapazitäten im Durchschnitt am meisten bewähren. Die Tatsache, dass die randomisierte Wahl gerade für die größte Instanz das beste Resultat liefert, ist auf den ersten Blick überraschend. Die Laufzeitlimitierung von drei Stunden ist jedoch für eine derart große Instanz sehr restriktiv. Da die Matheuristik sukzessive Verbesserungen der Lösung durchführt, werden die Fortschritte im Verlauf des Lösungsprozess in der Regel langsamer. Es ist anzunehmen, dass der Lösungsverlauf bei einer komplexen Instanz wie *DB-107-1584-10* nach drei Stunden Laufzeit noch nicht sehr weit fortgeschritten ist und die Intelligenz des Verfahrens erst im späteren Verlauf ihre volle Wirkung entfaltet, während in der Anfangsphase des Verfahrens auch mit zufälligen Nachbarschaften Fortschritte gemacht werden. Im weiteren Verlauf wird nur noch die Nachbarschaftswahl auf Basis freier Zugkapazitäten betrachtet.

6.5.2 Parametertuning

Eine sinnvolle Wahl von γ , d.h. der Anzahl der Relationen, die in jeder Iteration der Matheuristik umgeroutet werden, ist wichtig für den Algorithmus. Je kleiner der Parameter γ gewählt wird, desto schneller können die Subprobleme gelöst werden. Dafür ist die potentielle Verbesserung in einem Schritt größer, wenn γ groß gewählt wird. Um zwischen beiden Eigenschaften einen Kompromiss zu finden, wurde die Matheuristik für Werte von γ zwischen 10 und 150 für vier Instanzen getestet.

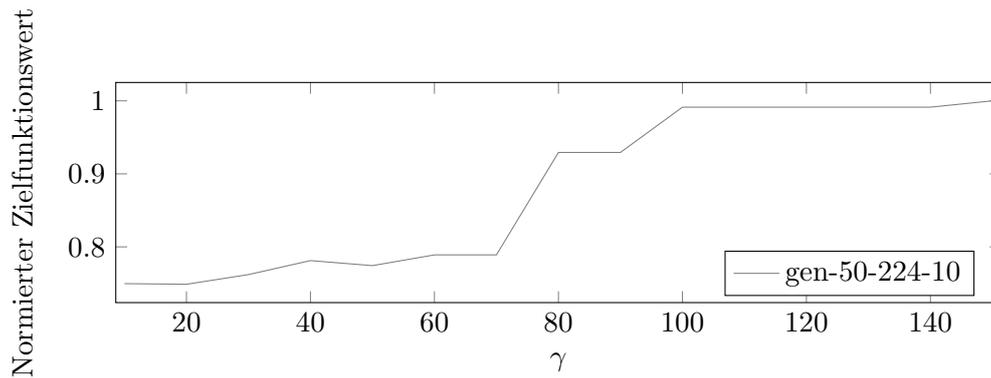
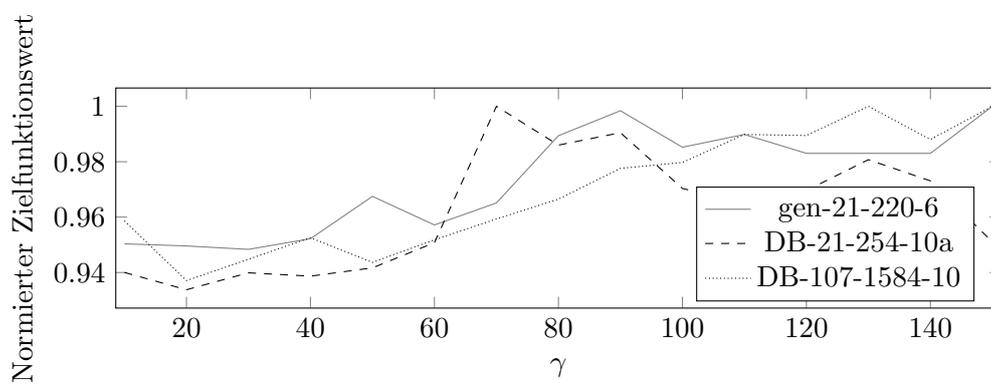
Abbildung 6.2: Parametertuning für die Matheuristik für Instanz *gen-50-224-10*

Abbildung 6.1: Parametertuning für die Matheuristik für drei Instanzen

In Abbildung 6.1 sind die Zielfunktionswerte für drei der vier Instanzen aus Abschnitt 6.1, normiert um das jeweils schlechteste Ergebnis, in Abhängigkeit von γ abgetragen. Es kann beobachtet werden, dass bei zu großem γ die Lösungsqualität des Verfahrens schlechter wird. Bei der vierten Instanz ist dieser Effekt so stark sichtbar, dass der Verlauf für Instanz *gen-50-224-10* in Abbildung 6.2 separat illustriert wurde. Es ist festzuhalten, dass eine Wahl von $\gamma > 50$ keine guten Ergebnisse liefert. Dies ist mit der zunehmenden Komplexität der Teilprobleme in Abhängigkeit von γ zu erklären. Die Ergebnisse der Experimente legen die Vermutung nahe, dass ein besonders kleines γ die besten Ergebnisse liefert. Ein zu kleines γ kann allerdings speziell im späten Verlauf des Verfahrens problematisch werden: Die Kostenfunktion forciert in den meisten Fällen die Bündelung von Wagenströmen. Je mehr Relationen parallel umgeroutet werden dürfen, desto größer ist die Wahrscheinlichkeit, Bündelungseffekte erzielen zu können. Um im späten Verlauf des Verfahrens, wenn bereits eine nahezu optimale Lösung erreicht ist, Verbesserungen zu erreichen, kann es daher wichtig sein, dass γ nicht zu klein ist. Die in den Algorithmus integrierte VDNS-Struktur, die bei Stagnation die Nachbarschaftsgröße

erhöht, sorgt dafür, dass ein zu kleines γ im Laufe des Verfahrens adaptiv angepasst wird, wenn die VNS keine Wirkung zeigt. Dennoch ist es sinnvoll, den Parameter γ nicht zu klein zu initialisieren, wie zahlreiche weitere Tests gezeigt haben. Unter diesem Gesichtspunkt ist ein Wert von $\gamma = 30$ zu empfehlen. Die Abbildungen 6.1 und 6.2 zeigen, dass diese Wahl von γ zumindest zu Beginn des Verfahrens zu guten Lösungen führt.

6.6 Finale Resultate der Matheuristik

In diesem Abschnitt werden die finalen Resultate der Matheuristik zur Lösung von (Opt_{arc}) präsentiert. In Tabelle 6.7 sind das vorgegebene Laufzeitlimit T_{max} , der Umwegfaktor Δ , das Zeitlimit für die Lösung des Subproblems in jeder Iteration T_{iter} sowie der Parameter γ gegeben. Der Parameter γ wurde dabei uniform als 30 initialisiert.

Tabelle 6.7: Finale Resultate der Matheuristik

	$T_{max}[h]$	Δ	T_{iter}	γ	Optimalitätslücke
gen-21-100-10	12	1,5	150	30	29,56%
gen-21-220-6	12	1,5	150	30	11,36%
DB-21-254-10a	12	1,5	150	30	5,30%
DB-21-254-10b	12	1,5	150	30	9,64%
gen-50-224-10	12	1,5	150	30	16,54%
gen-50-525-6	12	1,5	300	30	21,38%
gen-100-547-7	12	1,5	300	30	49,70%
DB-107-1584-10	24	1	600	30	8,44%
gen-107-5763-10	24	1	600	30	65,73%
DB-147-3272-5	48	1	600	30	8,10%

Die Resultate in Tabelle 6.7 weisen große Qualitätsunterschiede aus. Bei einer differenzierten Betrachtung fällt auf, dass die Qualität der Lösungen der Realinstanzen (DB) erheblich besser ist als bei den künstlichen Instanzen (gen). Für alle Realinstanzen können zulässige Lösungen mit Optimalitätslücken im einstelligen Prozentbereich erreicht werden. Bei den künstlichen Instanzen gelingt dies nicht. Es ist nicht mit endgültiger Sicherheit zu sagen, ob die schwachen Resultate für die künstlichen Instanzen auf eine schlechte ganzzahlige Lösung oder nur auf schwache untere Schranken zurückzuführen sind. Die höhere Qualität für die Realinstanzen lässt sich dadurch erklären, dass die Warenaufkommen und die Netzstruktur Abhängigkeiten aufweisen, die von den künstlichen Instanzen nicht sinnvoll abgebildet werden. So ist die Netzstruktur ein Ergebnis historischer Aufkommen. Gleichzeitig ist der SGV in Regionen mit starker Netzstruktur für den Versender attraktiver, wodurch potentiell mehr Aufkommen entstehen kann. Werden diese Abhängigkeiten nicht von den Instanzen abgebildet, ist die Anwendung von Modellen und Lösungsverfahren, die auf diese Strukturen ausgerichtet sind, weniger erfolgversprechend.

Eine weitere Auffälligkeit der in Tabelle 6.7 dargestellten Ergebnisse ist ein nicht eindeutiger Zusammenhang zwischen Instanzgröße und Lösungsqualität. Für eine Instanz geringerer Größe ist in der Regel auch eine geringere Optimalitätslücke zu erwarten. Es gibt zwei Erklärungen für die Tatsache, dass dieser Zusammenhang hier nicht erkennbar

ist: Einerseits wurde für größere Instanzen eine längere Lösungszeit vorgegeben. Grund hierfür ist die Fokussierung auf die Aufgabe, für Instanzen realer Größe eine Lösung hoher Güte zu finden und daher die Laufzeit für kleinere Instanzen proportional geringer zu halten, um sinnvolle Rückschlüsse auf die Effizienz der Lösungsverfahren zu erhalten. Andererseits hängt die in Tabelle 6.7 gezeigte Optimalitätslücke nicht nur von der Qualität der ganzzahligen Lösung, sondern auch von der unteren Schranke (vgl. Abschnitt 6.3) ab. Wie in Unterkapitel 4.2.2 diskutiert wurde, ist das Auffinden enger unterer Schranken im vorliegenden Fall sehr schwierig. Ein Vergleich der Optimalitätslücken, die aus diesen Schranken resultieren, erlaubt daher keinen direkten Vergleich der Qualität der ganzzahligen Lösungen, da die Optimalitätslücke nur ein theoretisches Verbesserungspotential aufzeigt, das nicht notwendigerweise im ganzzahligen Fall erreichbar ist.

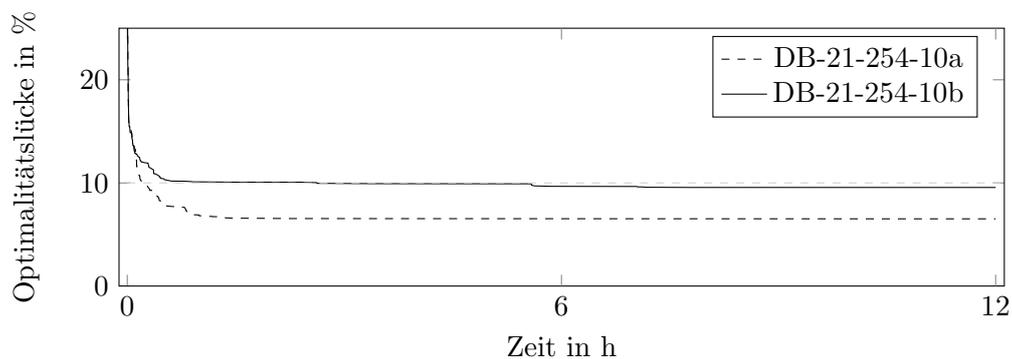


Abbildung 6.3: Lösungsverlauf für *DB-21-254-10a* und *DB-21-254-10b*

Der Versuch, ein Lösungsverfahren für reale Instanzen des Modells (Opt_{arc}) zu entwickeln, kann indes als erfolgreich angesehen werden. In Abbildung 6.3 ist der Verlauf der Optimalitätslücke der zwei kleineren Realinstanzen im Laufe der Lösungszeit abgebildet. Für beide wird bereits nach drei Stunden eine Lösung mit einer Optimalitätslücke von weniger als 10 % gefunden.

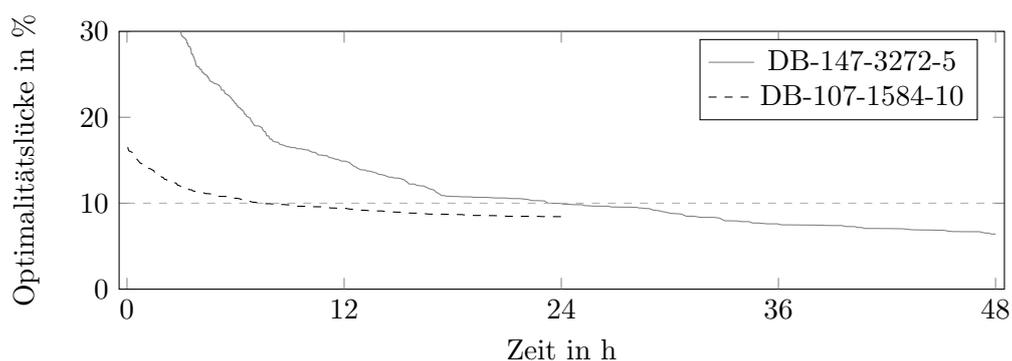


Abbildung 6.4: Lösungsverlauf für die Instanzen *DB-147-3272-5* und *DB-107-1584-10*

Für die Instanz *DB-107-1584-10* gelingt dies nach etwa zwölf Stunden (vgl. Abbildung 6.4). Obwohl anschließend noch Verbesserungen stattfinden, wäre auch nach zwölf Stunden eine akzeptable Lösung erreicht. Der Lösungsverlauf der Instanz *DB-147-3272-5* ist ebenfalls in Abbildung 6.4 illustriert. Hier dauert es etwa 24 Stunden, um eine Optimalitätslücke von 10 % zu erreichen. Auch hier sind im weiteren Verlauf Verbesserungen zu erkennen. Ob dieser Qualitätsgewinn den zeitlichen Aufwand rechtfertigt, ist bei Nutzung in der Praxis vom Anwender zu entscheiden.

Die hier vorgestellte Matheuristik wurde im Rahmen des Forschungsprojektes, aus dem diese Dissertation entstand, beim Forschungspartner DB Mobility Logistics implementiert. Die Resultate weiterer Optimierungsläufe auf einem dort verfügbaren, leistungsstarken Server zeigten eine starke Verbesserung der Laufzeit gegenüber den hier vorgestellten Ergebnissen.

Abbildung 6.6 enthält eine Darstellung der Lösung der Instanz *DB-107-1584-10* nach zwölf Stunden Laufzeit. In der Darstellung steht jede Farbe für eine eigene Relation. Die Koordinaten der Bahnhöfe entsprechen aus Datenschutzgründen nicht den realen Koordinaten.

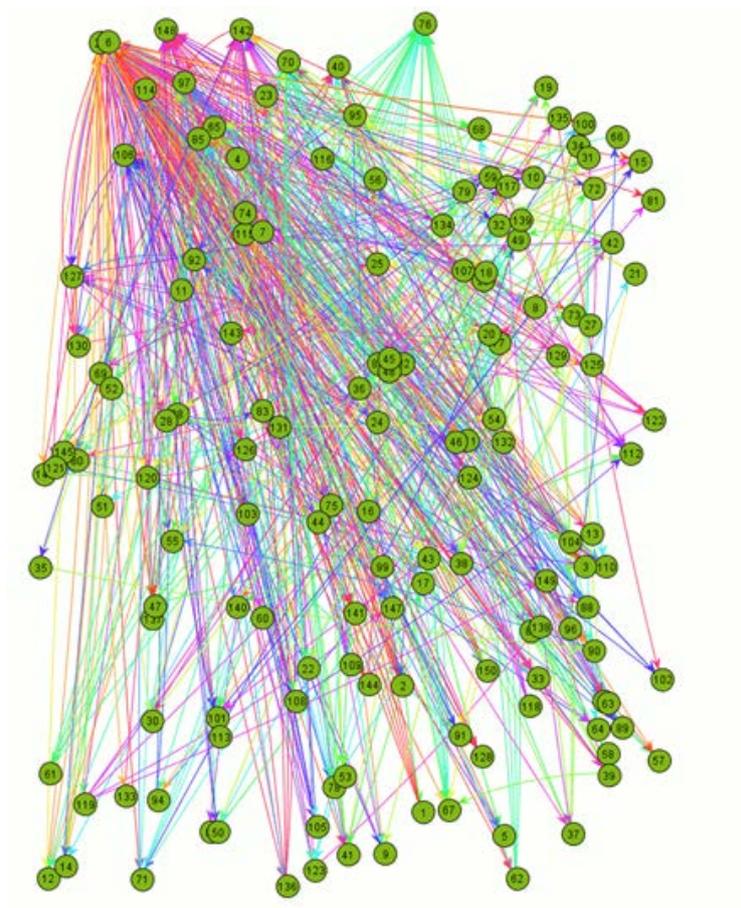


Abbildung 6.5: Darstellung einer Lösung der Instanz *DB-107-1584-10*

7 Validierung und Schlussfolgerungen

Die Lösungen des robusten Routenplanungsmodells (Opt_{arc}) werden auf Grund der Modellierung allen Zielsetzungen der Systembeschreibung aus Unterkapitel 2.3.2 gerecht. Überdies wurden in Unterkapitel 3.2.2 drei strukturelle Eigenschaften definiert, welche in den Lösungen sichtbar sein sollen: die Bündelung von Wagenströmen, die Robustheit gegenüber volatilen Wagenaufkommen und eine Entlastung der zentralen Infrastruktur. Zur Validierung der gewünschten Lösungseigenschaften ist ein Vergleich mit einem auf Basis der heutigen Leitwege erstellten Routenplan sinnvoll, um die Vorteile einer Planung ohne Leitwege aufzuzeigen. Wagenrouten realer Transporte der Deutschen Bahn dürfen nicht veröffentlicht werden. Für eine von DB Mobility Logistics zur Verfügung gestellte *Referenzinstanz* ist jedoch eine Leitwegelösung für einen Referenztag übergeben worden. Die Aufkommensdaten dieses Tages sind auch Teil der zuvor verwendeten Instanz *DB-21-254-10*. Der Datensatz enthält 21 Knoten und 254 Relationen. Damit ist die Instanzgröße zu gering, um das gesamte deutsche EWW-Netzwerk abzubilden. Die geringe Größe hat jedoch den Vorteil, dass die Lösungen übersichtlich genug sind, um grafische Darstellungen und genaue Auswertungen zu erlauben. Die Referenzinstanz wurde auf zehn Tage erweitert, indem die Tagesaufkommen randomisiert aus der Referenzinstanz erzeugt wurden. Dabei betragen die Tagesaufkommen jeder Relation zwischen 0 und 130 % des Aufkommens am Referenztag. Mit der Matheuristik war es möglich, die betreffende Instanz in drei Stunden mit einer Optimalitätslücke von 3,49 % zu lösen.

7.1 Bündelung und Transportkosteneffizienz

Die bisher in Deutschland verwendeten Leitwegepläne forcieren eine Bündelung der Verkehre auf zentrale Korridore (vgl. Unterkapitel 3.2.2). Die Leitwegepläne sind durch ihre Hierarchie- und Leitwegebedingungen bewusst so gestaltet, dass insbesondere zwischen den großen Rbf starke Bündelung stattfindet. Die im Modell (Opt_{arc}) verwendete Transportkostenstruktur auf Basis von Zugkilometern induziert ebenfalls einen Bündelungseffekt, verzichtet aber im Vergleich zu Leitwegeplänen auf die Hierarchie- und Leitwegebedingungen. Daher wird in diesem Unterkapitel analysiert, ob durch die Optimierung des Modells (Opt_{arc}) weniger effektiv gebündelt wird als in der gegebenen Lösung mit Hierarchie- und Leitwegebedingungen. Dazu werden drei Lösungen der genannten Referenzinstanz verglichen:

1. die vom Projektpartner übergebene Leitwegelösung
2. die Lösung des Modells (Opt_{arc})
3. die Lösung einer modifizierten Version von (Opt_{arc}), bei der die Summe der Wagenkosten nach nordamerikanischem Vorbild minimiert wird¹ (im Folgenden bezeichnet als *Blockinglösung*).

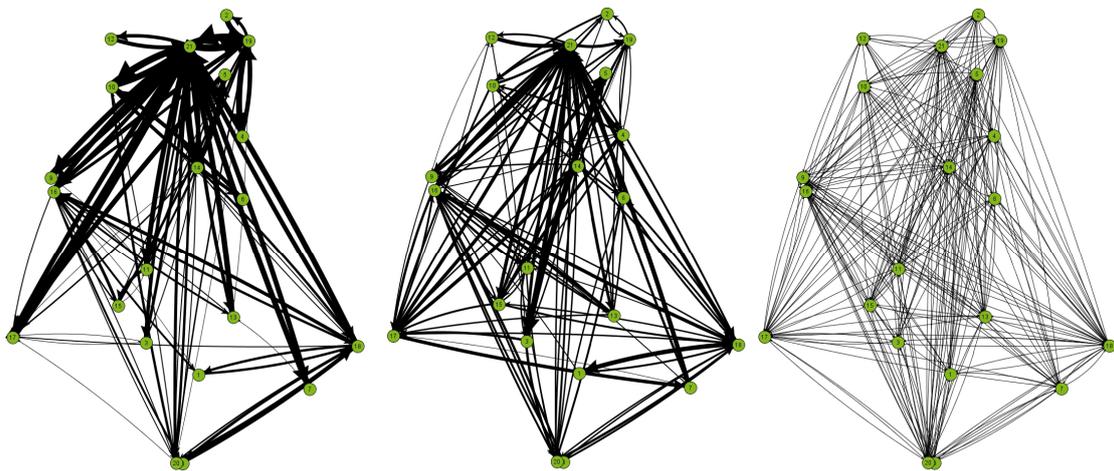


Abbildung 7.1: Visualisierungen der Lösungen (v.l.n.r.): Leitwegelösung, (Opt_{arc}), Blockinglösung

In Abbildung 7.1 sind die aggregierten Wagenströme für alle drei Lösungen visualisiert. Dabei ist die Dicke der Linien proportional zur Anzahl der Relationen, die gebündelt über diese Kante transportiert werden. Die Bündelung ist am deutlichsten sichtbar für die Leitwegelösung, während der reine Blockingansatz praktisch keinerlei Bündelungseffekte erzeugt. Im Folgenden wird analysiert, wie effektiv die Bündelung der Relationen in den einzelnen Lösungen stattfindet. Dazu werden vier Kennzahlen definiert, die einen Vergleich der drei Lösungen ermöglichen:

1. die Summe der Wagenkilometer
2. die Summe der Zugkilometer
3. die Anzahl der Umstellvorgänge
4. die Gesamtkosten, d.h. die Summe aus Transport- und Umstellkosten

Tabelle 7.1 enthält die vier genannten Kennzahlen für die drei Lösungen der Referenzinstanz. Der Kostenvergleich findet dabei auf Basis der zugbasierten Kostenstruktur aus

¹Ein Wagenkm wurde als $\frac{1}{25}$ eines Zugkm bewertet. CPLEX fand in Sekunden eine Optimallösung.

(OPT_{arc}) statt. Darüber hinaus sind die Kennzahlen in Abbildung 7.2 dargestellt. Die Werte wurden für die Abbildung bzgl. des jeweils größten der drei Werte normiert.

Tabelle 7.1: Kennzahlen dreier Lösungen der Referenzinstanz

	Leitwegelösung	Lösung von (Opt_{arc})	Blockinglösung
\sum Wagenkilometer	14.032.760	13.978.240	12.928.170
\sum Zugkilometer	730.200	625.100	1.315.100
\sum Umstellungen	61.149	59.007	44.770
\sum Gesamtkosten	5.909.925	5.225.775	9.009.850

Die Summe der gefahrenen Wagenkilometer ist für die Blockinglösung um ca. 8 % niedriger als für die beiden anderen Lösungen, deren Unterschied bei dieser Kennzahl marginal ist. Bei Betrachtung der Zugkilometer fallen gegenüber der Blockinglösung deutliche Vorteile für die Leitwegelösung und die Lösung von (Opt_{arc}) auf, die bei dieser Kennzahl den besten Wert generiert. Auch bei den Umstellungen liefert die Lösung von (Opt_{arc}) einen besseren Wert als die Leitwegelösung. In diesem Fall wird der geringste Wert jedoch beim Blockingansatz erreicht. Dies ist nicht überraschend, da Umstellungen in der Regel zu einer längeren Fahrtstrecke für den einzelnen Wagen führen und die Wagenkilometer beim Blockingansatz minimiert werden sollen. Die Gesamtkosten werden von (Opt_{arc}) am besten minimiert. Die Blockinglösung führt zu sehr hohen Gesamtkosten.

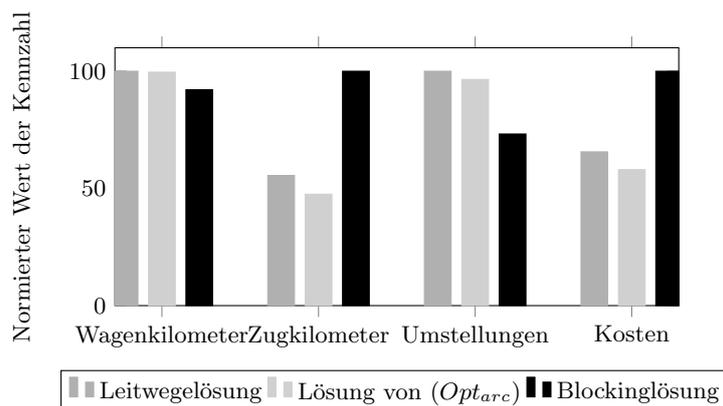


Abbildung 7.2: Normierte Kennzahlen dreier Lösungen der Referenzinstanz

Zusammenfassend ist festzustellen, dass die Minimierung von Wagenkosten, d.h. der Blockingansatz, bei dieser Instanz nicht zielführend ist. Wie bereits in Unterkapitel 3.3 erläutert, ist die bloße Minimierung von Wagenkilometern für Deutschland nicht sinnvoll, da eine Modellierung der ohne Zugkosten nicht die realen Kostenstrukturen abbildet. Eine reale Anwendung der Blockinglösung wäre im Vergleich zu den Lösungen der beiden übrigen Planungsansätze sehr unwirtschaftlich.

Ein Vergleich zwischen der Leitwegelösung und der Lösung von (Opt_{arc}) führt zu folgenden Ergebnissen (vgl. Tabelle 7.1): die Lösung von (Opt_{arc}) induziert 14,39 % weniger

Zugkilometer, 3,50 % weniger Umstellungen und damit eine Gesamtkostenersparnis von 11,58 % gegenüber der aktuell implementierten Leitwegelösung. Die Auswertung lässt für die hier betrachtete Instanz die Aussage zu, dass der mittels (Opt_{arc}) und der Matheuristik erzeugte Plan dem Leitwegeplan hinsichtlich der definierten Kennzahlen überlegen ist.

7.2 Robustheit

Die Integration multipler Szenarien in (Opt_{arc}) stellt eine Innovation bei der Erstellung von Umstellplänen dar. Das erklärte Ziel dieser Integration sind robuste Umstellpläne, die nicht auf eine einzige Aufkommensmatrix zugeschnitten sind. Es stellt sich die Frage, wie die Robustheit einer Lösung bewertet werden kann. Dazu werden in diesem Unterkapitel zwei Alternativen zur Lösung der Referenzinstanz mittels (Opt_{arc}) betrachtet und verglichen. Des Weiteren wird der *Preis der Robustheit* analysiert, d.h. die Kosten, die durch eine robuste Planung entstehen, werden mit den Kosten für Bestlösungen für jeden Einzeltag verglichen.

7.2.1 Analyse der Auswirkung multipler Szenarien in der Optimierung

Die Referenzinstanz beinhaltet zehn Aufkommensszenarien. Außer der Optimierung von (Opt_{arc}) ist kein Verfahren bekannt, das mehrere Aufkommensmatrizen bei der Planung berücksichtigen kann (vgl. Kapitel 3) und eine konsistente Lösung für alle Einzelszenarien erzeugt². Simple Heuristiken, die zulässige Lösungen für jedes einzelne Szenario erzeugen und diese dann zu einer Gesamtlösung für alle Szenarien zusammenfügen, wären zwar denkbar, aber unter dem Gesichtspunkt der Gesamtkostenoptimierung wenig erfolgversprechend. Ein Vergleich mit alternativen Planungsansätzen, die Robustheit auf andere Weise definieren, d.h. auf andere Weise multiple Aufkommensszenarien berücksichtigen, kann dennoch aufschlussreich sein. Die Erstellung eines Umstellplanes auf Basis eines einzigen Aufkommensszenarios kann mittels (Opt_{arc}) durchgeführt werden, indem nur eine Aufkommensmatrix verwendet wird. Zwei Alternativen bieten sich an, um die zehn Aufkommensmatrizen der Referenzinstanz in einer einzigen Matrix abzubilden: ein Worst-Case-Szenario und ein Szenario auf Basis gemittelter Aufkommen.

Worst-Case-Szenario

Das Worst-Case-Szenario ist in Form einer Aufkommensmatrix so zu definieren, dass die resultierende Lösung von (Opt_{arc}) für jedes Szenario $d \in \mathcal{D}_{Ref}$ der Referenzinstanz zulässig ist. Die Zulässigkeit für alle Szenarien kann nur garantiert werden, wenn für jede Relation der Instanz das jeweils maximale Aufkommen aller Szenarien der Referenzinstanz angenommen wird. Das Worst-Case-Aufkommen, gegeben als aggregierte Längen l_{WC}^k , Gewichte w_{WC}^k und Wagenzahlen v_{WC}^k , wird daher folgendermaßen defi-

²Die Zulässigkeit einer Lösung für (Opt_{arc}) impliziert die Zulässigkeit für jedes Einzelszenario

niert:

$$l_{WC}^k = \max_{d \in \mathcal{D}_{Ref}} l_d^k \quad \forall k \in \mathcal{K} \quad (7.1)$$

$$w_{WC}^k = \max_{d \in \mathcal{D}_{Ref}} w_d^k \quad \forall k \in \mathcal{K} \quad (7.2)$$

$$v_{WC}^k = \max_{d \in \mathcal{D}_{Ref}} v_d^k \quad \forall k \in \mathcal{K} \quad (7.3)$$

Die Umstellkapazitäten der Knoten der Referenzinstanz sind ausreichend, um eine zulässige Lösung der Worst-Case-Instanz mit Hilfe der Matheuristik aus Kapitel 5 zu erzeugen. Diese Voraussetzung muss im Allgemeinen nicht immer erfüllt sein, wodurch die Lösung des Worst-Case-Szenarios unmöglich würde.

Um die Wirkung des Worst-Case-Ansatzes sinnvoll bewerten zu können, wird wie folgt vorgegangen: Mit Hilfe der Matheuristik wird das Modell (Opt_{arc}) für das Worst-Case-Szenario gelöst. Die Lösung wird als *Worst-Case-Routing* bezeichnet. Des Weiteren wird (Opt_{arc}) unter Verwendung aller Szenarien der Referenzinstanz gelöst. Die Lösung wird als *robustes Routing* bezeichnet. Es wird nun verglichen, welche Kennzahlen sich ergeben, wenn das Worst-Case-Routing und das robuste Routing jeweils auf die Szenarien der Referenzinstanz angewendet werden.

Für die Worst-Case-Instanz erzeugte die Matheuristik in drei Stunden ein *Worst-Case-Routing* mit einer Optimalitätslücke von 1,28 %. Die Optimalitätslücke für das robuste Routing betrug nach der gleichen Laufzeit 3,49 %. Ein Vergleich der Kennzahlen des Worst-Case-Routings mit dem robusten Routing findet sich in Tabelle 7.2. Die Verwendung der Worst-Case-Lösung führt bei Anwendung auf die Referenzinstanz zu 10,67 % weniger Umstellungen, aber zu 14,57 % mehr Zugkilometern als die robuste Lösung. In Folge dessen erzeugt die Lösung auf Basis des Worst-Case-Szenarios um 7,44 % höhere Gesamtkosten. Die geringe Zahl der Umstellungen ist darauf zurückzuführen, dass im Worst-Case-Szenario hohe Aufkommen auf den meisten Relationen angenommen werden. Je höher das Aufkommen einer Relation ist, desto weniger lohnenswert sind Umstellungen im Vergleich zu einer direkteren Fahrt, weil dadurch Umstellkosten entstehen.

Wie bereits erklärt wurde, kann bei der Betrachtung des Worst-Case für Instanzen, in denen die Umstellkapazitäten der Knoten restriktiver sind, nicht garantiert werden, dass eine zulässige Lösung existiert. Neben den Kennzahlen spricht dieses Argument deutlich für den robusten Planungsansatz mittels (Opt_{arc}).

Tabelle 7.2: Vergleich der Kennzahlen des robusten und des Worst-Case-Routings

	robustes Routing	Worst-Case-Routing
\sum Wagenkilometer	14.032.760	13.416.310
\sum Zugkilometer	625.100	716.200
\sum Umstellungen	59.007	52.711
\sum Gesamtkosten	5.225.775	5.614.975

Mittelwertszenario

Alternativ zur Betrachtung eines Worst-Case-Szenarios ist es möglich, alle Aufkommensmatrizen der Referenzinstanz durch Mittelung der Werte in einer Matrix zu aggregieren. Daher wird in Analogie zum Worst-Case ein Mittelwertszenario mit aggregierten Längen l_{MW}^k , Gewichten w_{MW}^k und Wagenzahlen v_{MW}^k definiert:

$$l_{MW}^k = \frac{\sum_{d \in \mathcal{D}_{Ref}} l_d^k}{|\mathcal{D}_{Ref}|} \quad \forall k \in \mathcal{K} \quad (7.4)$$

$$w_{MW}^k = \frac{\sum_{d \in \mathcal{D}_{Ref}} w_d^k}{|\mathcal{D}_{Ref}|} \quad \forall k \in \mathcal{K} \quad (7.5)$$

$$v_{MW}^k = \frac{\sum_{d \in \mathcal{D}_{Ref}} v_d^k}{|\mathcal{D}_{Ref}|} \quad \forall k \in \mathcal{K} \quad (7.6)$$

Auf Basis dieser Aufkommen wurde mit Hilfe der Matheuristik in drei Stunden ein *Mittelwertrouting* mit einer Optimalitätslücke von 2,93 % erzeugt. Die Kennzahlen bei Anwendung auf die Szenarien der Referenzinstanz wurden analog zur Worst-Case-Instanz berechnet und in Tabelle 7.3 zusammengefasst.

Tabelle 7.3: Vergleich der Kennzahlen des robusten Routings und des Mittelwertroutings

	robustes Routing	Mittelwertrouting
\sum Wagenkilometer	14.032.760	13.847.430
\sum Zugkilometer	625.100	697.800
\sum Umstellungen	59.007	55.709
\sum Gesamtkosten	5.225.775	5.579.525

Ähnlich den Ergebnissen der Worst-Case-Instanz fällt für das Mittelwertrouting die Anzahl der Umstellungen um 5,59 % geringer aus, während die gefahrenen Zugkilometer um 11,63 % ansteigen. Damit führt die Berechnung auf Basis von Mittelwerten zu 6,77 % höheren Gesamtkosten als das robuste Routing. Wird die Optimierung für das Mittelwertszenario, d.h. für ein aus mehreren Einzelszenarien gemitteltes Aufkommen, durchgeführt, kann es passieren, dass die Lösung nicht für jedes dieser Einzelszenarien zulässig ist. Dieser Nachteil wird bei der Optimierung des robusten Modells (Opt_{arc}) vermieden, die überdies zu geringeren Gesamtkosten führt.

7.2.2 Preis der Robustheit

Die Erstellung eines robusten Umstellplanes ist notwendig, weil die Aufkommensschwankungen weder exakt prognostiziert noch tagesaktuell in der Planung berücksichtigt werden können (vgl. Unterkapitel 2.3.2). Hauptgrund hierfür ist aus planerischer Sicht die dynamische Auftragsannahme im EWW (vgl. Unterkapitel 2.2.2). Nichtsdestotrotz ist es

eine interessante Fragestellung, ob Umstellpläne, die exakt auf das Aufkommen einzelner Tage zugeschnitten sind, eine große Verbesserung gegenüber robusten Plänen für mittelfristige Planungszeiträume darstellen. Auf diese Weise wäre der *Preis der Robustheit* zu quantifizieren. Obwohl nach heutigem Stand eine tagesgenaue Anpassung der Umstellpläne an das Aufkommen nicht möglich ist, könnten in Zukunft bspw. verbesserte Prognoseverfahren oder ein weniger dynamisches Auftragsmanagement diesen Umstand ändern [DB 13].

Zur Beantwortung obiger Fragestellung wurde (Opt_{arc}) für die Referenzinstanz für jedes der zehn Aufkommensszenarien unabhängig voneinander gelöst. Dazu wurde erneut die Matheuristik eingesetzt. Das Zeitlimit lag für jedes Szenario bei einer Stunde. Das Ergebnis zeigt, dass die Robustheit der Lösung einen hohen Preis hat: Würde jeder Tag einzeln optimiert, wären die Gesamtkosten für die Referenzinstanz 21,09 % geringer als bei Lösung mittels (Opt_{arc}) für den gesamten Zeitraum. Dieser hohe Unterschied relativiert sich jedoch dadurch, dass bei einer langfristigen Planung, wie (Opt_{arc}) sie durchführt, in der Realität wesentlich geringere Trassenkosten entstünden als bei einer tagesgenauen Planung, die das Anmelden teurer Adhoc-Trassen erfordert (vgl. Abschnitt 2.3.1 und [SS12]). Da Trassenkosten einen großen Anteil der Gesamtkosten einer Zugfahrt ausmachen, ist das Ergebnis obiger Rechnung demnach mit Vorsicht zu bewerten. Es kann jedoch festgehalten werden, dass täglich angepasste Umstellpläne hohes Potential böten.

Zusammenfassung Robustheit

Die Verwendung des Modells (OPT_{arc}) erlaubt die Integration multipler Aufkommensmatrizen und generiert für die hier betrachtete Referenzinstanz erheblichen Mehrwert gegenüber der Verwendung gemittelter Aufkommen oder einer Planung auf Basis von Spitzenaufkommen (Worst-Case-Szenario). Dies wurde durch den Vergleich der Lösungen von (Opt_{arc}), des Worst-Case-Szenarios und des Mittelwertszenarios deutlich. Obwohl eine praktische Implementierung tagesgenauer Umstellpläne derzeit nicht möglich ist, wurde mit Hilfe der Matheuristik eine tagesgenaue Planung durchgeführt und mit der robusten Lösung verglichen. Die tagesgenaue Planung kann zu deutlichen Einsparungen führen, wenn technische und organisatorische Probleme ausgeräumt werden könnten. Der Ansatz, eine robuste Lösung auf Basis verschiedener Aufkommen zu generieren, stellt eine wissenschaftliche Neuerung dar, die praxisnähere Umstellpläne erzeugt als bisherige Verfahren.

7.3 Entlastung der zentralen Infrastruktur

Das heutige System der Leitwege im deutschen EWV führt zu einer starken Zentralisierung der Verkehre und Umstellarbeiten auf zentrale Korridore und Umstellbahnhöfe. Mit Blick auf potentiell steigende Verkehrsaufkommen im EWV (vgl. Abschnitt 2.2.1) ist zu erwarten, dass hierdurch Nadelöhre entstehen. In diesem Unterkapitel werden daher verschiedene Kennzahlen der Referenzinstanz untersucht, die Aufschluss über die Verteilung der Verkehre über das Netzwerk geben. Dazu wird zunächst in Abschnitt 7.3.1 die Auslastung von Zügen und Streckenabschnitten bei Verwendung der Leitwegelösung

bzw. der Lösung von (Opt_{arc}) untersucht. In Abschnitt 7.3.2 wird die Auslastung der Umstellanlagen in beiden Lösungen verglichen.

7.3.1 Auslastung von Zügen und Streckenabschnitten

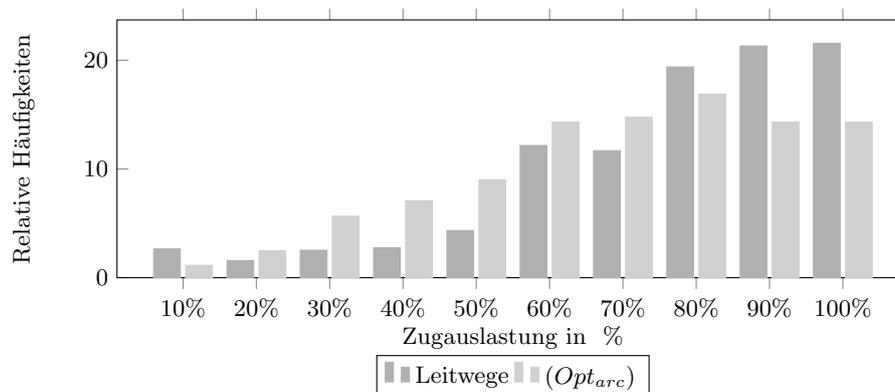


Abbildung 7.3: Dezile der relativen Zugauslastungen

Die Bündelung von Verkehren ist das erklärte Ziel des hierarchischen Leitwegesystems, das die Auslastung der Züge auf Hauptstrecken maximieren soll (vgl. Abschnitt 3.2.2). Daher soll in diesem Abschnitt zunächst überprüft werden, ob die Zugauslastung eine zielführende Kennzahl bei der Bewertung von Umstellplänen darstellt. In Abbildung 7.3 sind die Dezile der relativen Häufigkeiten der Zugauslastungen für die Leitwegelösung und die Lösung von (Opt_{arc}) abgetragen. Es ist zu erkennen, dass die Anzahl stark ausgelasteter Züge ($> 80\%$) im Falle der Leitwegelösung erheblich größer ist als bei der Lösung von (Opt_{arc}). Auch die Medianauslastung liegt mit $76,57\%$ gegenüber $67,74\%$ deutlich höher. Dafür verkehren in der Lösung von (Opt_{arc}) mehr Züge im mittleren Auslastungsbereich zwischen 30 und 70% . Leitwegepläne führen somit in der Regel zu einer sehr viel höheren Zugauslastung. Die Auslastung allein ist jedoch nur in bedingtem Maße ein sinnvolles Kriterium, wenn die Gesamtkosten zu minimieren sind: Wie bereits in Unterkapitel 7.1 gezeigt, führt die Leitwegelösung zu $13,09\%$ höheren Gesamtkosten und einer Erhöhung der Zugkilometer um $16,81\%$. Es ist somit festzustellen, dass die hohe Auslastung der Züge mit großen Umwegen erkauft wird, die in der Summe höhere Kosten als weniger gut ausgelastete Züge verursachen.

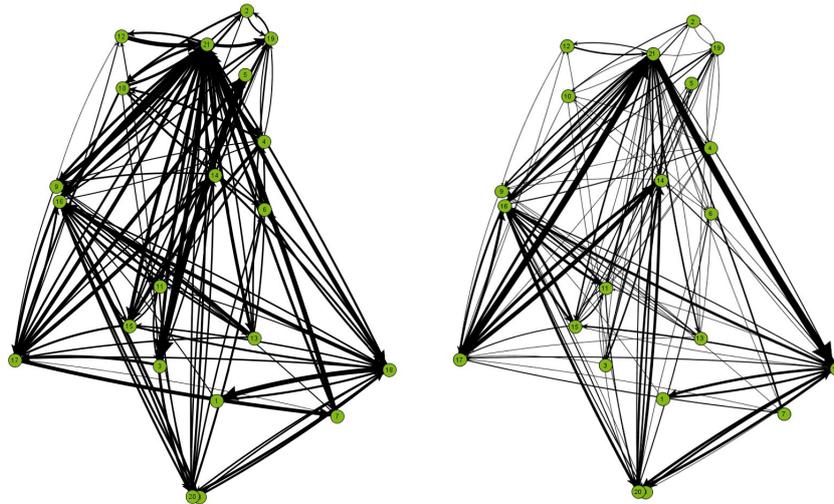


Abbildung 7.4: Züge je Kante (links: Leitwegelösung, rechts: Lösung von (Opt_{arc}))

Auch bei der Verteilung der Zugfahrten über das Netzwerk ergeben sich Unterschiede: Der Leitwegeplan führt zu täglich 304 Zügen auf 83 unterschiedlichen Kanten, d.h. Streckenabschnitten des Netzwerkes. Im Falle der Lösung von (Opt_{arc}) entstehen mit 268 Zugverbindungen nicht nur 11,84 % weniger Fahrten, sie verteilen sich überdies auf 36 % mehr, nämlich auf 113, verschiedene Kanten. Diese Zahlen sind ein Beweis für eine bessere Verteilung der Verkehre über das Netzwerk. Dieser Sachverhalt ist in Abbildung 7.4 illustriert, in der die Kantenstärke proportional zur Anzahl der auf der Kante verkehrenden Züge ist.

7.3.2 Auslastung der Umstellanlagen

In diesem Abschnitt wird die Auslastung der Knoten der Referenzinstanz für die Leitwegelösung und die Lösung von (Opt_{arc}) verglichen. In Tabelle 7.4 sind die Knotenauslastungen für beide Lösungen sowie deren Differenz aufgeführt. Zum Vergleich wurde zusätzlich die Anzahl der Richtungsgleise im jeweiligen Knoten vermerkt, die eine sinnvolle Größenangabe für die Knoten darstellt. Es fällt auf, dass die durchschnittliche Auslastung über alle Knoten in beiden Fällen sehr gering ausfällt. Dies ist dem vergleichsweise geringen Gesamtaufkommen, respektive zu hohen Anlagenkapazitäten, der Referenzinstanz geschuldet.

Der Spalte *Differenz* ist zu entnehmen, dass die Auslastung der größten Knoten, mit Ausnahme von Knoten 17, bei Verwendung der (Opt_{arc}) -Lösung geringer ist als für die Leitwegelösung. Die Umstellarbeiten werden zu einem größeren Anteil in kleineren und weniger zentralen Knoten durchgeführt. Es findet folglich für die Referenzinstanz eine Entlastung der im Leitwegesystem zentralen Knoten statt.

Tabelle 7.4: Auslastung der Knoten für Leitwege- und (Opt_{arc})- Lösung

Knoten	Richtungsgleise	Auslastung Leitwege	Auslastung (Opt_{arc})	Differenz
1	5	25,00%	25,00%	0,00%
2	5	15,00%	25,00%	10,00%
3	5	25,00%	35,00%	10,00%
4	5	10,00%	20,00%	10,00%
5	5	15,00%	10,00%	-5,00%
6	5	5,00%	20,00%	15,00%
7	10	12,50%	12,50%	0,00%
8	10	5,00%	5,00%	0,00%
9	10	10,00%	12,50%	2,50%
10	10	15,00%	10,00%	-5,00%
11	10	35,00%	35,00%	0,00%
12	10	7,50%	10,00%	2,50%
13	10	12,50%	15,00%	2,50%
14	15	31,67%	31,67%	0,00%
15	20	12,50%	18,75%	6,25%
16	40	24,37%	24,37%	0,00%
17	35	17,86%	20,71%	2,85%
18	30	25,00%	22,50%	-2,50%
19	30	17,50%	8,33%	-9,17%
20	25	20,00%	18,00%	-2,00%
21	25	56,00%	44,00%	-12,00%

Zusammenfassung zur Entlastung zentraler Infrastruktur

In diesem Unterkapitel wurde die Verteilung der Verkehre und Umstellarbeiten der Referenzinstanz analysiert, wenn diese mittels (Opt_{arc}) gelöst wird. Dazu wurde zunächst die Zugauslastung der (Opt_{arc})-Lösung und der Leitwegelösung verglichen. Dabei konnte festgestellt werden, dass Leitwege zwar zu einer höheren Zugauslastung führen, wodurch jedoch größere Umwege entstehen. Des Weiteren wurde festgestellt, dass die Anzahl der Zugkanten gegenüber der Leitwegelösung gesenkt werden kann und dass sich die Zugfahrten überdies besser über das Netzwerk verteilen lassen. Die nicht hierarchische Struktur der Lösungen von (Opt_{arc}) führt überdies zu einer gleichmäßigeren Verteilung der Umstellarbeiten, d.h. zu einer Entlastung der großen Anlagen.

Schlussfolgerungen aus der Validierung

In diesem Kapitel wurden die Resultate der Optimierung des Modells (Opt_{arc}) an einer exemplarisch ausgewählten Instanz ausgewertet. Die betrachtete Instanz eignete sich für die Untersuchung, weil eine Referenzlösung auf Basis des heute angewendeten Leitwegesystems für diese Instanz gegeben ist. Dabei konnte festgestellt werden, dass die zusätzlichen Freiheitsgrade, die bei Vernachlässigung der Leitwege- und Hierarchiebedingungen in Erscheinung treten, eine deutliche Verringerung der Zugkilometer zulassen, ohne eine größere Anzahl von Umstellvorgängen erforderlich zu machen. Die Verringerung der Zugkilometer lag bei etwa 14 % und resultierte in 11 % geringeren Gesamt-

kosten. Dieses Ergebnis lässt die Schlussfolgerung zu, dass Leitwege und insbesondere die Netzhierarchie ein zu starres System sind und, wie im Falle der Referenzinstanz, zu höheren Kosten führen als ein flexibleres System.

Im Rahmen der Analysen konnte gezeigt werden, dass die Auslastung von Zügen nicht zielführend für die Minimierung der Gesamtkosten hinsichtlich der gegebenen Kostenfunktion ist. Obwohl bei Anwendung der Leitwege im Vergleich zur Lösung von (Opt_{arc}) sehr viel mehr Züge eine Auslastung von mehr als 80 % aufweisen und überdies die Medianauslastung erheblich größer war, wird kein Kostenvorteil erreicht. Die hohe Auslastung wird mit Umwegen erkauft.

Eine Auswertung der resultierenden Knotenauslastungen in beiden Lösungen zeigte auf, dass die Lösung von (Opt_{arc}) zu einer weniger starken Auslastung der (gemessen in Rangiergleisen) größten Anlagen führt. Darüber hinaus findet eine stärkere Verteilung der Zugfahrten auf mehr Kanten des Netzwerkes statt. Die Folge ist eine gleichmäßigere Auslastung der netzweiten Infrastruktur.

Für eine Anwendung in der Realität wurde bereits in vorangegangenen Kapiteln dieser Arbeit die Robustheit der Lösung gegenüber Aufkommensschwankungen, repräsentiert durch die Zulässigkeit für multiple Szenarien, als wichtige Lösungseigenschaft identifiziert. Zur Validierung dieser Lösungseigenschaft wurde die Lösung von (Opt_{arc}) mit einem Worst-Case-Fall und mit einer Lösung für gemittelte Aufkommen verglichen. Die Lösung von (Opt_{arc}) ist im Gegensatz zu den Vergleichslösungen des Mittelwert-Szenarios und des Worst-Case-Szenarios nicht nur zulässig für jedes Einzelszenario, sondern zeigt überdies klare Kostenvorteile.

Zusammenfassend kann festgestellt werden, dass das Modell (Opt_{arc}) die drei gewünschten Lösungseigenschaften der kosteneffizienten Bündelung von Wagenflüssen, der Entlastung zentraler Infrastruktur und der Robustheit gegenüber multiplen Aufkommensszenarien vereint.

8 Zusammenfassung und Ausblick

In dieser Arbeit wurde die taktische Planung von Wagenrouten im Einzelwagenverkehr mit Hilfe mathematischer Methoden untersucht. Im Zuge einer allgemeinen Einführung in den Schienengüterverkehr in Deutschland lag der Fokus daher auf dem Einzelwagenverkehr. Nach Betrachtungen zur Zukunftsfähigkeit dieser Produktionsform fand eine Darstellung der Prozesse zur Auftragsdurchführung und der dazu notwendigen Planungsschritte statt. Eine Vertiefung der Thematik der Erstellung von Umstellplänen und der dazu äquivalenten Wagenroutenplanung führte zu einer Systembeschreibung und zu einer Definition der weiteren Problemstellung. Dabei wurde neben technischen Randbedingungen vor allem die Robustheit gegenüber Aufkommensschwankungen als wichtige Eigenschaft guter Umstellpläne herausgearbeitet.

Im Kapitel *Stand von Wissenschaft und Technik* wurde eine kurze Einführung in graphentheoretische Begriffe und Verfahren der mathematischen Optimierung gegeben. Des Weiteren wurde wissenschaftliche Literatur zu allgemeinen Planungsproblemen im Schienengüterverkehr vorgestellt und die weiteren Betrachtungen auf das Problem der Routenplanung eingegrenzt. Bei der Bewertung der wissenschaftlichen Literatur zur Routenplanung im Einzelwagenverkehr wurden drei Klassen von Umstellplänen identifiziert: individuelle Pläne für jeden Wagen, Blockingpläne und Leitwegepläne. Die Vor- und Nachteile der einzelnen Varianten wurden ebenfalls identifiziert: ein höherer Grad der Flexibilisierung von Wagenrouten steht im scheinbaren Widerspruch zu einer kosteneffizienten Konsolidierung von Wagenflüssen. Die Möglichkeit, diesen Widerspruch durch eine mathematische Herangehensweise aufzulösen, wurde als eines der Untersuchungsziele definiert.

Anschließend wurde ein mathematisches Modell zur robusten Planung von Wagenrouten im Einzelwagenverkehr vorgestellt. Es wurden alle Bedingungen hinsichtlich Zug- und Knotenkapazitäten aus der Systembeschreibung übernommen und darüber hinaus Variablen- und Kostenstrukturen so gewählt, dass die gewünschten Konsolidierungseffekte entstehen. Dazu wurden zugabhängige Transportkosten und wagenabhängige Umstellkosten berücksichtigt. Die Robustheit von Lösungen wurde durch die Integration multipler Aufkommensszenarien im Modell abgebildet. Das resultierende Netzwerkdesignmodell kann auf zweierlei Weise, kanten- oder pfadbasiert, dargestellt werden. Eine Komplexitätstheoretische Analyse zeigte auf, dass bereits das Finden einer zulässigen Lösung ein \mathcal{NP} -vollständiges Problem darstellt. Die Erzeugung unterer Schranken mit Hilfe linearer Relaxierungsmethoden wurde polyedrisch betrachtet und Annahmen zur Modellvereinfachung diskutiert.

Die Entwicklung von Lösungsverfahren für die zuvor genannten Modelle war Thema eines eigenen Kapitels. Eine Analyse der Problemstruktur motivierte den Einsatz von Column-Generation-Verfahren, einer Dekompositionsmethode für lineare Optimierungs-

probleme. Ihre Anwendung lässt das Problem der Wagenroutenplanung in viele Teilprobleme zerfallen, die sehr effizient gelöst werden können. Die Einbindung von Column Generation in ein Branch-and-Bound-Schema brachte algorithmische Probleme mit sich, die jedoch ausgeräumt werden konnten. Das Ergebnis war ein effizientes Branch-and-Price-Verfahren, das in numerischen Tests der kommerziell verfügbaren Optimierungsoftware überlegen war. Die Lösung von Realinstanzen praxisnaher Größe gelang jedoch nicht. Daher wurde die Anwendung von metaheuristischen Ansätzen ins Auge gefasst. Ihre Entwicklung orientierte sich an wissenschaftlichen Designprinzipien für Metaheuristiken und führte zu einer neuen Matheuristik, d.h. einer Kombination von Elementen metaheuristischer und exakter Optimierungsmethoden. Es entstand ein Verfahren, das eine zulässige Startlösung erzeugt und diese sukzessive verbessert.

Im darauf folgenden Kapitel wurden die Resultate aller entwickelten Verfahren für eine Reihe von Instanzen vorgestellt, die in Kooperation mit Experten bei DB Mobility Logistics erzeugt wurden und damit praxisnahen Ansprüchen genügen. Dabei konnte gezeigt werden, dass die entwickelte Metaheuristik auch für Instanzen realer Größe in praxistauglicher Laufzeit zulässige Lösungen mit einer für praktische Maßstäbe zufriedenstellenden Optimalitätslücke erzeugt.

Anhand einer Referenzinstanz, für die eine Lösung im Sinne des bisher verwendeten Produktionsverfahrens gegeben ist, wurden die Resultate von Modellierung und Algorithmen im letzten Kapitel ausgewertet. Dabei konnte validiert werden, dass durch das neue Modell die Konsolidierung von Verkehren trotz hoher Flexibilität bei der Routenführung Erfolg haben kann. Eine deutliche Kostenersparnis konnte trotz einer gleichmäßigeren Auslastung der Infrastruktur realisiert werden. Auch der Ansatz, Robustheit über multiple Szenarien abzubilden, stellte sich als sinnvolles Vorgehen heraus. Für die gegebene Referenzinstanz konnte der neue Optimierungsansatz seine Überlegenheit gegenüber älteren Produktionsverfahren qualitativ und quantitativ beweisen.

Transfer der Resultate in Forschung und Praxis

Die vorgestellten Modelle und Algorithmen fokussieren ein konkretes Planungsproblem des deutschen Einzelwagenverkehrs. Daher stellt sich die Frage, ob die Ergebnisse verallgemeinert werden können und praktischen Mehrwert stiften.

Hinsichtlich wissenschaftlicher Verwertungszwecke der Modelle und Algorithmen ist es sinnvoll, die Problematik zunächst zu abstrahieren. Das definierte Modell berechnet Routen für einzelne Sendungen, die durch Kapazitäten in Umschlagknoten und eine zugbezogene Kostenstruktur voneinander abhängen. Dabei werden die Zugkapazitäten in den zwei Dimensionen Länge und Gewicht berücksichtigt. Vergleichbare Planungsprobleme treten bei jeder Form des Teilladungsverkehrs auf: im Straßengüterverkehr, bei der Konsolidierung von Sendungen durch KEP-Dienstleister und im Luft- und Seefrachtverkehr. Die jeweiligen Anforderungen der genannten Einsatzbereiche können in Form geeigneter Nebenbedingungen im Modell abgebildet werden.

Obwohl die entwickelten Lösungsansätze auf die speziellen Modelle angepasst sind, können viele Ideen übernommen werden. Die Anwendung des Branch-and-Price-Ansatzes

mit seinen Besonderheiten hinsichtlich des Brachingschemas wird derzeit in einem Forschungsprojekt zur Lenkung von Sendungsflüssen im Straßengüterverkehr an der TU Dortmund untersucht. Die Matheuristik kann durch ihren modularen Aufbau ebenfalls an neue Nebenbedingungs- und Kostenstrukturen angepasst werden.

Die in dieser Arbeit vorgestellte Modellierung der Routenplanung könnte um weitere praktische Aspekte erweitert werden, sofern adäquate Parameter ermittelt werden können: Die Bündelung von Wagen ist nur möglich, wenn sie zur gleichen Zeit am gleichen Ort sind. Eine untertägige Planungsgenauigkeit würde aus dieser Perspektive Sinn machen. Da die Prognosen über die Wagenaufkommen selbst nur tagesgenau sind, war die Einbindung bisher nicht möglich. Ein aus praktischer Sicht sehr interessanter Punkt sind zeitliche Restriktionen für die Gesamttransportzeit einer Relation. In dieser Arbeit wurde auf explizite Zeitrestriktion zu Gunsten einer Begrenzung der Anzahl der Umstellungen jedes Wagens verzichtet. Wenn hinreichend gute Parameter für Transport- und Umstellzeiten zur Verfügung ständen, könnte dieser Aspekt modellseitig einfach integriert werden, ohne signifikante Anpassungen in den vorgestellten Algorithmen zu erfordern. Auf diese Weise könnten auch Sendungen mit unterschiedlicher zeitlicher Priorisierung abgebildet werden. Die genannten Erweiterungsmöglichkeiten der Modellierung offenbaren den Nutzen weiterer Verbesserungen der Prognoseverfahren für künftige Verkehrsaufkommen. Weiterer Forschungsbedarf besteht in verschiedenen Themenfeldern, die an die taktische Routenplanung angrenzen. Die hohe Komplexität der Planungsprozesse im Schienengüterverkehr führte bislang häufig zu sequentiellen Planungsmodellen. Die integrierte Betrachtung verschiedener Probleme bietet jedoch große Chancen und wird durch stetige Fortschritte der Wissenschaft zunehmend praxistauglich.

Der Ergebnistransfer in die Praxis gelingt für die hier vorgestellten Ergebnisse maßgeblich durch die Projektkooperation mit DB Mobility Logistics. Im Rahmen des *Effizienzclusters LogistikRuhr* wurde das Forschungsprojekt *Ein benutzergeführtes Planungsinstrument zur Bündelung von SGV-Produkten unter Nutzung dezentraler Übergabepunkte* durchgeführt und Teilergebnisse immer wieder gemeinsam bewertet. Eine Version der in dieser Arbeit entwickelten Matheuristik zur Planung von Wagenrouten ist beim Praxispartner bereits lauffähig implementiert. Bisher existierte dort keine Softwarelösung, die für Verkehrsaufkommen mehrerer Tage ein Wagenrouting erzeugen kann.

Bis zur operativen Nutzung der Algorithmen in der Planung des Einzelwagenverkehrs sind jedoch noch Entwicklungsschritte erforderlich: Die Eingabeparameter der Algorithmen stellen prognostizierte Aufkommensdaten dar, deren Vorhersage sehr zeitaufwändig ist. Es wäre somit eine stärkere Automatisierung dieses Prognoseprozesses und eine Anbindung an die in dieser Arbeit vorgestellten Algorithmen erforderlich. Die optimierten Routen, die von den Algorithmen bestimmt werden, ziehen, wie in früheren Kapiteln bereits erwähnt, zahlreiche Implikationen in der Triebfahrzeug-, Personal- und Trassenplanung nach sich. Die Einbindung der Optimierungsergebnisse in diese Planungsschritte erfordert somit Anpassung im übergeordneten Planungsprozess.

Die Anpassung von Wagenrouten hat netzweite Auswirkung auf die Umstellanlagen. Die aus der Optimierung der Routen resultierende Anpassung von Umstellplänen zieht somit die Notwendigkeit eines Informationsflusses in Richtung der Umstellanlagen nach sich, die auf die Anpassungen adäquat zu reagieren hätten. Diese Propagation von Rou-

teninformationen durch alle Hierarchien der Planung bis zur operativen Durchführung der Umstellarbeiten wirft erneut die Frage der Wechselfrequenz der Routenpläne, d.h. die Häufigkeit der Reoptimierung der Routenpläne auf. Die Wahl der in dieser Arbeit als taktische Planungszeiträume bezeichneten Zeitintervalle, in denen Routenpläne angepasst werden, sind daher in der Praxis zu bewerten. Dabei darf nicht außer Acht gelassen werden, dass diese Arbeit auf der Annahme kurzfristig eingehender Transportaufträge basiert. Die Anforderungen an die Servicequalität hinsichtlich der Voranmeldung von Transporten wird jedoch im Rahmen der Diskussion um künftige Produktionsverfahren neu bewertet:

Bei DB Schenker Rail findet aktuell eine Zusammenlegung von Einzelwagenverkehr und Ganzzugverkehr zu einer sogenannten *Netzwerkbahn* statt, die flexiblere Wagenrouten durch Blockingansätze ermöglichen soll [DB 13, HH13]. Dabei sollen nicht vollständig ausgelastete, regelmäßig verkehrende Ganzzüge mit Einzelwagen aufgefüllt werden. Diese Ankerblöcke genannten Ganzzüge können als fixierte Variablen problemlos in die hier vorgestellten Modelle integriert werden. Die Entwicklung des neuen Produktionsverfahrens Netzwerkbahn verdeutlicht die praktische Relevanz der in dieser Dissertation gewonnenen Erkenntnisse und den Wert kooperativer Projekte zwischen Theorie und Praxis.

Literaturverzeichnis

- [AEOP02] AHUJA, R.K. ; ERGUN, Ö. ; ORLIN, J.B ; PUNNEN, A.P: A survey of very large-scale neighborhood search techniques. In: *Discrete Applied Mathematics* 123 (2002), Nr. 1, S. 75–102
- [AJL07] AHUJA, R.K. ; JHA, K. C. ; LIU, J.: Solving real-life railroad blocking problems. In: *Interfaces : the INFORMS journal on the practice of operations research* 37 (2007), Nr. 5, S. 404–419
- [AMO93] AHUJA, R.K. ; MAGNANTI, T.L ; ORLIN, J.B: *Network flows: Theory, algorithms, and applications*. Englewood Cliffs and NJ : Prentice-Hall, 1993. – ISBN 9780136175490
- [All07] ALLIANZ PRO SCHIENE: *Mehr Bahn wagen: 13 erfolgreiche Verlagerungsbeispiele aus dem Güterverkehr*. (2007)
- [Ass80] ASSAD, A.A: Modelling of Rail Networks: Towards a Routing/Makeup Model. In: *Transportation Research Part B: Methodological* (1980), Nr. 14, S. 101–114
- [BHV00] BARNHART, C. ; HANE, C. ; VANCE, P.: Using Branch-and-Price-and-Cut to Solve Origin- Destination Integer Multicommodity Flow Problems. In: *Operations Research* 48 (2000), Nr. 2, S. 318–326
- [BJN⁺98] BARNHART, C. ; JOHNSON, E.L ; NEMHAUSER, G.L ; SAVELSBERGH, M.W.P ; VANCE, P.: Branch-And-Price: Column Generation For Solving Huge Integer Programs. In: *Operations Research* 46 (1998), Nr. 3, S. 316–329
- [Ber01] BERNDT, T.: *Eisenbahngüterverkehr*. 1. Auflage. Stuttgart : Teubner, 2001. – ISBN 9783519063872
- [Bia13] BIALLAS, J.: Im Bundestag notiert: Verlagerung von Verkehren auf Schiene und Wasserstraße. (30.8.2013). http://www.bundestag.de/presse/hib/2012_05/2012_249/07.html
- [BGS80] BODIN, L.D ; GOLDEN, B.L ; SCHUSTER, A.D: A Model for the Blocking of Trains. In: *Transportation Research Part B: Methodological* 14 (1980), S. 115–120

- [Bru06] BRUCKMANN, D.: *Entwicklung einer Methode zur Abschätzung des containerisierbaren Aufkommens im Einzelwagenverkehr und Optimierung der Produktionsstruktur*. Essen : Universität Duisburg–Essen, 2006. – ISBN 3922602274
- [BWZ97] BUSSIECK, M. ; WINTER, T. ; ZIMMERMANN, U.: Discrete optimization in public rail transport. In: *Mathematical Programming* (1997), Nr. 79, S. 415–444
- [CLPR06] CAMPETELLA, M. ; LULLI, G. ; PIETROPAOLI, U. ; RICCIARDI, N.: *Freight Service Design for the Italian Railways Company*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik and OASiCs - OpenAccess Series in Informatics. 6th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'06), 2006
- [CGL⁺08] CESELLI, A. ; GATTO, M. ; LÜBBECKE, M. E. ; NUNKESSER, M. ; SCHILLING, H.: Optimizing the Cargo Express Service of Swiss Federal Railways. In: *Transportation Science* 42 (2008), Nr. 4, S. 450–465
- [CG13] CLAUSEN, U. ; GEIGER, C.: *Handbuch der Verkehrs- und Transportlogistik*. Berlin : Springer, 2013. – ISBN 9783540342984
- [CV11] CLAUSEN, U. ; VOLL, R.: A Blocking Model with Bundling Effects Respecting Multiple OD-Matrices. In: *Proceedings of the 4th IC-EPSMSO* (2011), Nr. 1, S. 5–12
- [CV13] CLAUSEN, U. ; VOLL, R.: A comparison of North American and European railway systems. In: *European Transport Research Review* (2013), S. 1–5
- [CTV98] CORDEAU, J.-F ; TOTH P. ; VIGO D.: *A survey of optimization models for train routing and scheduling*. Montréal : Groupe d'études et de recherche en analyse des décisions, 1998
- [CLRS10] CORMEN, T. H. ; LEISERSON, C. E. ; RIVEST, R. ; STEIN, C.: *Algorithmen - eine Einführung*. 3. Auflage. München : Oldenbourg, 2010. – ISBN 9783486590029
- [CFR83] CRAINIC, T.G ; FERLAND, J.A ; ROUSSEAU, J.-M: *A tactical planning model for rail freight transportation*. Rév. Montréal : Université de Montréal, Centre de recherche sur les transports, 1983
- [CSC12] CREVIER, B. ; SAVARD, G. ; CORDEAU, J.-F: Integrated operations planning and revenue management for rail freight transportation. In: *Transportation Research Part B: Methodological* (2012), S. 100–119
- [DB 11] DB MOBILITY LOGISTICS AG: Zug um Zug. In: *logistics Kundenmagazin* (2011), Nr. 3, S. 42–47

- [DB 13] DB NETZE: *Trassenpreissystem (TPS): Liste der Entgelte der DB Netz AG 2013 für Trassen, Zusatz- und Nebenleistungen*. 2013
- [DB 12] DB NETZE: *835m lange Güterzüge zwischen Padborg (DK) und Maschen geplant*. www.dbnetze.com/laengeregueterzuege. Version: 25.09.2012
- [DB 13] DB SCHENKER RAIL: Innovation: Für mehr Qualität und Transparenz. In: *Railways* (2013), Nr. 2, S. 17–21
- [DDS02] DESAULNIERS, G. ; DESROSIERS, J. ; SOLOMON, M. M.: Accelerating Strategies in Column Generation Methods for Vehicle Routing and Crew Scheduling Problems. In: *Essays and Surveys in Metaheuristics* Bd. 15. Springer, 2002, S. 309–324
- [Des06] DESAULNIERS, G. et a.: *Column Generation*. Dordrecht : Springer, 2006. – ISBN 0387254862
- [DL10] DESROSIERS, J. ; LÜBBECKE, M.: Branch–Price–and–Cut Algorithms. In: *Wiley Encyclopedia of Operations Research and Management Science* (2010). ISBN 9780470400531
- [Deu13] DEUTSCHE BAHN AG: *Wettbewerbsbericht*.(2013)
- [DM07] D’INCÀ, J. ; MOSIMANN, U.: Internationalizing for European Rail Freight Growth. In: *Oliver Wyman Consulting* (2007)
- [EKJ06] EISENKOPF, A. ; KIRCHNER, C. ; JARZEMBOWSKI, G.: The liberalisation of rail transport in the EU. In: *Intereconomics* 41 (2006), Nr. 6, S. 292–313
- [Far90] FARLEY, A. A.: A note on bounding a class of linear programming problems, including cutting stock problems. In: *Operations Research* 38 (1990), Nr. 5, S. 922–923
- [FR95] FEO, T. A. ; RESENDE, M. G.: Greedy randomized adaptive search procedures. In: *Journal of global optimization* 6 (1995), Nr. 2, S. 109–133
- [Fie05] FIEDLER, J.: *Bahnwesen: Planung, Bau und Betrieb von Eisenbahnen, S-, U-, Stadt- und Strassenbahnen*. 5. Auflage. Düsseldorf : Werner, 2005. – ISBN 3804116124
- [Fra12a] FRANKFURTER ALLGEMEINE ZEITUNG: Die Güterbahn DB Schenker denkt europäisch: Interview mit dem Vorstandsvorsitzenden von DB Schenker Rail Alexander Hedderich. In: *Frankfurter Allgemeine Zeitung* (02.03.2012)
- [Fra12b] FRAUNHOFER INSTITUT INTEGRIERTE SCHALTUNGEN (IIS) UND KPMG: *Wirtschaftliche Rahmenbedingungen des Güterverkehrs: Studie zum Vergleich der Verkehrsträger im Rahmen des Logistikprozesses in Deutschland*. 2012

- [FHS09] FÜGENSCHUH, A. ; HOMFELD H. ; SCHÜLLDORF H.: Single Car Routing in Rail Freight Transport. In: *Dagstuhl Seminar Proceedings - Models and Algorithms for Optimization in Logistics* (2009)
- [FAPU02] FUKASAWA, R. ; ARAGÃO, M.V de P. ; PORTO, O. ; UCHOA, E.: Solving the Freight Car Flow Problem to Optimality. In: *Electronic Notes in Theoretical Computer Science* 66 (2002), Nr. 6, S. 42–52
- [GJ97] GAREY, M.R ; JOHNSON, D. S.: *Computers and intractability: A guide to the theory of NP-completeness*. 19. New York and NY : Freeman, 1997. – ISBN 0716710455
- [GC94] GENDROM, B. ; CRAINIC, T.G: Relaxations for Multicommodity Capacitated Network Design Problems. In: *Publication CRT-965, Centre de recherche sur les transports, Université de Montréal* (1994)
- [Gor98] GORMAN, M.F: An application of genetic and tabu searches to the freight railroad operating plan problem. In: *Annals of Operations Research* (1998), Nr. 78, S. 51–69
- [HGC⁺10] HARROD, S. ; GORMAN, M.F ; COCHRAN, J.J ; COX, L.A ; KESKINOCAK, P. ; KHAROUFEH, J.P ; SMITH, J. C.: Operations Research for Freight Train Routing and Scheduling. In: *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc, 2010. – ISBN 9780470400531, S. 1–10
- [HH13] HEDDERICH, A. ; HEINRICI, T.: "Man macht das Richtige, wenn man auf unser System vertraut": Interview mit dem Vorstandsvorsitzenden von DB Schenker Rail. In: *Eisenbahntechnische Rundschau* (2013), Nr. 1+2, S. 40–43
- [HNS10] HEWITT, M. ; NEMHAUSER, G.L ; SAVELSBERGH, M.W.P: Combining Exact and Heuristic Approaches for the Capacitated Fixed Charge Network Flow Problem. In: *INFORMS Journal on Computing* 22 (2010), Nr. 2, S. 314–325
- [Hom12] HOMFELD, H.: *Consolidating car routes in rail freight service by discrete optimization*. München : Dr. Hut, 2012. – ISBN 3843904685
- [HBSM95] HUNTLEY, C. L. ; BROWN, D. E. ; SAPPINGTON, D. E. ; MARKOWICZ, B. P.: Freight Routing and Scheduling at CSX Transportation. In: *Interfaces* 25 (1995), Nr. 3, S. 58–71
- [ICF⁺04] IRELAND, P. ; CASE, R. ; FALLIS, J. ; DYKE, C. van ; KUEHN, J. ; MEKE-TON, M.: The Canadian Pacific Railway Transforms Operations by Using Models to Develop Its Operating Plans. In: *Interfaces* 34 (2004), Nr. 1, S. 5–14

- [JAG07] JHA, K. C. ; AHUJA, R.K. ; GÜVENÇ, S.: New Approaches for Solving the Block-to-Train Assignment Problem. In: *Networks* 51 (2007), Nr. 1, S. 48–62
- [Jin98] JIN, H.: *Designing robust railroad blocking plans*. Cambridge : Massachusetts Institute of Technology, 1998
- [JZL13] JIN, J.G ; ZHAO, J. ; LEE, D.-H: A column generation based approach for the Train Network Design Optimization problem. In: *Transportation Research Part E: Logistics and Transportation Review* (2013), Nr. 50, S. 1–17
- [Kar84] KARMARKAR, N.: A new Polynomial-Time Algorithm for linear programming. In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing* (1984)
- [Kea89] KEATON, M.H: Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches. In: *Transportation Research B: Methodological* 23 (1989), Nr. 6, S. 415–431
- [Kea92] KEATON, M.H: Designing railroad operating plans: A dual adjustment method for implementing Lagrangian relaxation. In: *Transportation Science* 26 (1992), Nr. 4, S. 263–279
- [KGV83] KIRKPATRICK, S. ; GELATT, C. ; VECCHI, M.: *Research reports / IBM*. Bd. 9355: *Optimization by simulated annealing*. 1983
- [KV12] KORTE, B. ; VYGEN, J.: *Kombinatorische Optimierung: Theorie und Algorithmen*. 2. Auflage. Berlin : Springer Spektrum, 2012. – ISBN 9783642254017
- [KY97] KOUVELIS, P. ; YU, G.: *Nonconvex optimization and its applications*. Bd. 14: *Robust discrete optimization and its applications*. Dordrecht : Kluwer Acad. Publ, 1997. – ISBN 0792342917
- [KHA⁺09] KROON, L. ; HUISMAN, D. ; ABBINK, E. ; FIOOLE, P.-J ; FISCHETTI, M. ; MAROTI, G. ; SCHRIJVER, A. ; STEENBEEK, A. ; YBEMA, R.: The New Dutch Timetable: The OR Revolution. In: *Interfaces* 39 (2009), Nr. 1, S. 6–17
- [Las70] LASDON, L. S.: *Optimization theory for large systems*. London : MacMillan, 1970. – ISBN 9780486419992
- [LWJ⁺12] LIN, B.-L ; WANG, Z.-M ; JI, L.-J ; TIAN, Y.-M ; ZHOU, G.-Q: Optimizing the freight train connection service network of a large-scale rail system. In: *Transportation Research B: Methodological* 46 (2012), Nr. 5, S. 649–667
- [LD05] LÜBBECKE, M. ; DESROSIERS, J.: Selected Topics in Column Generation. In: *Operations Research* 53 (2005), Nr. 6, S. 1007–1023

- [MW83] MAGNANTI, T. L. ; WONG, R. T.: *Network design and transportation planning: Models and algorithms*. Cambridge : Operations Research Center, Massachusetts Institute of Technology, 1983
- [MSV09] MANIEZZO, V. ; STÜTZLE, T. ; VOSS, S.: *Matheuristics: Hybridizing metaheuristics and mathematical programming*. New York : Springer, 2009. – ISBN 144191305X
- [MD10] MCKINSEY & COMPANY ; DEUTSCHE BAHN AG: *Zukunftsperspektiven für Mobilität und Transport: Eisenbahn in Deutschland 2025*. Berlin : DB Mobility Logistics AG, 2010
- [Mon11] MONOPOLKOMMISSION: *Bahn 2011: Wettbewerbspolitik unter Zugzwang: Sondergutachten der Monopolkommission gemäß §36 AEG*. 1. Auflage. Baden-Baden : Nomos, 2011. – ISBN 3832971084
- [NA10] NEMANI, A. K. ; AHUJA, R.K.: OR Models in Freight Railroad Industry. In: *Wiley Encyclopedia of Operations Research and Management Science* (2010), S. 3925–3944
- [NW99] NEMHAUSER, G. L. ; WOLSEY, L. A.: *Integer and combinatorial optimization*. New York and NY : Wiley, 1999 (A Wiley-Interscience publication). – ISBN 0471359432
- [New97] NEWTON, H. N.: *Network Design Under Budget Constraints With Application To The Railroad Blocking Problem*. Ft. Belvoir : Defense Technical Information Center, 1997
- [NBV98] NEWTON, H.N ; BARNHART, C. ; VANCE, P.: Constructing Railroad Blocking Plans to Minimize Handling Costs. In: *Transportation Science* 32 (1998), Nr. 4
- [NP09] NIKUTTA, S. ; PAHL M.: Schlanke Produktionsstrukturen in vernetzten Systemen: Das Produktionssystem im Schienengüterverkehr. In: *Deine Bahn* (2009), Nr. 2
- [OD12] OLIVER WYMAN CONSULTING: VAN DYKE, C.; D’INCA, G. J.; R. J.; Roucolle: *Multirail Planning Suite*. 2012
- [PR10] PISINGER, D. ; ROPKE, S.: Large Neighborhood Search. In: *Handbook of metaheuristics* (2010), S. 399–420. ISBN 9781441916631
- [Pos08] POSNER, H.: *Rail Freight in the USA : Lessons for Continental Europe*. (2008)
- [PR08] PUCHINGER, Jakob ; RAIDL, Günther: Combining metaheuristics and exact algorithms in combinatorial optimization: A survey and classification. In: *Hybrid Metaheuristics* (2008), S. 31–62

- [Sch00] SCHRIJVER, A.: *Theory of linear and integer programming*. Reprinted. Chichester : Wiley, 2000 (A Wiley-Interscience publication). – ISBN 0471982326
- [Sch08] SCHÜLLDORF H.: *Optimierung der Leitwegeplanung im Schienengüterverkehr, Diplomarbeit*. Universität Darmstadt, 2008
- [SK07] SCHWARZ, F. ; KOCHSIEK, J.: *Software für EVU im Schienengüterverkehr*. Dortmund : Fraunhofer Institut für Materialfluss und Logistik, 2007
- [Sie97] SIEGMANN, J.: *Wege zu einer anforderungsgerechten und wirtschaftlichen Güterbahn: Habilitationsschrift*. Hannover : Institut für Verkehrswesen, Eisenbahnbau und -betrieb, 1997. – ISBN 00132845
- [Sie12] SIEGMANN, J.: Betriebsdurchführung im Einzelwagenverkehr. In: *Bundesministeriums für Verkehr, Bau und Stadtentwicklung* (27.04.2012)
- [SS12] SIEGMANN, J. ; STUHR, Helge: Hat der Einzelwagenverkehr (EV) in Europa noch eine Chance? In: *Eisenbahntechnische Rundschau (ETR)* 3 (2012), S. 10–18
- [SG10] SILBERHOLZ, J. ; GOLDEN, B.: Comparison of Metaheuristics. In: *Handbook of metaheuristics* (2010), S. 399–285. ISBN 9781441916631
- [Sör13] SÖRENSEN, K.: Metaheuristics: The metaphor exposed. In: *International Transactions in Operational Research* (2013)
- [SG10] SÖRENSEN, K. ; GLOVER, F.: Metaheuristics. In: *Encyclopedia of Operations Research and Management Science* (2010)
- [Sta13] STATISTISCHES BUNDESAMT (INTERNETAUFTRITT): <https://www.destatis.de/>. 2013
- [SD10] STÜTZLE, T. G. ; DORIGO, M.: Ant colony optimization. In: *Handbook of metaheuristics* (2010), S. 227–264. ISBN 9781441916631
- [Van11] VANDERBECK, F.: Branching in Branch-and-Price: a Generic Scheme. In: *Mathematical Programming, Series A* 130 (2011), Nr. 2, S. 249–294
- [Vog11] VOGT, A.: *Studien zur Mobilitäts- und Verkehrsforschung*. Bd. 24: *Renaissance des Einzelwagenverkehrs nach Liberalisierung des Schienengüterverkehrs in Deutschland? Mit 33 Tabellen*. Mannheim : Verlag MetaGIS, 2011. – ISBN 9783936438390
- [VC11] VOLL, R. ; CLAUSEN, U.: Column Generation for Multi-Matrix Blocking Problems. In: *Operations research proceedings 2011: Selected papers of the International Conference on Operations Research* (2011), S. 281–286. ISBN 9783642292101

- [VC13a] VOLL, R. ; CLAUSEN, U.: Branch-and-Price for a European variant of the Railroad Blocking Problem. In: *Electronic Notes in Discrete Mathematics* 41 (2013), S. 45–52
- [VC13b] VOLL, R. ; CLAUSEN, U.: Heuristic approaches for robust consolidation in wagonload traffic. In: *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis* (2013)
- [Weg03] WEGENER, I.: *Komplexitätstheorie: Grenzen der Effizienz von Algorithmen*. Berlin : Springer, 2003. – ISBN 3540001611
- [YABS13] YAGHINI, M. ; AHADI, H. ; BARATI, E. ; SAGHIAN, Z.: Tabu Search Algorithm for the Railroad Blocking Problem. In: *Journal of Transportation Engineering* 139 (2013), Nr. 2, S. 216–222
- [YFN11] YAGHINI, M. ; FOROUGHI A. ; NADJARI B.: Solving Railroad Blocking Problem Using Ant Colony Optimization Algorithm. In: *Applied Mathematical Modelling* (2011), Nr. 35, S. 5579–5591
- [YZYF11] YUE, Y. ; ZHOU, L. ; YUE, Q. ; FAN, Z.: Multi-route railroad blocking problem by improved model and ant colony algorithm in real world. In: *Computer & Industrial Engineering* (2011), Nr. 60, S. 34–42
- [ZCG11] ZHU, E. ; CRAINIC, T.G ; GENDREAU, M.: *Scheduled service network design for freight rail transportation*. Montréal : CIRRELT, 2011

Stichwortverzeichnis

Symbole

(Opt_{arc})	46
(Opt_{path})	47
\mathcal{NP} -Vollständigkeit	30
\mathcal{O} -Notation	30
(IMCFP)	32

B

Bündelungseffekt	28
Bellman-Ford-Algorithmus	33
Blockingnetzwerk	23
Blockingplan	23
Branch-and-Bound	35
Branch-and-Price (B&P)	67
Branch-and-Price-and-Cut (B&P&C)	70

C

Column Generation (CG)	59
------------------------------	----

D

Dualität	34
----------------	----

E

Einzelwagenverkehr (EWV)	7
Eisenbahninfrastrukturunternehmen (EIU)	6
Eisenverkehrsunternehmen (EVU)	7

G

Ganzzahliges Lineares Programm (ILP)	31
Ganzzahliges Mehrgüterflussproblem (IMCFP)	32
Ganzzugverkehr (GV)	7
Gap	36

K

Kürzeste-Wege-Problem	33
Knotenpunktbahnhof (Kbf).....	9
Komplementärschlupfbedingung	35

L

Lasdonschranke	63
Leitwegeplan	26
Lineare Relaxierung	36
Lineares Programm (LP)	31
Lokale Suche	38
Lower Bound (LB)	36

M

Marktorientiertes Angebot Cargo (MORA-C)	10
Matheuristik	39, 72
Metaheuristik	38

O

Optimalitätslücke	36
-------------------------	----

P

Polyeder	33, 52
polynomielle Laufzeit	30

R

Rangierbahnhof (Rbf)	9
Realinstanz	95
Reduzierte Kosten	35
Referenzinstanz	95
Relation	16, 43
Relaxierung	36
Robustheit	17, 29

S

s-t-Fluss	19
Satellit mit Rangiermitteln (SmR)	9
Satellitenbahnhöfen (Sat)	9
Schienengüterverkehr (SGV)	5
Schwacher Dualitätssatz	34

Solver 37

T

Tailing-off-Effekt 67

Transportkosten 15, 44

Trasse 7

U

Umstellpläne 14, 21

Umstellplan 14

Untere Schranke 36, 51

V

Variable Depth Neighborhood Search (VDNS) 39

Variable Neighborhood Search (VNS) 38

Vorgehensmodell der Optimierung 2

Abbildungsverzeichnis

1.1	Schaubild zum Vorgehensmodell der mathematischen Optimierung	2
2.1	Entwicklung der Verkehrsleistung im SGV seit 2000	6
2.2	Schematische Darstellung eines Rangierbahnhofes	9
2.3	Verkürzte Darstellung des Angebotsprozesses	12
2.4	Darstellung des regionalen Wagenaufkommens für 10 Werktage	16
3.1	Beispiel eines Blockingnetzwerkes als virtuelles Teilnetz	22
3.2	Schematische Darstellung von Leitwegen im EWW	27
3.3	Optimallösungen in Abhängigkeit der Kostenstruktur	28
3.4	Beispiele beschränkter Polyeder in 2 und 3 Dimensionen	33
3.5	Branch-and-Bound-Baum	36
3.6	Visualisierung des Schnittebenenverfahrens	37
3.7	Vergleich von VNS und VDNS	39
4.1	Vergleich zwischen relaxierter und nicht-relaxierter Kostenstruktur	52
4.2	Illustration heuristischer Preprocessings	55
5.1	Struktur der Nebenbedingungsmatrix von (Opt_{arc})	60
5.2	Schaubild zu Column-Generation-Verfahren	61
5.3	Verlauf von LB_{Lasdon} und UB für $DB-21-254-10a$	66
5.4	Verlauf von LB_{Lasdon} und UB für $gen-50-224-10$	67
5.5	Visualisierung des Subproblems für $q = 1$ und $q = 2$	69
6.2	Parametertuning für die Matheuristik für Instanz $gen-50-224-10$	91
6.1	Parametertuning für die Matheuristik für drei Instanzen	91
6.3	Lösungsverlauf für $DB-21-254-10a$ und $DB-21-254-10b$	93
6.4	Lösungsverlauf für die Instanzen $DB-147-3272-5$ und $DB-107-1584-10$	93
6.5	Darstellung einer Lösung der Instanz $DB-107-1584-10$	95
7.1	Visualisierungen von Leitwegelösung, (Opt_{arc}) und Blockinglösung	98
7.2	Normierte Kennzahlen dreier Lösungen der Referenzinstanz	99
7.3	Dezile der relativen Zugauslastungen	104
7.4	Züge je Kante (links: Leitwegelösung, rechts: Lösung von (Opt_{arc}))	105

