

Einfluss von Preissystemen auf die Verbindungsauswahl im Eisenbahnfernverkehr

vorgelegt von
Master of Science
Johannes Friedrich
geb. in Stuttgart

von der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Markus Hecht
Gutachter: Prof. Dr.-Ing. habil. Jürgen Siegmann
Gutachter: Prof. Dr.-Ing. Frank Lademann

Tag der wissenschaftlichen Aussprache: 10.04.2017

Berlin 2017

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

(Ort, Datum / Unterschrift)

Danksagung

Diese Arbeit wäre ohne die großartige Liebe, die Zuneigung und Unterstützung, die bewundernswerte Geduld und stetige Motivation, durch meinen Lebenspartner Eric nie möglich gewesen.

Meine Familie, allen voran meine Eltern Elisabeth und Michael, hat mich zu jeder Zeit unterstützt und gefördert. Mein Studium und diese Promotion wurden allein durch sie ermöglicht.

Ein besonderer Dank gebührt Herrn Professor Siegmann, dessen fachliche Beratung mir jederzeit offenstand, und der meine Promotion stets befördert hat, sowie Herrn Professor Lademann, der mir mit seinem fachlichen Feedback stets zur Seite stand.

Außerdem möchte ich meinen Kollegen für ihr herausragendes Engagement für meine Pflichten und diese Arbeit sowie ihr großartiges Verständnis danken. Diese Arbeit hätte ohne ihre Unterstützung nicht erstellt werden können.

Last, but not least: Thanks to the awesome LaTeX community!

— Created with L^AT_EX

Zusammenfassung

Die vorliegende Arbeit ergründet den Zusammenhang zwischen der Tarifierung und der Anzahl für Fahrgäste relevanter Verbindungen im Eisenbahnfernverkehr.

Nach einer Abgrenzung des Untersuchungsraums auf Deutschland und wesentliche Strecken des angrenzenden europäischen Auslands sowie des Untersuchungsgegenstands auf Eisenbahn-Fernverkehre werden die wesentlichen Merkmale der Auswahl von Verbindungen durch Fahrgäste erläutert: Die Kriterien Preis, Zeit und Komfort (aufgeteilt in den Fahrt- und den Umsteigekomfort) spielen dabei eine wesentliche Rolle. In Abgrenzung zu einschlägiger Literatur waren die Zielgruppen detailliert zu betrachten, neben Privatreisenden und Geschäftsreisenden waren daher auch Fernpendler und deren jeweilige Unterzielgruppen einzubeziehen.

In der Monetarisierung werden den Präferenzen der Zielgruppen Geldwerte zugeordnet, um darauffolgend das Modell der verwendeten Nutzenfunktion auf Basis von Verkehrswiderständen erläutern und diskutieren zu können. Eine Definition des Modells erfolgt ebenfalls in diesem Rahmen, um die Umsetzung der Nutzenfunktion in das Modell erläutern zu können. Die Vorstellung des Datenmodells ist dann ein letzter Schritt vor der konkreten Verbindungserzeugung, dabei werden alle vorgestellten Datenmodelle sowie der Aufbau der Datenbankumgebung erläutert.

Darauffolgend werden die zur Verbindungserzeugung verwendeten Algorithmen entwickelt, der Prozess ist dabei darauf angelegt, die Nachvollziehbarkeit und Vollständigkeit aus wissenschaftlicher Sicht durchgehend zu gewährleisten und gleichzeitig eine effiziente Algorithmik zu verwenden. Weitere Betrachtungen dienen speziell der Aussortierung von Verbindungen, die für keine Zielgruppe vorteilhaft sind.

Die anschließende Auswahl der Verbindungen für jede Zielgruppe schließt den Verbindungssuchprozess ab. Nach einer kurzen Diskussion verbreiteter Tarifmodelle werden dann Auswirkungen der Tarifsystemwahl auf die Anzahl der zu bildenden vorteilhaften Routen berechnet und vorgestellt: im Ergebnis stellt sich heraus, dass ein Wechsel von einer entfernungsabhängigen Tarifierung in ein Relationspreissystem mit einer massiven Abnahme von Verbindungen einhergeht, die Bahnhöfe mit wenigen Direktverbindungen stärker betrifft als Bahnhöfe mit besserer Vernetzung.

Die Arbeit schließt dann mit einer wertenden Zusammenfassung aller Erkenntnisse im Fazit und einem Ausblick auf mögliche weitere Forschungsvorhaben.

Abstract

This thesis deals – within a railway long distance route setup – with the impact of changes of tariff system on the number of connections passengers may choose from.

Basis of the study are railway lines in Germany and main railway routes in surrounding European countries. The subject of the study is limited to railway long distance traffic. The main elements that influence the choice of possible passengers are price, duration and route comfort (split into train-related comfort and the comfort of interchanges). The target groups are derived from a literature study and augmented by a new target group, long-distance commuters; in addition, these target groups are split up into separate sub-target groups to consider their specific needs.

First each target group is considered in detail, showing their set of preferences, after which the modelling of these passenger groups is shown. Then all utility preferences are monetarized. This is necessary for the further processing in the route preference choice model. Next, a definition of the utility function and the passenger model is given and discussed. The final step before the routes can be created, is the introduction and discussion of the data model and the database system environment.

Subsequently, the algorithms used to create the set of useful routes are developed and the finalized algorithms are discussed concerning their efficiency. Later – as part of an iterative process – they are gradually improved. The process improving the algorithms' efficiency is designed to yield great traceability and exactness in order to guarantee the scientific correctness of the results. Further investigation is made to find connections that are of no use to any target group and can therefore be removed.

The target group related route choice completes the route finding process, followed by a short discussion of commonly used tariffs systems. Impacts of the choice of a tariff system on the number of useful routes (for passengers of all target groups to choose from) are then calculated and discussed. As a main result, the change from a distance-based tariff system to a relation-based tariff system lowers the number of advantageous connections greatly. This applies especially to stations serving few destinations directly, more so than to stations with direct connections to a greater number of destinations.

The thesis finishes with an evaluation and recapitulation and gives an outlook on further possible research based on this thesis.

Inhaltsverzeichnis

Danksagung	I
Zusammenfassung	II
Abstract	III
Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
Algorithmenverzeichnis	IX
Abkürzungsverzeichnis	X
Technische Hinweise	XI
1 Einleitung	1
2 Abgrenzung der Problemstellung	3
3 Passagiertheorie	6
3.1 Auswahlproblem	6
3.2 Nutzentheorie	9
3.3 Zielgruppen im Schienenpersonenverkehr	11
3.3.1 Geschäftsreisende	13
3.3.2 Privatreisende	19
3.3.3 Fernpendler	21
3.3.4 Zusammenfassung der Ergebnisse	23
4 Modellierung der Passagiere	25
4.1 Monetarisierung der Auswahlpräferenzen	25
4.1.1 Zeitkosten	26
4.1.2 Umsteigekosten	30
4.1.3 Komfortkosten	32
4.2 Auswahlgrundlage virtueller Preis	36
4.3 Modelldefinition	38
4.3.1 Geographische Eingrenzung	38
4.3.2 Fahrten über den Nachtsprung	39
4.3.3 Passagierverhalten im Modell	40
5 Datenmodell	42
5.1 Grundsätzliche Erwägungen	42
5.1.1 Voraussetzungen des Modellaufbaus	42
5.1.2 Technisches Setup	44
5.2 Datenmodell	46

5.2.1	Primärdaten	47
5.2.2	Sekundärdaten	54
6	Algorithmische Berechnung der Verbindungen	63
6.1	Algorithmusdefinition	63
6.2	Verbindungserzeugung	72
6.2.1	Vollständige Erzeugung des Verbindungstupels	72
6.2.2	Identifikation unvorteilhafter Verbindungen	75
6.2.3	Ausschluss unvorteilhafter Verbindungen mit vollständiger Ver- kehrstageüberdeckung	79
6.2.4	Ausschluss unvorteilhafter Verbindungen mit teilweiser Ver- kehrstageüberdeckung	81
6.2.5	Ausschluss unvorteilhafter Verbindungen mit weniger Umstiegen	84
6.3	Spezielle Betrachtung der Verbindungen des ersten Umstiegs	86
6.3.1	Aufbau der Daten des ersten Umstiegs	87
6.3.2	Ausschluss versetzt parallel verkehrender Züge desselben Start- bahnhofs und Ziels	90
6.3.3	Ausschluss versetzt parallel verkehrender Züge mit unterschied- lichen Zielen	94
6.3.4	Ausschluss versetzt parallel verkehrender Züge mit unterschied- lichen Startbahnhöfen	97
6.3.5	Verbindungserzeugung des ersten Umstiegs	100
6.4	Zusammenfassung der Algorithmen	105
7	Verbindungsauswahl und Auswirkungen des Tarifs	107
7.1	Auswahl der optimalen Verbindungen	107
7.2	Vorstellung der Tarifsysteme	110
7.2.1	Entfernungsabhängige Tarifierung	110
7.2.2	Relationstarif	112
7.2.3	Kombinierte Tarifsysteme	114
7.3	Auswirkungen der Tarifsysteme auf die Routenauswahl	115
7.3.1	Veränderungen im Algorithmus	115
7.3.2	Beispielhafte Betrachtung von Stationen und Verbindungen . .	116
7.3.3	Gesamthafte Betrachtung von Abhängigkeiten	117
7.3.4	Wertung der Ergebnisse	121
8	Fazit und Handlungsempfehlung	122
	Literatur	XII

Abbildungsverzeichnis

1	Intermodale Entscheidungsfindung	7
2	Intramodale Entscheidungsfindung	7
3	Unterscheidung der Passagiere in Primärzielgruppen	11
4	Differenziertere Aufteilung der Primärzielgruppen	12
5	Darstellung einer Linie mit vier Halten im Ausschnitt	65
6	Darstellung von zwei Linien mit vier Halten im Ausschnitt	66
7	Darstellung verschiedener Fahrmöglichkeiten entlang einer Relation	67
8	Erzeugung der Verbindungen des ersten Umstiegs	105
9	Erzeugung der Verbindungen ab dem zweiten Umstieg	106
10	Darstellung der Verbindungsquote nach Tarifsystemänderung gegen das Verhältnis der Umsteigeverbindungen je Direktverbindung	118
11	Darstellung der Verbindungsquote nach Tarifsystemänderung gegen die Anzahl der Direktverbindungen	119
12	Darstellung der Verbindungsquote nach Tarifsystemänderung gegen die Anzahl der direkt erreichbaren Stationen	120

Tabellenverzeichnis

1	Darstellung der Umrechnung von Fremdwährungen	XI
2	Datentypen in SQL	XI
3	Passagiere und Umsatzanteile des InterCity von British Rail im Jahr 1992	14
4	Darstellung des value of time für verschiedene Reisezwecke und Verkehrsarten	26
5	Zeitkosten der verschiedenen Zielgruppen im Modell	30
6	Umsteigekosten und Kosten der Umsteige-Wartezeit für den ersten Umstieg	32
7	Flexpreise der Deutschen Bahn für die Strecke Stuttgart – Ulm	33
8	Flexpreise der Deutschen Bahn für die Strecke Mannheim – Saarbrücken .	34
9	Flexpreise der Deutschen Bahn für die Strecke Berlin – Stralsund	34
10	Übersicht der Primärdatenquellen	47
11	Setup der Datenbanktabelle bp_fp_bf	48
12	Setup der Datenbanktabelle einwohner	48
13	Setup der Datenbanktabelle bp_fp_zug	49
14	Setup der Datenbanktabelle zugpreis	50
15	Setup der Datenbanktabelle bp_fp_zeiten	50
16	Setup der Datenbanktabelle entfernungen	51
17	Setup der Datenbanktabelle zeitenpreis	53
18	Setup der Datenbanktabelle relevante_bahnhoefe	54
19	Übersicht der Sekundärdatenquellen	55
20	Setup der Datenbanktabelle direktverbindungen	56
21	Setup der Datenbanktabelle umstiege_zeiten	57
22	Setup der Datenbanktabelle parallele_zuege	58
23	Setup der Datenbanktabelle parallelverbindungen	58
24	Setup der Datenbanktabelle verbindungen	59
25	Setup der Datenbanktabelle umstiege_dv_mem	60
26	Setup der Datenbanktabelle verbindungen_zielgruppe	61
27	Darstellung eines Auswahlproblems vorzugswürdiger Verbindungen . . .	68
28	Anteil der zu berücksichtigenden Verbindungen ab Stuttgart Hbf mit einem Umstieg	69
29	Anteil der zu berücksichtigenden Verbindungen ab Stuttgart Hbf mit zwei Umstiegen	70
30	Darstellung eines Umsteige-Kombinationsproblems	70
31	Auswahlproblem verschiedener Verbindungen, erstes Beispiel	76
32	Auswahlproblem verschiedener Verbindungen, zweites Beispiel	77
33	Auswahlproblem verschiedener Verbindungen, drittes Beispiel	78

34	Auswahlproblem verschiedener Verbindungen, viertes Beispiel	78
35	Abtrag der Verbindungszahlen für Stuttgart Hbf mit einer Auswahlstufe .	80
36	Abtrag der Verbindungszahlen für Stuttgart Hbf mit zwei Auswahlstufen .	86
37	Datensumme des Kreuzprodukts (bezogen auf Datensätze desselben Bahnhofs mit höchstens 124 Minuten Umsteigezeit, ohne Eigenanschlüsse) . .	87
38	Verschiedene Modelle der entfernungsabhängigen Tarifierung	110
39	Abhängigkeit zwischen der Anzahl direkt verbundener Ziele und den Auswirkungen eines Tarifsysteemwechsels	116

Algorithmenverzeichnis

1	Erzeugung von Umsteigeverbindungen mit mehreren Umstiegen	73
2	Aussortierung von Verbindungen desselben Umstiegs bei kompletter Überdeckung der Verkehrstage	81
3	Aussortierung von Verbindungen desselben Umstiegs bei partieller Überdeckung der Verkehrstage	82
4	Entfernung von Verbindungen ohne vorteilhafte Verkehrstage	83
5	Aussortierung von Verbindungen mit weniger Umstiegen bei partieller Überdeckung der Verkehrstage	85
6	Erzeugung der Umsteigezeiten des ersten Umstiegs	88
7	Erzeugung der Datenbanktabelle <code>parallele_zuege</code>	91
8	Entfernung nicht streng paralleler Zugpaare aus der Datenbanktabelle <code>parallele_zuege</code>	92
9	Löschen von Umsteigealternativen aus der Tabelle <code>umstiege_zeiten</code>	93
10	Erzeugung der Datenbanktabelle <code>parallele_zuege</code>	95
11	Löschen von Datensätzen überlappender Züge aus der Datenbanktabelle <code>parallele_zuege</code>	96
12	Löschen von nicht relevanten Datensätzen aus der Datenbanktabelle <code>umstiege_zeiten</code>	97
13	Erzeugung der Datenbanktabelle <code>parallele_zuege</code>	98
14	Löschen von Datensätzen überlappender Züge aus der Datenbanktabelle <code>parallele_zuege</code>	99
15	Löschen von nicht relevanten Datensätzen aus der Datenbanktabelle <code>umstiege_zeiten</code>	100
16	Erzeugen der Verbindungen des ersten Umstiegs	101
17	Löschen von quasi-doppelten Verbindungen derselben Relation	103
18	Löschen von Verbindungen, die teilparallel sind	104
19	Löschen von Verbindungen über den Verkehrstageswechsel	105
20	Auswahl von Verbindungen für eine Startstation und Zielgruppe . . .	108
21	Aussortieren von Verbindungen für eine Startstation und Zielgruppe mit vollständiger Überdeckung der Verkehrstage	109

Glossar

AEG	Allgemeines Eisenbahngesetz
D	Schnellzug (Zuggattung der Bundesbahn)
EBO	Eisenbahn-Bau- und Betriebsordnung
F	Fernschnellzug (Zuggattung der Bundesbahn)
GUI	Graphical User Interface (Grafische Benutzerschnittstelle)
ID	eindeutiger Identifikator
ÖPNV	Öffentlicher Personennahverkehr
PHP	Hypertext Preprocessor
RAM	Random Access Memory: Arbeitsspeicher in Personal Computern [Kir16]
SPFV	Schienenpersonenfernverkehr
SQL	Structured Query Language
TEE	Trans Europ Express (Zuggattung)
value of time	Wert der Reisezeit; Indikator für Zahlungsbereitschaft einer Zielgruppe
virtueller Preis	Der virtuelle Preis dient zur Darstellung der Präferenzen eines durchschnittlichen Fahrgasts zum Zweck der Optimierung der Berechnung

Technische Hinweise

Die Arbeit verwendet im Regelfall die männliche Form von Bezeichnungen. Dies geschieht allein aus Gründen der Lesbarkeit und impliziert selbstverständlich keine Wertung.

Die in Tabelle 1 dargestellten Wertefaktoren wurden zur Konvertierung fremder Währungen in Eurowerte genutzt (Werte angelehnt an Umrechnungskurse mit Stand vom 22.09.2016 nach [exc16a] und [exc16b] unter Rundung auf für die Methodik der Arbeit günstige Werte):

Fremdwährung	Umrechnungsfaktor
1,00 \$ (USD)	0,90 €
1,00 £ (GBP)	1,17 €

Tabelle 1: Darstellung der Umrechnung von Fremdwährungen
Quelle: Eigene Darstellung

Quellen mit Jahresangaben vor 2016 wurden dabei aufgezinst, um eine Vergleichbarkeit aller Angaben innerhalb der Arbeit zu gewährleisten. Als Aufzinsungsfaktor wurden dabei 1,5 % als durchschnittliche Inflation nach [Pel16] angenommen und der entstehende Zinseszinsseffekt berücksichtigt.

Alle verwendeten SQL-Datentypen und deren spezifische Eigenschaften sind in der folgenden Tabelle 2 abgetragen:

Datentyp	Verwendung	Wertebereich
VARCHAR	Wortwerte	<i>beliebig</i>
TIME	Zeitwerte	<i>von SQL festgelegt</i>
INTEGER	Ganzzahlen	-2.147.483.647 bis +2.147.483.647
MEDIUMINT	Ganzzahlen	-8.388.608 bis +8.388.607
SMALLINT	Ganzzahlen	-32.768 bis +32.767
TINYINT	Ganzzahlen	-128 bis +127
DOUBLE	Gleitkommawerte	<i>beliebig</i>

Tabelle 2: Datentypen in SQL
Quelle: Eigene Darstellung nach [Pel99] und [myS16a]

1 Einleitung

„Eine Fahrt mit der Eisenbahn kann ich beim besten Willen nicht als Reise bezeichnen. Man wird ja lediglich von einem Ort zum anderen befördert und unterscheidet sich damit nur sehr wenig von einem Paket.“

— JOHN RUSKIN, 1819–1900¹, zitiert nach [Mat15, S.84]

Das Reisen mit der Eisenbahn brachte im 19. Jahrhundert eine neue Qualität bezogen auf Geschwindigkeit und Komfort und damit eine völlig neue Art des Reisens mit sich. Im Zuge dieses Fortschritts gingen natürlich auch alte Traditionen – wie der Aufenthalt an Posthaltereien zum Wechsel des Gespanns und zur Versorgung der Reisenden – verloren, sodass dem Wandel der Technik eine gewisse Fortschrittsskepsis entgegenschlug. Gleichzeitig beleuchtet das Zitat von JOHN RUSKIN aber auch einen Vergleich, der das Kernthema dieser Arbeit widerspiegelt: *wie* der Fahrgast vom Start zum Ziel kommt, ist im Bezug auf die Eigenschaften der Fahrt (wie Preis und Komfort) zwar wichtig, jedoch im Bezug auf die konkrete Reiseroute für den Fahrgast im Regelfall nicht relevant.

Die Möglichkeit der Routenwahl ist damit für den Fahrgast in vielen Fällen unerheblich oder vielleicht sogar lästig (weil damit unweigerlich in vielen Systemen auch eine Fahrpreisentscheidung verknüpft ist), doch kann die Wahl einer Route auch vom Nutzer gewollt sein, beispielsweise bei einer Fahrt durch eine besonders schöne Landschaft oder die Wahl einer besonders günstigen Reiseverbindung.

In dieser Arbeit wird nun ergründet, warum die Möglichkeit der Routenwahl insbesondere bei elektronischen Auskunftssystemen zu einem erheblichen Teil durch die Tarifierung des Unternehmens bedingt ist, und welchen Einfluss die Tarifgestaltung auf die Anzahl der Verbindungen hat, aus denen der potenzielle Fahrgast wählen kann. An einem kleinen Beispiel lässt sich diese Abwägung bereits darstellen: wenn zwischen zwei Stationen ein Einheitspreis unabhängig von der gefahrenen Route vom Unternehmen verlangt wird, aber zwei Routen mit unterschiedlicher Dauer angeboten werden, so werden sich die Fahrgäste – vorbehaltlich weiterer Entscheidungskriterien wie Komfort und Auslastung des Verkehrsmittels – mehrheitlich für die schnellere Verbindung entscheiden. Die Beeinflussung dieser Routenauswahl fördert daher insbesondere die

¹ Im Original: „Going by railroad I do not consider as travelling at all; it is merely "being sent" to a place, and very little different from becoming a parcel.“

Auslastungssteuerung innerhalb von Systemen, ohne die preisliche Integrität des Systems anzutasten.

Außerdem werden die Auswirkungen von Tarifierungswechseln auf die Anzahl der erzeugten Verbindungen – und damit die Wahlmöglichkeiten der Fahrgäste – dargestellt und diskutiert. Hierbei ist insbesondere fraglich, welchen Anteil die Tarifierung unter welchen Umständen und für welche Art oder Gattung von Stationen besitzt. Beim Aufbau eines Relationspreissystems mit der Bepreisung konkreter Relationen statt der zwischen Start- und Zielstation zurückgelegten Entfernung werden dabei die Wahlmöglichkeiten der Reisenden in erster Linie erhöht; Algorithmisch jedoch könnte sich auch eine Senkung der Anzahl der Alternativen ergeben, da jetzt zuvor noch vorteilhafte – weil beispielsweise preisgünstigere, aber langsamere – Verbindungen ihre Vorteilhaftigkeit verlieren könnten, da der Preisvorteil gegenüber schnelleren und komfortableren Verbindungen nivelliert wird.

Dazu wird zu Beginn dieser Arbeit eine Abgrenzung der Problemstellung von anderen Arbeiten desselben Fachgebiets dargestellt. Im darauffolgenden Abriss zur Passagiertheorie werden dann wesentliche wissenschaftliche Theoreme vorgestellt und kritisch diskutiert, um anschließend aus diesen Theoremen ein eigenes Routenwahlmodell für den im Rahmen dieser Arbeit zu entwickelnden Algorithmus zur Routenauswahl aus zur Laufzeit erzeugten Routen innerhalb eines Netzwerks zu erstellen. Dazu werden auch die verschiedenen wesentlichen Zielgruppen des Schienenpersonenfernverkehrs (SPFV) erläutert und in Abgrenzung zu existierender Literatur differenzierter gegliedert.

In den weiteren Kapiteln wird der Algorithmus zur Auswahl der Routen durch die Passagiere nach einer Erläuterung der technischen Grundlagen vorgestellt und ausgiebig besprochen. Eine weitreichende Datenanalyse wird sodann genutzt, um die Ergebnisse des Algorithmus zu verifizieren und die Auswirkungen der Tarifierung auf diese Routenwahl aufzuzeigen. Die Arbeit schließt mit einem Fazit und Verweisen auf weitere Forschungsansätze.

Die tatsächliche Verteilung einer Menge von Passagieren auf die mithilfe des entwickelten Algorithmus-Systems erstellten und schließlich ausgewählten Routen ist dabei kein Bestandteil der Arbeit; Im Rahmen der Arbeit werden jedoch an vielen Stellen Hinweise entwickelt, wie eine solche Ergänzung aus dieser Arbeit entwickelt werden könnte bzw. wie diese Arbeit als Basis einer weiteren Studie zur Verteilung von Passagieren und zur Prognose konkreter Verkehrsströme genutzt werden könnte.

2 Abgrenzung der Problemstellung

Die grundlegende Problematik bezieht sich auf das Verhalten von Reisenden bei der Buchung einer Fahrkarte bzw. bei der Fahrplansuche. Die dem Reisenden angebotene Verbindung beinhaltet im Regelfall auch die Empfehlung einer passenden Fahrkarte oder – falls der Reisende bereits eine Fahrkarte besitzt – den Einbezug dieser Fahrkarte in die Fahrplansuche unter Berücksichtigung eines geringstmöglichen zu zahlenden Preisaufschlags. Insbesondere bei der Auswahl aus verschiedenen Routen zu einem Ziel ist daher zu erörtern, welche Eigenschaften die verschiedenen Routen besitzen, um sie zu vergleichen und die aus Sicht des Reisenden möglichst beste Lösung herauszufinden.

Das impliziert im Anwendungsfall eine Aufteilung des grundsätzlichen Problems „Finden der besten Route“ in zwei Teilprobleme: die Auswahl des dargestellten Angebots von Seiten des Fahrkartenverkäufers respektive der Informationsplattform als Vermittler der Reisedienstleistung, sowie die Auswahl aus diesem dargestellten Angebot durch den Reisenden. Diese Trennung ist insbesondere deswegen erforderlich, weil die Darstellung aller möglichen Routen im Regelfall den (potenziellen) Fahrgast völlig überfordern würde: Psychologische Studien gehen davon aus, dass Menschen Entscheidungen am besten aus einem Bündel mit wenigen Alternativen treffen können, eine Entscheidungsfindung aus einer großen Anzahl an Alternativen bedingt daher schon aus Gründen der Übersichtlichkeit eine Vorauswahl, die im Falle des hier konkreten Problems der Routenauswahl im Schienenpersonenverkehr durch den Anbieter des Eisenbahnverkehrsdienstes bzw. den Fahrkartenverkäufer erfolgt.[Gre07]

Diese Arbeit wird jedoch über diesen – rein rationalen – Umfang hinaus die Entscheidungsfindung des Fahrgasts nicht weiter beleuchten, das heißt insbesondere Unternehmenspräferenzen und Präferenzen für besondere Fahrten oder besondere Fahrtwege nicht einbeziehen. Der Ansatz auf makroskopischer Ebene wird dadurch nicht beeinträchtigt, da keine Beleuchtung der konkreten Situation in Deutschland und dem angrenzenden Ausland auf quantitativ korrekter Ebene erfolgen soll, sondern die geografische Eingrenzung lediglich der Nutzung eines bereits vorhandenen Datenmodells des Autors aus vorheriger Forschungstätigkeit (derzeit unveröffentlicht) sowie der besseren Visualisierung für die Ergebnisse der Arbeit dient. Im verwendeten Datenmodell sind dabei insgesamt 1.917 Verkehrsstationen und 21.265 werktägliche Zugleistungen integriert, von denen 2.181 Züge des Fernverkehrs und 19.084 Züge des Nahverkehrs darstellen.

Aus den gewonnenen Daten sind daher insbesondere keine konkreten Rückschlüsse auf die Situation für Anbieter des Schienenpersonenverkehrs oder Verkäufer von Fahrkar-

ten in Deutschland zu ziehen. Die gewonnenen Erkenntnisse sind rein akademischer Natur und sind nach Anpassung der lokalen Gegebenheiten und Eingangsdaten auf jede identisch oder ähnlich aufgebaute Modellumgebung verwendbar. Es erscheint praktikabel, das Modell auch mit konkreten Werten für einzelne Länder zu bestücken und daraus folgend zur Gewinnung von Realdaten für diese Länder zu verwenden. Weiterhin werden auch andere mikroskopische Effekte aufgrund fehlender Datengrundlagen nicht detailliert betrachtet werden können, so beispielsweise die Effekte besonderer eingesetzter Fahrzeuge,² oder die Spezifika bestimmter Regionen, Städte und Länder, die eine Routenauswahl beeinflussen oder eingrenzen könnten.

Im Rahmen der Arbeit werden auch keine Algorithmen vorgestellt, die die (präferierte) Routenauswahl aus Sicht des Anbieters darstellen oder dazu geeignete Steuerungsanreize (hier: insbesondere eine Preissteuerung) vorstellen. Ein solcher Ansatz ist als Vertiefung der vorliegenden Arbeit denkbar und erscheint auf Basis des hier zu entwickelnden Algorithmus der Routenfindung für die Passagiere als passendes Gegenstück auf Unternehmensseite, quasi als Simulator der Wirkungsweise verschiedener Grundsatz- und Einzelfallentscheidungen des Preis- und Auslastungs-Steuerungsmanagements.

Die Entscheidungsfindung bei der Routenauswahl wird in dieser Arbeit ausschließlich auf den Fernverkehr des Verkehrsträgers Schiene bezogen, und hier insbesondere auf Schienenbahnen, die in Deutschland nach den Regelungen des Allgemeinen Eisenbahngesetzes (AEG) betrieben werden. Dies ermöglicht im Rahmen dieser Arbeit insbesondere die Abgrenzung zu Straßenbahnen, die in Deutschland im Regelfall auch abweichend tarifiert werden.

Die Abgrenzung des Schienenpersonenfernverkehrs erfolgt analog der negativen normativen Definition des Allgemeinen Eisenbahngesetzes:

„Schienenpersonennahverkehr ist ein Verkehrsdienst, dessen Hauptzweck es ist, die Verkehrsbedürfnisse im Stadt-, Vorort- oder Regionalverkehr abzudecken. Das ist im Zweifel der Fall, wenn in der Mehrzahl der Beförderungsfälle eines Zuges die gesamte Reiseweite 50 Kilometer oder die gesamte Reisezeit eine Stunde nicht übersteigt.“

— §2 (12) Allgemeines Eisenbahngesetz [AEG16]

² Die Werte für den Fahrtkomfort werden im Modell grundsätzlich auf die Zuggattung bezogen, unabhängig auch von weiteren Eigenschaften wie der gefahrenen Geschwindigkeit oder dem Halteabstand.

Die ausschließliche Berücksichtigung von Fahrten im Fernverkehr (auch mit Verkehrsmitteln des Eisenbahnnahverkehrs) ist zum einen algorithmisch bedingt, da insbesondere für kürzere Fahrten weitere Faktoren wie der schnelle Zugang zur Verkehrsstation sowie eine verstärkte intermodale Konkurrenz durch weitere Verkehrsträger wie Bus und Straßenbahn eine große Rolle spielen, Zum anderen ist die fehlende Berücksichtigung von Nahverkehrsmitteln auch modellbedingt, da Eingangsdaten für diese Verkehrsmittel in den Grunddaten nicht zur Verfügung stehen. Eine detailliertere Auseinandersetzung mit der Problematik der Verkehrsträgerauswahl wird in der Zielgruppendefinition in Kapitel 3.3 vorgenommen.

3 Passagiertheorie

In diesem Kapitel werden die wesentlichen Theorien bezogen auf die Auswahl und Buchung einer Reise als Grundlage der Arbeit vorgestellt.

3.1 Auswahlproblem

Grundsätzlich steht jeder Reisende zu Beginn seiner Anfrage vor zwei grundsätzlichen Abwägungen, nämlich erstens der nach dem Verkehrsmittel und zweitens der nach dem Suchmodus. Typischerweise wird der potenzielle Reisende dazu vor der Reise ein oder mehrere Unternehmen um eine Verbindungs- und Preisauskunft ersuchen, und sich dann zwischen den gegebenen Alternativen entscheiden. Elektronische Portale wie Qixxit³ bieten heute auch die Suche einer Reiseverbindung mit der optionalen Kombination mehrerer Verkehrsmittel und Verkehrsträger an; Jeder Suche (auch verkehrsmittelübergreifend) muss jedoch aufgrund der vorzunehmenden Implikation der Fahrgastwünsche zur Reduktion der Alternativen auf eine beherrschbare Menge an Entscheidungsoptionen bereits eine Abwägung zwischen Preis, Komfort und Reisedauer vorangestellt werden. Entsprechend nehmen diese Reiseportale statt der sonst üblichen Auswahl eines Unternehmens bzw. des Verkehrsträgers durch den Kunden den Suchmodus vorweg, abgesehen von der Priorisierung einzelner Optionen wie Verbindungen mit besonders wenigen Umstiegen durch den Reisenden.

Portale mit einer kombinierten Routensuche über mehrere Verkehrsmittel stellen jedoch noch eine recht neue Entwicklung dar. Die überwiegende Mehrzahl der Reisenden trifft zuerst die Entscheidung für ein oder mehrere Verkehrsmittel, bevor dann eine gründliche Auseinandersetzung mit den konkreten Auswahloptionen des oder der gewählten Verkehrsmittel stattfindet.[Boe14]

Im Folgenden wird daher ein Entscheidungsbaum für die intermodale Entscheidung in Abbildung 1 dargestellt und dann auch für den weiteren Verlauf der Arbeit als vollständig zutreffend angenommen:

³ online unter <https://www.qixxit.de/>.

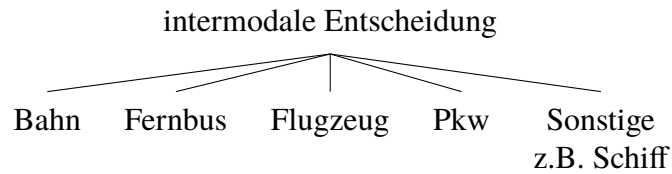


Abbildung 1: Intermodale Entscheidungsfindung
Quelle: Eigene Darstellung

Um zu dieser ersten Entscheidung zu kommen, wird der Fahrgast die verschiedenen infrage kommenden Optionen zumindest grob prüfen, zum Beispiel im Hinblick auf die zeitliche und räumliche Verfügbarkeit des Verkehrsmittels, sowie das grundsätzliche Zusammentreffen mit den eigenen Präferenzen, zum Beispiel in Bezug auf Komfort und Reisedauer.

Ist die Entscheidung für ein bzw. mehrere Verkehrsmittel getroffen, wird innerhalb des Prozesses der Routenfindung eine intramodale Entscheidungsfindung notwendig: Welche der vorgeschlagenen Optionen ist für die Fahrt vorzuziehen? Es ist dabei im Übrigen zu beachten, dass die Entscheidungsfindung zwar intramodal stattfindet (die Entscheidung bezieht sich dabei nur auf die Optionen eines Haupt-Verkehrsträgers), aber die Reisekette selbst auch intermodale Bestandteile beinhalten kann, insbesondere im Vor- und Nachlauf zur Eisenbahnfahrt.

Die intramodale Entscheidungsfindung (dargestellt in der folgenden Abbildung 2) selbst wird dann insbesondere unter Einbezug der zeitlichen Verfügbarkeit, der Dauer der Reise, des zu erwartenden Reisekomforts sowie des Beförderungspreises getroffen:

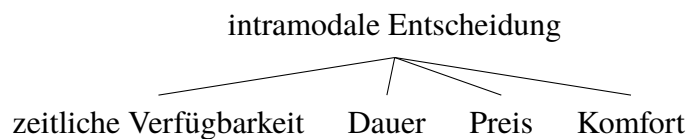


Abbildung 2: Intramodale Entscheidungsfindung
Quelle: Eigene Darstellung

Hierbei sind für jeden Parameter verschiedene Aspekte gesondert zu beachten, auch im Bezug auf die Person des Reisenden. Je nach Zweck der Reise kann sich beispielsweise die Zahlungsbereitschaft deutlich unterscheiden; Abhängig von der Dauer der Reise wird auch der Reisekomfort unterschiedliche Gewichtungen erfahren. Daher ist es zuerst notwendig, die Zielgruppen im Schienenpersonenverkehr zu definieren und detailliert zu betrachten, bevor danach mit der Ursprungsfrage fortgefahren werden kann.

Es stellt sich weiterhin die grundsätzliche Frage, welche Basistheorie zur Auswahl einer Route aus einem gegebenen Bündel verwendet wird: eine rein präferenzbasierte Auswahl oder das Konzept der Nutzentheorie. Bei einer präferenzbasierten Auswahl würde die Entscheidung über die Wichtigkeit der einzelnen Kriterien prozentuiert und diese danach miteinander verrechnet (zum Beispiel „Preis 70 %, Dauer 10 %, Komfort 20 %“). Im Rahmen der Nutzentheorie werden den einzelnen Präferenzen Gewichtungsfaktoren zugewiesen und diese anschließend miteinander verrechnet. Das führt dazu, dass die Gewichtungsfaktoren zwar ebenfalls eine Verteilung der einzelnen Gewichtungen vornehmen, diese aber anders als bei der präferenzbasierten Auswahl nicht starr ist, sondern von der Höhe des jeweiligen Kriteriums abhängt. Bei Zielgruppen mit starken Präferenzen für bzw. gegen eines der Kriterien wirkt die Nutzentheorie daher eher ausgleichend, insbesondere bei gleichzeitig stark abweichenden anderen Kriterien.[Jan95, S.254ff.]

Die Nutzentheorie unterscheidet sich aber auch noch in einem weiteren Detail von der präferenzbasierten Auswahl: bei der Nutzentheorie können den einzelnen Kriterien zielgruppenweise andere Gewichtungsfaktoren zugeordnet werden, sodass jeder Zielgruppe eigene Werte für den Nutzen (zum Beispiel für einen zu vermeidenden Umstieg) zugeordnet werden können, während bei der präferenzbasierten Auswahl lediglich die Gewichtung der Kriterien untereinander verändert werden würde, bedingungsgemäß aber nicht die Gewichtungsfaktoren der einzelnen Kriterien.

Eine Änderung des Gewichtungsfaktors der einzelnen Gruppen in der präferenzbasierten Auswahl würde darüber hinaus deren eigentlichen Ansatz konterkarieren, da dann zwei verschiedene Faktoren – die Gewichtung der einzelnen Präferenz sowie die Verrechnung der dann gewichteten Präferenzen insgesamt – sich gegenseitig ergänzen oder kompensieren würden; Mithin wäre das Auswahlzenario nicht klar, da sich die Wertebereiche des möglichen Entscheidungspfads grafisch als Fläche darstellen würden. Die Entscheidung des einzelnen Reisenden für oder gegen eine einzelne Route wäre dann nicht mehr ausreichend nachvollziehbar, weil das in jedem Fall auch von den Details der Route und den gewählten Präferenzen abhängen würde.

Aus den vorgenannten Gründen, insbesondere aufgrund der Möglichkeit der Nachvollziehbarkeit der Entscheidungen und der präferenzbasierten Entscheidung bei starken Präferenzen für bestimmte Kriterien, wird daher die Nutzentheorie der präferenzbasierten Auswahl vorgezogen. Die Nutzentheorie wird damit als Basis des algorithmischen Modellaufbaus verwendet und im Folgenden auch detaillierter erläutert.

3.2 Nutzentheorie

Die Nutzentheorie besagt allgemein, dass Menschen bestrebt sind, mit rationalen Entscheidungen einen Auswahlvorgang gegebener Elemente möglichst so zu bestreiten, dass sie ihren individuellen Nutzen maximiert haben. Die Kernidee der Nutzentheorie ist also die Rationalisierung von Entscheidungen und deren Abbildung im Rahmen eines wissenschaftlichen Modells, sodass für eine breite Masse auf makroskopischer Ebene Entscheidungen nachgebildet werden können, um Verhaltensprognosen erstellen zu können.[Pie16] Innerhalb der betrachteten Menge an Konsumenten (in diesem Modell: Fahrgästen) werden individuelle Präferenzen dann nur noch durch die Zugehörigkeit zu bestimmten Zielgruppen ausgedrückt. Das ermöglicht die Nachbildung von Entscheidungen in Modellen, da jetzt Effekte von verschiedenen Systemeingriffen wie Preis- oder Angebotsmengenänderungen dargestellt und deren Auswirkungen möglichst exakt berechnet werden können.

Die Nutzentheorie ist im Gegensatz dazu nicht zur Entscheidungsnachbildung auf mikroskopischer Ebene, also beispielsweise für die Routenwahl eines konkreten Fahrgasts, geeignet, da im Einzelfall zu viele, schwer nachzubildende Faktoren wie weitere starke individuelle Präferenzen eine Rolle spielen können (zum Beispiel die Präferenz für eine gewisse Umsteigestation, weil dort aus Sicht des Fahrgasts ein besseres gastronomisches Angebot als an anderen möglichen Umsteigestationen besteht).

Weiterhin ist grundsätzlich zwischen der kardinalen Nutzentheorie und der ordinalen Nutzentheorie zu unterscheiden. Da es in einer Vielzahl von Fällen schwierig ist, probate kardinale Umrechnungsskalen zu finden, bestehen starke Bestrebungen in der jüngeren Literatur, nicht mehr die kardinale Nutzentheorie zu verwenden, sondern die ordinale Nutzentheorie. Beide unterscheiden sich im wesentlichen in der Frage der Skalierung der Endwerte und damit der Interpretierbarkeit des ermittelten Ergebnisses: Bei der ordinalen Nutzentheorie drückt das Ergebnis hierbei nur eine Rangfolge der verschiedenen Auswahloptionen aus, bei der kardinalen Nutzentheorie drückt der Ergebniswert auch die Größe des Unterschiedes aus. [Bat08, S.39ff.][II03, S.45ff.]

Die grundsätzliche Entscheidungsfindung der ordinalen und der kardinalen Nutzentheorie ist identisch, da aber im Rahmen dieses Modells ohnehin alle Werte kardinalisiert werden (durch die Monetarisierung drückt sich unmittelbar auch die Stärke des Bedürfnisses aus), ist infolge der ausschließlichen Verrechnung kardinaler Werte auch das Ergebnis kardinaler Natur; Mithin ist das in der Literatur typisch dargestellte Problem der Konvertierung von ordinalen Werten in eine kardinale Skala in diesem Modell nicht

von Belang und stellt kein Hindernis bei der Erzeugung der Werte dar. Im Folgenden wird daher die kardinale Nutzentheorie verwendet werden, sodass das Ergebnis auch die Stärke der Präferenz bemisst.[II03, S.45ff.]

Dafür wird nun zuerst der zu ermittelnde Gesamtnutzen U_{gesamt} in mehrere handhabbare Teilnutzenwerte U_i aufgeteilt; Im Fall einer Bahnreise basiert die intramodale Entscheidungsfindung⁴ dabei auf den Teilnutzenwerten der (gefühlten) Dauer, des Preises und des Komforts (bedingungsgemäß ist die zeitliche Verfügbarkeit ein reines Ausschlusskriterium und würde – sofern benötigt – ansonsten in die gefühlte Reisedauer mit einfließen):

$$U_{\text{gesamt}} = U_{\text{Dauer}} + U_{\text{Preis}} + U_{\text{Komfort}} \quad (1)$$

Hierfür ist allen für den Nutzen maßgeblichen Werten ein neutraler kardinalisierter Nutzenwert N zuzuordnen, und die einzelnen Teilnutzenwerte sind dann jeweils mit einem Gewichtungsfaktor ω zu versehen, sodass die „neutralen“ Nutzenwerte mit der jeweiligen Präferenz des Fahrgastes versehen werden:

$$U_{\text{gesamt}} = N_{\text{Dauer}} \cdot \omega_{\text{Dauer}} + N_{\text{Preis}} \cdot \omega_{\text{Preis}} + N_{\text{Komfort}} \cdot \omega_{\text{Komfort}} \quad (2)$$

Um das Nutzenmodell besser an die Verhältnisse im SPFV anzupassen, wird der Nutzenwert des Komforts der gesamten Verbindung aufgeteilt in den Anteil des Komfortnutzens, der aus den Zugfahrten selbst resultiert (also durch den Komfort an Bord, beispielsweise durch Verpflegungsmöglichkeiten und den Sitzkomfort), und denjenigen Teil des Komfortnutzens, der durch Umsteigen bedingt ist – diese beiden Anteile können bei verschiedenen Zielgruppen unterschiedliche Gewichtungen erfahren (abhängig von den jeweiligen Präferenzen). Ohne das Ergebnis der kommenden detaillierten Betrachtung vorwegnehmen zu wollen, erscheint intuitiv klar, dass bereits die Skalenwerte beider Kriterien in die entgegengesetzte Richtung zeigen: mehr Komfort an Bord ist tendenziell gut, mehr Umstiege sind tendenziell schlecht.

Das Nutzenmodell wird jetzt dieser Idee folgend um die Aufteilung des Komfortnutzens in zwei Teilnutzenwerte erweitert und präsentiert sich danach wie folgt:

$$U_{\text{gesamt}} = N_{\text{Dauer}} \cdot \omega_{\text{Dauer}} + N_{\text{Preis}} \cdot \omega_{\text{Preis}} + N_{\text{Reise}} \cdot \omega_{\text{Reise}} + N_{\text{Umstieg}} \cdot \omega_{\text{Umstieg}} \quad (3)$$

⁴ intermodale Entscheidungsfindungen sind bedingungsgemäß kein Teil des Modells.

Um die programmtechnische Umsetzung zu erleichtern, wird im Folgenden ausschließlich mit negativen Nutzenwerten gearbeitet, das heißt die Nutzenwerte werden insgesamt zu einer Zahl, die kumuliert angibt, wie „teuer“ eine Alternative ist, zusammengefasst – dem sogenannten Reisewiderstand. In Analogie zu im Eisenbahnbereich verbreiteten physikalischen Widerständen wie dem Fahrwiderstand wird der Reisewiderstand im Rahmen dieser Arbeit mit W_{gesamt} bezeichnet. Jedes Teilwiderstandskriterium W_i bekommt dabei abhängig vom Reiseverlauf Werte zugewiesen und wird in Abhängigkeit von den persönlichen Präferenzen des Reisenden entsprechend gewichtet, und die einzelnen gewichteten Reisetilwiderstände werden dann zu einem Gesamtwiderstand aufsummiert:

$$W_{\text{gesamt}} = W_{\text{Dauer}} \cdot \omega_{\text{Dauer}} + W_{\text{Preis}} \cdot \omega_{\text{Preis}} + W_{\text{Reise}} \cdot \omega_{\text{Reise}} + W_{\text{Umstieg}} \cdot \omega_{\text{Umstieg}} \quad (4)$$

Die im jeweiligen Einzelfall beste Verbindung bestimmt sich dann durch den geringsten Widerstand, es wird also diejenige Verbindung durch den Nutzer präferiert, die den geringsten Gesamtaufwand verspricht.

3.3 Zielgruppen im Schienenpersonenverkehr

In der Literatur, zum Beispiel bei WALTHER ET AL. in [Val97], werden die Zielgruppen im Schienenpersonenverkehr regelmäßig nach ihrem Reisezweck unterschieden, denn diese Unterscheidung nimmt andere Merkmale – wie zum Beispiel den Anlass oder den Veranlasser – bereits vorweg. Die Unterscheidung erscheint daher auch insbesondere für diese Arbeit zweckmäßig, da im Folgenden die Veranlassung der Reise keine übergeordnete Rolle spielt, sondern lediglich die durch den Reisezweck und persönliche Vorlieben bestimmte Präferenzen Eingang in das Modell finden.

Die Aufteilung wird hierbei in der Literatur regelmäßig wie folgt vorgenommen:

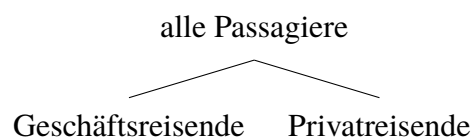


Abbildung 3: Unterscheidung der Passagiere in Primärzielgruppen
Quelle: Eigene Darstellung

Intuitiv erscheint bereits klar, dass sich Geschäfts- und Privatreisende in ihrer Zahlungsbereitschaft unterscheiden, aber gegebenenfalls auch in ihrem Komfortwunsch.

Während hierbei die Zahlungsbereitschaft von der persönlichen Situation und dem Status des Einzelnen abhängt (bei Geschäfts- wie Privatreisen), stehen beim Komfort auch andere Interessen im Vordergrund: So hat der Arbeitgeber ein Interesse daran, dass der Arbeitnehmer nach der Reise wieder die bestmögliche Arbeit leistet und wird daher auch aus diesem Grund dem Komfortwunsch des Reisenden eher nachgeben können, als dies bei einer Privatreise regelmäßig der Fall ist.

Besonders bei mittleren Distanzen im Fernverkehr tritt jedoch auch eine „Mischgruppe“ von Reisenden auf, die die Eigenschaften beider Gruppen zu vermischen scheinen und eine Relation besonders regelmäßig befahren: Fernpendler. Pendler sind aufgrund der Häufigkeit der Nutzung einer spezifischen Reiseverbindung besonders mit der Umgebung und den jeweiligen Einsatzeigenschaften der verkehrenden Züge (zum Beispiel der Belegung des Zuges in verschiedenen Reiseabschnitten und der Komfortausstattung des Zuges) vertraut und entwickeln daher auch ein eigenes Präferenzmodell, das zum Beispiel auch die regelmäßige Fahrplanstabilität der gewählten Verbindung mit betrachtet. Die Motivation für Fernpendler, zu pendeln, kann dabei geschäftlicher oder privater Natur sein. In beiden Fällen treten jedoch die genannten Phänomene in Abgrenzung zu den bisher abgetragenen Zielgruppen auf. Das in Abbildung 3 dargestellte Entscheidungsmodell ist also um die neu entwickelte Zielgruppe zu ergänzen:

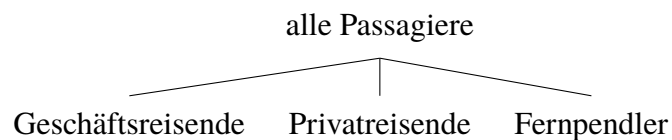


Abbildung 4: Differenziertere Aufteilung der Primärzielgruppen
Quelle: Eigene Darstellung

Jede Zielgruppe besitzt eigene Präferenzen, die auch innerhalb der Zielgruppe durchaus differieren können. Als Beispiel wird die Gruppe der Privatreisenden zuerst gedanklich aufgeteilt in verschiedene Altersgruppen, betrachtet werden sollen dabei Senioren und Jugendliche. Nun soll eine längere Strecke als Fahrt zu einem Urlaubsaufenthalt gewählt werden, zum Beispiel vom Süden Deutschlands zur Insel Rügen. Während mutmaßlich der Anspruch der Senioren in einer möglichst umsteigefreien Verbindung als erster Präferenz bestehen wird, achten die Jugendlichen vielleicht besonders auf den Preis oder besondere Fahrzeugeigenschaften (zum Beispiel Unterhaltungsmöglichkeiten im Zug, Internetzugang etc.). Während hier also schon innerhalb einer Zielgruppe – hier: der der Privatreisenden – Unterscheidungen vorgenommen werden müssen, um den spezifischen Präferenzen jeder Teilzielgruppe gerecht werden zu können, stellt sich insbesondere

auch die Frage nach der Generalisierung der Sensibilitätseigenschaften und Verteilung von Teilzielgruppen innerhalb aller Zielgruppen.[Rüg05]

Bezogen auf jeden aufgeworfenen Parameter hat jede Zielgruppe – oder, wie bereits gezeigt, fallweise jede Teilzielgruppe – unterschiedliche Präferenzen. Um diese Präferenzen separat ausweisen und später auch bewertend vergleichen zu können, ist die Umrechnung dieser Präferenzen in eine uniforme Einheit notwendig; Im Regelfall werden dabei in der Literatur (beispielsweise in [Nor12, S.273f.]) Geldwerte benutzt. Alle Präferenzen bzw. genauer alle Zuschreibungen von Wertungseigenschaften zu Parametern werden daher monetarisiert, um eine Vergleichbarkeit innerhalb des Systems vornehmen zu können. Die Monetarisierung der Präferenzen ist auch deshalb gegenüber anderen Einheitskonvertierungen vorzuziehen, da die Mehrzahl der Studien zum Thema ebenfalls eine Monetarisierung vornimmt, und außerdem ein Gutteil der Parameter – die Verbindungspreise sowie alle Nebenleistungspreise – bereits per definitionem monetarisiert sind.

Die Monetarisierung der Präferenzwerte dient auch der programmtechnischen Umsetzung des Modells, vgl. Kapitel 3.2. Die hier gewonnenen monetarisierten Werte entsprechend der in der Nutzentheorie verlangten kardinalen Skala und können daher direkt und ohne weitere – gegebenenfalls mit Unschärfen belegte – Konvertierungen zur Berechnung der Reisetilwiderstände genutzt werden.

Um die Zielgruppen nun genauer untersuchen zu können, werden sie im Folgenden einzeln vorgestellt und mit weiteren Eigenschaften und ihren Präferenzen bezogen auf die Reiseparameter versehen, beginnend mit der Gruppe der Geschäftsreisenden.

3.3.1 Geschäftsreisende

Geschäftsreisende sind einer der größten Umsatzbringer im SPFV der Deutschen Bahn AG, obgleich zahlenmäßig in der Minderheit: Ihr Anteil an der gesamten Reisendenzahl in den Zügen der DB Fernverkehr AG beträgt nach Berichten des Handelsblatts lediglich 38 %.[Tho15] In Großbritannien ist dieser Wert mit über 50 % angegeben,[Off14, S.24] beinhaltet allerdings explizit genannt auch Pendler, das ist daher analog auch bei dem für die Fernzüge der Deutschen Bahn angegebenen Wert anzunehmen. In ihrer Umfrage unter britischen Eisenbahnreisenden kommen HOLLEY ET AL. im Jahr 2004 zu dem Ergebnis, dass der reine Geschäftsreisen-Verkehr etwa 16 % der Stichprobe ausmacht, und weitere 43 % der Reisen dem Pendeln dienen.[Hol05]

Im Folgenden soll nun die Studie von COLE aus 1992 [Col05] vorgestellt und deren Relevanz erörtert werden. Vor der Verwendung ist ob des Datums der Studie die Aktualität und damit die Übertragbarkeit der Ergebnisse der Studie auf die derzeitige Situation zu diskutieren. COLE erhebt in seiner Studie Daten der British Rail, die zum Schutz der Unternehmensgeheimnisse so aufbereitet wurden, dass eine Nachvollziehbarkeit auf konkrete Ergebnisse einzelner Streckenabschnitte, Verkehrsmittel oder Verkehrstage nicht gegeben sein soll. Die Daten stellen allerdings die Anzahl und den Umsatzanteil von Fahrgästen in den verschiedenen Produktklassen (1. und 2. Wagenklasse) und Buchungsarten (Standard und Discount) dar und ermöglichen daraus Rückschlüsse auf Deckungsanteile der Reisendengruppen.

Daraus ist – unabhängig vom Alter der Studie – eine Erläuterung des Reisendenverhaltens möglich, denn obgleich die Daten der Studie nicht mehr im Detail aktuell sind, bleiben die Zusammenhänge nach Meinung des Autors sachlich identisch, da sich das Reisendenverhalten in der Zwischenzeit nicht wesentlich verändert hat.⁵ Hinzu kommt, dass die von COLE erhobenen Daten in ihrer Gesamtheit den heutigen Produktklassen und Buchungsarten in ihren Grundzügen entsprechen, sodass Rückschlüsse auf das Verteilungsverhalten und die Umsatzrelevanz einzelner Zielgruppen bezogen auf die Art der Verteilung (nicht zwingend jedoch auch auf die Richtigkeit einzelner Daten) zulässig erscheinen.

In Tabelle 3 ist dargestellt, wie sich die prozentualen Anteile der durchgeführten Reisen und des damit generierten Umsatzes (bezogen auf den Fahrkartenpreis) innerhalb der Zielgruppen der Geschäftsreisenden sowie der Privatreisenden verteilen. Um die Zahlen innerhalb der Zielgruppen besser vergleichen zu können und die Umsatzrelevanz zu verdeutlichen, wurde vom Autor dieser Arbeit noch der im Folgenden als Faktor bezeichnete Umsatzfaktor (berechnet durch das Verhältnis zwischen Umsatz einer Zielgruppe und deren Anzahl der Reisen) der Darstellung in Tabelle 3 hinzugefügt:

Produktklasse (Anteile in %)	Geschäftsreisende			Privatreisende		
	Reisen	Umsatz	Faktor	Reisen	Umsatz	Faktor
First Class	21,7	44,5	2,05	2,3	7,2	3,13
Standard	29,0	31,8	1,10	12,3	19,7	1,60
Discount	49,3	23,7	0,48	85,4	73,1	0,86

Tabelle 3: Passagiere und Umsatzanteile des InterCity von British Rail im Jahr 1992
Quelle: Eigene Darstellung auf Basis der Daten von [Col05, S.109]

⁵ Vergleiche dazu auch die Zusammenfassung der Werte unterschiedlicher Daten von TROTTEBERG und BELENKY in [Bel11].

Wie Tabelle 3 nun zu entnehmen ist, reisen 21,7 % der Geschäftsreisenden in der First Class und tragen damit 44,5 % zum Umsatz aller Geschäftsreisenden im InterCity bei (mit einem Faktor von 2,05 Umsatzanteilen pro Reise mehr als doppelt so viel wie der durchschnittliche Geschäftsreisende). Die etwas ungünstige Aufteilung von COLE mit Zusammenfassung der Umsätze nach Reisendengruppen ermöglicht darüber hinaus allerdings keinen Vergleich nach der tatsächlichen Relevanz (da die Verhältniszahlen Geschäfts-/Privatreisende sowie die für die spezifische Kombination aus Ticketart und Reisendencluster durchschnittliche Reiseweite bzw. der durchschnittliche Ticketpreis fehlen), zeigt aber die faktische Umsatzmehrung durch die höhere Anzahl der First-Class-Kunden bei den Geschäftsreisenden gegenüber den Privatreisenden.⁶[Col05]

Bei der Segmentierung der Kunden in First Class und Standard Class handelt es sich mithin um eine Form der Produktdifferenzierung; Dabei erhalten Kunden der ersten und der zweiten Klasse zwar dieselbe Hauptleistung (die Beförderung vom Start- zum Zielort), aber unterschiedliche Nebenleistungen (beispielsweise im Sitzplatzkomfort oder dem Bordservice). Mit dieser Produktdifferenzierung ist es besonders einfach, Preise zu differenzieren, da nun die unterschiedliche Zahlungsbereitschaft durch die differenzierte Leistung mittelbar abgeleitet werden kann.[Mar16d]

Die Preisdifferenzierung über andere Faktoren wie zum Beispiel lediglich über den zeitlichen Zugang zu Preisen abzurufen, ist zwar denkbar, wird jedoch im Transportsektor eher selten praktiziert.⁷ Sie stellt jedoch den Regelfall in Branchen dar, bei denen eine Preisdifferenzierung anhand anderer Merkmale kaum oder gar nicht möglich ist, zum Beispiel im Lebensmittel-Einzelhandel (dort wird der Zugang zu vergünstigten Preisen klassischerweise an einen gewissen Angebotszeitraum gekoppelt).[Sch07, S.111ff.]

Andere Formen der Produkt- und Preisdifferenzierung unterscheiden die Fahrkartenkäufer nach ihrer Flexibilität bzw. nach den Fahrkartenbedingungen für Umtausch und Rückgabe. Hierbei werden zwar einzelne Gruppen gezielt angesprochen, es kann jedoch auch Mischformen der Kundengruppen geben: So können Geschäftsreisende beispielsweise bei sicher feststehenden Terminen auch Fahrkarten mit eingeschränkter Stornierbarkeit buchen.

⁶ COLE weist in seiner Arbeit auch darauf hin, dass die Nachvollziehbarkeit nicht gegeben sein soll, dies war seiner Darstellung nach Bedingung für die Nutzung von Realdaten der British Rail.

⁷ Vergleiche dazu allerdings verschiedene Promotion-Aktionen, die jeweils nur sonntags durchgeführt werden – und damit für Geschäftsreisende als Buchungszeitraum regelmäßig nicht infrage kommen; im August 2016 beispielsweise bei airberlin der „Super Sunday“ mit Buchungszeitraum jeweils Sonntags zwischen 13 und 18 Uhr.[Air16]

Mithin wird daher mit flankierenden Maßnahmen wie Firmenkundenrabatten für spezielle Fahrausweise versucht, die Preisdifferenzierung möglichst exakt anhand der Kundengruppen vorzunehmen. Neben der Senkung der Attraktivität der Rabattpreise für Geschäftsreisende wird die Preisdifferenzierung auch anders flankiert: Unternehmen versuchen mit sogenannten Upsell-Maßnahmen, Kunden von günstigeren Produkten zum Kauf teurerer Produkte zu bewegen, ggf. unter Verzicht auf einen Teil des ursprünglich für das teurere Produkt geforderten Preises.⁸ Hierbei sind diese Upsell-Angebote oft auf die Person des Kunden bezogen und werden aus spezifischen Gesichtspunkten bewusst durch das Unternehmen ausgewählt, beispielsweise Vielfahrer oder als wenig preisbewusst bekannte Kunden.⁹

Bei der Verteilung der Ticketarten ist insbesondere zu beachten, dass im Segment der Geschäftsreisenden 50,7 % der Reisen mit Fahrkarten zum Regelpreis durchgeführt wurden (Sparpreise für die 1. Klasse gab es zum Zeitpunkt der Erhebung von COLE bei British Rail noch nicht), bei den Privatreisenden sind es lediglich 14,6 %. Setzt man dieselbe durchschnittliche Reiseweite und dieselben Kilometerpreise für beide Gruppen an, so ergibt sich, dass die Privatreisenden innerhalb ihrer Gruppe mit Fahrkarten zum Regelpreis etwa 50 % mehr zur durchschnittlichen Deckung beitragen als Geschäftsreisende innerhalb ihrer Gruppe. Es sei jedoch darauf verwiesen, dass die Reiseweite vermutlich gruppenweise differiert: so ist das Verhältnis aus Erlösen der First-Class- und der Standard-Tickets bei den Geschäftsreisenden 1,87, bei den Privatreisenden hingegen 1,95. Bei einer einheitlichen Tarifierung müsste jedoch das Verhältnis der Umsätze gleich sein, da das Preisverhältnis – unabhängig von der verkauften Menge – identisch wäre.

Die große Menge der verkauften Tickets zu günstigeren Preisen als dem Regelpreis (die in Tabelle 3 als „Discount“ bezeichnete Kategorie) deutet darauf hin, dass der Regelpreis insbesondere für Privatreisende entweder preislich oder aufgrund der mit ihm verbundenen Bedingungen nicht attraktiv erscheint; Im vorliegenden Fall waren zwar die Discount-Tarife mit Restriktionen bzgl. der Buchungsmöglichkeiten und der Reiseflexibilität versehen, nicht jedoch der Regelpreis. Die festzustellende hohe Nutzungsquote der Discount-Tarife statt des Standard-Tarifs durch die Privatreisenden stellt daher vermutlich lediglich eine Ausweichreaktion auf die höheren Preise des Standard-Tarifs dar. Zu den Fahrgästen mit Discount-Fahrscheinen zählen in den Daten

⁸ Die Deutsche Bahn beispielsweise bietet für ausgewählte Fahrkarten ein Upsell der 2. Klasse auf die 1. Klasse für einen Festpreis von 12€ per E-Mail an. Der Festpreis gilt nach Erfahrung des Autors hierbei auch, wenn der Aufpreis für die 1. Klasse eigentlich mehr als 100€ betragen würde.

⁹ Bei der Deutschen Bahn AG ist daher der Bezug des Newsletters – und damit eine Teilnahme am Marktforschungsprogramm – verpflichtend zum Bezug von Upsell-Angeboten.

von COLE jedoch auch die Besitzer von Vergünstigungskarten, die nach dem Kauf Rabatt gewähren (im deutschen SPFV beispielsweise die BahnCard 25 der Deutschen Bahn AG), sodass die Abgrenzung auch bezüglich der einzelnen Statusgruppen (zum Beispiel Vielfahrer oder Gelegenheitsfahrer) nicht zwingend sauber herleitbar ist.[Col05]

COLE stellt in seiner Studie weiterhin fest, dass die Preiselastizität der Geschäftsreisenden im von ihm betrachteten Marktbereich besonders niedrig ist: mit einem Faktor von -0,4 (s. [Col05, S.63]) würden mit einer Preiserhöhung um 10 % lediglich 4 % Nachfrageverlust einhergehen; mithin sind bei Geschäftsreisenden aufgrund der geringen Preiselastizität offensichtlich andere Elastizitäten (und damit andere Präferenzen) prioritär. Zum Vergleich: die Preiselastizität sonstiger Reisender bewegt sich in der Studie von Cole in der Spanne zwischen -0,3 und -1,1. Im aus Sicht des Eisenbahnverkehrsunternehmens ungünstigsten Fall würden daher bei einer Preiserhöhung von 10 % insgesamt 11 % weniger Fahrgäste mitfahren. Umgekehrt bedeutet das für Eisenbahnverkehrsunternehmen, dass die Preise für Geschäftsreisende (insbesondere auch im Vergleich zu anderen Reisenden) tendenziell eher hoch zu wählen sind. Der kritische Punkt könnte dabei jedoch die Unterscheidung des Fahrgasts nach seinem Reiseanlass sein, um die Preiselastizität abschätzen zu können. Entsprechend der geringen Preiselastizität ist die Auslastungssteuerung mit Geschäftsreisenden im Übrigen nur sehr beschränkt (oder sehr teuer) machbar.[Col05, S.63ff.]

OWEN und PHILLIPS stellen dabei in [Phi87, S.233] auch fest, dass die Preiselastizität insbesondere im Eisenbahnverkehr nicht nur speziellen Einflüssen aufgrund der Marktsituation ausgesetzt ist (viele Relationen sind intramodale Monopolmärkte, da nur ein Eisenbahnverkehrsunternehmen eine Verbindung anbietet), sondern aufgrund des Bündels an möglichen Nebenleistungen und zusätzlichen Bedingungen beim Ticketkauf (beispielsweise Stornierungskonditionen und Angebotszeiträume) auch für einzelne Zielgruppen keine dauerhaft gültigen Werte unabhängig von der Art der Fahrkarte angebracht werden können, und daher jeweils eine differenzierte Betrachtung der speziellen Gegebenheiten des Einzelfalls durchgeführt werden muss.

Weiterhin ist zu beachten, dass die Preiselastizität nach COLE mit den bereits hergeleiteten Elastizitätswerten nur zur Anwendung gebracht werden kann, sofern die Rede von einem Monopolmarkt ist, der Anbieter mithin keiner intramodalen Konkurrenz ähnlicher Qualität ausgesetzt ist. Im Fall einer Reise im SPFV wird dieses Modell also nur zur Anwendung zu bringen sein, wenn die Reise in dieser oder einer ähnlichen Weise nur durch einen Anbieter erbringbar ist, wobei sich die Unterscheidung nur auf den intramodalen Wettbewerb bezieht, intermodale Wettbewerber sind in der Elastizitätsbe-

rechnung bereits berücksichtigt. Es ist jedoch zu beachten, dass sich insbesondere die intermodale Wettbewerbssituation je nach Reiserelation naturgemäß stark unterscheidet, die Elastizität mithin eigentlich relationsscharf zu betrachten wäre.[Col05]

Innerhalb eines intramodalen Wettbewerbsspektrums werden die Wanderungsbewegungen zwischen den Anbietern um ein Vielfaches stärker ausgeprägt sein als bei einer intermodalen Betrachtung, da wesentliche Entscheidungskriterien wie Komfort und Zugang identisch oder einfach vergleichbar sind; diese Annahme ist einfach zu veranschaulichen, wenn für eine gegebene Reiserelation ein Modell angenommen wird, in dem zwei Anbieter exakt parallel verkehren, zu denselben Fahrzeiten sowie mit gleichwertigem Fahrzeugmaterial und Serviceangebot. Abgesehen von einzelnen persönlichen Präferenzen werden die Fahrgäste vermutlich schlicht den günstigeren Anbieter auswählen; entsprechend geht die Preiselastizität gegen unendlich (= alle Fahrgäste wandern ab), wenn der Konkurrent im Falle einer eigenen Preiserhöhung günstiger wäre, und gegen null (= keine Fahrgäste wandern ab), wenn der Konkurrent trotz einer eigenen Preiserhöhung immer noch teurer ist für dieselbe Leistung.

In der von COLE angegebenen, relativ geringen Preiselastizität spiegelt sich indirekt auch die Zahlungsbereitschaft der Zielgruppe wieder: Die Geschäftsreisenden bezahlen entsprechend für Mehrleistungen auch entsprechende Mehrpreise, und belohnen (aus Sicht eines Eisenbahnverkehrsunternehmens) schnelle Verkehre mit Preisauflägen. Umgekehrt ist die Zahlungsbereitschaft der Zielgruppe ein Hinweis darauf, dass Geschäftsreisende im Regelfall überdurchschnittlich gut verdienen oder die Leistung nicht selbst bezahlen, weil Dienstreisen vom Arbeitgeber bezahlt werden, und kein spezifischer Anreiz zur Sparsamkeit bei der Durchführung von Dienstreisen gesetzt wurde.[Col05]

Hierbei spielt auch das Statusdenken von Geschäftsreisenden eine Rolle; eine Vorzugsbehandlung zu genießen (beispielsweise durch eine Fahrt in der 1. Klasse oder einen Zugang zur Business-Lounge der Eisenbahngesellschaft) ist ein auch aus Sicht des Unternehmens bewusst einzusetzender Motivationsfaktor für den Arbeitnehmer. Neben der Abgrenzung gegenüber anderen Arbeitnehmern als Statusmerkmal spielt der Komfortgewinn selbst sicher eine Rolle in der Entscheidung für eine Fahrt in der 1. Wagenklasse. Gleichzeitig kann auch das Unternehmen gewinnen: neben der verhältnismäßig kostengünstigen Motivation des Arbeitnehmers lassen sich so unter Umständen auch andere Kosten vermeiden, zum Beispiel bei der Bereitschaft des Geschäftsreisenden, in der 1. Klasse auch frühmorgens anzureisen, um eine Hotelübernachtung zu umgehen.

Zusammenfassend lässt sich sagen, dass Geschäftsreisende keine in sich homogene Gruppe darstellen: Fernpendler werden deswegen noch separat betrachtet. Innerhalb der Gruppe der Geschäftsreisenden (ohne Fernpendler) lässt sich als Tendenz insbesondere in Bezug auf die Feststellungen von COLE festhalten, dass die Zahlungsbereitschaft recht hoch ist – umgekehrt lässt sich also sagen, dass andere Faktoren (Komfort, Status) die Entscheidung tendenziell überlagern und die Preiselastizität folgerichtig relativ gering ist, eine Preiserhöhung also nur verhältnismäßig wenige Fahrgäste ausschließt.[Col05]

Entsprechend steht auch zu vermuten, dass die weiteren Elastizitäten (beispielsweise die Umsteigeelastizität) höher sind und damit für Geschäftsreisende im Sinne der Entscheidungsfindung kritischer sind als der Preis. Eisenbahnverkehrsunternehmen können folglich von Geschäftsreisenden überdurchschnittliche Erlöse erzielen, wenn sie sich hinreichend auf einen zielgruppenspezifischen Fahrkartenverkauf und passende Reiseangebote fokussieren.

3.3.2 Privatreisende

Innerhalb der Zielgruppe der Privatreisenden bietet es sich an, zu Beginn ebenfalls eine Gliederung von Teilzielgruppen vorzunehmen, und zwar in Analogie zur Untergliederung der Geschäftsreisenden nach dem Reisezweck. Hierbei überlagern sich die Bedürfnisse der nach soziodemografischen Merkmalen eingeteilten Zielgruppen allerdings mit der Unterteilung nach dem Reisezweck: So spielt beispielsweise für die Bedeutung des Umsteigewiderstands nicht nur das Alter der Zielperson eine Rolle, sondern auch der Reiseumstand – bei einer Urlaubsreise mag der Umsteigewiderstand oftmals höher sein als bei einem Tagesausflug.[Rüg05]

Bei der Unterteilung in die Teilzielgruppen wird nachfolgend der in vielen Untersuchungen genutzte Reisezweck *Leisure* bzw. *Freizeit* unter Berücksichtigung der Feststellungen von Rüger in [Rüg05] unterteilt in Fahrten, die für Tagesausflüge und Kurzurlaube von wenigen Tagen Aufenthalt durchgeführt werden, und Fahrten, die längeren Urlauben dienen. Entsprechend der Dauer des Urlaubs unterscheiden sich auch die Anforderungen dieser Teilzielgruppen: während bei kürzeren Aufenthalten die gesamte Reisezeit einen nicht unbedeutenden Teil der Urlaubszeit einnimmt und somit vom Urlaubenden eher minimiert werden möchte, steht bei längeren Urlaubsreisen im Regelfall der Reisekomfort und dabei insbesondere die Vermeidung von Umstiegen im Vordergrund. Bezogen auf den Reisepreis ergibt sich, dass für kürzere Urlaubsdauern der Reisepreis zu einem wesentlichen Teil vom Fahrkartenpreis der Bahnreise determiniert

wird: Mithin ist der Preis bei kürzeren Reisen entscheidender als bei langen Reisen, die Preiselastizität nimmt entsprechend mit der zunehmenden Reisezeit ab.[Cle96, S.140ff.]

Sofern die Unterteilung hinsichtlich soziodemografischer Merkmale vorgenommen wird, ergibt sich ein ähnliches Bild, allerdings nicht vorrangig bezogen auf die Dauer der Reise. Während Jugendliche tendenziell eine höhere Preiselastizität besitzen und daher bereit sind, den Reisekomfort entsprechend geringer zu gewichten, steht mit zunehmendem Alter für Berufstätige aufgrund der limitierten Urlaubszeit die Reisezeit im Fokus und für Senioren ist auch aufgrund körperlicher Einschränkungen tendenziell der Reisekomfort relevant.¹⁰[Rüg05, S.8ff.]

Insbesondere die bei Kurzreisen im Verhältnis hohe Preiselastizität sorgt für gute Möglichkeiten der Eisenbahnverkehrsunternehmen, die Auslastung der Züge zu steuern. Aus den in Tabelle 3 nach COLE dargestellten Daten ergibt sich, dass Privatreisende im Standard-Fahrkartensegment im Schnitt fast 87 % mehr Einnahmen Erlösen als im Discount-Segment; Dies mag auf eine geringere Verfügbarkeit von Sparpreisen für längere Strecken hindeuten (und erlaubt aufgrund des Fehlens von Absolutzahlen zur Reiseweite nur Mutmaßungen), jedoch zeigt auch der Vergleich der absolut verkauften Fahrkarten, bei denen das Discount-Segment jenes der zum Regeltarif verkauften Fahrkarten um das Sechsfache übertrifft, dass eine Vielzahl der an Privatreisenden verkauften Fahrkarten also solche sind, die in irgendeiner Weise rabattiert sind. Zu den Rabatten zählen hierbei neben Rabattierungskarten wie der BahnCard auch besondere Rabatte für Hin- und Rückfahrten sowie spezielle Sparpreise für bestimmte Regionen oder Angebotszeiträume.[Col05]

Es besteht mit diesem Wissen die von Eisenbahnverkehrsunternehmen zunehmend auch genutzte Möglichkeit, dass die Preiselastizität von Privatreisenden dazu verwendet wird, die Auslastung der Züge in absehbar schwächeren Nachfrageperioden zu erhöhen, zum Beispiel mit Kombinationsangeboten zusammen mit weiteren Anbietern wie Museen und Hotels, aber auch in Form von werbewirksamen Sparpreisen zu günstigeren Ticketpreisen als die Fahrkarten zum Regelpreis. Dies ist für das Unternehmen dann eine gute Wahl, wenn möglichst wenig bisherige Standard-Fahrkarten-Käufer auf die neuen rabattierten Angebote umsteigen und gleichzeitig möglichst viele zusätzliche Kunden gewonnen werden können, sodass die durch die Rabattierung der Fahrkarten

¹⁰Eisenbahnunternehmen können hierbei den Reisekomfort durch flankierende Maßnahmen wie einen Gepäckträger-Service am Umsteigebahnhof oder die Möglichkeit des Gepäckversands per Kurierdienst erhöhen.

entstehende Umsatzminderung durch den höheren Absatz überkompensiert wird und damit insgesamt mehr Umsatz erzielt wird.

Weiterhin ist in den in Tabelle 3 dargestellten Daten von COLE noch eine andere Tendenz sichtbar: Privatreisende bezahlen im Durchschnitt weniger als Geschäftsreisende. Das ist am Verhältnis zwischen der Anzahl der Reisen und dem daraus erzeugten Umsatz der First-Class-Tickets erkennbar: bei Geschäftsreisenden beträgt er 2,05, bei den Privatreisenden beträgt er 3,13. Gesezt, beide Zielgruppen hätten dieselbe Reiseweite, ergibt sich für die Privatreisenden ein wesentlich niedriger Durchschnittspreis über alle verkauften Fahrkarten, da die Fahrkartenpreise beider Zielgruppen in der ersten Klasse relativ zur Reiseweite identisch sind.[Col05]

Zusammengefasst lässt sich für die Zielgruppe der Privatreisenden festhalten, dass eine hohe Diversität bezogen auf die soziodemografischen Merkmale und daraus abgeleitet auch auf die Kundenbedürfnisse besteht. Die beiden Hauptteilzielgruppen unterscheiden sich dann in eher preissensible Kurzurlauber und eher komfortensible Reisende bei längeren Urlaubsreisen (Fern- und Wochenendpendler als dritte wesentliche Teilzielgruppe werden im folgenden Kapitel separat betrachtet). Eisenbahnverkehrsunternehmen können jedoch die hohe Preissensibilität von Kurzurlaubern auch zur Generierung neuer Umsätze nutzen, indem durch günstige Preise Nachfrage induziert wird und von anderen Verkehrsträgern Reisende zur Eisenbahn wechseln.

3.3.3 Fernpendler

Die in den vorherigen Betrachtungen ausgeklammerten Fernpendler stellen in ihrer Anlage eine Mischgruppe zwischen den Privatreisenden und den Geschäftsreisenden dar. Da sie mit dem Pendeln einen erheblichen Anteil ihrer Zeit verbringen, sind sie in der Wahrnehmung und Wertschätzung als sensibler anzusehen als typische Privatreisende im Freizeitverkehr. Da sie jedoch im Regelfall den Ticketpreis selbst entrichten müssen, sind sie gleichzeitig preissensibler als Geschäftsreisende bzw. verzichten zugunsten des Preises auf Komfortmerkmale wie die 1. Klasse (vergleiche dazu die Erhebungen von RÜGER in [Rüg05] und die Zusammenfassung von CLEVER in [Cle96]). Die Gruppe der Fernpendler kann jetzt noch unterschieden werden in Wochenendpendler, die jeweils zum Wochenende pendeln, und Tagespendler, die arbeitstäglich pendeln; Beide Gruppen weisen in einigen Belangen ein tendenziell ähnliches Fahrtverhalten auf und werden daher im Folgenden teilweise gemeinsam betrachtet.

Untersuchungen (zum Beispiel vom herausgegeben vom Bundesministerium für Verkehr, Bau und Stadtentwicklung, dargestellt in [Ins06]) zeigen, dass die zurückgelegte Reiseweite von Pendlern in den zurückliegenden Jahrzehnten stetig angestiegen ist, die für das Pendeln aufgewandte Reisezeit jedoch nicht. Die Menschen reisen also nicht weiter, weil sie mehr Zeit aufwenden, sondern lediglich, weil sie mit ihrem vorhandenen Zeitbudget durch Beschleunigungen des Verkehrs jetzt weiter fahren können. Gleichzeitig ist belegt, dass auch der Anteil der Fernpendler zunimmt, insbesondere, weil die Arbeitsverhältnisse heute eine weitergehende Flexibilität erfordern als früher.[Ins06, S.13ff.]

Sofern die wesentlichen Widerstände – im Folgenden zuerst für Tages-Fernpendler – beurteilt werden sollen, ist beginnend beim Zeitwiderstand insbesondere davon auszugehen, dass das jeweils als „Ziel“ gesetzte Budget des jeweiligen Pendelnden nicht überschritten werden darf. Die Zeitelastizität ist also relativ hoch. Eine geringe Zeiteinsparung erzeugt damit einen überproportional hohen Nachfragezuwachs im betroffenen Marktsegment, weil eine Vielzahl von potenziellen Pendelnden bereits mit einer geringen Zeitersparnis zum Umstieg bewegt werden kann.

Gleichzeitig ist der Reisekomfort besonders in Form einer unterbrechungsfreien und daher möglichst umsteigefreien Fahrt wichtig, damit die Zeit im Zug optimal genutzt werden kann (beispielsweise zur Erholung oder zum Arbeiten). Das überwiegt den Komfortgewinn, der ansonsten aus komforterhöhenden Faktoren wie der Nutzung der 1. Klasse gewonnen werden würde. Die Nachfrageelastizität bezogen auf den Komfort ist also gespalten zu sehen: Für Umstiege ist sie besonders hoch, für den restlichen Komfort ist sie nicht besonders entscheidend.[Cle96, S.140]

Bei Wochenend-Fernpendlern steht aufgrund der tendenziell höheren Distanzen auch die generelle Erreichbarkeit des Ziels zur Debatte: Dabei können Umstiege eher in Kauf genommen werden, wenn im Gegenzug dafür der Preis der Reise gesenkt werden kann oder sich weitere Vorteile wie ein erhöhter Reisekomfort während der Fahrt ergeben, insbesondere ist hierbei der Fokus auf die Bordunterhaltung zu legen. Entsprechend ist für Wochenendpendler der Umsteigewiderstand etwas geringer als für Tagespendler, dafür aber das Fahrtkomfortempfinden im Verhältnis höher.[Rüg05, S.19f.]

Als weitere Dimension ist die Preiselastizität zu nennen, zur Darstellung der Nachfrageänderung bei einer Preisanpassung. Grundsätzlich stellt sich hierbei die Frage, inwieweit die Entscheidung des Pendelns tatsächlich durch Anbieter von Schienenpersonenverkehr bzw. insbesondere den Preis dieser Angebote beeinflusst werden kann; Im Speziellen jedoch wird zielgruppenspezifisch innerhalb der Fernpendler eine Preiselastizität als

Reaktion auf einen ansonsten notwendigen Wohnortwechsel bzw. die Einrichtung einer weiteren Wohnung anzunehmen sein.

Entsprechend ist von einer Preiselastizität auszugehen, die im Fall rationaler Entscheider dann zum Tragen kommt, wenn die Auswahl zwischen den Kosten der Einrichtung einer weiteren Wohnung (oder alternativ den Kosten eines Umzugs) und den Kosten des Pendelns „kippt“, also das Pendeln im spezifischen Fall gerade noch günstiger ist oder bereits teurer ist als die jeweilige individuelle Alternative.

Insbesondere für Wochenendpendler sind jedoch oftmals äußere Umstände entscheidend, die zum Pendeln beitragen, wie zum Beispiel die familiäre Situation oder eine lediglich zeitlich befristete Verlagerung der Hauptarbeitsstätte. Hierbei ist das Pendeln dann nur bedingt Ergebnis einer rationalen Abwägung, sondern mehr Ausdruck der Notwendigkeit der Anpassung an veränderte Lebensumstände. Entsprechend verschieben sich abhängig von der persönlichen Situation die Präferenzen: Studierende beispielsweise werden im Regelfall preissensibler sein als Berufstätige.

Abseits der Frage, in welchem Anteil die Bereitschaft, weitere Strecken zu pendeln, tatsächlich durch den Preis der Angebote im SPFV determiniert wird (also preiselastisch ist), ist über alle Pendlergruppen grundsätzlich von einer mittleren Preiselastizität auszugehen; der genaue Anteil ist im Rahmen dieser Untersuchung nicht von überragender Bedeutung, es ist jedoch festzuhalten, dass sich die Preiselastizität von Fernpendlern von den Elastizitäten der anderen Zielgruppen unterscheidet: während Geschäftsreisende im Regelfall weniger preissensibel sind, sind Privatreisende mit dem Reisezweck „Freizeit“ als preissensibler einzustufen. Für Fernpendler scheinen daher andere Faktoren wie der Reisekomfort und insbesondere die Reisezeit im Vordergrund zu stehen.

3.3.4 Zusammenfassung der Ergebnisse

Die Gruppe der Geschäftsreisenden verhält sich im Sinne des Modells homogen: Mit hoher Preisbereitschaft, aber auch einer hohen Leistungserwartung und damit einhergehend hohen Ansprüchen an den Fahrt- und Umsteigekomfort stellen sie eine besonders anspruchsvolle Zielgruppe dar. Gleichzeitig können Eisenbahnverkehrsunternehmen mit Geschäftsreisenden besonders hohe Deckungsbeiträge erzielen, wenn die hohe Preisbereitschaft dieser Zielgruppe spezifisch adressiert wird (beispielsweise mit einer Produkt- oder zeitlichen Preisdifferenzierung).

Privatreisende werden zuerst in zwei separat zu betrachtende Teilgruppen aufgeteilt: Urlaubsreisende und Pendler. Urlaubsreisende sind dabei nach der Reisedauer zu gliedern: Langzeiturlaubende haben vorrangig Komfortansprüche und bevorzugen aufgrund der Mitführung von Reisegepäck umsteigearme Verbindungen, Kurzzeiturlaubende nutzen die Urlaubszeit möglichst optimal und haben daher eine stärkere Präferenz für die zeitliche Verfügbarkeit und die Dauer der Reise als auf die weiteren Parameter. Pendler sind dann vorrangig unter dem Gesichtspunkt der Verbindungsdauer und weniger stark mit ihrer Preiselastizität zu betrachten, da Pendler bereit sind, auf ein gewisses Maß an Verbindungskomfort (wie die Fahrt in der 1. Klasse) zu verzichten, um den Preis der Fahrkarte zu reduzieren.

Als wesentliche Parameter der Entscheidungsfindung für alle Zielgruppen ergeben sich zusammengefasst

- die Dauer der Reise,
- der Preis der Reise,
- der Komfort des Reiseverkehrsmittels sowie
- der Umsteigekomfort der Verbindung.

4 Modellierung der Passagiere

Zur Modellierung der Passagiere werden nun die zur Verbindungssuche und -auswahl notwendigen Präferenzen monetarisiert. Zur Verbesserung der Performance des Modells wird dann folgend eine algorithmische Hilfsvariable erläutert und diskutiert, bevor anschließend noch die Grenzen des Modells dargestellt werden, damit danach in einem weiteren Kapitel die Berechnungsmethodik vorgestellt werden kann.

4.1 Monetarisierung der Auswahlpräferenzen

Um die Auswahlpräferenzen, die im vorangegangenen Kapitel 3.3 zielgruppenspezifisch erörtert wurden, modellieren zu können, wird folgend die jeweilige Präferenz monetarisiert; Hierbei wird bei den einzelnen Zielgruppen auch die Priorität und Gewichtung der Eigenschaften untereinander als Monetarisierungskriterium genutzt, obwohl die Reihenfolge der Eigenschaft exakt betrachtet lediglich ein ordinales und kein kardinales Merkmal darstellt. Die Monetarisierungswerte in diesem Kapitel sind dabei die Basis des Modells zur Auswahl von Verbindungen; Insoweit haben sie einen Einfluss auf die entstehenden und für die Zielgruppen ausgewählten Routen, sind jedoch nicht kritisch im Sinne des Modells, da im Rahmen dieses Modells lediglich die Menge der Verbindungen, nicht jedoch die Menge der auf den Verbindungen verkehrenden Passagiere bestimmt wird. Entsprechend ist die Validität der ermittelten Daten auch von geringerer Relevanz als dies bei einem Passagierzuzuordnungsmodell der Fall wäre.

Die zu monetarisierenden Präferenzen wurden hierbei in Kapitel 3.2 (Nutzentheorie) festgelegt: es handelt sich um die Kosten der Reisezeit, die Kosten des Komforts (aufgeteilt in Umsteigekosten und die Kosten des Fahrtkomforts) sowie die Preiswahrnehmung. Die Preiswahrnehmung wird dabei nicht separat betrachtet, da die Referenzierung der anderen drei Präferenzen eine Gewichtung der Preiswahrnehmung überflüssig macht, weil sich diese aus der Gewichtung der anderen Präferenzen automatisch als „übrige Größe“ ergibt. An einem Beispiel illustriert ergibt sich bei der Betrachtung der Zielgruppe der Geschäftsreisenden, dass deren hohe Zeitsensibilität im Verhältnis hohe Zeitkosten verursacht. Die Wahrnehmung der Fahrtkosten ist demgegenüber also niedriger als bei Gruppen mit einer niedrigen Zeitsensibilität, die bereits dadurch automatisch preissensibler werden, da das Gesamtsystem – und damit die Summe der Präferenzen – unverändert bleibt.

4.1.1 Zeitkosten

Im Folgenden wird nun zuerst die Monetarisierung der Reisezeit für alle abgegrenzten Zielgruppen vorgenommen. Grundsätzlich stellt sich hierbei die Frage nach der generellen monetären Bewertung von Reisezeit; Bei Geschäftsreisen mag für die Firma noch ein Anreiz bestehen, die Reisezeit zu minimieren, beziehungsweise die Auswirkungen der Reisezeit ökonomisch zu minimieren: Wenn mit dem durch die geringere Reisezeit eingesparten Gehalt der höhere Fahrpreis der schnelleren Reise überkompensiert wird, ist die Wahl des schnelleren Verkehrsmittels direkt ökonomisch vorteilhaft. Bei diesem Ansatz werden jedoch weitere Aspekte wie die Zufriedenheit des Mitarbeiters und die weiteren impliziten Bewertungen (mithin der oft höhere Komfort eines schnelleren Verkehrsmittels) nicht berücksichtigt.

Es erscheint jedoch zielführend, Analysen zum *value of time* zu verwenden, da der value of time als Wert der Reisezeit Ausdruck der möglichen Kosten der Reisezeit ist und daraus die Möglichkeit bietet, zielgruppenspezifische Reisezeitkosten zu bestimmen. Gleichzeitig ist eine breite Datenverfügbarkeit gegeben, was unabdingbare Voraussetzung zur Diskussion verschiedener Modellansätze ist.

In einer Handlungsempfehlung von 2011 gibt das U.S. Department of Transportation dabei mögliche Spannen an, hier aufgezinst und umgerechnet:

Reisezweck	Intercity-Verkehr		Hochgeschwindigkeitsverkehr	
	von	bis	von	bis
Privatreise	14,28 €/h	21,48 €/h	27,37 €/h	40,95 €/h
Geschäftsreise	18,28 €/h	27,47 €/h	45,75 €/h	68,52 €/h
alle Reisen	15,18 €/h	22,77 €/h	34,76 €/h	52,14 €/h

Tabelle 4: Darstellung des value of time für verschiedene Reisezwecke und Verkehrsarten
Quelle: Eigene Darstellung (Umrechnung, Aufzinsung)
auf Basis von [Bel11, S.23f.]

Bei den in Tabelle 4 dargestellten Werten ist zur Verrechnung der Werte in der Zeile 'alle Reisen' zu bemerken, dass die Zusammensetzung differiert: im Intercity-Verkehr sind nach der Annahme von TROTTENBERG und BELENKY 78,6 % des Verkehrs Privatreisen und 21,4 % Geschäftsreisen, im Hochgeschwindigkeitsverkehr wird von 59,6 % Privatreisen und 40,4 % Geschäftsreisen ausgegangen, mithin einem fast doppelt so hohen Wert der Geschäftsreisen. Entsprechend ist auch der value of time aller Rei-

sen im Hochgeschwindigkeitsverkehr in Tabelle 4 überproportional höher als beim Intercity-Verkehr.[Bel11]

Die in Tabelle 4 abgetragenen Werte sind jedoch deswegen von besonderem Interesse für das Modell dieser Arbeit, weil sie sich lediglich auf das Gehalt des Reisenden beziehen; Mithin ist die Aussage von TROTTENBERG und BELENKY, dass Privatreisende im Hochgeschwindigkeitsverkehr durchschnittlich 90 % mehr Gehalt erzielen als Reisende im Intercity-Verkehr (im Geschäftsreisendenverkehr beträgt die Differenz sogar 150 %). Die Werte beziehen sich also nicht auf eine persönliche Wahrnehmung des Reisenden oder erfahren eine steuernde Einwirkung durch Präferenzmodelle, sondern basieren rein rational auf dem tatsächlichen Gehalt, das während einer Reise Stunde erzielt werden könnte (TROTTENBERG und BELENKY lassen hierbei im Übrigen außer Acht, dass die Reisenden auch während der Fahrt arbeiten könnten, das ist jedoch bedingungsgemäß nicht Ziel der Schlussfolgerungen der Autoren).[Bel11, S.23f.]

Dabei ist fraglich, ob Privatreisende bereit sind, für einen schnelleren Transport einen höheren Fahrpreis in Höhe ihres durchschnittlichen Gehalts zu bezahlen, gleichwohl wird eine Präferenz von Gutverdienern auch von TROTTENBERG und BELENKY angenommen. Entsprechend sind diese Werte nicht direkt zu verwenden, geben aber bereits Richtlinien vor, in welchem Rahmen sich die möglichen Werte befinden können. TROTTENBERG und BELENKY geben außerdem einen Überblick über Werte von Nahpendlern und bemerken, dass die meisten Länder für diese Pendler einen Wert zwischen 25 % und 42 % des durchschnittlichen Einkommens annehmen. Die Werte betrachteter Untersuchungen bewegen sich (auf den Bearbeitungsstand der Arbeit aufgezinst und umgerechnet) zwischen 6,76 € und 18,52 € pro Stunde.[Bel11, S.11f.]

Weitere Erkenntnisse der von TROTTENBERG und BELENKY untersuchten Studien beziehen sich auf den value of time unter verschiedenen Umständen: in einigen der Studien wird hervorgehoben, dass das Warten auf das Verkehrsmittel besonders unangenehm ist und aus diesem Grund die Zeitkosten des Wartens gegenüber dem regulären value of time mit Aufschlägen zwischen 50 % und 150 % versehen werden.[Bel11, S.11f.]

WARDMAN ET AL. halten in [Str01] zu den Wartezeitkosten fest, dass die Empfindung der Wartezeit neben demografischen Daten auch von der Zielgruppe des Wartenden abhängt und sich grob im Rahmen von etwa der doppelten Zeitwahrnehmung gegenüber normaler Fahrzeit bemisst. Gerade Pendler bemessen WARDMAN ET AL. jedoch mit einem um 29 % höheren Wert der Wartezeit.[Str01, S.64f.]

COLE stellt in [Col05] zur Bemessung der Zeitkosten zuerst fest, dass verschiedene Kundengruppen auch verschiedene Zahlungsbereitschaften (und damit auch verschiedene Zeitkosten) besitzen; Er unterscheidet dazu in Arbeitszeit und Nicht-Arbeitszeit, also Fahrten, die keine Geschäftsreisen sind oder nicht während der Arbeitszeit stattfinden. In beiden Fällen werden hierbei – da seine Beispiele auf den Autoverkehr referenzieren – verschiedene Besetzungszahlen und Fahrzeuggrößen angenommen. Bezogen auf den Fahrer und damit normiert auf einen Nutzer erreicht er Werte nach Aufzinsung und Umrechnung von 20,01 € pro Stunde für Geschäftsreisende und 4,89 € pro Stunde für Privatreisende. [Col05, S.317ff.]

Interessant und auf den ersten Blick widersprüchlich ist hierbei die große Differenz zu den Ergebnissen von TROTTEBERG und BELENKY, insbesondere aufgrund der von ihnen durchgeführten breiten Literaturanalyse. Bei genauerer Betrachtung stellt sich jedoch heraus, dass sich zwar die Literaturanalyse von TROTTEBERG und BELENKY auf Durchschnittswerte bezieht, die Verrechnung der Werte auf Geschäfts- und Privatreisende hingegen auf deren eigens entwickelter Systematik basiert. Entsprechend weichen die Ergebnisse von COLE ab, da er – wohl auch aufgrund eines kleineren Untersuchungsraums – von anderen Prozentzahlen der tatsächlichen Ersparnis für Privatreisende vom Durchschnittslohn ausgeht, als dies bei den Langstreckenverkehre analysierenden Untersuchungen von TROTTEBERG und BELENKY der Fall war.[Col05, S.317ff.][Bel11, S.4ff.]

NASH ET AL. wählen einen anderen Ansatz: Sie unterscheiden nach der Reisemotivation des Fahrgasts als Grundlage der Höhe des value of time und machen sich dabei zunutze, dass die Präferenzen dabei je nach Reisemotivation des Fahrgasts differieren: in ihrem Modell ergeben sich (inflationbereinigt und aufgezinst) Werte zwischen 17,52 € und 26,57 €. Dabei unterscheiden sich die Werte je nach Motivation um etwa 30 %, was jedoch auch bei den anderen vorgestellten Studien (dort im Rahmen von Korridorwerten ohne explizite Motivationsbindung) angenommen wurde. Mithin fügen sich die Werte von NASH ET AL. also in die bereits früher zitierten Wertebereiche ein und untermauern damit die Verlässlichkeit der bereits vorgestellten weiteren Studien.[Nas86]

Bei der Interpretation der vorgestellten Ergebnisse anderer Autoren ist zuerst zu bedenken, dass die Auftraggeber bzw. intendierte Nutzer der Studien in der Regel Staaten oder staatliche Organisationen sind, die Ergebnisse mithin eher einen volkswirtschaftlichen Hintergrund besitzen als den Anspruch an einen wissenschaftlich korrekten Ausdruck eines value of time einer oder mehrerer Zielgruppen. Dafür spricht auch, dass die Werte im Regelfall in den Studien gemittelt und dann nach Verkehrsart aufgeteilt

wurden, jedoch keine zielgruppenscharfe Zuordnung stattfand. Wenn eine Zuordnung stattfand, wurde sie nach Art der Reisenden vorgenommen, nicht nach Reisezweck oder -motivation. Es entsteht daher ein anderes – im Vergleich zur Nutzung in dieser Arbeit verzerrtes – Bild der tatsächlichen Verkehrswiderstände, die in den vorangegangenen Untersuchungen zusätzlich durch weitere Merkmale wie den Komfort des Verkehrsmittels überlagert wurden, ohne diese Überlagerungen zu quantifizieren oder anderweitig separat auszuweisen.

Des Weiteren ist eine Interpretation der dargestellten Resultate dahingehend notwendig, als der in den Studien regelmäßig vorgestellte Fall von Reisenden nicht nur den value of time betrachtet, sondern die kombinierten Zeitkosten mit inkludierter Verrechnung anderer Bezugsgrößen wie dem Reisekomfort, und außerdem die Bezugsgrundlage des Vergleichs der Studien unklar ist. Wenn Reisende zum Beispiel alternativ mit dem Auto fahren müssten, wäre der Komfort als Selbstfahrer vermutlich geringer und als Mitfahrer vielleicht höher, abhängig auch von den persönlichen Präferenzen.

Die Überlagerung mit weiteren Präferenzen und insbesondere die unternehmensbezogene Beeinflussung der Kundenwahrnehmung durch die Unternehmensreputation, Kundenbindungsprogramme und deren Annehmlichkeiten (beispielsweise Zugang zu speziellen Aufenthaltsbereichen an der Start- oder Zielstation für die Reisenden) oder weitere Komfortmerkmale wie ein Bordunterhaltungsprogramm, das entsprechend die Zeitwahrnehmung der Reisenden beeinflusst, sind zu berücksichtigen. Es ist daher unverzichtbar, die Werte der Studien für das Modell anzupassen. Die Werte der Umsteigewartezeitkosten müssen im Übrigen im Kontext mit dem Umsteigewiderstand betrachtet werden, der zusätzlich zu den Wartekosten entsteht, und im Folgekapitel 4.1.2 behandelt wird.

Um für das Modell entsprechend einfach verarbeit- und handhabbare Werte zu erhalten, werden die Werte passend justiert. Die konkrete Ausgestaltung der Werte ist für ein stimmiges Gesamtbild nicht von überragender Bedeutung, entscheidender ist die Stimmigkeit des Gefüges, also der Zahlen untereinander. Zuerst wird der Wert des 'value of time' für Geschäftsreisende bemessen: Dieser bewegt sich den Studien zufolge bei etwa 25–30 € pro Stunde. Weitere Kosten, welche die Studie von TROTTEBERG und BELENKY anführt, werden im Folgenden in den anderen Kategorien (Umsteigekomfort, Komfort) berücksichtigt.[Bel11] Der Umsteigezuschlag wird dazu mit einem Drittel angenommen, die Werte werden aus Modellierungsgründen im Übrigen grundsätzlich auf die nächste ganze Zahl gerundet.

Im Folgenden werden daher im Modell für Geschäftsreisende Zeitkosten von 30,00 € pro Stunde und Umsteige-Wartezeitkosten von 40,00 € pro Stunde angenommen. Für tägliche Fernpendler werden aufgrund des werktäglichen Pendelns tendenziell ähnliche, aber in Summe etwas geringere Zeitkosten angesetzt, zur besonderen Gewichtung des Komfortempfindens der Umstieg jedoch prozentual verteuert: Für die Zeitkosten werden 20,00 € pro Stunde angenommen, für die Umsteige-Wartezeitkosten 30,00 € pro Stunde.

Für die Gruppen der Privatreisenden inklusive der Wochenendpendler ergeben sich aufgrund des veränderten Präferenzspektrums niedrigere Werte.[Rüg05, S.18ff.] Zuerst werden die Wochenendpendler behandelt: Ihre Präferenzen ähneln denen der täglichen Fernpendler, sie sind jedoch deutlich preissensibler. Entsprechend sind die Zeitkosten der Wochenendpendler deutlich geringer als die der werktäglichen Pendler und werden mit 12,00 € pro Stunde angenommen; Die Umsteige-Wartezeitkosten sind ebenfalls geringer und betragen 15,00 € pro Stunde. Für Urlaubsreisende unterscheiden sich die Werte besonders der Höhe nach: Für Kurzaurlauber ergeben sich 16,00 € pro Stunde als reguläre Zeitkosten und 20,00 € pro Stunde als Umsteige-Wartezeitkosten, für Langzeiturlauber 10,00 € pro Stunde und 13,00 € pro Stunde als Umsteige-Wartezeitkosten.

Zusammengefasst ergeben sich folgende Werte:

Zielgruppe	Zeitkosten (Fahrt)	Umsteige-Wartezeitkosten
Geschäftsreisende	30,00 €/h	40,00 €/h
Fernpendler (täglich)	20,00 €/h	30,00 €/h
Wochenend-Fernpendler	12,00 €/h	15,00 €/h
Kurzaurlauber	16,00 €/h	20,00 €/h
Langurlauber	10,00 €/h	13,00 €/h

Tabelle 5: Zeitkosten der verschiedenen Zielgruppen im Modell

Quelle: Eigene Darstellung

4.1.2 Umsteigekosten

Die Monetarisierung des Umsteigewiderstands ist im Gegensatz zu den Zeitkosten nicht direkt quantifizierbar, sondern hängt in hohem Maße nicht nur von der Zielgruppe, sondern auch von den Begleitumständen ab. GEURS ET AL. haben in einer Studie für die niederländische Staatsbahn (Nederlandse Spoorwegen) verschiedene Ansätze zusammengeführt, insbesondere beim Vergleich mehrerer ähnlicher Situationen. Dabei hatten sich die Präferenzen der über 700 teilnehmenden Kunden als schwer einschätzbar gestaltet, insbesondere bei der Wahl der Kriterien zeigt sich schon die Diversität der

Anliegen von Wartenden: Neben der Wartezeit und den Warteumständen (bezogen auf die Ausstattung der Station) spielen auch die Zuverlässigkeit und Frequenz der beteiligten Züge sowie die Möglichkeit bahnsteiggleicher Transfers und die tatsächlichen Umstände der Umsteigenden (beispielsweise die Menge an Gepäck) eine Rolle. GEURS ET AL. haben als „Zeitstrafe“ eine durchschnittliche Zeit von 23 Minuten ermittelt (eine Direktverbindung mit 23 Minuten höherer Fahrzeit wäre einer Umsteigeverbindung bezogen auf die Fahrgastpräferenz damit ebenbürtig). [Kei12, S.2ff.]

GEURS ET AL. stellen in ihrer Untersuchung weiterhin fest, dass der Umsteigewiderstand jedes weiteren Umstiegs höher ist als die des vorangegangenen Umstiegs, daher muss insbesondere davon ausgegangen werden, dass Kunden mehrfache Umstiege mit größerem Interesse vermeiden wollen. Gleichzeitig können Umstiege innerhalb von Stadtverkehrsmitteln oder zwischen Stadt- und Fernverkehrsmitteln das Empfinden verändern, da diese anders als Umstiege innerhalb der Fernverkehrsreisekette weder von der Zuverlässigkeit und Frequenz der Züge noch von ihrer Komplexität bezogen auf das Verkehrsmittel und den Umsteigevorgang miteinander vergleichbar sind. Mithin muss daher darauf geachtet werden, im Rahmen des Modells dieser Ausarbeitung einen Kompromiss zu finden zwischen der zunehmenden Aversion der Reisenden gegenüber einer steigenden Umsteigehäufigkeit und der Möglichkeit innerhalb des Modells, solche Zwänge und insbesondere die Zusammenhänge innerhalb von Stadtverkehrsmitteln abzubilden. [Kei12, S.2ff.]

WARDMAN ET AL. halten dann auch in ihrer Studie fest, dass der Umsteigewiderstand neben anderen Kriterien wie dem Geschlecht und der Zielgruppenzugehörigkeit der Reisenden auch mit der Reishäufigkeit der Kunden zusammenhängt: Wer weniger als einmal wöchentlich fährt, besitzt der Aussage von WARDMAN ET AL. zufolge einen um 85 % höheren Umsteigewiderstand. [Str01, S.64f.]

Es erscheint nicht trivial lösbar, dieses Problem umfänglich im Rahmen dieser Ausarbeitung zu behandeln, da verschiedene Einflussfaktoren nicht oder nicht direkt durch die Verbindungswahl beeinflusst werden. Dadurch, dass in den meisten Studien statt eines direkt monetarisierten Umsteigewiderstands die Zeitkosten (durch die gefühlte Reisezeitverlängerung aufgrund des Umstiegs) für die Darstellung des durch einen Umstieg entstehenden Widerstands genutzt wurden, ist der durch den Umstieg selbst (und nicht die Wartezeit) bedingte Malus nicht sauber trennbar, sondern in die Zeitkosten integriert. Um für das Modell dieser Arbeit die Trennung der Effekte zu erhalten, wurde bereits im vorangegangenen Kapitel eine eher konservative Annahme der Umsteige-Wartezeitkosten getroffen, und an dieser Stelle werden die weiteren, noch nicht berücksichtigten, Kos-

ten als Umsteigekosten hinzugefügt. Um der dargestellten besonderen Aversion von Geschäftsreisenden, täglichen Fernpendlern und Langzeiturlaubenden gegenüber Umstiegen Geltung zu verleihen, sollen für diese Zielgruppen die Umsteigekosten höher sein als für die anderen Zielgruppen.

Die reinen Umsteigekosten ohne Berücksichtigung von Umsteigewartekosten berechnen sich im Folgenden zielgruppenweise aus den Umsteigezeitkosten des Modells nach GEURS ET AL. (in der Tabelle als „Zeitkosten“ bezeichnet) und wurden nach den dargestellten Kriterien (beispielsweise der besonderen Umsteige hemmung von Langzeiturlaubern) dann zielgruppenspezifisch erhöht und aus Handhabbarkeitsgründen gerundet:

Zielgruppe	Zeitkosten	Umsteigekosten
Geschäftsreisende	11,50 €	14,00 €
Fernpendler (täglich)	7,67 €	12,00 €
Wochenend-Fernpendler	4,60 €	6,00 €
Kurzzeiturlauber	6,13 €	6,50 €
Langzeiturlauber	4,98 €	14,00 €

Tabelle 6: Umsteigekosten und Kosten der Umsteige-Wartezeit für den ersten Umstieg
Quelle: Eigene Darstellung

Die in Tabelle 6 dargestellten Kosten sind für den ersten und den zweiten Umstieg anzusetzen, ab dem dritten Umstieg werden die Kosten dann verdoppelt, um den bereits dargestellten erhöhten Umsteigewiderstand nachzubilden. Das erscheint auch ein gangbarer Kompromiss für Vor- und Nachläufe des Stadtschnellbahnverkehrs: handelt es sich lediglich um eine Direktverbindung im Vor- und Nachlauf, steigen die Kosten nicht überproportional (insbesondere nicht im Vergleich der Kosten untereinander); Handelt es sich um komplexere Verbindungen, erscheint der Abzug höherer Kosten gerechtfertigt.

4.1.3 Komfortkosten

Die monetarisierte Komfortwahrnehmung bildet im Sinne der Widerstandstheorie den negativen Komfort ab und dient entsprechend dazu, das Fehlen von Komfort auszudrücken. Daher bekommt die höchste Komfortstufe auch keine Komfortkosten zugewiesen, denn sie entspricht dem erreichbaren Maximum und damit genau keinen (zusätzlichen) Komfortkosten. Die Komfortkosten werden für die jeweiligen Zielgruppen ansonsten über Faktoren gemäß des Stellenwerts definiert, den Komfort für die jeweilige

Zielgruppe besitzt; Mithin ist also wieder dem Umstand Rechnung zu tragen, dass die verschiedenen Zielgruppen unterschiedliche Ausprägungen der Präferenzen besitzen.

Als Bezugsgröße zur Monetarisierung der Komfortwerte wird die Zeit verwendet. Es wäre natürlich auch möglich, beispielsweise die zurückgelegte Distanz in einem Verkehrsmittel als Indikator der Komferteinschränkung zu verwenden, das menschliche Empfinden richtet sich aber nach der verbrachten Zeit, und nicht nach der Distanz (die der Fahrgast nicht unmittelbar erlebt). Dieses Berechnungsverfahren führt natürlich auch dazu, dass unkomfortablere Verkehrsmittel, die oftmals auch langsamer verkehren, einen zusätzlichen Malus erleiden; diesem Fakt muss dann bei der konkreten Ausgestaltung der Werte Rechnung getragen werden.

Da dem Autor keine empirischen Referenzen von Komfortkosten bekannt sind, soll an dieser Stelle als Methodik die real existierende Bepreisung als Referenz herangezogen werden. Dazu wurden Streckenprofile verwendet, bei denen die Zuggattungen möglichst ähnliche Fahrzeiten besitzen, sodass sich die für den Fahrgast relevanten Unterschiede zwischen den Zuggattungen, sofern möglich, nicht auf die Reisegeschwindigkeit beziehen, sondern ganz oder überwiegend auf den Komfort des Verkehrsmittels. Dazu werden drei Beispiele mit Fahrten auf Basis der Flexpreise der Deutschen Bahn AG angeführt, im Folgenden beginnend mit den Preisen und Eigenschaften von Fahrten auf der Strecke Stuttgart – Ulm:

Zuggattung	Fahrzeit	Preis	Komfortkosten
ICE	0:54 h	26,00 €	0,00 €/h
EC	0:57 h	22,50 €	3,68 €/h
IRE	1:00 h	20,20 €	5,80 €/h

Tabelle 7: Flexpreise der Deutschen Bahn für die Strecke Stuttgart – Ulm
Quelle: [DB 16c], abgerufen am 05.10.2016 für „Stuttgart Hbf nach Ulm Hbf, am 05.10.2016, ab 12:00 Uhr, ohne weitere Präferenzen“

Tabelle 7 können die Preise und Dauern von Fahrten verschiedener Zuggattungen auf der Strecke Stuttgart – Ulm entnommen werden. Dabei ist zuerst zu beachten, dass Züge der Zuggattungen EC und IRE etwas länger benötigen als die Züge der Zuggattung ICE, der Unterschied zwischen den Zuggattungen beträgt jedoch jeweils nur drei Minuten und erscheint damit vernachlässigbar gering. Der entstehende Kostenvorteil, der von der Deutschen Bahn AG für den Komfortverzicht gewährt wird, beträgt im EC 3,50 € und im IRE 5,80 € für die gesamte Strecke, mithin ergibt sich ein Stundenkostensatz von 3,68 €/h im EC und 5,80 €/h im IRE. Es sei jetzt dahingestellt, inwieweit die

Entscheidung, auf dieser Strecke einen ICE zu benutzen, tatsächlich nur von einer Komfort-Kosten-Abwägung geprägt ist; Auch die Vor- und Nachläufe der Züge sowie die Möglichkeit von Anschlüssen entlang der Strecke bzw. die generelle Fahrlage können bei der Entscheidungsfindung der Fahrgäste eine Rolle spielen.

Als weiteres Beispiel wird nun die Strecke Mannheim – Saarbrücken angeführt:

Zuggattung	Fahrzeit	Preis	Komfortkosten
ICE	1:20 h	34,00 €	0,00 €/h
IC	1:31 h	30,00 €	2,63 €/h
RE	1:32 h	27,30 €	4,37 €/h

Tabelle 8: Flexpreise der Deutschen Bahn für die Strecke Mannheim – Saarbrücken
Quelle: [DB 16c], abgerufen am 05.10.2016 für „Mannheim Hbf nach Saarbrücken Hbf, am 05.10.2016, ab 15:00 Uhr, ohne weitere Präferenzen“

Wie Tabelle 8 entnommen werden kann, sind auf der Strecke Mannheim – Saarbrücken die Züge der Zuggattung IC trotz einer ähnlich langen Fahrt wie die Züge der Zuggattung RE noch 2,70 € teurer. Dies stellt mittelbar (neben den bereits erwähnten weiteren Faktoren wie Durchbindungen) den Ausgleich für den Komfortwiderstand dar. Die Züge der Zuggattung ICE sind 11 und 12 Minuten schneller als die Züge der anderen beiden Zuggattungen, sodass der Aufpreis hier auch etwas mit einem Zeitgewinn zu tun haben könnte; Da die Preissteigerung gegenüber der letzten verglichenen Strecke Stuttgart – Ulm jedoch proportional niedriger ist (ausgedrückt im niedrigeren Komfortkosten-Stundensatz), hat dieser Fakt vermutlich keinen Eingang in die Preisgestaltung auf der Strecke gefunden.

Schließlich werden noch Verbindungen auf der Strecke Berlin – Stralsund betrachtet und verglichen:

Zuggattung	Fahrzeit	Preis	Komfortkosten
ICE	2:50 h	56,00 €	0,00 €/h
IC	2:41 h	53,00 €	1,12 €/h
RE	3:06 h	44,00 €	3,87 €/h

Tabelle 9: Flexpreise der Deutschen Bahn für die Strecke Berlin – Stralsund
Quelle: [DB 16c], abgerufen am 05.10.2016 für „Berlin Hbf nach Stralsund Hbf, am 07.10.2016, ab 15:00 Uhr, ohne weitere Präferenzen“

Bei der Durchsicht der Daten aus Tabelle 9 fällt auf, dass die Züge der Zuggattung IC schneller sind als jene der höheren Zuggattung ICE. Das mag den verhältnismäßig geringen Preisaufschlag der an sich weniger komfortablen Zuggattung IC auf dieser Strecke erklären, kann im Einzelfall aber natürlich auch mit der konkreten Fahrlage der

Zuggattungen, eventuell weiteren beeinflussenden Zügen sowie einer unterschiedlichen Haltepolitik zusammenhängen. Der Preisaufschlag des Regionalverkehrs fügt sich in die Werte der vorangegangenen Erhebungen aus Tabelle 7 und Tabelle 8 ein, sodass dieser Wert konsistent erscheint.

Zusammengefasst lässt sich anhand der zusammengetragenen Streckendaten plausibel schlussfolgern, dass die Deutsche Bahn AG eine Art Komfortzuschlag für hochwertige Züge erhebt, der sich in den Beispielen mit etwa 2,50 €/h für IC-Züge und etwa 5,00 €/h für Züge des Regionalverkehrs bemisst. Weitere Unterschiede innerhalb des Regionalverkehrs sowie der durch die Deutsche Bahn AG nicht mehr verwendeten Zuggattung InterRegio (die im Modell als günstiges Bindeglied zwischen Fern- und Nahverkehr integriert wurde) werden jetzt nur innerhalb des Algorithmus dieser Arbeit definiert. Für InterRegio-Züge werden hierbei Werte von 4,00 €/h angenommen, entsprechend der intendierten Positionierung dieser Züge als günstiges Bindeglied zwischen Regional- und Fernverkehr und des tatsächlichen Ausstattungskomforts.[Bod05, S.15ff.]

Im Regionalverkehr wird der bereits ermittelte Wert von 5,00 €/h für InterRegioExpress (IRE)-Züge genutzt; Im Falle der Untersuchungen aus Tabelle 8 und Tabelle 9 wurde der RE dargestellt, da jeweils kein IRE auf den Gesamtstrecken verkehrt, sodass der RE bezogen auf den Verkehr zwischen den ausgewählten Bahnhöfen das hochwertigste und damit komfortabelste Verkehrsmittel im Regionalverkehr darstellt. Da IRE-Züge auch weniger Halte (und damit weniger Störungen im Fahrtverlauf) als die anderen Zuggattungen im Regionalverkehr aufweisen, wurde für die weiteren Zuggattungen entsprechend noch etwas höhere Komfortkosten angesetzt, nämlich 5,50 €/h für Regional-Express-Züge und 6,00 €/h für Regionalbahnen und S-Bahnen. Bezogen auf die weitere Verwendung der Daten auf die Spezifika dieses Modells sind nun noch die zielgruppenspezifische Faktoren (und damit die tatsächliche Auswirkung) dieser Werte zu bestimmen.

Dabei ist plausibel anzunehmen, dass Geschäftsreisende tendenziell dem vorgestellten Modellansatz insoweit entsprechen, Flexpreise zur Reise zu verwenden, und ein hohes Komfortbedürfnis und damit entsprechend eine hohe Komfortzahlungsbereitschaft zu besitzen. Da im Rahmen dieses Modellansatzes keine Trennung in die verschiedenen Klassen im Zug erfolgt, wird dieser Tatsache Rechnung getragen, indem die ermittelten Stundenkostensätze in doppelter Höhe zum Ansatz kommen. Wie in Kapitel 3.3 eruiert, besitzen alle weiteren Zielgruppen eine erheblich geringere Zahlungsbereitschaft für Reisekomfort, sodass für alle anderen Gruppen pauschal ein Ansatz der ermittelten Stundenkostensätze erfolgt.

4.2 Auswahlgrundlage virtueller Preis

Wie bereits erläutert wurde, sind die Verkehrswiderstände, die im Rahmen des Modells dieser Arbeit verwendet werden, zielgruppenspezifische Werte auf Basis der Präferenzen der einzelnen Fahrgastgruppen. Die Definition eines gemeinsamen Kanons an mindestens zu erfüllenden Kriterien ermöglicht dabei, die Nachfrager für einzelne Teile des Modells – im Wesentlichen bei der Verbindungserzeugung – als eine homogene Masse zu behandeln, innerhalb derer spezielle Präferenzen durch separate algorithmische Behandlung berücksichtigt werden können, jedoch ein gemeinsames geringstes Widerstandslevel die einheitliche Verbindungserzeugung ermöglicht. Auf Basis der so gebildeten Verbindungen kann dann die Routenauswahl unter Berücksichtigung der Präferenzen der jeweiligen Zielgruppe erfolgen.

Der *virtuelle Preis* stellt dabei nun eine Art Vergleichswährung dar, die direkt einen (monetarisierten) Mindeststandard der Verbindung ermittelt, was notwendig ist, um die später erfolgenden Vergleiche zu vereinfachen: Zum einen sollten aus datenbanktechnischen Gründen die Vergleiche auf möglichst wenigen Kriterien basieren, sodass die Zusammenfassung vieler Faktoren (Dauer, Preis und Komfort) zu einem einzigen Kriterium die Performance deutlich erhöht; Zum anderen wird nur so die gewünschte zielgruppenneutrale Verbindungsauswahl möglich, ansonsten müsste für jede Zielgruppe eine separate Betrachtung auf die Vorteilhaftigkeit der betrachteten Verbindungen erfolgen, und schließlich alle diejenigen Verbindungen behalten werden, die für eine beliebige Zielgruppe vorteilhaft sind. Das führte schließlich zum selben Ergebnis wie die Rechnung mit dem virtuellen Preis, wäre jedoch mit deutlich mehr Rechenschritten verbunden, und wurde aus diesem Grund im Modell dieser Ausarbeitung vermieden.

Die Bildung des virtuellen Preises muss daher so vorgenommen werden, dass keine optimale Lösung einer Zielgruppe ausgeschlossen wird, sondern im Zweifel eher noch zu viele Verbindungen in der Datenbank verbleiben um dann, sofern sie für keine Zielgruppe vorteilhaft sind, bei der konkreten Routenauswahl schlicht unberücksichtigt zu bleiben, oder vorher durch andere Kriterien als unvorteilhaft bestimmt und aussortiert werden zu können.

Gleichzeitig kann aber nicht ein einziges Merkmal alle Dimensionen der verschiedenen Zielgruppen abbilden; Aus diesem Grund werden bei den später vorgestellten Algorithmen zur Verbindungsauswahl neben dem virtuellen Preis auch die Dauer der Verbindung und die Anzahl der Umstiege mit einbezogen, um die besonderen Fälle einzelner, besonders starker, Präferenzen (bezogen auf die Umsteigearmut und die besonders

geringe Dauer einzelner, ansonsten für den „Durchschnittsfahrgast“ unvorteilhafter, Verbindungen) separat zu berücksichtigen.

Der virtuelle Preis muss folglich so beschaffen sein, dass durch die Anwendung dieses Kriteriums möglichst viele unvorteilhafte Verbindungen ausgeschlossen werden und gleichzeitig keine vorteilhaften Verbindungen ausgeschlossen werden (sofern dieser Ausschluss nicht durch die separate Berücksichtigung der Umsteigearmut und der Dauer bereits abgefangen ist). Dazu sind die verschiedenen Bestandteile der Entscheidungsfindung bei der Routenauswahl separiert zu betrachten und werden danach zum virtuellen Preis zusammengeführt:

$$P_{\text{virt}} = P_{\text{Reise}} + P_{\text{Fahrzeit}} + P_{\text{Umstieg}} + P_{\text{Komfort}} \quad (5)$$

Nachfolgend werden die einzelnen Bestandteile des in Formel 5 gewählten Ansatzes zur Berechnung des virtuellen Preises separat beschrieben: Ein besonders wichtiges Element der Entscheidungsfindung ist dabei die Reisezeit. Nach der Konvertierung in Geldwerte wird im Folgenden vom Reisezeitpreisäquivalent oder dem monetarisierten Reisezeitwiderstand P_{Fahrzeit} gesprochen. Der monetarisierte Reisezeitwiderstand ist dabei bedingungsgemäß so anzusetzen, dass eine Vielzahl nicht vorteilhafter Verbindungen eliminiert werden kann und gleichzeitig für beliebige Zielgruppen unter bestimmten Bedingungen vorteilhafte Verbindungen grundsätzlich beibehalten werden; Der Reisezeitwiderstand sollte sich daher den niedrigsten Zielgruppenwerten annähern bzw. ihnen entsprechen.

Bei der Monetarisierung des Reisezeitwiderstands ist darauf zu achten, dass bei dieser Variable lediglich die Fahrzeit (genauer: die Aufenthaltszeit in den Zügen, jeweils ab dem Einstieg bis zum Ausstieg) in die Betrachtung mit eingeht, da die Umsteigezeit aufgrund der speziellen Relevanz in einem späteren Kriterium dargestellt wird.

Mithin definiert sich der virtuelle Preis nun über die minimalen Anforderungen der einzelnen Zielgruppen, die als Widerstände monetarisiert wurden und in Formel 6 formell aufgliedert und mit separaten Monetarisierungswerten versehen aufsummiert werden:

$$P_{\text{virt}} = \begin{cases} P_{\text{Reise}} + t_{\text{Fahrt}} \cdot 12 \frac{\text{€}}{\text{h}} + t_U \cdot 13 \frac{\text{€}}{\text{h}} + U \cdot 5\text{€} + P_{\text{Komfort}} & \text{für } U \leq 2 \\ P_{\text{Reise}} + t_{\text{Fahrt}} \cdot 12 \frac{\text{€}}{\text{h}} + t_U \cdot 13 \frac{\text{€}}{\text{h}} + (U - 1) \cdot 10\text{€} + P_{\text{Komfort}} & \text{für } U > 2 \end{cases} \quad (6)$$

In Formel 6 wird die Berechnung des virtuellen Preises als Summe des vom Fahrgast zu entrichtenden Reisepreises, der mit einem Preisfaktor von 12 € je Stunde verrechneten Fahrtzeit (sie entspricht der Reise abzüglich der Summe der Umsteigezeiten), der mit 13 € pro Stunde berechneten Umsteigezeit, dem Preis für die Anzahl der vorgenommenen Umstiege U sowie dem Komfortaufschlag für den Komfortwiderstand der einzelnen Züge vorgenommen.

Um die Anzahl der Umstiege besser berücksichtigen zu können und gleichzeitig zur Performanceoptimierung weiterhin nur lineare Modelle zur Anwendung bringen zu müssen, wurde eine Fallunterscheidung innerhalb der Formel 6 getroffen, um die besondere Aversion von Fahrgästen gegenüber häufigeren Umstiegen auszudrücken, ohne mit einer stetig linearen Formel spätere Umstiege zu unattraktiv zu gestalten und damit preissensiblen Fahrgästen die Vielfalt an Reisealternativen zu stark einzuschränken.

Die einzelnen Werte der Formel wurden empirisch am Modell ermittelt und sind auf Basis der in Kapitel 4.1 beschriebenen Monetarisierungswerte der einzelnen Präferenzen erzeugt worden.

4.3 Modelldefinition

Nach der Monetarisierung aller spezifischen Zielgruppenwerte zur Routenauswahl wird in diesem Kapitel das der Arbeit zugrundeliegende Modell und die Rahmenparameter vorgestellt. Dabei wird neben einer kurzen Betrachtung des Modellumfangs auch auf Grenzen des Modells eingegangen, die Basis für notwendige Einschränkungen bei späteren Betrachtungen sind.

4.3.1 Geographische Eingrenzung

Das Datenmodell ist geographisch auf Deutschland und das direkt angrenzende Ausland reduziert, wobei außerhalb Deutschlands nur die wesentlichen Hauptstrecken, die als Radialverbindung von und nach Deutschland genutzt werden, implementiert sind. Im Wesentlichen ist das Datenmodell also auf in Deutschland stattfindende und in Deutschland beginnende bzw. endende Fahrten bezogen.

Innerhalb des Modells werden nicht alle Verkehrsstationen, sondern im Regelfall lediglich Bahnhöfe im bahnbetrieblichen Sinne¹¹ betrachtet. Die quasi-ausschließliche Nutzung von Bahnhöfen liegt insbesondere in der Datenverfügbarkeit begründet, wäre aber auch aus Performancegründen zu empfehlen, da Umstiege in der Regel nicht an Haltepunkten stattfinden müssen, sondern auch an den nächstgelegenen Bahnhöfen vollzogen werden können.¹²

Entsprechend müssen die Fahrgastpotentiale der Verkehrsstationen angepasst werden: Zum Einen werden sonst Strecken mit vielen Haltepunkten gegenüber Strecken ohne Haltepunkte schlechter gestellt in der Berechnung (und werden damit in der streckenweisen Routenbetrachtung verfälscht), zum Anderen müssen intermodalen Verknüpfungspunkten (zum Beispiel regionalen Mittelzentren mit koordiniertem Bus-Bahn-Umstieg oder Flughäfen) entsprechende Zusatzpotentiale zugeordnet werden, weil die Wahrscheinlichkeit, dass an diesen Punkten eine intermodale Fahrt einen Umstieg, der zum Verlassen des im Modell definierten Verkehrsraums führt, stattfindet, höher ist als an anderen, „gewöhnlichen“ Bahnhöfen. Diese Betrachtungsweise führt auch dazu, dass die zugeordneten Fahrgastpotentiale sehr stark von den Einwohnerzahlen der zugeordneten Gemeinde differieren, je nach den dort zur Verfügung stehenden intermodalen Verknüpfungsmöglichkeiten.

Für ein vollständiges Passagiererzeugungsmodell wäre auch die Angabe einer zielgruppenspezifischen Relevanz der Verkehrsstation als Start- bzw. Endpunkt einer Reise nötig; Diese Einordnung kann hier entfallen, da die Fahrgastpotentiale lediglich der Optimierung des Routenfindungsalgorithmus dienen, indem Fahrten, deren Relevanz für das System nicht erkennbar ist, nicht erzeugt werden. Eine nähere Betrachtung dieser Vorgehensweise erfolgt im Kapitel 6.2 (Verbindungserzeugung).

4.3.2 Fahrten über den Nachtsprung

Um das Modell im Rahmen von Performance und Datenhandling beherrschbar zu halten, wird aufgrund der im Verhältnis geringen räumlichen Distanz und der im Verhältnis zum sonstigen Verkehr geringeren ökonomischen Relevanz stehende Nachtzugverkehr

¹¹Gemäß §4 (2) EBO sind „Bahnhöfe [...] Bahnanlagen mit mindestens einer Weiche, wo Züge beginnen, enden, ausweichen oder wenden dürfen.[...]“.[EBO16]

¹²Ausnahmen sind beispielsweise Haltepunkte, an denen Züge des höherwertigen Verkehrs halten. Dies ist in Deutschland im Jahresfahrplan 2016 beispielsweise in Jena Paradies und Hamburg Dammtor der Fall, solche Haltepunkte sind entsprechend auch in die Modellumgebung mit einbezogen worden.[DB 16c]

aus dem Datenmodell herausgenommen. Entsprechend wird im Modell ein Nachtsprung definiert, der im Einklang mit gängigen Modellen im Öffentlichen Personennahverkehr (ÖPNV) um 3 Uhr stattfindet. Alle Fahrten, die vor 3 Uhr stattfinden, gelten damit im Sinne des Modells als Fahrten des Vortags; Fahrten über 3 Uhr hinaus finden nicht statt bzw. werden aus dem Modell vor der eigentlichen Berechnung entfernt.

Dies dient insbesondere auch der Vereinfachung der Darstellung der Verkehrstage der Verbindungen innerhalb des Modells, da die Verkehrstage ansonsten bei Nachtfahrten aufwändig korrigiert werden müssten: Bei Umsteigeverbindungen müsste stets geprüft werden, ob der anschlussabgebende und der anschlussaufnehmende Zug am selben – ggf. um den Nachtsprung korrigierten – Verkehrstag verkehren; Um diese etwas aufwändige Prüfung (bzw. die ansonsten notwendige Datenvorhaltung des Verkehrstags für jede einzelne Fahrplanzeit) unterlassen zu können, wurden Fahrten, die über den Nachtsprung verkehren, technisch ausgeschlossen.

Im Rahmen der Modellierung würde eine Aufnahme von Nachtzugverbindungen eine weitere Herausforderung mit sich bringen: sowohl die Zielgruppen als auch deren individuelle Präferenzen weichen im Nachtverkehr von den festgestellten Regeltypen des Tagverkehrs ab: Besonderes Merkmal ist dabei die Verschiebung der Zeitpräferenzen, weg von der reinen Fahrzeit hin zu einer komfortbeeinflussten Ruhezeit, um während der Fahrt ausreichend Schlaf zu finden. Auch die zeitliche Verfügbarkeit spielt zielgruppenspezifisch eine bestimmende Rolle, insbesondere für den Geschäftsreiseverkehr sind zum Beispiel Ankünfte am Morgen wichtig, während im Urlauberverkehr die Ankunft später liegen kann.[Höd06, S.66ff.]

Darüber hinaus ist die Wahl des Fahrziels im Nachtreisezugverkehr von vielen extern induzierten Faktoren abhängig, insbesondere für Privatreisende beispielsweise der Verfügbarkeit von Freizeitangeboten in der betreffenden Region. Da diese Faktoren nicht in ausreichendem Maße quantifizierbar und für die Berücksichtigung der Routenanzahl darüber hinaus nicht von großer Bedeutung sind, wird im Rahmen dieser Arbeit von der Möglichkeit, Fahrten über den Nachtsprung zu betrachten, Abstand genommen.

4.3.3 Passagierverhalten im Modell

Die Passagiere innerhalb des Modells verhalten sich im Bezug auf ihre Eigenschaften über alle Zielgruppen identisch. So ist beispielsweise allen Passagieren ein Umstieg in fünf Minuten möglich, und die Passagiere sind nicht auf eine spezielle Ausrüstung

der Verkehrsstationen (zum Beispiel mit Fahrtreppen oder Aufzügen) angewiesen. Entsprechend werden diese Annahmen im Modell fest berücksichtigt, ansonsten wäre es notwendig, die Algorithmenkette entweder fallscharf auszurichten (das würde bedeuten, jeweils mehrere mögliche optimale Lösungen zu berechnen), oder den Algorithmus für jede Zielgruppe separat in der Ablaufsteuerung zu betreiben. In beiden Fällen stünde der entstehende Zusatzaufwand an Ressourcen in keinem angemessenen Verhältnis zu der entstehenden geringfügig höheren Präzision bzw. höheren Spezifität der Ergebnisse.

Auch die spezifische Wahrnehmung der Eigenschaften von Verbindungen (zum Beispiel Fahrkomfort und Serviceangebot im Zug) ist identisch, lediglich die Präferenzen und damit die monetarisierte Wertung der wahrgenommenen Eigenschaften (respektive der zugrundeliegenden Fahrtwiderstände) sind zielgruppenspezifisch.

5 Datenmodell

In der Hinführung zur algorithmischen Berechnung wird an dieser Stelle zuerst die Erzeugung der zur Verteilung von Fahrgästen notwendigen Fahrtverbindungen vorgestellt, bevor auf die Fahrtenauswahl und deren relative Häufigkeit bei der Nutzung eingegangen werden wird. Die grundsätzliche Problemstellung dieser Arbeit dreht sich um die Frage der Suche nach der optimalen Verbindung eines Fahrgasts mit gegebenem Start und Ziel. Die Routensuche dieses Modells kann also nicht dazu verwendet werden, eine Route wie „die schönste Fahrt Deutschlands“ oder „eine Rundreise durch Ost-Thüringen“ zu erzeugen; Die algorithmische Darstellung dieser Probleme ist zwar lösbar, aber nicht Bestandteil dieser Arbeit.

Zwischen Start- und Zielstation ergibt sich möglicherweise ein Tupel optimaler Verbindungen, das zuerst kumuliert erzeugt wird, bevor eine Einteilung nach der Relevanz der Verbindungen vorgenommen werden kann und vorteilhafte Verbindungen gegebenenfalls aussortiert werden. In einem späteren Schritt können dann auch jeder Zielgruppe die für sie optimalen Verbindungen zugeordnet werden und in einer weiteren Arbeit daraus auch das Aufkommen und die Verteilung der Passagiere zwischen Start und Ziel anhand der Reisewiderstände sowie des Aufkommens im Quelle-Ziel-Verkehr berechnet werden.

5.1 Grundsätzliche Erwägungen

Vor einer Klärung des konkreten technischen Setups und der dazu notwendigen Softwareumgebung wird im Folgenden die Erläuterung der Voraussetzungen des Algorithmus vorangestellt, um daraus mögliche geeignete Umgebungen zu identifizieren.

5.1.1 Voraussetzungen des Modellaufbaus

In der Literatur gilt die Verbindungssuche auf gegebenen Verkehrskanten als Vorzugsbeispiel des DIJKSTRA-Algorithmus und des Knoten-Kanten-Modells. Das Knoten-Kanten-Modell setzt hierbei auf eine Umsetzung der gesamten vorgefundenen Situation (zum Beispiel des Fahrplans sowie aller beteiligten Infrastrukturelemente) mittels Knoten und Kanten, wobei die Knoten bei einer Routensuche-Anwendung für Eisenbahnverkehre jeweils einzelne Elemente von verkehrlicher oder betrieblicher Relevanz sind (zum Beispiel Haltepunkte oder Bahnhöfe) und die Kanten als zwischen den Knoten definierte Verbindungsstücke die Strecke darstellen. Knoten und Kanten werden nun mit für

sie spezifisch relevanten Eigenschaften versehen, die Knoten zum Beispiel mit einem Namen, die Kanten zum Beispiel mit der Streckenlänge und der Fahrzeit für verschiedene Züge.[Zar02][Gol76, S.1164ff.]

Der DIJKSTRA-Algorithmus dient grob vereinfachend dargestellt dazu, für eine gegebene Start-Ziel-Beziehung den Weg des kürzestmöglichen Widerstands zu finden. Dazu muss für jede Kante eine eindeutige Zuordnungszahl, beispielsweise die Dauer oder der Preis, bestimmt und genutzt werden, da der DIJKSTRA-Algorithmus ein globales eindimensionales Minimum zu erreichen sucht. Das erschwert im DIJKSTRA-Algorithmus die multidimensionale Routensuche, also die Routensuche nach mehreren Optima wie zum Beispiel einer möglichst kurzen und gleichzeitig möglichst komfortablen Route. Der DIJKSTRA-Algorithmus ist für diese Arbeit insbesondere nicht geeignet, da er der Auswahl *einer* Route für eine konkrete Start-Ziel-Relation dient. Er findet damit lediglich *eine* beste Lösung, nicht jedoch alle zum Beispiel über einen Tag verteilten Lösungen.[Jip09][Gol76, S.1164ff.]

Um mehrere Lösungen finden zu können, deren Gewichtung erst nach ihrer Erzeugung vorgenommen werden soll, wurde im Rahmen dieser Arbeit ein eigener Algorithmus entworfen, der sich an verschiedenen algorithmischen Modellen (insbesondere an den in [Jip09], [Zha97] und [Mie00] vorgestellten Modellen und Modellierungsansätzen) orientiert. Grundsatz der Entwicklung war einerseits, eine ausreichende Berücksichtigung der Wissenschaftlichkeit zu gewährleisten und damit der Möglichkeit, Einfluss auf die Parameter der Verbindungserzeugung und -auswahl nehmen zu können, sowie andererseits eine möglichst performanceschonende Routensuche zu entwickeln, die es ermöglicht, auch größere Modellstufen (zum Beispiel ein deutschlandweites Modell) berechnen zu können, ohne auf Supercomputer oder Rechenzentren ausweichen zu müssen. Auch aus wirtschaftlichen Gründen war daher eine sinnvolle Abwägung zwischen der Menge der zu verarbeitenden Informationen und der Reduktion der Ausgabemenge geboten; umso wichtiger erscheint daher die Maßgabe, die Wissenschaftlichkeit durchgehend gewährleisten zu können.

Auch aus Gründen der Nachvollziehbarkeit wurde daher entschieden, kein bereits existierendes Modell zu verändern oder vorhandene Programmbibliotheken abzuändern, sondern im Rahmen dieser Arbeit ein komplett neues algorithmisches Modell zu entwickeln, das den aufgeworfenen Bedingungen genügen kann und ausreichend performant ist, um auch auf handelsüblichen Bürocomputern berechnet zu werden.

5.1.2 Technisches Setup

Das generelle technische Setup des Modells besteht aus einem Datenbankmodell. Basis der Erwägungen ist hierbei, dass die anzuliefernden Daten eine solche Menge umfassen (die modellierte Datenbank umfasst einen erheblichen Teil des deutschen Schienenverkehrsangebots und besteht aus ca. 25.000 Zügen mit etwa 177.000 Halteplansätzen), dass sich eine Datenbank als Speichermedium anbietet und die Primärdaten auch in Form einer Datenbank angeliefert werden. Eine Speicherung auch der erzeugten Routen in der für die Zulieferung genutzten Datenbank bzw. im selben Datenbankmodell erscheint daher zielführend.

Zudem bietet die Nutzung einer Datenbank aus programmertechnischer Sicht den Vorteil, die Ergebnisse zu jeder Zeit ohne Systemeingriffe begutachten zu können, was auch für die Programmierung und die unvermeidliche Fehlersuche (das „Debugging“) von Vorteil ist. Aus technischen Gründen ist dabei die Ergebnissichtung von Programmen zur Laufzeit nicht trivial und erfolgt im Regelfall mit einer Vielzahl von Hilfskonstrukten wie speziellen Programmier-Laufzeitumgebungen, die ihrerseits aufgrund der notwendigen Programmüberwachung zur Datenbereitstellung die Laufzeit des Programms enorm erhöhen. Bei Nutzung einer Datenbank kann durch das Datenbankverwaltungsprogramm der aktuelle Stand der Datenbank zu jeder Zeit abgefragt werden, nur während eines einzelnen Erzeugungsschritts ist der Inhalt – abhängig von der verwendeten Datenbanktechnik – gegebenenfalls nicht immer abrufbar.[Ste14, S.170f.]

Die von vielen Datenbanksystemen verwendete Programmiersprache Structured Query Language (SQL) dient der Erstellung und Bedienung von Datenbanksystemen. SQL-Datenbanken sind eine Art relationaler Datenbanksysteme, welche die rationale elektronische Datenverwaltung auch großer Mengen ermöglichen. Die Daten werden hierbei in einzelnen Tabellen abgelegt, wobei innerhalb jeder Tabelle ein Zeilen-Spalten-Modell verwendet wird, mit definierten Spalteneinträgen (zum Beispiel definiert als Zahlenfeld oder als Textfeld), definierten Spaltenlängen (um die Größe des zu reservierenden Speichers im Vorhinein zu determinieren) und ggf. definierten Beziehungen zwischen den einzelnen Tabellen (abhängig vom Datenbanksystem).[Ste14, S.15ff.]

Als zugrundeliegende Datenbank wurde ein auf dem MySQL-Datenbank-Verwaltungssystem (die MySQL Community Edition) von Oracle basierendes System genutzt. MySQL-Datenbanksysteme werden weltweit sehr häufig eingesetzt, MySQL selbst spricht hierbei bei der MySQL Community Edition von „the world’s most popular

open source database”¹³[myS16b]; als Open-Source-Software kann sie von jedermann frei heruntergeladen und auch verändert werden, was die Verbreitung vereinfacht und weitere Einsätze ermöglicht.[myS16b]

Die im Rahmen dieser Arbeit verwendete Datenbanksoftware MariaDB ist hierbei ein durch Abspaltung von MySQL entstandenes Datenbankverwaltungssystem, das besonderen Fokus auf die Geschwindigkeit unter speziellen Einsatzszenarien legt und viele Erweiterungen der MySQL-Datenbanksysteme analog umsetzt. Obwohl beträchtliche technische Unterschiede bestehen, ist das Verhalten gegenüber dem Nutzer ähnlich, um eine einfache Migration von Projekten zu ermöglichen. Eine Umstellung von MySQL auf MariaDB wird dadurch erheblich erleichtert.[Mar16b]

Der wesentliche Anreiz, das Projekt mit MariaDB und nicht mit einer anderen Datenbank-Engine umzusetzen, besteht in einer besonders effizienten MEMORY-Engine bei der MariaDB-Umgebung. Diese Engine hält Datenbanktabellen vollständig im Arbeitsspeicher des Computers, der um Vielfaches schneller an den Hauptprozessor angebunden ist und über schnelle Datentransferraten verfügt als konventionelle Festplattenspeicher.¹⁴ Die entstehende Verkürzung der Laufzeiten ist besonders bei einem entstehenden großen Datenbearbeitungsvolumen aus einer an sich geringen Menge an Primärdaten (wie im Rahmen dieser Arbeit) von Relevanz; negativ bleibt dabei aber festzuhalten, dass der Arbeitsspeicher beim Herunterfahren des Computers gelöscht wird, die dauerhaft festzuhaltenden Ergebnisse also weiterhin klassisch auf Festplattenspeichern gesichert werden müssen. MEMORY-Tabellen eignen sich aus diesem Grund besonders für Daten, die nur während der Laufzeit des Programms relevant sind.[Mar16c]

Zusammengefasst ist MariaDB als Datenbankverwaltungssystem insbesondere aus Performancesicht anderen SQL-Datenbanksystemen wie der MySQL Community Edition vorzuziehen, und dafür für Projekte wie das Modell dieser Arbeit besser geeignet als gängige andere SQL-Derivate oder Standard-SQL-Umgebungen.[Mar16a; Bel14]

Zur Abfragesteuerung und Programmierung wird die Skriptsprache Hypertext Preprocessor (PHP) eingesetzt, die in ihrer Ausgabe HTML-Output erzeugt, der mit Textprogrammen oder – besser formatiert – mit Internet-Browsern angezeigt werden kann. Dadurch wird ein rudimentäres Setup einer grafischen Oberfläche (Graphical User Interface – GUI) möglich, die zur Steuerung des Programmablaufs dient. Mit PHP wird hierbei

¹³ etwa: „die populärste Open-Source-Datenbank weltweit“

¹⁴ Auch heute verbreitete Festplattenspeicher auf Basis von Flash-Speichern (sogenannte SSD-Speicher) sind zwar herkömmlichen Festplatten in ihrer Datentransferrate weit überlegen, sind aber noch wesentlich langsamer als im Computer verbauter Arbeitsspeicher (Random Access Memory, RAM).

eine weit verbreitete und zu MySQL und MySQL-Derivaten kompatible Skriptsprache genutzt. Ein großer Vorteil von PHP ist, dass die Dateien – wie bei Skriptsprachen üblich – während der Laufzeit in Maschinensprache übersetzt (kompiliert und assembled) werden. Eine spezielle Kompilierung abhängig vom Zielsystem muss daher nicht vor der Laufzeit (wie ansonsten bei Programmen üblich) vorgenommen werden; die Dateien sind mithin ohne Konvertierungsprobleme unter allen PHP unterstützenden Plattformen frei übertragbar und jeweils einsetzbar.[Wal08, S.189ff.]

5.2 Datenmodell

In Referenz auf das Knoten-Kanten-Modell (vgl. Kapitel 5.1.1) werden für das Datenbanksetup alle Daten zur Definition der Knoten und Kanten sowie deren spezifischer Eigenschaften als Eingangsdaten benötigt. Als Knoten dienen dabei alle Verkehrsstationen, die Teil des Modells sind; mithin über 2.000 Bahnhöfe und Haltepunkte. Jeder dieser Knoten besitzt als Eigenschaften seinen Namen und das zugeordnete Fahrgastpotential (vgl. hierzu die erläuternden Ausführungen in Kapitel 4.3).

Weitere Eigenschaften der Knoten sind für dieses Modell im akademischen Einsatz nicht notwendig; es wäre jedoch vorstellbar, das Modell für einen Produktiveinsatz dahingehend zu erweitern. Für einen Einsatz im Schienenpersonennahverkehr und insbesondere bei der intermodalen Verknüpfung mit Verkehren von Bussen und Straßenbahnen wäre es denkbar, die Barrierefreiheit der Verkehrsstation und der einzelnen Umsteigevorgänge sowie die Distanz (und damit eine Mindesttransferzeit) zwischen den Halten der einzelnen Verkehrsmittel ebenfalls in das Modell zu integrieren. Außerdem könnte überlegt werden, das Modell auf einzelne Bahnsteige zu beziehen, da im gegenwärtigen Modell jeder Knoten nur eine einzelne Eigenschaft zum Beispiel für die Umsteigezeit oder die Distanz zu anderen Verkehrsmitteln besitzen kann.¹⁵ Entsprechend können auch Umstiege am selben Bahnsteig an der gegenüberliegenden Kante nicht separat berücksichtigt werden, was eine gewisse Unschärfe im Produktiveinsatz erzeugen würde oder anderweitig zu behandeln wäre (beispielsweise mit separat festgelegten schnelleren Umstiegen).

Auch weitere für die Präferenz entscheidende Faktoren wie die Infrastruktur der Verkehrsstation oder die gefühlte Sicherheit beim Wartevorgang werden im vorgelegten Status

¹⁵Insbesondere in weitläufigen Bahnhöfen und Verkehrsstationen mit mehreren betrieblichen Bahnhofsteilen kann diese Distanz relevant sein, so wird die Umsteigezeit in München Hbf von der DB Netz AG abhängig von der konkreten Ankunft mit zwischen 3 und 15 Minuten angegeben.[DB 15b]

mangels thematischer Relevanz für das Kernthema der Arbeit nicht berücksichtigt, sind aber in die Datenstruktur integrierbar. Eine solche Integration würde ermöglichen, den Algorithmus der Entscheidungsfindung auf andere Fragestellungen zu erweitern, zum Beispiel auf die Verbindungssuche für Nachtfahrten im ÖPNV unter Berücksichtigung einer Sicherheitskomponente und Minimierung der entstehenden Wartezeiten.

5.2.1 Primärdaten

Die Primärdaten werden in mehreren Tabellen strukturiert, die vor der detaillierten Beschreibung in Tabelle 10 kurz vorgestellt werden:

Name der Tabelle	Speicherung der Daten
bp_fp_bf	Name der Verkehrsstationen
einwohner	Fahrgastpotential der Verkehrsstationen
bp_fp_zug	Eigenschaften der Züge
zugpreis	Preis der Züge
bp_fp_zeiten	Haltestatensätze der Züge
entfernungen	Entfernungen zweier Verkehrsstationen
zeitenpreis	Preis eines Zugs zwischen zwei Stationen
relevante_bahnhoeefe	Relevante Bahnhöfe im Modell

Tabelle 10: Übersicht der Primärdatenquellen
Quelle: Eigene Darstellung

Aufgrund der speziellen Zulieferung der Rohdaten erfolgt eine Aufteilung der Knotendaten in die Bahnhöfe mit ihrem Namen und einem eindeutigen Identifikator (einer ID) in der Tabelle `bp_fp_bf` sowie in die Tabelle `einwohner`, die zu jeder Verkehrsstation das Fahrgastpotential speichert. Knoten sind dabei im Modell grundsätzlich Verkehrsstationen, da rein betriebliche Halte für die Verbindungserzeugung nicht von Bedeutung sind (ein Ein- bzw. Ausstieg wäre sowieso nicht zugelassen, mithin wären von/zu diesen Halten also keine Verbindungen konstruierbar).¹⁶

¹⁶Hinweis: Sollen beispielsweise für die Dienstplanung eines Unternehmens Routen für Personal geplant werden, kommt eine Aufnahme dieser Stationen in Betracht, da dort selbstverständlich Betriebshalte stattfinden können.

Das Setup der Datenbanktabellen ist in den folgenden Tabellen 11 und 12 abgetragen:

Name des Felds	Inhalt und Datentyp des Felds	Größe des Felds
ID	Ganzzahlenwerte [INTEGER]	6 Zeichen
Name	Wortwerte [VARCHAR]	255 Zeichen

Tabelle 11: Setup der Datenbanktabelle `bp_fp_bf`
Quelle: Eigene Darstellung

Name des Felds	Inhalt und Datentyp des Felds	Größe des Felds
ID	Ganzzahlenwerte [INTEGER]	6 Zeichen
Bhf	Ganzzahlenwerte [INTEGER]	6 Zeichen
Einwohner	Ganzzahlenwerte [INTEGER]	7 Zeichen

Tabelle 12: Setup der Datenbanktabelle `einwohner`
Quelle: Eigene Darstellung

Die Datenbanktabelle `bp_fp_bf`, die die Namen der beteiligten Verkehrsstationen im Feld `Name` speichert, enthält im hier verwendeten Setup keine Informationen über den geographischen Standort oder die Höhenlage der Verkehrsstation. Beide Daten könnten als Erweiterung des Datenmodells bereitgestellt werden, um beispielsweise die erreichte mittlere Reisegeschwindigkeit bezogen auf die Luftlinie zwischen den Verkehrsstationen zu ermitteln, und darauf aufbauend die Erzeugung von Passagierzahlen von der erreichten Durchschnittsgeschwindigkeit bezogen auf die direkte Entfernung (als Indikator für die intermodale Konkurrenzfähigkeit) abhängig zu machen, oder auch, um eine grafische Anzeige in Form einer Stationskarte im Produktiveinsatz zu ermöglichen. In der Tabelle `einwohner` ist neben der Bahnhofs-ID (aus `bp_fp_bf`, gespeichert im Feld `Bhf`) jeweils das zugeordnete Fahrgastpotential als Ganzzahl im Feld `Einwohner` gespeichert.

Als Kanten fungieren alle Verbindungen zwischen den Knoten, wobei diese im Modell nicht direkt referenziert sind, sondern über Zugläufe definiert werden. Da die konkreten Fahrwege damit nicht hinreichend präzise bestimmt sind (weil zwischen zwei verkehrlichen Halten unter Umständen mehrere konkrete betriebliche Fahrwege genutzt werden können), ist das Modell im aktuellen Setup auch nicht zur Auswahl spezifischer Verbindungswünsche auf Basis der Fahrtroute einsetzbar, sondern nur zur Auswahl einer Verbindung zwischen zwei Haltebahnhöfen eines Zuges. Die zwingend notwendigen Eigenschaften der Zugläufe im Bezug auf die Kanten sind die Fahrtdauer im Abschnitt sowie der vom Verkehrsunternehmen vorgegebene Preis für diesen Abschnitt. Die weite-

ren Faktoren wie beispielsweise Komfortfaktoren sind nicht zwingend kantenabhängig, sondern hängen vom genutzten Zug ab.

In den Daten ist diese Teilung ebenfalls sichtbar: Die Züge werden mit ihren Eigenschaften in der Tabelle `bp_fp_zug` verwahrt (abgetragen in Tabelle 13), die Haltebahnhöfe und Fahrzeiten der Züge in der Datenbanktabelle `bp_fp_zeiten` (dargestellt in Tabelle 15). Die den Abschnitten zugeordneten Reiseweiten werden in der Tabelle `entfernungen` gespeichert, um im Zuglauf nicht zu viele Daten doppelt vorzuhalten (die meisten Züge befahren Abschnitte mit jeweils derselben Länge). Folgend wird zuerst die Struktur der die Stammdaten der Züge verwaltenden Datenbanktabelle `bp_fp_zug` vorgestellt:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>typ</code>	Textwerte [VARCHAR]	3 Zeichen
<code>zugnr</code>	Ganzzahlenwerte [INTEGER]	7 Zeichen
<code>start</code>	Ganzzahlenwerte [INTEGER]	6 Zeichen
<code>ziel</code>	Ganzzahlenwerte [INTEGER]	6 Zeichen
<code>mo</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>di</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>mi</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>do</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>fr</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>sa</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>so</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen

Tabelle 13: Setup der Datenbanktabelle `bp_fp_zug`
Quelle: Eigene Darstellung

In der Tabelle `bp_fp_zug` sind neben den Primärdaten des Zuges (die Zuggattung im Feld `typ`, die Zugnummer im Feld `zugnr` sowie Start- und Zielbahnhof in den entsprechenden Feldern) auch die Verkehrstage angegeben (in den Feldern `mo` ... `so`, jeweils mit 0 für nicht verkehrend und 1 für verkehrend vermerkt).

Die Länge der Felder kann insbesondere bei der Zuggattung und den Zugnummern vom verwendeten Grunddatenmodell abhängen, sodass hier Anpassungen auf Basis der ggf. verwendeten abweichenden Primärdaten beim Einsatz in anderen Systemen bzw. mit anderen Daten vorgenommen werden müssten. Die Länge anderer Felder – hier sind die festen Werte der Verkehrstage zu nennen – ist algorithmisch determiniert (daher wird in diesem Fall bei den Verkehrstagen als Datentyp der Zahlenfelder ein

TINYINT verwendet, der aufgrund des eingeschränkten Wertebereichs gegenüber einem INTEGER weniger Speicherplatz verbraucht).

Das Preismerkmal der Züge wird nicht innerhalb der Stammdatentabelle `bp_fp_zug`, sondern in einer eigenen Tabelle `zugpreis` verwahrt:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
id	Ganzzahlenwerte [INTEGER]	10 Zeichen
Zug	Ganzzahlenwerte [INTEGER]	10 Zeichen
Preis	Gleitkommawerte [DOUBLE]	5,2 Zeichen

Tabelle 14: Setup der Datenbanktabelle `zugpreis`
Quelle: Eigene Darstellung

Dabei wird in der in Tabelle 14 dargestellten Datenbanktabelle `zugpreis` jedem Zug ein Preis, gemessen in Cent pro Kilometer, zugeordnet (die Angabe „7,2“ als Parameter des DOUBLE-Felds drückt hierbei ein zehnstelliges Feld mit sieben Stellen vor dem Komma, zwei Nachkommastellen und einer Stelle für das Komma aus). Um auch Kommawerte darstellen zu können, ist als Datenfeld ein DOUBLE-Feld ausgewählt worden. Mit dieser Tabellenstruktur wäre es auch möglich, einzelne Linien bzw. ganze Zuggattungen mit einem einheitlichen Preis zu versehen. Die Wahl dieser Struktur gründet auch auf dem Gedanken, Preise nicht nur zugscharf, sondern gruppeneinheitlich zu setzen, was im Rahmen dieser Arbeit genutzt wird, um Preise für ganze Linien zu setzen.

Die Speicherung und Zuordnung von Haltezeitdatensätzen der Züge sowie deren jeweiliger Eigenschaften wird folgend in der Tabelle `bp_fp_zeiten` vorgenommen:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
id	Ganzzahlenwerte [INTEGER]	10 Zeichen
zug_nr	Ganzzahlenwerte [INTEGER]	7 Zeichen
halt_nr	Ganzzahlenwerte [INTEGER]	3 Zeichen
an	Zeitwerte [TIME]	<i>SQL-Vorgabe</i>
ab	Zeitwerte [TIME]	<i>SQL-Vorgabe</i>
bf_id	Ganzzahlenwerte [INTEGER]	6 Zeichen
vorh_id	Ganzzahlenwerte [INTEGER]	10 Zeichen
nae_id	Ganzzahlenwerte [INTEGER]	10 Zeichen

Tabelle 15: Setup der Datenbanktabelle `bp_fp_zeiten`
Quelle: Eigene Darstellung

Die Datenbanktabelle `bp_fp_zeiten` verwahrt alle Haltedaten der jeweiligen Zugläufe. Dazu gehören neben der Zugnummer (`zug_nr`) als Referenz und dem Identifikator

der Verkehrsstation (`bf_id`) die Nummer des Halts als Position im Zuglauf (`halt_nr`; dabei wird dem Startbahnhof die niedrigste und dem Zielbahnhof die höchste Nummer zugewiesen), sowie die Verkehrszeiten mit Ankunft (`an`) und Abfahrt (`ab`) für jede Verkehrsstation (die Länge für `TIME`-Datenfelder wird hierbei von SQL explizit zugewiesen und ist nicht selbst zuordenbar). Ankünfte oder Abfahrten, die dabei übergangen werden sollen (zum Beispiel, weil an Bahnhöfen kurz vor dem Ziel kein Einstieg mehr zugelassen ist), bekommen dabei den Sekundenwert `00:00:01` zugewiesen – alle anderen Werte sind fahrplantypisch nur minutengenau aufgeführt und damit durch den Algorithmus eindeutig abgrenzbar. Zur Beschleunigung der Suche der nächsten und der vorherigen Verkehrsstation im Zuglauf werden die Felder `vorh_id` und `nae_id` genutzt, die den vorherigen bzw. den folgenden Identifikator des entsprechenden Zuglaufabschnitts speichern.

Zur Bestimmung grundlegender Verbindungsdaten wie der zurückgelegten Länge zwischen Start- und Zielstation sowie dem Preis der Verbindung muss die Länge aller Zuglaufabschnitte jeweils als Primärdatum vorgegeben sein und in der Datenbank gespeichert werden; dies wird mit der Datenbanktabelle `entfernungen` realisiert:

Name des Felds	Inhalt und Datentyp des Felds	Größe des Felds
ID	Ganzzahlenwerte [MEDIUMINT]	6 Zeichen
von	Ganzzahlenwerte [MEDIUMINT]	6 Zeichen
nach	Ganzzahlenwerte [MEDIUMINT]	6 Zeichen
Entfernung	Ganzzahlenwerte [MEDIUMINT]	6 Zeichen

Tabelle 16: Setup der Datenbanktabelle `entfernungen`
Quelle: Eigene Darstellung

In Tabelle 16 ist das Setup der Datenbanktabelle `entfernungen` dargestellt; dort wird für ein gegebenes Tupel aus Startstation (im Feld `von`) und Zielstation (im Feld `nach`) jeweils die Entfernung festgehalten. Um spätere Algorithmen zu beschleunigen, werden die Daten für jedes Bündel jeweils doppelt vorgehalten, also in Hin- und Rückrichtung separat notiert. Die Entfernung ist hierbei im Tabellenfeld `Entfernung` in 100-Meter-Schritten angegeben, das ermöglicht eine auf ein Datenmodell auf Basis von Staaten bezogene ausreichende Genauigkeit, ohne durch zu große Zahlenwerte die entstehenden Daten unhandlich zu machen. Im Übrigen sind auch die Daten von Eisenbahnverkehrsunternehmen im Regelfall auf dieses oder ein gröberes Entfernungsintervall geeicht, so werden Entfernungen beispielsweise in den Kursbuchtabellen der Deutschen Bahn AG nur in ganzen Kilometern angegeben (vgl. [DB 15a] und [DB 15b]).

Dabei ist zu beachten, dass dieser Ansatz keine Wahl für unterschiedliche Längen verschiedener Verbindungen bzw. verschiedener Zugläufe zulässt. Dies hat zwei unterschiedliche Folgen: Zum Einen ist zu gewährleisten, dass entweder alle Züge einer gegebenen Relation zwischen zwei Verkehrsstationen dieselbe Fahrtroute wählen (bzw. mindestens eine Route derselben Entfernung) oder es ist ansonsten die entstehende Unschärfe hinzunehmen. Zum Anderen kann an dieser Stelle keine Spreizung zwischen tatsächlicher Entfernung und der Tarifentfernung vorgenommen werden. Das letztgenannte Mittel wird bei Strecken angewandt, die nach Fertigstellung von Um- oder Neubaumaßnahmen direkter befahren werden können, die direkte Befahrbarkeit sich aber aufgrund der höheren Geschwindigkeit oder anderweitigen Vorteilen nicht in einem günstigeren Preis niederschlagen soll; In diesem Fall wird dann die Tarifentfernung der kürzeren Strecke erhöht, um einen finanziellen Ausgleich für die „bessere“ Leistung bei ansonsten starrem Preissystem verlangen zu können.

Die Spreizung der Preise entfaltet hierbei außerdem eine Lenkungswirkung: Preissensible Fahrgäste können die direkte Strecke meiden (und dafür die zeitlich längere Fahrt wählen), während zeitsensible Fahrgäste den Aufschlag bezahlen können. Mithin kann innerhalb eines starren Preissystems mit einem rein entfernungsbasierten Ansatz unter Zuhilfenahme der Tarifentfernungen eine Lenkungswirkung erzielt werden, die das Preissystem definitionsgemäß gar nicht vorsieht.

Eine solche Maßnahme wurde durch die Deutsche Bahn AG zum Beispiel auf der Schnellfahrstrecke Mannheim – Stuttgart angewandt: Die Entfernung im Abschnitt Vaihingen (Enz) – Stuttgart-Zuffenhausen ist über die Schnellfahrstrecke (Streckenummer 4080) deutlich kürzer als die Streckenführung über die Altbaustrecke Vaihingen (Enz) – Bietigheim-Bissingen – Stuttgart-Zuffenhausen (unter Nutzung der Strecken 4842 und 4800). Um nun eine gewisse Lenkungswirkung zu erzielen und die Fahrt über die Schnellfahrstrecke nicht günstiger zu machen als die Fahrt über die Altbaustrecke, wurde allerdings die Tarifentfernung der neuen Strecke der physikalischen Länge der alten Strecke angepasst; so ist die Neubaustrecke physikalisch in diesem Abschnitt 22,340 km lang, während die Altbaustrecke über Bietigheim-Bissingen 32,179 km lang ist – beide Strecken werden allerdings auf Basis der längeren Entfernung bepreist. [DB 15c; DB 15a]

Nach der Speicherung des Preises einzelner Züge wird aus ressourcenökonomischen Gründen noch eine weitere Datenbanktabelle mit Preisen gebildet: In der Tabelle `zeitenpreis` ist neben der Länge des Abschnitts das jeweilige Produkt aus Preis des Zuges und Länge des Abschnitts (also der zu entrichtende Teilfahrpreis) zwischen einem Haltedatensatz und dem folgenden Haltedatensatz abgetragen:

Name des Felds	Inhalt und Datentyp des Felds	Größe des Felds
ID	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>zeiten_id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>laenge</code>	Ganzzahlenwerte [MEDIUMINT]	6 Zeichen
<code>preis</code>	Gleitkommawerte [DOUBLE]	7,2 Zeichen

Tabelle 17: Setup der Datenbanktabelle `zeitenpreis`
Quelle: Eigene Darstellung

Die Speicherung der Länge des Abschnitts und des Abschnittspreises dienen der schnelleren Berechnung der bei Direktverbindungen ansonsten jeweils separat zu berechnenden Abschnittssummen. Diese Datenbanktabelle ist im engen Sinne keine Primärdatentabelle, da sie während der Laufzeit aus den gegebenen Datentabellen berechnet wird; ihr Ergebnis dient allerdings der Bestimmung der Eigenschaften aller aufzustellenden Verbindungen, sodass sie an dieser Stelle im Rahmen der Vorstellung der Primärdatentabellen erwähnt wird.

Zur Konzentration der Ressourcen auf eher bedeutende Verbindungen wird außerdem eine Tabelle `relevante_bahnhoeefe` eingeführt, in der alle Verkehrsstationen festgehalten sind, die entweder Anschlüsse bieten können oder mindestens ein Fahrgastpotential von 20.000 Einwohnern zugeordnet bekommen haben. Verbindungen werden dabei hergestellt zwischen allen Verkehrsstationen, die relevant im Sinne dieser Definition sind, sowie zwischen allen Verkehrsstationen, wenn mindestens eine der beteiligten Verkehrsstationen relevant im Sinne dieser Definition ist.

In Summe sind etwa 60 % der Verkehrsstationen relevante Verkehrsstationen im Sinne des Modells, mit den damit zu bildenden Verbindungen werden jedoch über 95 % der Passagiere des Gesamtmodells abgedeckt, da die übrigen, nicht-relevanten, Verkehrsstationen im Verhältnis nur wenige Einwohner abdecken und nur die Verbindungen zwischen zwei als nicht relevant anzusehenden Verkehrsstationen ausgeschlossen werden. Das ermöglicht eine Konzentration auf die wichtigeren Verbindungen zwischen Zentren sowie zwischen Zentrum und ländlicher Region, was damit auch die Aussagekraft des Endergebnisses erhöht, da die Anzahl der erzeugten Verbindungen mit dieser Einteilung auch die Relevanz der Station widerspiegelt.

Das Setup der Datenbanktabelle `relevante_bahnhoefe` findet sich in der folgenden Tabelle 18:

Name des Felds	Inhalt und Datentyp des Felds	Größe des Felds
ID	Ganzzahlenwerte [INTEGER]	6 Zeichen
bf	Ganzzahlenwerte [INTEGER]	6 Zeichen

Tabelle 18: Setup der Datenbanktabelle `relevante_bahnhoefe`
Quelle: Eigene Darstellung

Die Datenbanktabelle `relevante_bahnhoefe` besteht dabei lediglich aus zwei Feldern: dem eindeutigen Identifikator sowie dem Identifikator derjenigen Verkehrsstationen, die als relevant im Sinne des Systems erachtet werden. Die aufgrund der Anwendung der Datenbanktabelle `relevante_bahnhoefe` zwischen nicht als relevant eingestuftem Verkehrsstationen nicht erzeugten Verbindungen verzerren das Datenbild natürlich in dem Sinne, dass eine Anzahl möglicher Verbindungen von vornherein ausgeschlossen wird.

Gleichzeitig kann man sich die umgekehrte Argumentation zu Eigen machen: Eine nicht in dieser Größe relevante Bevölkerungszahl hätte eine überproportionale Verbindungsmenge erzeugt und – dadurch, dass Verkehre in ländlichen Regionen oftmals in geringerer Taktung stattfinden – das Ergebnis der Verbindungssuche allgemein bezogen auf die jeweils betroffene Menge an Passagieren erheblich verzerrt. Insoweit ist nach einer gründlichen Abwägung die Verzerrung durch wegfallende Verbindungen zwischen ländlichen Regionen als weniger relevant und damit selbige als verzichtbar einzustufen.

5.2.2 Sekundärdaten

Im Gegensatz zur Speicherung der Primärdaten, die im Wesentlichen zugeliefert werden bzw. als Eingangsdaten zur Verfügung stehen müssen, um das gesamte Datenmodell bilden zu können, werden die Sekundärdaten zur Laufzeit neu erzeugt und unterscheiden sich daher je nach den Umgebungsvariablen und definierten Anforderungen des zu berechnenden Szenarios. Eine Übersicht der Datenquellen von Sekundärdaten ist in Tabelle 19 abgetragen:

Name der Tabelle	Speicherung der Daten
direktverbindungen	Direkte Verbindungen zwischen Stationen
umstiege_zeiten	Zeitendatsätze für mögliche Umstiege
parallele_zuege	Parallel verkehrende Züge
parallelverbindungen	Parallelverbindungen
verbindungen	vorteilhafte Verbindungen
umstiege_dv_mem	Verbindungen während der Erzeugung
verbindungen_zielgruppe	für eine Zielgruppe vorteilhafte Verbindungen

Tabelle 19: Übersicht der Sekundärdatenquellen
Quelle: Eigene Darstellung

Allen Szenarien ist gemein, dass eine grundlegende Anzahl von Verbindungen direkt ermittelt werden kann, nämlich Verbindungen entlang eines Zuglaufs ohne Umstieg. Sie werden in der Tabelle `direktverbindungen` gespeichert und ihnen werden direkt eine Vielzahl von Eigenschaften mitgegeben, die in der späteren Berechnung von Umsteigeverbindungen nützlich sind, wie die Dauer der Verbindung. Die Verbindungsdauer könnte auch jedes Mal in der Laufzeit der Verbindungssuche berechnet werden. Bei der Abwägung zusätzlicher Rechnerkapazität zur Live-Berechnung dieser Daten gegen den Speicherplatz einer solchen Information für alle dieser Direktverbindungen fällt die Abwägung allerdings zugunsten der Speicherung aus, da insgesamt während der Laufzeit eine Vielzahl von Verbindungserzeugungsvorgängen vorgenommen wird, und die gesamte Durchführungsdauer erheblich reduziert werden kann, wenn solche Daten dauerhaft vorgehalten werden.

Die Datenstruktur der Datenbanktabelle `direktverbindungen` ist in Tabelle 20 festgehalten:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>Zug</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>Zeile_eins</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>Zeile_zwei</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>bf_von</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>bf_nach</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>Dauer_Min</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>ab</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>an</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>Preis</code>	Gleitkommawerte [DOUBLE]	7,2 Zeichen
<code>Laenge</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>Komfort</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>virtPreis</code>	Gleitkommawerte [DOUBLE]	10,2 Zeichen
<code>Mo</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Di</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Mi</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Do</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Fr</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Sa</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>So</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Vergleich</code>	Textwerte [VARCHAR]	25 Zeichen

Tabelle 20: Setup der Datenbanktabelle `direktverbindungen`
Quelle: Eigene Darstellung

Die Datenbanktabelle `direktverbindungen` ist ihrem Zweck entsprechend neben dem in jedem Fall notwendigen Identifikator (`id`, als Ganzzahl) mit den für Verbindung erwartbar notwendigen Feldern ausgestattet: der zugrundeliegenden Fahrplanleistung (im Feld `Zug`), dem Start- und Zielbahnhof (in den Feldern `bf_von` und `bf_nach`) sowie den Verkehrstagen der Leistung (in den Feldern `Mo` bis `So`, analog zur Notation in der Tabelle `bp_fp_zug` wieder in binärer Notation, wobei der Wert 1 eine Leistung ausweist, die am betreffenden Verkehrstag angeboten wird).

Die Abfahrts- und Ankunftszeit des Zuges wird in der Datenbanktabelle `direktverbindungen` notiert (in den Feldern `ab` und `an`), dies ist jedoch nicht zwingend notwendig, da auch die Identifikatoren der Datensätze aus der Haltedatensatztabelle `bp_fp_zug` in den Feldern `zeile_eins` (für die Startzeile der Direktverbindung) und `zeile_zwei` (für die Zielzeile der Direktverbindung) gespeichert. Die dargestellte Doppelhaltung der Daten dient Algorithmen, die nur die Tabelle `direkt-`

verbindungen nutzen, ohne auf die Zeitendatensätze in `bp_fp_zeiten` zu referenzieren; ansonsten müssten auch in diesem Fall die Zeitdaten erst aus der Datenbanktabelle der Zeitendatensätze mit einer zusätzlichen Verknüpfung entnommen werden.

Das Datenbankfeld `Vergleich` ist dabei ein Textfeld und wird zur Verbesserung der Laufzeit der Algorithmen bei der Entfernung unvorteilhafter Verbindungen verschiedentlich genutzt, ist aber für die Erzeugung der Direktverbindungen sowie der anfänglichen Umstiege nicht weiter von Belang und wird daher erst im weiteren Verlauf der Arbeit bei der ersten tatsächlichen Verwendung erläutert.

Weitere Eigenschaften der Verbindung wie der Verbindungspreis (im Feld `Preis`, hier als `DOUBLE`-Datenfeld, also als Gleitkommazahlenfeld), die Länge der Verbindung (gespeichert in 100-Meter-Intervallen, daher als `INTEGER` im Feld `Laenge`) sowie der virtuelle Preis (zur Definition des virtuellen Preises siehe Kapitel 4.2) sind ebenfalls Eigenschaften der gespeicherten Verbindungen und werden daher ebenfalls in der Datenbanktabelle `direktverbindungen` gespeichert.

Da die Erstellung der Daten der Tabelle `direktverbindungen` zur Laufzeit vor der Berechnung der Umsteigeverbindungen und der Auswahl der Routen für alle Zielgruppen vorgenommen wird, wird auf die Darstellung der Erzeugung der Direktverbindungen verzichtet.

Im Zusammenhang mit der Erstellung der Datenbanktabelle `direktverbindungen` werden zur Bestimmung von Umstiegen weitere Datenbanktabellen benötigt: Die Datenbanktabelle `umstiege_zeiten` wird dabei genutzt, um für die Verbindungen des ersten Umstiegs eine Verbindung zwischen je zwei Datensätzen der Datenbanktabelle der Zeitendatensätze, `bp_fp_zeiten`, als Umstieg in einer Verkehrsstation herzustellen:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [<code>INTEGER</code>]	10 Zeichen
<code>bf</code>	Ganzzahlenwerte [<code>SMALLINT</code>]	5 Zeichen
<code>id_1</code>	Ganzzahlenwerte [<code>MEDIUMINT</code>]	8 Zeichen
<code>id_2</code>	Ganzzahlenwerte [<code>MEDIUMINT</code>]	8 Zeichen
<code>wartezeit</code>	Ganzzahlenwerte [<code>SMALLINT</code>]	5 Zeichen

Tabelle 21: Setup der Datenbanktabelle `umstiege_zeiten`
Quelle: Eigene Darstellung

Neben dem Bahnhof im Feld `bf` werden in der Datenbanktabelle `umstiege_zeiten` daher, wie in Tabelle 21 dargestellt, die beteiligten Zeitendatensätze in den Feldern `id_1` und `id_2` festgehalten, was der Bildung von Umsteigeverbindungen des ersten Umstiegs zwischen diesen beiden Datensätzen dient. Die in dieser Tabelle erzeugten Datensätze werden danach noch mit verschiedenen Routinen zur Reduktion der Erzeugung unvorteilhafter Anschlüsse behandelt, dazu dient die im Folgenden vorgestellte Datenbanktabelle `parallele_zuege`:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>zug_1</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>zug_2</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen

Tabelle 22: Setup der Datenbanktabelle `parallele_zuege`
Quelle: Eigene Darstellung

In der Datenbanktabelle `parallele_zuege` werden dabei Züge gespeichert, die jeweils eine partielle oder komplette Überdeckung besitzen, die also auf einem bestimmten Abschnitt oder auf dem kompletten Zuglauf parallel verkehren, und die in anderen Tabellen – insbesondere in der bereits vorgestellten Tabelle `umstiege_zeiten` – nur insoweit berücksichtigt werden, als ob nur einer dieser Züge verkehren würde. Diese Tabelle wird dabei in mehreren Szenarien unterschiedlich verwendet und ist daher nicht spezifisch vorgerüstet. Ob dabei der Eintrag im Feld `zug_1` oder im Feld `zug_2` den vorzugswürdigen Eintrag darstellt, ist je nach Szenario unterschiedlich und wird in Kapitel 6.3 näher erläutert. Eine ähnliche Funktion nimmt die Tabelle `parallelverbindungen` ein, sie speichert beteiligte Verbindungen, die sich überdecken:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>v1</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>v2</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen

Tabelle 23: Setup der Datenbanktabelle `parallelverbindungen`
Quelle: Eigene Darstellung

Im Unterschied zur bereits vorgestellten Datenbanktabelle `parallele_zuege` sind in dieser Tabelle jedoch keine Züge, sondern konkrete Verbindungen gespeichert und in die Nutzung der Felder eindeutig zugeordnet: Die im Feld `v2` notierte Verbindung ist jeweils die eindeutig nachteilige, die im Feld `v1` notierte Verbindung jeweils die eindeutig vorteilhafte.

Da in den bisherigen Datenbanktabellen nur Verbindungen ohne Umstieg gespeichert werden konnten, muss jetzt noch eine Speicherung aller Umsteigeverbindungen erfolgen. Um die Verbindungsauswahl nach Erstellung aller Verbindungen möglichst ressourcenschonend zu gestalten, werden in der Tabelle `verbindungen` alle Verbindungen – also auch solche ohne Umstieg – notiert. Diese Tabelle dient auch Sortierprozessen bei der Erstellung von Umsteigeverbindungen mit mehreren Umstiegen; ihre Datenstruktur ist in der folgenden Tabelle 24 abgetragen:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>dv1</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>dv2</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>dv3</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>dv4</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>dv5</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>dv6</code>	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
<code>bf_von</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>bf_nach</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>virtPreis</code>	Gleitkommawerte [DOUBLE]	10,2 Zeichen
<code>Dauer</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>Preis</code>	Gleitkommawerte [DOUBLE]	10,2 Zeichen
<code>ab</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>an</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>Mo</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Di</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Mi</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Do</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Fr</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Sa</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>So</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Vergleich</code>	Textwerte [VARCHAR]	25 Zeichen
<code>Umstiege</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen

Tabelle 24: Setup der Datenbanktabelle `verbindungen`
Quelle: Eigene Darstellung

Die Datenstruktur der Tabelle `verbindungen` unterscheidet sich von der bereits vorgestellten Datenstruktur der Tabelle `direktverbindungen` nur in zwei wesentlichen Punkten: Zum einen wird jetzt die Notation der beteiligten Direktverbindungen notwendig (Datenfelder `dv1` . . . `dv6`), die ihrerseits alle im Rahmen der Verbindungserstellung benötigten Zugdaten beinhalten; zum anderen wird die Zahl der Umstiege im gleichnamigen Feld notiert, um danach filtern zu können. Gleichzeitig konnten viele nicht

direkt benötigten Daten der Direktverbindungen entfallen, neben den Haltedatensätzen des beteiligten Zuges zum Beispiel auch die Länge des Verbindungsweges oder der tatsächlich vorhandene Fahrtkomfort. Beide letztgenannten Daten können problemlos über die notierten Direktverbindungen abgerufen und aufsummiert werden, für die weitere algorithmische Auswahlbetrachtung spielen sie jedoch keine Rolle, daher besteht keine Notwendigkeit, diese Daten permanent vorzuhalten.

Diverse Tabellen werden zur Beschleunigung des Ablaufs für einzelne Algorithmen kurzzeitig oder konstant im Arbeitsspeicher (RAM) gespeichert; dies wird im Folgenden durch den Zusatz „_mem“ zum Namen der Tabelle ausgedrückt. Beispielhaft wird die Tabelle `direktverbindungen` dann als Tabelle `direktverbindungen_mem` bezeichnet, wenn sie im Arbeitsspeicher und nicht auf der Festplatte bereitgehalten wird.

Während der Erzeugung der Verbindungen wird eine in ihrem Aufbau der Tabelle `verbindungen` ähnliche Datenbanktabelle genutzt:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
id	Ganzzahlenwerte [INTEGER]	10 Zeichen
dv1	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
dv2	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
dv3	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
dv4	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
dv5	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
dv6	Ganzzahlenwerte [MEDIUMINT]	8 Zeichen
bf_von	Ganzzahlenwerte [SMALLINT]	5 Zeichen
bf_nach	Ganzzahlenwerte [SMALLINT]	5 Zeichen
virtPreis	Gleitkommawerte [DOUBLE]	10,2 Zeichen
Dauer	Ganzzahlenwerte [SMALLINT]	5 Zeichen
Preis	Gleitkommawerte [DOUBLE]	10,2 Zeichen
ab	Ganzzahlenwerte [SMALLINT]	5 Zeichen
an	Ganzzahlenwerte [SMALLINT]	5 Zeichen
Mo	Ganzzahlenwerte [TINYINT]	1 Zeichen
Di	Ganzzahlenwerte [TINYINT]	1 Zeichen
Mi	Ganzzahlenwerte [TINYINT]	1 Zeichen
Do	Ganzzahlenwerte [TINYINT]	1 Zeichen
Fr	Ganzzahlenwerte [TINYINT]	1 Zeichen
Sa	Ganzzahlenwerte [TINYINT]	1 Zeichen
So	Ganzzahlenwerte [TINYINT]	1 Zeichen
Vergleich	Textwerte [VARCHAR]	25 Zeichen

Tabelle 25: Setup der Datenbanktabelle `umstiege_dv_mem`

Quelle: Eigene Darstellung

Der große Vorteil bei der Nutzung der Tabelle `umstiege_dv_mem` mit quasi identischem Aufbau zur Datenbanktabelle `verbindungen` – lediglich das Feld `umstiege` fehlt – ist die einfache Transformation der optimalen Verbindungen, die in `umstiege_dv_mem` erzeugt wurden, in die Tabelle `verbindungen`, in der die Verbindungen dauerhaft gespeichert werden: Da die Spalten beider Datenbanktabellen identisch belegt sind, ist der Übertrag sehr effizient.

Eine weitere Datenbanktabelle, `verbindungen_zielgruppe`, wird dafür genutzt, jeweils die vorteilhaften Verbindungen einer bestimmten Zielgruppe festzuhalten. Dazu wird für eine Zielgruppe ein Teil der Daten der Datenbanktabelle `verbindungen` gespiegelt; die notwendigen und daher festzuhaltenden Informationen werden in der Datenbanktabelle `verbindungen_zielgruppe` unter derselben ID wie in der Datenbanktabelle `verbindungen` gespeichert und sind in der folgenden Tabelle 26 abgetragen:

Name des Felds	Inhalt und Datentyp des Felds	Länge des Felds
<code>id</code>	Ganzzahlenwerte [INTEGER]	10 Zeichen
<code>bf_von</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>bf_nach</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>virtPreis</code>	Gleitkommawerte [DOUBLE]	10,2 Zeichen
<code>ab</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>an</code>	Ganzzahlenwerte [SMALLINT]	5 Zeichen
<code>Mo</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Di</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Mi</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Do</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Fr</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Sa</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>So</code>	Ganzzahlenwerte [TINYINT]	1 Zeichen
<code>Vergleich</code>	Textwerte [VARCHAR]	25 Zeichen

Tabelle 26: Setup der Datenbanktabelle `verbindungen_zielgruppe`
Quelle: Eigene Darstellung

Zur kurzfristigen Speicherung der bereits erzeugten Verbindungen ab einer Startstation vor der Erzeugung neuer Verbindungen ab dieser Station wird die Datenbanktabelle `verbindungen_mem_gesamt` verwendet; sie gleicht in ihrem Aufbau der bereits vorgestellten Tabelle `verbindungen`, wird jedoch zur Minimierung der Abfragezeiten komplett im Arbeitsspeicher gehalten (und ist daher auf Verbindungen einer einzelnen Startstation limitiert). Auch die Datenbanktabelle `verbindungen_mem_umstiege` besitzt einen identischen Aufbau mit der bereits vorgestellten Datenbanktabelle `verbindungen`.

Sie dient dabei der Speicherung der Gesamtzahl aller Verbindungen eines niedrigeren Umstiegs, im Regelfall aller Verbindungen mit einem Umstieg.

6 Algorithmische Berechnung der Verbindungen

Nach der Vorstellung der Datenstruktur und der verwendeten Technik wird nun die zur Erstellung und Auswahl der Verbindungen verwendete Logik vollständig dargestellt und erläutert. Die ausführliche Vorstellung dient dabei insbesondere der Vorbereitung des in späteren Kapiteln vorzunehmenden Vergleichs der Auswirkung unterschiedlicher Tarifsysteme; Dazu ist eine genaue Kenntnis der verwendeten Algorithmik und der daraus entstehenden Verbindungen essentiell, um die entstehende Auswirkung der Tarifierung plausibilisieren und einordnen zu können.

6.1 Algorithmusdefinition

Bevor eine algorithmische Definition der Problemstellung vorgenommen werden kann, muss zuerst eine verbale Definition gefunden werden, um das Problem gedanklich vollständig durchdringen zu können und danach die algorithmische Umsetzung durchführen zu können. Bei diesem Prozess ist insbesondere auf eine präzise Formulierung der verbalen Definition abzustellen, um den Algorithmus exakt nach den ausformulierten Vorgaben erstellen zu können.[Web05, S.54ff.]

Die grundsätzliche Herausforderung besteht darin, ein Bündel an Verbindungen zwischen Start und Ziel zu finden, das möglichst alle optimalen Lösungen abdeckt, aber dennoch handhabbar bleibt. Es muss während bzw. direkt nach der Erstellung aller Verbindungen darauf geachtet werden, dass diejenigen Verbindungen, die erkennbar keinen kompetitiven Vorteil gegenüber einer anderen Verbindung für eine Zielgruppe besitzen, nicht weiter zu berücksichtigen sind. Es bleiben lediglich diejenigen Verbindungen in der Datenbank, die zum Beispiel preislich oder in ihrer zeitlichen Verfügbarkeit einen Vorteil gegenüber allen anderen Verbindungen für eine oder mehrere Zielgruppen besitzen.

In der Konsequenz ergibt sich, dass zwischen allen Startbahnhöfen und allen Zielbahnhöfen ein Routensetup erzeugt werden muss, das (mangels konkreter Infrastrukturvorgabe bzw. unbekannter Fahrtrouten) auf den Fahrplänen und damit der Verbindung von im Zuglauf aufeinanderfolgenden Verkehrsstationen beruht. Diese Verkehrsstationen müssen dabei nicht zwingend lokal benachbart sein: Gerade bei Fernverkehrszügen wird der Regelfall sein, dass sie gegenüber Nahverkehrszügen Halte auslassen, sodass Verkehrsstationen zwar immer im Sinne des Modells (durch die Verknüpfung zweier

aufeinanderfolgenden Halte eines Zuglaufs) benachbart sind, nicht zwingend jedoch auch lokal benachbart sind.

Ein Routensetup, das infrastrukturbasiert erfolgen würde, hätte im Übrigen beim hier gegebenen Datenmodell einen größeren Nachteil: Es wäre eine Vielzahl von betrieblichen Durchfahrten zu ignorieren, die das Datenmodell zwar erweiterten, aber keine zusätzlichen Verbindungen erzeugen würden (da dort nicht ein-/aus-/umgestiegen werden kann). Es wäre also bei einer infrastrukturbasierten Darstellung zuerst eine Glättung vorzuschalten, bei der die Infrastrukturmerkmale (bezogen auf den konkreten Laufweg zwischen zwei Verkehrshalten) aus dem Modell entfernt werden müssten. Das Ergebnis dieser Glättung wäre das bereits erläuterte Datenmodell dieser Arbeit unter ausschließlicher Verwendung von Verkehrshalten.

Es ergibt sich nun eine erste verbale Definition des Algorithmus der Verbindungssuche als Menge aller Verbindungen zwischen zwei Verkehrshalten, ausgenommen sind an dieser Stelle lediglich Züge über den Nachtsprung (vgl. Kapitel 4.3). In diesem Zustand des Algorithmus sind allerdings lediglich Direktverbindungen modelliert.

An allen Verkehrshalten sollen zwischen Zügen Umstiege (unter Einhaltung einer Mindestumsteigezeit, im Modell einheitlich fünf Minuten an allen Stationen) ermöglicht werden, um die Routenfindung für Verbindungen anzustoßen, für die keine oder nur ungünstige Direktverbindungen existieren. Unter Berücksichtigung der ökonomisch zielführenden Auslastung der zur Verfügung stehenden Rechnerkapazitäten ist auch eine Maximalumsteigezeit vorzusehen. Es ist daher durchaus möglich, dass Verbindungen existieren, die trotz Überschreitung der vorgegebenen Höchstumsteigezeit noch optimal sind; Die Höchstumsteigezeit ist deswegen so zu setzen, dass möglichst keine sinnvollen Verbindungen ausgeschlossen werden. Da die in das Modell eingehenden Primärdaten im Wesentlichen auf Taktverkehren basieren, die im Regelzustand zweistündlich beplant werden, wird die Höchstumsteigezeit mit 124 Minuten angesetzt – damit sind alle möglichen Umstiege eines Zwei-Stunden-Rasters abgedeckt.¹⁷

Die nächste verbale Definition des Algorithmus der Verbindungssuche besteht aus der Summe aller Verbindungen zwischen zwei Bahnhöfen (Direktverbindungen) sowie dem Bündel aus denjenigen Verbindungen, die an einem beliebigen Bahnhof auf einen beliebigen anderen Zug Anschluss haben und daher mit allen Verbindungen dieses anschließenden Zuges kombiniert werden können und sollen. Ausgeschlossen sind

¹⁷Aufgrund der Mindestumsteigezeit von fünf Minuten müssen Übergänge mit bis vier Minuten im nächsten Taktintervall – im Höchstfall damit nach 124 Minuten – gewährleistet werden.

weiterhin alle Verbindungen über den Nachtsprung – jetzt natürlich auch solche, bei denen der Anschluss über den Nachtsprung erfolgt.

Diese Stufe muss aber zwingend eingeschränkt werden, denn schon bei einer überschaubaren Zahl an Zügen und Zughaltnen ergibt sich durch die an dieser Stelle nicht eingeschränkten Umsteigebeziehungen eine Vielzahl von Verbindungstupeln, die nicht zielführend sind und Ressourcen kosten. Im Beispiel soll jetzt eine Linie mit einem Ausschnitt von vier Halten betrachtet werden:

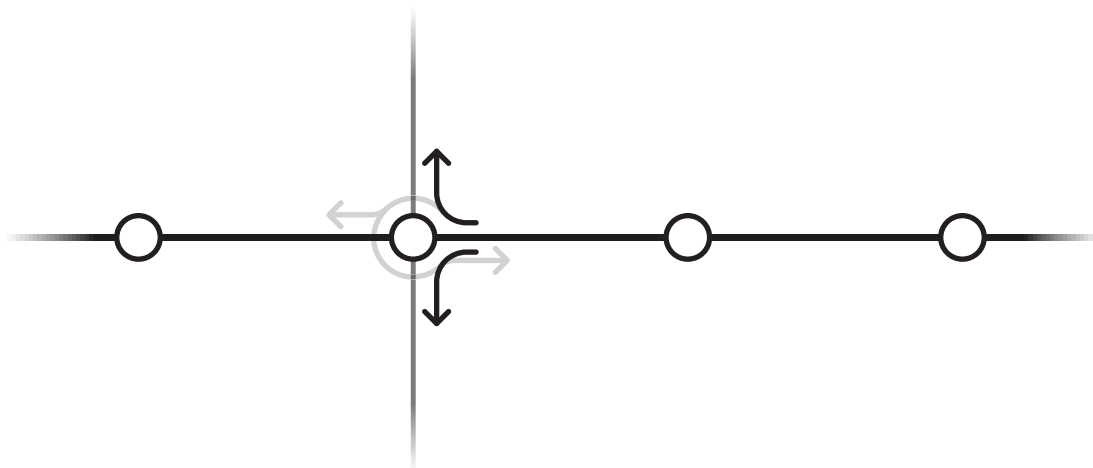


Abbildung 5: Darstellung einer Linie mit vier Halten im Ausschnitt
Quelle: Eigene Darstellung

An jedem dieser Halte werden jetzt die Anschlüsse geprüft, dargestellt ist die Situation in Abbildung 5 für die Anschlusssituation am dritten Halt. Da bislang bedingungsgemäß zu allen Zügen Anschlüsse gewährt werden, würde bei einer Standzeit im Bahnhof von mehr als fünf Minuten auch ein Anschluss an den eigenen Zug gewährt werden. Zuerst auszuschließen sind also Eigen-Anschlüsse (in Abbildung 5 ist der auszuschließende Eigenanschluss grau gekennzeichnet). Außerdem würde Anschluss gewährt an Züge, die direkt wieder zurückfahren. Folglich sind Anschlüsse zu Zügen auszuschließen, deren nächster Halt gleich mit dem letzten Halt des anschlussgebenden Zuges ist (in Abbildung 5 ebenfalls grau dargestellt).

Schließlich kann es einen „Anschluss“ an einen Folgezug mit demselben nächsten Bahnhof geben. Dabei gestaltet sich die Frage nach dem Ausschluss der Anschlüsse nicht mehr ganz so trivial, denn bei zwei parallelen Linien, die sich nach einem gemeinsamen Liniensegment verzweigen, muss es (mindestens) einen Anschlusspunkt geben. Diese Situation ist in Abbildung 6 dargestellt:

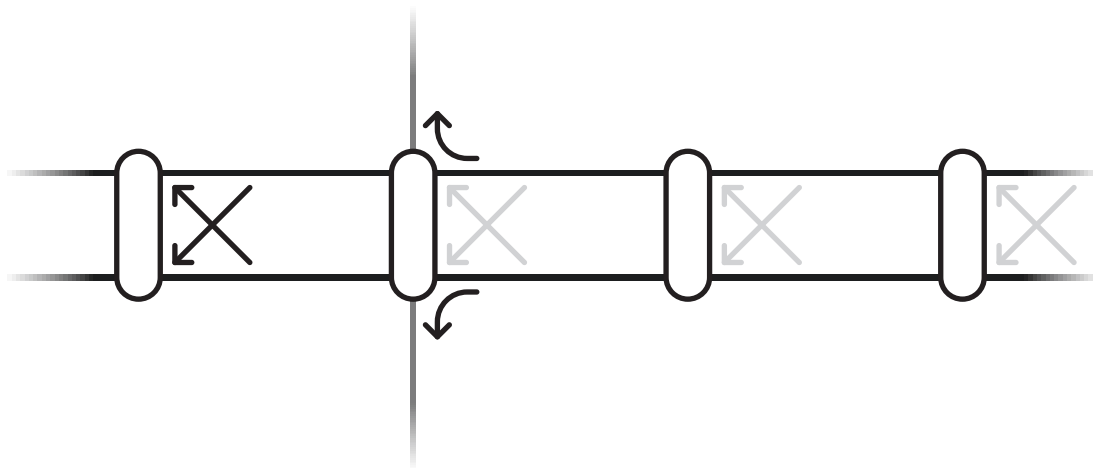


Abbildung 6: Darstellung von zwei Linien mit vier Halten im Ausschnitt
Quelle: Eigene Darstellung

Nutzt man nun für Anschlüsse zweier parallel verlaufender Linien grundsätzlich den letzten gemeinsamen Bahnhof (wie in Abbildung 6 dargestellt, ohne direkten Bezug auf andere Anschlüsse, die an anderen Verkehrsstationen stattfinden), so würde bei einer stattfindenden Überholung der beiden Linien untereinander der Anschluss nie gewährt werden, denn nach der Überholung kann der Anschluss nicht mehr gewährleistet werden, und vor der Überholung war er bedingungsgemäß nicht zu gewährleisten. Eine ähnliche Problematik stellt sich auch, wenn eine der beiden Linien in manchen Verkehrsstationen nicht hält, aber insgesamt mehrere Anstoßpunkte zwischen den beiden Linien existieren: Hier ergibt sich je nach gewünschter Verbindung eine differierende, jeweils ideale Anschlussstation zwischen den Linien, insbesondere bei zusätzlich erfolgenden Überholungen.

Im Folgenden wird diese Bedingung definiert als Ausschluss einer Anschlussgewährung, sofern der Anschluss am folgenden Haltebahnhof unter denselben Bedingungen noch gewährt werden kann. Das impliziert insbesondere eine Gleichrangigkeit der Züge, da ansonsten sowieso eine eindeutige Entscheidung des Trade-Offs Preis vs. Komfort zu treffen ist (sollten die Züge den Anschluss an verschiedenen Stationen gewährleisten können, ist eine der möglichen Lösungen bei unterschiedlichen Zugattungen komfortabler, aber vermutlich auch teurer).

In dieser Ausgestaltung kann der Algorithmus aber noch nicht für mehrere aufeinanderfolgende Anschlüsse verwendet werden, denn dabei ist besonders zu berücksichtigen, dass kein Haltebahnhof doppelt vorkommt. Doppelt befahrene Haltebahnhöfe deuten auf eine Lösung hin, die ohne den Abschnitt zwischen der doppelten Befahrung des

Haltebahnhofs erzeugt werden könnte (ggf. ist das aber regelungstechnisch unterbunden worden, zum Beispiel, weil die Höchstumsteigezeit überschritten wurde), jedoch sollen diese Verbindungen nicht weiter behandelt werden, da sie keinen bestmöglichen im Sinne der Systemlösung darstellen (können). Obwohl diese Verbindungen ggf. sogar optimal sein könnten, müssen sie algorithmisch unterbunden werden, weil es nicht die Aufgabe des Algorithmus sein darf, die ihm gesetzten Regeln zu brechen; In einem solchen Fall wären zuerst die Regeln zu ändern, sodass die optimalen Lösungen zugelassen sind, der Algorithmus aber weiterhin optimal arbeiten kann (ohne doppeltes Befahren einer Kante bzw. eines Knotens).

Technisch wird hierbei ausgenutzt, dass jede optimale Umsteigeverbindung ihrerseits die Kombination mehrerer in sich optimaler Verbindungen darstellt. Diese Kombination ergibt sich nicht zwingend, da zum Beispiel eine langsamere Verbindung unter Umständen denselben Anschluss wie eine schnelle Verbindung erreicht und damit nur dazu führt, dass sich die Umsteigezeit verändert, nicht jedoch die gesamte Reisezeit ändert; daher ist beim Ausschließen von Verbindungen insbesondere darauf zu achten, dass keine sinnvollen Anschlussverbindungen wegfallen. Wenn jedoch die optimalen Verbindungen als Setup aller optimalen oder möglichen optimalen Verbindungen aufgestellt werden, ergibt sich das Tupel aller Anschlussverbindungen definitionsgemäß bei der Nutzung des DIJKSTRA-Algorithmus als Kombination der besten Verbindungen.

Im folgenden Beispiel soll diese Überlegung mit zwei verschiedenen Verbindungen veranschaulicht werden:

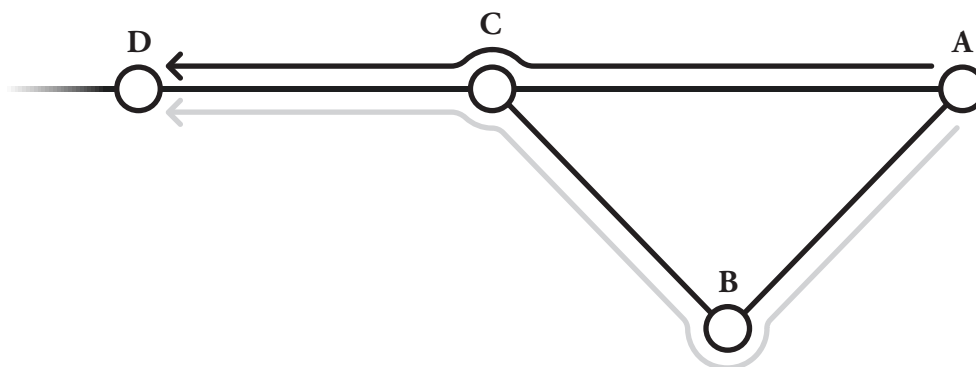


Abbildung 7: Darstellung verschiedener Fahrtmöglichkeiten entlang einer Relation
Quelle: Eigene Darstellung

In Abbildung 7 sind verschiedene Verbindungsmöglichkeiten zwischen den Stationen A und D abgetragen, wobei ein direkter Weg zwischen A und C dargestellt ist und eine andere Verbindung zur Fahrt A – B – C – D. Hierbei können die dargestellten Wege zwischen A und B sowie zwischen B und C für sich genommen optimal sein; der Weg A – B – C – D ist jedoch gegenüber der Fahrt A – C – D nicht vorzugswürdig (und daher ausgegraut dargestellt). Alle ab D weiterführenden Anschlüsse zu anderen Stationen müssen nur noch auf der Verbindung A – C – D aufbauen, nicht jedoch auf der Verbindung A – B – C – D: Da sie für diese Teilverbindung nicht optimal ist, kann auch eine daran anschließende Fahrtmöglichkeit nach einem Umstieg keine optimale Fahrtmöglichkeit darstellen.

Eine anschließende Überlegung könnte sein, zu verhindern, dass die Ausgrenzung von Verbindungen zu streng vorgenommen wird, um weitere Anschlussverbindungen herzustellen, die insgesamt weiterhin vorteilhaft sein könnten; Im folgenden Beispiel ist daher ein exemplarischer Fahrplan dieser Verbindungen und eine mögliche Anschlussverbindung in D dargestellt:

Station	Zug #1	Zug #2	Zug #3
A	12:00	12:02	
B	12:10		
C	12:23	12:20	
D	12:35	12:32	
D			14:29
E			15:00

Tabelle 27: Darstellung eines Auswahlproblems vorzugswürdiger Verbindungen
Quelle: Eigene Darstellung

Bei dem in Tabelle 27 dargestellten Auswahlproblems wäre theoretisch noch nach den weiteren Eigenschaften der Verbindungen zu unterscheiden: Obwohl rein zeitmäßig die Verbindung von Zug #2 der Verbindung von Zug #1 zwischen A und D vorzuziehen wäre, ist der Unterschied zwischen A und E bezogen auf die gesamte Reisedauer de facto vernachlässigbar: Beträgt der Unterschied der Fahrdauer zwischen A und D noch fast 17 % der Fahrdauer, so beträgt der Unterschied zwischen A und E nur noch 1,1 % der gesamten Dauer der Reise. Entsprechend könnten in dieser Relation andere Faktoren wie der Preis der Gesamtverbindung oder der Komfort bezogen auf die erste Teilreise noch eine Veränderung der Vorzugswürdigkeit der ersten Verbindung herbeiführen.

Dieser Ansatz greift gleichwohl wesentlich zu kurz: Die Monetarisierung aller Präferenzen führt bereits bei der Auswahl der Verbindungen zwischen A und D dazu, dass

alle Unterschiede nivelliert werden und eine insgesamt vorteilhafte Verbindung für jede Zielgruppe bestimmt werden kann. Diese Vorteilhaftigkeit ändert sich durch die zweite Teilverbindung nicht mehr, da die zweite Teilverbindung für beide ersten Verbindungen identisch ist und die Auswahl nach der absoluten Höhe der Kriterien erfolgt. Sofern jetzt weitere Verbindungen nach E existieren, ist eine Betrachtung dieser weiteren Bedingungen zwar theoretisch in Verbindung mit beiden ersten Teilverbindungen bis D zu bringen, praktisch ändert sich jedoch auch dort nichts an der Vorzugswürdigkeit der Gesamtverbindung als Tupel der Auswahl der Teilverbindungen.

Der Ausschluss der Verbindung A – E über B bedeutet dabei allerdings nicht, dass die Verbindung B – D – E unter Nutzung von Zug #1 nicht hergestellt werden kann. Ein solcher Ausschluss bezieht sich bedingungsgemäß jeweils lediglich relationsgebunden auf die Nutzung von Zug #1 in der Start-Ziel-Relation A – D.

Die Nutzung der Umsteigeverbindungen mehrerer optimaler Verbindungen reduziert bereits die grundsätzlich zu berücksichtigende Verbindungsanzahl um einen großen Anteil nicht zu berücksichtigender Verbindungen (Verbindungen, die beispielsweise wieder zurückfahren nach einem Umstieg, würden sowieso in einem weiteren Schritt aussortiert, weil sie keinen zeitlichen Vorteil bieten und teurer sind); die Auswirkungen einer solchen Optimierung bezogen auf die beispielhaft gewählte Startstation Stuttgart Hbf ist in der folgenden Tabelle 28 dargestellt:

Szenario	Anzahl der Verbindungen	Anteil
Erzeugung aller Verbindungen	734.527	100,00 %
Erzeugung der Verbindungen aus einem Tupel optimaler Verbindungen	26.827	3,65 %

Tabelle 28: Anteil der zu berücksichtigenden Verbindungen ab Stuttgart Hbf mit einem Umstieg

Quelle: Eigene Darstellung

Im Ergebnis ist die Optimierung als erfolgreich anzusehen: über 96 % der Verbindungen konnten von der Verwendung ausgeschlossen werden, was den Algorithmus entsprechend beschleunigt. Zu erwarten ist außerdem, dass sich dieser Anteil mit jedem weiteren Umstieg erhöht, weil die Anzahl der nicht zielführenden Verbindungen, die wechselseitig aufeinander aufbauen, mit der gesamten Menge der Verbindungen ebenfalls ansteigt.

In der folgenden Tabelle 29 sind die Anteile der Anzahl an Verbindungen mit zwei Umstiegen ab Stuttgart Hbf dargestellt (ohne die Direktverbindungen und ohne Verbindungen mit nur einem Umstieg):

Szenario	Anzahl der Verbindungen	Anteil
Erzeugung aller Verbindungen	85.285.504	100,00 %
Erzeugung der Verbindungen aus einem Tupel optimaler Verbindungen	26.428	0,03 %

Tabelle 29: Anteil der zu berücksichtigenden Verbindungen ab Stuttgart Hbf mit zwei Umstiegen

Quelle: Eigene Darstellung

Die dabei erreichte Quote von lediglich 0,03 % relevanten Verbindungen belegt die Wichtigkeit einer solchen Optimierung der Verbindungssuche. Einen Spezialfall innerhalb dieses Optimierungsszenarios stellt eine Kombination von zwei Anschlussverbindungen dar, bei denen Taktverkehre mit identischen Eigenschaften (Komfort, Preis, Dauer) verkehren, und an der Start- und Zielstation jeweils auf einen unterschiedlichen Taktzug der Linie ein optimaler Anschluss besteht, wie in Tabelle 30 dargestellt:

Bahnhof		Zug #1	Zug #2	Zug #3	Zug #4
Berlin Hbf	ab	08:20			
Hannover Hbf	an	10:15			
Hannover Hbf	ab		10:20	10:50	
Dortmund Hbf	an		11:20	11:50	
Dortmund Hbf	ab				11:55
Köln Hbf	an				13:05

Tabelle 30: Darstellung eines Umsteige-Kombinationsproblems

Quelle: Eigene Darstellung

In diesem Fall ergibt sich nie eine eindeutige optimale Lösung für die durchgehende Verbindung, da an einem der beiden Anschlussbahnhöfe – im Beispiel entweder in Hannover Hbf oder in Dortmund Hbf – ein Taktintervall zwischen Zug #2 und Zug #3 (in diesem Fall also 30 Minuten) als zusätzliche Anschlusswartedauer zur minimalen Anschlusszeit von 5 Minuten in beiden Stationen abgewartet werden muss. Dies gilt selbstverständlich nur unter der Voraussetzung, dass neben den in der Tabelle aufgeführten Verbindungen keine weiteren Verbindungen existieren, die eine optimale Lösung herleiten könnten; in diesem Sinne ist die Tabelle also für die Routenfindung als abschließend anzusehen.

Technisch werden daher neben den tatsächlich optimalen Anschlüssen auch alle „parallel“ zu diesen verkehrende Verbindungen betrachtet und vorgehalten (in der Tabelle parallelverbindungen); Parallel bezeichnet dabei Verbindungen, die zwischen denselben Verkehrsstationen verkehren, unterschiede Abfahrts- oder Ankunftszeiten aufweisen, ansonsten aber dieselben weiteren Eigenschaften (Preis, Dauer, Komfort) besitzen, denn lediglich in diesem speziellen Fall ergibt sich unter Umständen (abhängig von den Folgeanschlüssen) keine eindeutig beste Lösung, da dann keine der beiden Teillösungen absolut vorteilhaft ist.

Dabei ergibt sich zwar auch eine Vielzahl absolut unvorteilhafter Verbindungen (weil diese schlechter sind als die Optimalverbindung, ohne gleichzeitig für irgendeine Zielgruppe einen Vorteil zu bieten), diese können jedoch bereits zur Laufzeit bei der Erstellung abgefangen werden und stören somit weder im Sinne der Performance noch im Sinne der Nachvollziehbarkeit das Ergebnis nachhaltig. Insbesondere im Sinne des wissenschaftlichen Anspruchs der Vollständigkeit sind diese Verbindungen daher in einem ersten Schritt noch mit zu berücksichtigen.

6.2 Verbindungserzeugung

Zur Erzeugung aller nützlichen Verbindungen wird im Folgenden zuerst der Algorithmus zur Erzeugung einer breiten Menge an Verbindungen vorgestellt und danach Algorithmen entwickelt, um die Teilmenge der nicht zielführenden Verbindungen zu eliminieren.

6.2.1 Vollständige Erzeugung des Verbindungstupels

Um nun aus allen erzeugten Verbindungen lediglich diejenigen auszuwählen, die für die weitere Betrachtung relevant sind, werden Beurteilungskriterien aus den Eckdaten der Monetarisierung der Auswahlpräferenzen (siehe Kapitel 4.1) erzeugt. Grundsätzlich werden dazu alle Verbindungen behalten, die ein lokales zeitliches Optimum verglichen mit allen weiteren existenten Verbindungen darstellen, mithin alle Verbindungen, die während ihrer Laufzeit nicht durch eine weitere Verbindung überholt werden. Dies geschieht aufgrund der Tatsache, dass die Zielgruppe der Geschäftsreisenden besonders zeitsensibel ist und daher ggf. auch für nur kleine Zeitgewinne im Verhältnis größere Aufpreise zu zahlen bereit ist. Außerdem werden alle Verbindungen beibehalten, die weniger Umstiege haben als zeitlich und preislich vorteilhaftere Verbindungen mit mehr Umstiegen, da auch dieses Kriterium je nach Zielgruppe ggf. stark gewichtet wird.

Technisch wird so umgesetzt, dass während der Erzeugung der Verbindungen bereits geprüft wird, ob diese ausschlusswürdig wären; dazu wird im folgenden Algorithmus 1 neben der reinen Erzeugung der Umsteigeverbindungen (als Kreuzprodukt der Datenbanktabellen `verbindungen_mem` und `verbindungen_mem_umstiege`, deren Funktion noch erläutert wird) auch bereits eine Prüfung auf deren Verwendbarkeit durchgeführt:

Algorithmus 1 Erzeugung von Umsteigeverbindungen mit mehreren Umstiegen

```
INSERT INTO umstiege_dv_mem
    (dv1, dv2, dv3, bf_von, bf_nach, Mo, Di, Mi, Do, Fr,
     Sa, So, ab, an, Dauer, virtPreis, Vergleich)
SELECT
    u1.dv1, d2.id, d3.id, u1.bf_von,
    d3.bf_nach, d1.Mo*d2.Mo*d3.Mo,
    d1.Di*d2.Di*d3.Di, d1.Mi*d2.Mi*d3.Mi,
    d1.Do*d2.Do*d3.Do, d1.Fr*d2.Fr*d3.Fr,
    d1.Sa*d2.Sa*d3.Sa, d1.So*d2.So*d3.So,
    u1.ab, d3.an, (d3.an-u1.ab),
    (d1.virtPreis + d2.virtPreis + d3.virtPreis + 10 +
     (d3.an - u1.ab - d1.Dauer_Min - d2.Dauer_Min
      - d3.Dauer_Min) * 0.1),
    CONCAT(d3.bf_nach, '.', ROUND((u1.ab+d3.an)/30))
FROM verbindungen_mem u1
INNER JOIN verbindungen_mem_umstiege u2
    ON u2.dv1 = u1.dv2
INNER JOIN direktverbindungen_mem d1
    ON d1.id = u1.dv1
INNER JOIN direktverbindungen_mem d2
    ON d2.id = u2.dv1
INNER JOIN direktverbindungen_mem d3
    ON d3.id = u2.dv2
LEFT JOIN verbindungen_mem_gesamt u3
    ON u3.bf_von = d1.bf_von
    AND u3.bf_nach = d3.bf_nach
    AND u3.ab >= d1.ab
    AND u3.an <= d3.an
    AND u3.virtPreis <
        (d1.virtPreis + d2.virtPreis + d3.virtPreis + 10
         + (d3.an - u1.ab - d1.Dauer_Min - d2.Dauer_Min
          - d3.Dauer_Min) * 0.1)
    AND u3.Mo >= (d1.Mo*d2.Mo*d3.Mo)
    AND u3.Di >= (d1.Di*d2.Di*d3.Di)
    AND u3.Mi >= (d1.Mi*d2.Mi*d3.Mi)
    AND u3.Do >= (d1.Do*d2.Do*d3.Do)
    AND u3.Fr >= (d1.Fr*d2.Fr*d3.Fr)
    AND u3.Sa >= (d1.Sa*d2.Sa*d3.Sa)
    AND u3.So >= (d1.So*d2.So*d3.So)
WHERE
    u3.ID IS NULL
    AND d3.bf_nach <> d1.bf_von
```

Wenn man den Algorithmus rudimentär in Worte übersetzen wollte, ergibt sich ungefähr folgender Wortlaut: „Nimm alle bereits existierenden Verbindungen mit einem Umstieg ab einem bestimmten Bahnhof (gespeichert in `verbindungen_mem`, mit dem Kürzel `u1`), und füge an sie eine Verbindung mit einem Umstieg an (diese werden in `verbindungen_mem_umstiege` gespeichert), wenn die zweite Verbindung der ersten Tabelle (= der anschlussaufnehmende Zug in der ersten Tabelle) und die erste Verbindung der zweiten Tabelle (= der anschlussgebende Zug der zweiten Tabelle) dieselbe ist“ (das entspricht der Verknüpfung `„INNER JOIN [...] ON u2.dv1 = u1.dv2“`; der Punkt führt dabei die Zuordnung einer Datenbankspalte zu einer definierten Datenbanktabelle herbei und wird notwendig, wenn ein Spaltenname in verschiedenen Tabellen auftaucht oder Tabellen doppelt referenziert werden).

In der folgenden `WHERE`-Bedingung wird geprüft, dass sich zum einen Start- und Zielstation der neu gebildeten Verbindung unterscheiden (`„d3.bf_nach <> d1.bf_von“`) sowie, dass es keine Verbindung in der Tabelle `verbindungen_mem_gesamt` gibt (`„u3.ID IS NULL“` negiert hierbei die Verbindungsprüfung), die einen günstigeren virtuellen Preis hat (gebildet über die Summe der einzelnen virtuellen Preise der Routensegmente zzgl. des kumulierten Umsteigewiderstands), mindestens an den Verkehrstagen dieser Verbindung verkehrt sowie außerdem gegenüber der zu prüfenden Verbindung keine zeitlichen Nachteile besitzt.¹⁸ Dass zur Prüfung der Verbindungen die Tabelle `verbindungen_mem_gesamt` genutzt werden kann, liegt daran, dass dort alle bereits erstellten Verbindungen ab einem bestimmten Bahnhof festgehalten werden; Da sie bereits erstellt sind, haben sie grundsätzlich einen Umstieg weniger als die aktuell zu erstellenden Verbindungen.

Mit dieser Einschränkung auf mindestens dieselben Verkehrstage kann insbesondere der häufig vorkommende Fall, dass Züge am Wochenende trotz derselben Zeitlage unter einer abweichenden Zugnummer verkehren (zum Beispiel, da sie zu einem abweichenden Endbahnhof fahren), nicht abgefangen werden – denn in `u3` wird darauf geprüft, dass die alternative Verbindung mindestens an allen Fahrtagen der zu erzeugenden Verbindung verkehrt. Dieser Fall muss weiterhin nach der Erstellung der Umsteigeverbindungen aussortiert werden.

Technisch wäre es auch möglich, diesen Fall bereits zur Laufzeit des Algorithmus bei der Erstellung der Verbindungen abzufangen; dazu wäre ein weiterer `JOIN` auf die Tabelle `verbindungen_mem_gesamt` vorzunehmen, wobei dann die Verkehrstage

¹⁸Dass bei der Verknüpfung der Tabellen (`„JOIN“`) auch auf den Startbahnhof geprüft wird (`„ON u3.bf_von = d1.bf_von“`), ist semantisch überflüssig, aber technisch bedingt.

der ersten und die Verkehrstage der zweiten Verbindung gemeinsam die Umsteige Verbindung abdecken müssten. Um alle Verbindungen abfangen zu können, müssten allerdings sieben Verknüpfungen verwendet werden – eine je Verkehrstag.

Im Sinne der Ressourcenökonomik ist die Berücksichtigung kombinierter Verbindungen mehrerer Verkehrstage an diesem Punkt jedoch abzulehnen, da die Performance durch die Integration gesamter Tabellen erheblich leidet und hier (aufgrund der Multiplikation der Daten) eine Vervielfachung der Daten in einer zur Berechnung hilfsweise aufgespannten Kreuztabelle erfolgen würde, während die Anzahl der nicht erzeugten Verbindungen demgegenüber nicht in dem Maße relevant erscheint, da viele Verbindungen täglich oder mindestens werktätig angeboten werden. Auch für den Fall, lediglich zwei sich ergänzende Verbindungen zu betrachten (beispielsweise eine werktägliche und eine am Wochenende verkehrende Verbindung), würde der kombinierte Algorithmus einer späteren Aussortierung unterlegen sein: Das liegt am dann erforderlichen weiteren JOIN auf die Tabelle `verbindungen_mem_gesamt`, die dadurch dann zwei Mal für alle infrage kommenden Verbindungen des Start-Ziel-Tupels im Arbeitsspeicher gehalten werden müsste und den Prozess aufgrund der bereits dargestellten SQL-Logik von Kreuztabellen erheblich verlangsamen würde.

Dem vorgestellten Algorithmus 1 kann jetzt noch die Prüfung auf die Erzeugung von Verbindungen mit mindestens einer beteiligten als relevant eingestuften Verkehrsstation hinzugefügt werden, dazu sind im Algorithmus zusätzlich die Zeilen

```
INNER JOIN relevante_bahnhoefe r1
  ON r1.bf = d3.bf_nach
```

einzubringen, mit denen darauf geprüft wird, dass der Zielbahnhof eine relevante Station darstellt. Die Integration ist dabei technisch so gelöst, dass dieser Algorithmusteil nur angewandt wird, sofern die beteiligte Startstation als nicht-relevant eingestuft wurde; Bedingung für die Erstellung von Verbindungen ist dabei, dass mindestens eine der beteiligten Stationen als relevant eingestuft werden muss, sodass sich bei einer als relevant eingestuften Startstation die Nutzung des Algorithmusteils erübrigt.

6.2.2 Identifikation unvorteilhafter Verbindungen

In einem zweiten Schritt werden die Verbindungen jetzt untereinander verglichen und jeweils dann ausgeschlossen, wenn es eine vorteilhaftere Verbindung gibt, die gleichzeitig überholt bzw. mindestens gleich schnell ist (also nicht früher abfährt und

nicht später ankommt; ansonsten könnte eine weitere Verbindung einen Vorteil im Sinne der zeitlichen Verfügbarkeit – ggf. auch nur für wenige Minuten – darstellen und wäre daher zu behalten). Eine vorteilhafte Verbindung wird hierbei über den sogenannten *virtuellen Preis* definiert, der die Mindestpräferenzen des Fahrpreises, von Komfort und Reisezeit monetarisiert und miteinander vereint. Mit diesem virtuellen Preis wird quasi eine gemeinsame „Mindestbasis“ geschaffen, die komfortseitig durch die Umsteigepräferenz von Verbindungen mit wenigen Umstiegen eingeschränkt wird sowie zeitenseitig durch die Fahrgastpräferenz von schnellen Verbindungen flankiert wird (eine ausführliche Definition des virtuellen Preises findet sich in Kapitel 4.2).

Dieses System wird nun anhand mehrerer folgender Beispiele erläutert, zuerst in Tabelle 31 anhand zweier beispielhafter Verbindungen zwischen Berlin Hbf und Bremen Hbf:

	Start	ab	an	Umsteigestation	ab	an	Ziel
#1	Berlin Hbf	09:02	(direkt)			12:15	Bremen Hbf
#2	Berlin Hbf	09:00	10:58	Hannover Hbf	11:04	12:20	Bremen Hbf

Tabelle 31: Auswahlproblem verschiedener Verbindungen, erstes Beispiel
Quelle: Eigene Darstellung

In Tabelle 31 sind zwei unterschiedliche Verbindungen abgetragen: Die erste Verbindung verkehrt direkt zwischen Berlin Hbf und Bremen Hbf (gegebenenfalls mit hier nicht dargestellten Zwischenhalten), die zweite Verbindung wird als Umsteigeverbindung mit Umstieg in Hannover Hbf angeboten (das entspricht der Situation, wie sie bei der Erzeugung von Daten mit Algorithmus 1 entsteht). Da die erste Verbindung die zweite überholt, entscheidet bei gleichen Verkehrstagen der virtuelle Preis beider Verbindungen über die Sinnhaftigkeit der zweiten Verbindung: ist der virtuelle Preis der ersten Verbindung gleich oder kleiner dem virtuellen Preis der zweiten Verbindung, so ist Verbindung #2 irrelevant (weil für alle Zielgruppen nicht mehr relevant) und kann gelöscht werden. Ist der virtuelle Preis der zweiten Verbindung jedoch geringer, kann es eine Zielgruppe geben, für die diese Verbindung relevant wird (zum Beispiel preissensible Fahrgäste), und sie ist zu behalten.

Die Verwendung des virtuellen Preises impliziert hier insbesondere, dass der reale Preis nur eine untergeordnete Rolle spielt: Wird er vom Komfortunterschied oder dem wegfallenden Umsteigehindernis kompensiert, kann der zu entrichtende Fahrpreis von Verbindung #1 auch höher sein als der von Verbindung #2; selbst preissensible Fahrgäste würden dann bei einem geringen Preisunterschied die komfortablere Verbindung wählen,

wenn der (hohe) Komfortunterschied den (geringen) Preisunterschied im Sinne des Nutzerempfindens überkompensiert.

Im folgenden Beispiel ändert sich nun die Abfahrtszeit der zweiten Verbindung so, dass keine zeitliche Überholung der Verbindungen mehr stattfindet:

	Start	ab	an	Umsteigestation	ab	an	Ziel
#1	Berlin Hbf	09:02		(direkt)		12:15	Bremen Hbf
#2	Berlin Hbf	09:05	10:58	Hannover Hbf	11:04	12:20	Bremen Hbf

Tabelle 32: Auswahlproblem verschiedener Verbindungen, zweites Beispiel
 Hinweis: Änderungen gegenüber der letzten Tabelle sind mit **Fettdruck** markiert.
 Quelle: Eigene Darstellung

Die in Tabelle 32 dargestellte Situation unterscheidet sich von der in Tabelle 31 dargestellten Situation nur insoweit, als die Startzeit der zweiten Verbindung um fünf Minuten verlegt wurde: Der Zug startet neu erst um 9:05. Unabhängig von den virtuellen Preisen der Verbindungen ergibt sich jetzt für beide Verbindungen ein optimales Einsatzfeld, abhängig von den persönlichen Präferenzen und Rahmenbedingungen der Reisenden (zum Beispiel feststehenden Terminen), aber auch abhängig von weiterführenden Umsteigeverbindungen. So wäre zum Beispiel denkbar, dass ein Anschlusszug Bremen Hbf um 12:20 Uhr verlässt (in diesem Fall wäre nur Verbindung #1 in der Lage, den Anschluss aufzubauen), oder, dass ein Anschlusszug Berlin Hbf erst um 9:00 Uhr erreicht, womit nur zu Verbindung #2 noch umgestiegen werden kann, da für Verbindung #1 die Umsteigezeit bedingungsgemäß zu knapp wäre. Mithin sind in diesem Szenario „Entscheidung während der Erstellung von Verbindungen“ beide Verbindungen zu behalten.

Nach erfolgter Erstellung aller Verbindungen ist jetzt noch eine weitere Abwägung zu treffen: Rechtfertigt der Unterschied der virtuellen Preise, beide Verbindungen zu behalten? Auch im Falle von feststehenden Terminen könnte eine (geringe) Möglichkeit bestehen, Zeiten zu justieren und daher um wenige Minuten flexibel zu sein. Insbesondere bei einem hohen Preisunterschied oder einem hohen Komfortunterschied könnte die Bereitschaft, die Zeitjustierung vorzunehmen und damit auf die unvorteilhaftere Verbindung in jedem Fall zu verzichten, vorhanden sein. Nach der Erstellung aller Verbindungen besteht also die Möglichkeit, abhängig vom Unterschied des virtuellen Preises die Anzahl der Verbindungen, die – weil für den Nutzer potentiell interessant – in der Routenauswahl berücksichtigt werden, nochmals zu reduzieren.

Im folgenden dritten Beispiel ändern sich die Zeiten der Direktverbindung um wenige Minuten:

	Start	ab	an	Umsteigestation	ab	an	Ziel
#1	Berlin Hbf	09:05		(direkt)		12:20	Bremen Hbf
#2	Berlin Hbf	09:05	10:58	Hannover Hbf	11:04	12:20	Bremen Hbf

Tabelle 33: Auswahlproblem verschiedener Verbindungen, drittes Beispiel
Quelle: Eigene Darstellung

Im Beispiel aus Tabelle 33 ergibt sich die Situation, dass beide Verbindungen zeitgleich starten und ankommen. In diesem Fall gibt es nur zwei potentielle Möglichkeiten: Entweder beide virtuellen Preise sind gleich (was aufgrund des Umsteigehindernisses der Verbindung #2 bedeuten müsste, dass Verbindung #1 entweder teurer oder unkomfortabler wäre) – dann wäre die Entscheidung ambivalent –, oder die virtuellen Preise unterscheiden sich. In diesem letzten Fall wäre theoretisch nur eine der beiden Verbindungen vorteilhaft, nämlich die mit dem günstigeren virtuellen Preis. Es muss dabei jedoch geprüft werden, ob die umsteigeärmere Verbindung dieses Kriterium erfüllt, denn bedingungsgemäß ist die Verbindung mit weniger Umstiegen auch dann zu behalten, wenn sie für den „Durchschnittskunden“ nicht die im Sinne des virtuellen Preises vorteilhaftere Verbindung darstellt, weil spezifische Kundengruppen einen hohen Umsteigewiderstand besitzen und damit die Verbindung #1 unter Umständen trotz eines höheren virtuellen Preises für einzelne Kundengruppen vorzugswürdig wäre.

Im folgenden vierten Beispiel ist noch der Fall zweier Verbindungen mit gleichen Eckdaten, aber unterschiedlichem Umsteigepunkt zu betrachten; hierzu wird der ersten Verbindung ebenfalls ein Umstieg zugewiesen, allerdings in einem von der zweiten Verbindung abweichenden Umsteigeort:

	Start	ab	an	Umsteigestation	ab	an	Ziel
#1	Berlin Hbf	09:05	10:45	Hamburg Hbf	11:30	12:20	Bremen Hbf
#2	Berlin Hbf	09:05	10:58	Hannover Hbf	11:04	12:20	Bremen Hbf

Tabelle 34: Auswahlproblem verschiedener Verbindungen, viertes Beispiel
Quelle: Eigene Darstellung

Im diesem Beispiel (dargestellt in Tabelle 34) stellt sich die Lage wie folgt dar: Beide Verbindungen starten zeitgleich und sind auch zur selben Zeit am Zielbahnhof, beide Verbindungen sind auch mit einem Umstieg behaftet. Sollte der virtuelle Preis gleich sein, wäre die Entscheidung wieder ambivalent. Die Dauer des Umstiegs kann dabei wie

auch die Eigenschaften des Umsteigebahnhofs einen Einfluss auf die Routenwahl des Fahrgasts ausüben. Grundsätzlich wird innerhalb des Datenmodells dieser Arbeit davon ausgegangen, dass Fahrgäste keine Präferenz für einen speziellen Umsteigebahnhof besitzen, also bezüglich des Umsteigeorts indifferent sind.

Die Umsteigedauer wird als Entscheidungskriterium genutzt, und einer als Umsteigezeit verbrachten Zeit höhere virtuelle Kosten zugeordnet als einer Fahrzeit im Fahrzeug. Dies spiegelt den Fakt wider, dass die in Umsteigebahnhöfen verbrachte Zeit in der Regel für den Fahrgast keine regelmäßig plan- und nutzbare Zeit darstellt, also zum Beispiel keine Arbeitsinfrastruktur bereitsteht. Es mag trotzdem im Einzelfall Fahrgäste geben, die einer längeren Umsteigezeit den Vorzug gegenüber einer knapperen Verbindung geben, zum Beispiel zur Erhöhung der Stabilität der Verbindung oder zur Nutzung der Umsteigezeit für weitere Zwecke wie der Einnahme von Mahlzeiten oder dem Erwerb von Reiseartikeln.

Zusammengefasst wird im Rahmen dieser Arbeit bei gleichdauernden Verbindungen mit denselben Eckdaten (Abfahrtszeit, Ankunftszeit, virtueller Preis) derjenigen Verbindung der Vorzug gegeben, die die kürzere Umsteigezeit besitzt, da davon ausgegangen wird, dass der Effekt der eingesparten Umsteigezeit komfortmäßig andere Optionen (wie zum Beispiel Einkaufsgelegenheiten) überwiegt, weil so die Zeit im Zug verbracht wird. Im Falle gleicher zeitlicher Eckdaten, aber differierender virtueller Preise wird ausschließlich diejenige Verbindung mit dem geringeren virtuellen Preis (mit dem geringeren Reisewiderstand) genutzt werden; Die jeweils andere Verbindung besitzt keinen im Modell zu berücksichtigenden Vorteil (beachte hierzu auch die einleitenden Ausführungen in Kapitel 5).

6.2.3 Ausschluss unvorteilhafter Verbindungen mit vollständiger Verkehrstageüberdeckung

Die mit dem bereits erläuterten Algorithmus 1 erzeugten Verbindungen müssen jetzt mit weiteren Algorithmen behandelt werden, um die Vielzahl erzeugter, jedoch in keiner Situation vorteilhafter Verbindungen wieder zu entfernen. Hierbei ist insbesondere die Verbindungsauswahl von Verbindungen desselben Umstiegs relevant, denn diese Verbindungen wurden ohne Rücksicht aufeinander erzeugt, sodass zu vermuten steht, dass eine Vielzahl dieser Verbindungen nicht zielführend ist. Der Blick auf die Zahlen kann bereits bestätigen, dass durch die Erzeugung nur selektiver Verbindungen schon eine erhebliche Einsparung eintritt; In Tabelle 35 sind die Verbindungszahlen für alle

Verbindungen Stuttgart Hbf mit drei Umstiegen mit und ohne Verbindungsprüfung abgetragen:

Szenario	Anzahl der Verbindungen	Anteil
ohne Entfernung von Verbindungen	76.310	100,0 %
mit Entfernung der Verbindungen nur bei der Erzeugung	34.047	44,6 %

Tabelle 35: Abtrag der Verbindungszahlen für Stuttgart Hbf mit einer Auswahlstufe
Quelle: Eigene Darstellung

Es wird deutlich sichtbar, dass bereits die Erstellung nur derjenigen Verbindungen, für die keine absolut vorteilhaftere Verbindung mit weniger Umstiegen existiert, im Beispiel schon über 55 % der zu erzeugenden Verbindungen einspart.¹⁹ Die bisher behaltene Verbindungen müssen nun zuerst um jene Verbindungen desselben Umstiegs bereinigt werden, die in keinem Fall vorteilhaft sind. Dazu werden zwei verschiedene Kategorien von Algorithmen genutzt; zum einen sind die Verbindungen desselben Umstiegs gegeneinander auszuspielen, zum anderen fehlt bedingungsgemäß noch der Vergleich der aktuell erzeugten Verbindungen mit den früher erzeugten Verbindungen mit weniger Umstiegen, sofern die Verkehrstage der neu erzeugten Verbindungen nicht mindestens denen der bereits existierenden Verbindung entsprechen (dann wäre die Verbindung aufgrund der WHERE-Klausel in Algorithmus 1 nicht erzeugt worden).

Der erste vorzustellende Algorithmus behandelt die Auswahl derjenigen Verbindungen, die gegenüber anderen Verbindungen desselben Umstiegs nicht vorteilhaft sein können, da sämtliche geprüften Eckdaten nicht vorteilhafter sind; Von den Eckdaten Dauer, virtueller Preis und Umstieg werden hier allerdings bedingungsgemäß nur die Dauer und der virtuelle Preis berücksichtigt, da die Anzahl der Umstiege bei allen zu vergleichenden Verbindungen gleich sein soll (nur auf diese wird referenziert). Alle neu gebildeten Verbindungen befinden sich in einer temporären Datenbanktabelle namens `umstiege_dv_mem` und werden innerhalb dieser direkt auf Basis der Kriterien „Überholung“ bzw. „selbe Reisezeit“ („WHERE [...] u1.Dauer >= u2.Dauer AND u1.ab <= u2.ab AND u1.an >= u2.an“) und „höherer virtueller Preis“ („WHERE u1.virtPreis > u2.virtPreis“) aussortiert:

¹⁹Der Anteil der vorteilhaften Verbindungen wird sich für jede Kombination aus Startstation und Anzahl an Umstiegen unterscheiden.

Algorithmus 2 Aussortierung von Verbindungen desselben Umstiegs bei kompletter Überdeckung der Verkehrstage

```
DELETE u1
FROM umstiege_dv_mem u1
INNER JOIN umstiege_dv_mem u2
    ON u1.Vergleich = u2.Vergleich
WHERE
    u1.virtPreis > u2.virtPreis
    AND u1.Dauer >= u2.Dauer
    AND u1.ab <= u2.ab
    AND u1.an >= u2.an
    AND u1.Mo <= u2.Mo
    AND u1.Di <= u2.Di
    AND u1.Mi <= u2.Mi
    AND u1.Do <= u2.Do
    AND u1.Fr <= u2.Fr
    AND u1.Sa <= u2.Sa
    AND u1.So <= u2.So
```

Aus Performancegründen wird der Vergleich der Verbindungen nicht über den (möglichen) Abgleich „ON u1.bf_nach = u2.bf_nach“ (in Worten: „alle Verbindungen, bei denen der Zielbahnhof übereinstimmt“; der Startbahnhof stimmt bei allen Verbindungen in der Tabelle `umstiege_dv_mem` bedingungsgemäß überein, da die Verbindungen jeweils startbahnhofsweise erstellt werden) durchgeführt, sondern über einen speziell erzeugten Vergleichsparameter (hier in der Spalte `Vergleich` gespeichert). Der Vergleichsparameter wird aus dem Zielbahnhof und der Ankunftszeit gebildet und in drei verschiedenen Stufen – 60 Minuten, 240 Minuten und ein ganzer Tag – in ein gröberes zeitliches Raster überführt. So werden die übereinstimmenden Verbindungen, die in ihrer Ankunft nahe zusammenliegen, zuerst schnell entfernt, ohne dass eine besonders große temporäre Datenbanktabelle gebildet werden muss, da die sich bildende temporäre Tabelle jeweils das Kreuzprodukt aus allen Einträgen beider Tabellen darstellt, bei denen die Bedingungen (zum Beispiel dasselbe 60-Minuten-Raster und dieselbe Zielstation) übereinstimmen.

6.2.4 Ausschluss unvorteilhafter Verbindungen mit teilweiser Verkehrstageüberdeckung

Nach der dargestellten ersten Sortierung werden nun in einem weiteren Algorithmus alle Verbindungen erfasst, die dieselben Rahmenbedingungen (selbe Zielstation, höherer

virtueller Preis und bessere zeitliche Verfügbarkeit) des ersten Algorithmus erfüllen würden, aber nicht an allen Verkehrstagen der weniger vorteilhaften Verbindung von der vorteilhaften Verbindung überdeckt werden:

Algorithmus 3 Aussortierung von Verbindungen desselben Umstiegs bei partieller Überdeckung der Verkehrstage

```
UPDATE umstiege_dv_mem u1
INNER JOIN umstiege_dv_mem u2
  ON u1.Vergleich = u2.Vergleich
SET
  u1.Mo = IF(u2.Mo=1, 0, u1.Mo),
  u1.Di = IF(u2.Di=1, 0, u1.Di),
  u1.Mi = IF(u2.Mi=1, 0, u1.Mi),
  u1.Do = IF(u2.Do=1, 0, u1.Do),
  u1.Fr = IF(u2.Fr=1, 0, u1.Fr),
  u1.Sa = IF(u2.Sa=1, 0, u1.Sa),
  u1.So = IF(u2.So=1, 0, u1.So)
WHERE
  u1.ab <= u2.ab
  AND u1.an >= u2.an
  AND u1.Dauer >= u2.Dauer
  AND u1.virtPreis > u2.virtPreis
  AND
  (
    u1.Mo = u2.Mo AND u1.Mo = 1
    OR u1.Di = u2.Di AND u1.Di = 1
    OR u1.Mi = u2.Mi AND u1.Mi = 1
    OR u1.Do = u2.Do AND u1.Do = 1
    OR u1.Fr = u2.Fr AND u1.Fr = 1
    OR u1.Sa = u2.Sa AND u1.Sa = 1
    OR u1.So = u2.So AND u1.So = 1
  )
```

Im Unterschied zum Algorithmus der vollständigen Überdeckung der Verkehrstage (Algorithmus 2) kann hier kein direkter Ausschluss mit DELETE angewandt werden, da die weniger vorteilhafte Verbindung (u1) nicht an allen Verkehrstagen von der vorteilhafteren Verbindung (u2) ersetzt werden kann (in jenem Fall wäre sie bereits im Algorithmus der vollständigen Überdeckung abgefangen worden); daher muss der Ausschluss hier verkehrstageweise erfolgen. Verkehrt die vorteilhaftere Verbindung beispielsweise montags, so wird über `u1.Mo = IF(u2.Mo=1, 0, u1.Mo)` die weniger vorteilhafte Verbindung dann nicht mehr angeboten. Die IF-Routine wird von MySQL mit folgendem Aufbau bereitgestellt: Der erste Ausdruck (hier: „u2.Mo=1“) innerhalb

der Klammer wird dabei auf die enthaltene Bedingung geprüft; ist er wahr (also verkehrt die Verbindung u2 montags), so wird der zweite Ausdruck weitergegeben, ist er nicht wahr (also wenn die Verbindung u2 montags nicht verkehrt), der dritte. Die Trennung der Ausdrücke innerhalb der Klammer erfolgt dabei jeweils mit einem Komma.

Dieser Vorgang der Aussortierung der Verbindungen über die Löschung der Verkehrstage muss iterativ erfolgen, da als 'worst case' für jede Verbindung sieben mögliche andere bessere Verbindungen bereitstehen könnten, der Algorithmus jedoch immer nur die Korrelation einer Verbindung auf genau eine andere Verbindung behandelt. Entsprechend wird der Algorithmus sieben Mal ausgeführt, um alle weniger vorteilhaften Verbindungen um die Verkehrstage zu reduzieren, an denen sie tatsächlich weniger vorteilhaft sind.

Die behandelten Verbindungen müssen jetzt noch auf ihre weitere Verwendbarkeit geprüft werden: Verbindungen, die keine vorteilhaften Verkehrstage mehr besitzen, müssen gelöscht werden. Dazu wird ein weiterer kleiner Befehl alle Verbindungen ohne zugeordnete vorteilhafte Verkehrstage aus der Datenbanktabelle `umstiege_dv_mem` entfernen:

Algorithmus 4 Entfernung von Verbindungen ohne vorteilhafte Verkehrstage

```
DELETE u1
FROM umstiege_dv_mem u1
WHERE
    Mo = 0
    AND Di = 0
    AND Mi = 0
    AND Do = 0
    AND Fr = 0
    AND Sa = 0
    AND So = 0
```

Wörtlich übertragen ist der Inhalt aus Algorithmus 4, dass alle Verbindungen der Tabelle `umstiege_dv_mem` zu entfernen sind, bei denen sich die Verbindung an allen Wochentage als unvorteilhaft herausgestellt hat (in diesem Fall wird bedingungsgemäß das Wochentagefeld mit dem Wert 0 belegt). Die mit diesem Algorithmus gelöschten Verbindungen können zu Testzwecken oder Zwecken der wissenschaftlichen Vollständigkeit natürlich auch im Modell verbleiben, wie das regelmäßig bei der Erstellung der Arbeit zur Generierung der Übersichtswerte zu löschender Verbindungen auch vom Autor genutzt wurde. Es wäre damit aber auch möglich, im Nachhinein Präferenzän-

derungen oder Änderungen der Verbindungseigenschaften zu berücksichtigen, sodass diese Verbindungen wieder an Relevanz gewinnen würden.

Die letztgenannte Situation (die Verhinderung des Löschens von Verbindungen) ist insbesondere zu berücksichtigen, wenn es eine bahnbetriebliche Störung zur Laufzeit des Realsystems gibt, also berechnete Verbindungen plötzlich nicht mehr genutzt werden können, weil beispielsweise Züge verspätet sind oder Strecken gesperrt werden und dadurch Fahrplanleistungen entfallen müssen. In diesem Fall kann aus einer eigentlich unvorteilhaften Verbindung aufgrund der geänderten Situation wieder eine vorteilhafte Verbindung werden; Es wäre allerdings darauf zu achten, dann auch alle Verbindungen bei der Erzeugung zu berücksichtigen, damit für die Umleitung der Fahrgäste im Störfall eine vollständige Verbindungsumgebung vorhanden ist. Es wäre aber auch möglich, in der Laufzeit der Simulation die Werte für einen Bahnhof nachträglich anzupassen und dann alle Verbindungen zu generieren, die nicht bereits generiert wurden; da das Datenmodell die Identifikatoren der Direktverbindungen als Werte übernimmt, sind doppelte Verbindungen unter Nutzung derselben Züge und Umsteigestationen schnell erkennbar.

6.2.5 Ausschluss unvorteilhafter Verbindungen mit weniger Umstiegen

In einem weiteren Schritt wird der zweite zur Aussortierung von Verbindungen vorgestellte Algorithmus (Algorithmus 3) darauf angepasst, alle Verbindungen zu berücksichtigen – außer denen des aktuellen Umstiegs, denn diese wurden bereits betrachtet. In diesem Fall wird die Tabelle `umstiege_dv_mem` durch die Tabelle `verbindungen_mem_gesamt` ersetzt, in der zur Laufzeit alle Verbindungen des Startbahnhofs gespeichert werden, die nicht im laufenden Zyklus (das heißt mit der aktuellen Zahl der Umstiege) erzeugt wurden, sondern dieser Erzeugung als Basis dienen:

Algorithmus 5 Aussortierung von Verbindungen mit weniger Umstiegen bei partieller Überdeckung der Verkehrstage

```
UPDATE umstiege_dv_mem u1
INNER JOIN verbindungen_mem_gesamt u2
  ON u1.Vergleich = u2.Vergleich
SET
  u1.Mo = IF(u2.Mo=1, 0, u1.Mo),
  u1.Di = IF(u2.Di=1, 0, u1.Di),
  u1.Mi = IF(u2.Mi=1, 0, u1.Mi),
  u1.Do = IF(u2.Do=1, 0, u1.Do),
  u1.Fr = IF(u2.Fr=1, 0, u1.Fr),
  u1.Sa = IF(u2.Sa=1, 0, u1.Sa),
  u1.So = IF(u2.So=1, 0, u1.So)
WHERE
  u1.ab <= u2.ab
  AND u1.an >= u2.an
  AND u1.Dauer >= u2.Dauer
  AND u1.virtPreis > u2.virtPreis
  AND
  (
    u1.Mo = u2.Mo AND u1.Mo = 1
    OR u1.Di = u2.Di AND u1.Di = 1
    OR u1.Mi = u2.Mi AND u1.Mi = 1
    OR u1.Do = u2.Do AND u1.Do = 1
    OR u1.Fr = u2.Fr AND u1.Fr = 1
    OR u1.Sa = u2.Sa AND u1.Sa = 1
    OR u1.So = u2.So AND u1.So = 1
  )
```

Da die Tabellen `umstiege_dv_mem` und `verbindungen_mem_gesamt` hinsichtlich der für diesen Algorithmus relevanten Spalten identisch aufgebaut sind, sind auch die Algorithmen insoweit gleich. Analog zu Algorithmus 2 könnte nun auch der Schritt der Aussortierung der Verbindung mit kompletter Überdeckung der Verkehrstage auf die Nutzung in `verbindungen_mem_gesamt` migriert werden; Es ergibt sich jedoch, dass bedingungsgemäß die Prüfung der vollständigen Überdeckung von Verbindungen mit weniger Umstiegen bereits bei der Erzeugung (also in Algorithmus 1) durchgeführt und Verbindungen, die nach einer Migration des Algorithmus noch von diesem betroffen wären und damit gelöscht werden könnten, aufgrund der `WHERE`-Bedingung in Algorithmus 1 gar nicht erzeugt worden sind. Mithin kann dieser Schritt also entfallen.

Der Blick auf die Zahlen bestätigt die ursprüngliche Annahme, dass eine Vielzahl der Verbindungen desselben Umstiegs keinesfalls vorteilhaft sind; in Tabelle 36 sind

die Verbindungszahlen für alle Verbindungen Stuttgart Hbf mit drei Umstiegen in verschiedenen Stufen abgetragen:

Szenario	Anzahl der Verbindungen	Anteile	
ohne Entfernung von Verbindungen <i>Nutzung des Algorithmus 1 ohne WHERE-Klausel</i>	76.310	100,0 %	
mit Entfernung der Verbindungen nur bei der Erzeugung <i>Nutzung des Algorithmus 1 mit WHERE-Klausel</i>	34.047	44,6 %	100,0 %
mit Entfernung der Verbindungen bei der Erzeugung sowie nachgeschaltet <i>Nutzung der Algorithmen 1–5</i>	8.064	10,6 %	23,7 %

Tabelle 36: Abtrag der Verbindungszahlen für Stuttgart Hbf mit zwei Auswahlstufen
Quelle: Eigene Darstellung

Nach dem ersten Schritt – der Erstellung lediglich derjenigen Verbindungen, für die keine absolut vorteilhaftere Verbindung mit weniger Umstiegen existiert – werden schon über 55 % weniger Verbindungen erzeugt. Der jetzt nachgeschaltete zweite Schritt mit der Entfernung von Verbindungen ohne relative Vorteile gegenüber anderen Verbindungen sorgt für eine weitere Reduktion um 76,3 %, sodass nach der gesamten nachgeschalteten Auswahl nur noch 11,4 % der originalen Verbindungen insgesamt vorteilhaft sind und damit für die Auswahl der abgehenden Passagierverbindungen ab Stuttgart Hbf betrachtet werden müssen.

6.3 Spezielle Betrachtung der Verbindungen des ersten Umstiegs

Eine besondere Situation ist bei den Verbindungen des ersten Umstiegs festzustellen: Das bereits vorgestellte Konzept der Verknüpfung optimaler Verbindungen mit den idealen Umstiegen ist hierbei nicht anwendbar, da zwei Direktverbindungen miteinander zu kombinieren sind, mithin die Informationen über die Relevanz von Anschlüssen noch nicht über die Verbindungsauswahl vorhanden sind. Entsprechend muss eine andere Abschätzung bezüglich der Relevanz einzelner Umstiege getroffen werden, um das ansonsten zu berechnende vollständige Kreuzprodukt der Tabelle *direktverbindungen* mit sich selbst auf geringere Beträge zu reduzieren. Um die Datenvorhaltung zu optimieren, wer-

den dabei in einer ersten Stufe nicht die Anschlüsse zwischen Verbindungen festgehalten, sondern lediglich zwischen Halte Datensätzen der Tabelle `bp_fp_zeiten`.

Entsprechend ist es natürlich möglich, dass bei der späteren Verbindungserzeugung Verbindungen erstellt werden, die unvorteilhaft sind; gleichzeitig aber wird die Datenvorhaltung entscheidend reduziert, wie Tabelle 37 entnommen werden kann:²⁰

Szenario	Anzahl der Datensätze	Anteil
Kreuzprodukt aller Zeitendatensätze	3.837.395	4,79 %
Kreuzprodukt aller Direktverbindungen	80.074.609	100,00 %

Tabelle 37: Datensumme des Kreuzprodukts (bezogen auf Datensätze desselben Bahnhofs mit höchstens 124 Minuten Umsteigezeit, ohne Eigenanschlüsse)

Quelle: Eigene Darstellung

Tabelle 37 zeigt die Einsparung bei Verwendung der den Direktverbindungen zugrundeliegenden Zeitdaten: Über 95 % der ansonsten zu bildenden Daten müssen jetzt nicht erzeugt werden. Trotz der entstehenden Unschärfe bzgl. entstehender unvorteilhafter Verbindungen ist die Weiterführung der Daten aus dem Kreuzprodukt der Zeiten zu bevorzugen, da die absolute Menge der vorzuhaltenden Daten deutlich geringer ist, die Daten handhabbarer sind und auch aufgrund der geringeren Tabellengröße auch im Arbeitsspeicher vorgehalten werden können, was die weitere Umsteigeberechnung deutlich beschleunigt.

6.3.1 Aufbau der Daten des ersten Umstiegs

Zur Vorhaltung der Umsteigedaten wird die Tabelle `umstiege_zeiten` genutzt, der Algorithmus zur Erzeugung der Daten dieser Tabelle ist im Folgenden abgetragen:

²⁰Für die Berechnung wurde bei den Direktverbindungen noch die Einschränkung verwendet, dass der Startbahnhof der zu erstellenden Umsteige Verbindung nicht ihr Endbahnhof sein darf; weitere Einschränkungen wurden aber – auch beim Kreuzprodukt der Zeitentabelle – nicht getroffen.

Algorithmus 6 Erzeugung der Umsteigedaten des ersten Umstiegs

```
INSERT INTO umstiege_zeiten
  (bf, id_1, id_2, wartezeit)
SELECT
  z1.bf_id,
  z1.id,
  z2.id,
  TIME_TO_SEC((IF(z2.ab>z1.an, TIMEDIFF(z2.ab, z1.an),
    TIMEDIFF(ADDTIME(z2.ab, '24:00:00'), z1.an))))/60
FROM bp_fp_zeiten z1
INNER JOIN bp_fp_zeiten z2
  ON z2.bf_id = z1.bf_id
  AND (IF(z2.ab>z1.an, TIMEDIFF(z2.ab, z1.an),
    TIMEDIFF(ADDTIME(z2.ab, '24:00:00'), z1.an))
    >= '00:05:00')
  AND (IF(z2.ab>z1.an, TIMEDIFF(z2.ab, z1.an),
    TIMEDIFF(ADDTIME(z2.ab, '24:00:00'), z1.an))
    <= '02:04:00')
INNER JOIN bp_fp_zeiten_mem z3
  ON z3.id = z1.vorh_id
INNER JOIN bp_fp_zeiten_mem z4
  ON z4.id = z2.nae_id
  AND z3.bf_id != z4.bf_id
LEFT JOIN bp_fp_zeiten_mem z5
  ON z5.id = z1.nae_id
LEFT JOIN bp_fp_zeiten_mem z6
  ON z6.id = z2.vorh_id
INNER JOIN bp_fp_zug g1
  ON z1.zug_nr = g1.zugnr
INNER JOIN bp_fp_zug g2
  ON z2.zug_nr = g2.zugnr
  AND (g1.Mo*g2.Mo+g1.Di*g2.Di+g1.Mi*g2.Mi+g1.Do*g2.Do
    +g1.Fr*g2.Fr+g1.Sa*g2.Sa+g1.So*g2.So) > 0
WHERE
  z1.an != '00:00:01'
  AND z2.ab != '00:00:01'
  AND z1.zug_nr != z2.zug_nr
  AND (z5.ID IS NOT NULL AND z6.ID IS NOT NULL AND
    (z4.bf_id != z5.bf_id OR z3.bf_id != z6.bf_id)
    OR z5.ID IS NULL OR z6.ID IS NULL)
GROUP BY
  z1.id,
  z2.id
```

In Worten ausgedrückt, werden mit Algorithmus 6 die möglichen Verbindungen des ersten Umstiegs festgehalten; dabei sind jedoch wichtige Einschränkungen zu beachten, die sich aus der Datengrundlage ergeben. Durch den Algorithmus wird zuerst die Kreuztabelle aus der Tabelle der Halte Datensätze `bp_fp_zeiten z1` mit sich selbst (Kürzel `z2`) gebildet, unter der Voraussetzung, dass für die Kreuztabelle (die Bildung erfolgt unter Nutzung der Routine `INNER JOIN`) nur solche Datensätze verwendet werden, bei denen die Bahnhöfe beider Zugläufe übereinstimmen („`ON z2.bf_id = z1.bf_id`“) und die Zeit zwischen Ankunft des ersten Zuges (`z1.an`) und Abfahrt des zweiten Zuges (`z2.ab`) mindestens 5 Minuten und höchstens 124 Minuten beträgt. In der `WHERE`-Bedingung ist festgehalten, dass nur solche Züge verwendet werden, für die überhaupt ein Anschluss existieren kann (also nicht der Zug aus `z1` erst in der als Anschlussstation vorgesehenen Betriebsstelle seine Fahrt beginnt oder der Zug aus `z2` die Fahrt dort beendet) und die Zugnummer der beiden Züge nicht dieselbe sein darf (ansonsten wäre bei Wartezeiten ab 5 Minuten in einem Bahnhof ein Eigenanschluss, also ein „Anschluss“ an den eigenen Zug, möglich).

Die weiteren Bedingungen dienen allesamt der groben Prüfung auf den Ausschluss möglicher Anschlüsse, die keine vorteilhaften Verbindungen erzeugen könnten: Mit dem `INNER JOIN` auf `bp_fp_zeiten_mem z3` und `z4` wird sichergestellt, dass der letzte Haltebahnhof des anschlussgebenden Zuges nicht dem nächsten Bahnhof des anschlussaufnehmenden Zuges entspricht (mit „`AND z3.bf_id != z4.bf_id`“) – denn dann wäre der Anschluss ja auch schon in diesem Bahnhof möglich, ohne die Umwegfahrt zum aktuell betrachteten Bahnhof.

Außerdem wird (sofern diese Daten existieren, dazu dient die Prüfung mit der Routine `LEFT JOIN` statt der Routine `INNER JOIN`) der nächste Haltebahnhof des anschlussgebenden Zuges (in der Tabelle `bp_fp_zeiten_mem z5`) und der letzte Haltebahnhof des anschlussaufnehmenden Zuges (in der Tabelle `bp_fp_zeiten_mem z6`) darauf geprüft, dass keine Parallelfahrten entstehen: entweder die beiden Folgebahnhöfe oder die beiden zuletzt angefahrenen Verkehrsstationen müssen sich unterscheiden („`WHERE [...] z4.bf_id != z5.bf_id OR z3.bf_id != z6.bf_id`“). Wie bereits angesprochen, müssen diese Daten dabei nicht zwingend existieren: Der anschlussgebende Zug kann beispielsweise an der aktuell betrachteten Verkehrsstation enden, sodass dann kein Folgehalt existiert (und damit diese Prüfung auch nicht notwendig wird).

Weiterhin ist für die Erstellung der Anschlussverbindungen darauf zu achten, dass parallel verkehrende Züge im sich überdeckenden Abschnitt Verbindungen doppelt erzeugen. Am

ersten und am letzten möglichen Umsteigepunkt muss jeweils eine Verbindung erzeugt werden. Daher muss für ansonsten gleiche Verbindungen noch eine separate Auswahl stattfinden, da sonst zwei quasi identische Verbindungen aufgenommen werden würden, die sich nur in ihrer Umsteigestation unterscheiden, nicht jedoch in ihren Eigenschaften (bedingungsgemäß würde zwischen Verbindungen mit unterschiedlichen Eigenschaften, aber denselben Start- und Zielzeiten, eine eindeutige Rangfolge ermittelt und nur die vorteilhaftere Verbindung würde behalten werden).

Schließlich wird in Algorithmus 6 noch darauf geprüft, dass die potentielle Anschlussverbindung mindestens einen gemeinsamen Verkehrstag beider Züge aufweist; Dazu wird auf die Zugstammdatensatztabellen `bp_zp_zug_g1` und `g2` verwiesen, wobei das Summenprodukt aller Verkehrstage größer null sein muss (im Algorithmus: „ $(g1.Mo * g2.Mo + g1.Di * g2.Di + g1.Mi * g2.Mi + g1.Do * g2.Do + g1.Fr * g2.Fr + g1.Sa * g2.Sa + g1.So * g2.So) > 0$ “), das heißt mindestens ein gemeinsamer Verkehrstag beider Leistungen existieren muss. Schließlich wird mit der GROUP BY-Bedingung erreicht, dass das Ergebnis nach den Zeitendatensätzen gegliedert wird, sodass jede Kombination zweier Zeitendatensätze nur einmal erzeugt wird.

6.3.2 Ausschluss versetzt parallel verkehrender Züge desselben Startbahnhofs und Ziels

Problematisch an der Nutzung von Algorithmus 6 ist die fehlende Prüfung auf die Vorteilhaftigkeit einer Verbindung: So wird während der Erstellung nicht geprüft, ob bereits ein Umstieg erzeugt wurde, bei dem von demselben anschlussgebenden Zug auf einen früher verkehrenden anschlussaufnehmenden Zug alle Haltebahnhöfe erreicht werden, die ein später verkehrender anschlussaufnehmender Zug erreicht. In diesem Fall wäre der frühere Zug grundsätzlich zu bevorzugen, da der spätere Zug – unter ansonsten unveränderten Bedingungen – keinen Vorteil erfahren könnte, da er versetzt parallel verkehrt: Man wird mit ihm immer später am Ziel sein. Dieser Fall tritt insbesondere bei Taktverkehren im suburbanen Raum häufig auf: Bei einer im 5-Minuten-Takt verkehrenden S-Bahn-Linie werden insgesamt 24 Anschlüsse innerhalb des zu betrachtenden Zwei-Stunden-Intervalls gebildet, von denen lediglich der erste sinnvoll wäre, die folgenden 23 jedoch nicht.

Entsprechend ist eine passende Auswahl und Elimination der unvorteilhaften Umstiege vorzunehmen; dies wird auf Basis einer Tabelle parallel verkehrender Züge

(parallele_zuege) vorgenommen. In dieser Tabelle werden alle Züge festgehalten, die dieselben Haltestationen bedienen und bei denen der zweite Zug zu jeder Zeit dem ersten Zug zeitlich nachgelagert ist. Die Tabelle wird dabei mit Algorithmus 7 so erzeugt, dass zuerst alle möglichen parallelen Züge festgehalten werden, die tatsächliche Prüfung auf einen komplett parallelen Zuglauf erfolgt in einem späteren Schritt:

Algorithmus 7 Erzeugung der Datenbanktabelle `parallele_zuege`

```
INSERT INTO parallele_zuege
    (zug1, zug2)
SELECT
    g1.zugnr,
    g2.zugnr
FROM bp_fp_zug g1
INNER JOIN bp_fp_zug g2
    ON g1.start = g2.start
    AND g1.ziel = g2.ziel
    AND g1.strecke_id = g2.strecke_id
    AND g1.id <> g2.id
    AND g1.Mo >= g2.Mo
    AND g1.Di >= g2.Di
    AND g1.Mi >= g2.Mi
    AND g1.Do >= g2.Do
    AND g1.Fr >= g2.Fr
    AND g1.Sa >= g2.Sa
    AND g1.So >= g2.So
INNER JOIN bp_fp_zeiten z1
    ON z1.zug_nr = g1.zugnr
    AND z1.halt_nr = 1
INNER JOIN bp_fp_zeiten z2
    ON z2.zug_nr = g2.zugnr
    AND z2.halt_nr = 1
    AND z2.ab > z1.ab
    AND TIMEDIFF(z2.ab, z1.ab) <= '02:00:00'
INNER JOIN bp_fp_zeiten z3
    ON z3.zug_nr = g1.zugnr
    AND z3.halt_nr = 2
INNER JOIN bp_fp_zeiten z4
    ON z4.zug_nr = g2.zugnr
    AND z4.halt_nr = 2
    AND z3.bf_id = z4.bf_id
```

In Algorithmus 7 werden alle Paare aus den Zugläufen der Züge der Tabellen `bp_fp_zug` `g1` und `g2` berücksichtigt, bei denen der erste und der zweite Zug auf derselben Linie verkehren („`g1.strecke_id = g2.strecke_id`“), die denselben Start- und Zielbahnhof aufweisen („`g1.start = g2.start AND g1.ziel = g2.ziel`“)

und bei denen der erste (also frühere) Zug g_1 mindestens alle Verkehrstage des späteren Zuges g_2 abdeckt („ $g_1.Mo \geq g_2.Mo$ AND $g_1.Di \geq g_2.Di$ AND $g_1.Mi \geq g_2.Mi$ AND $g_1.Do \geq g_2.Do$ AND $g_1.Fr \geq g_2.Fr$ AND $g_1.Sa \geq g_2.Sa$ AND $g_1.So \geq g_2.So$ “). Außerdem wird darauf geprüft, dass die Identifikatoren der Züge nicht identisch sind, es sich also trotz ggf. identischer Zugnummer um verschiedene Züge handelt („ $g_1.id <> g_2.id$ “).

Es erfolgt ferner eine Prüfung darauf, dass der zweite Zug am ersten Haltebahnhof später verkehrt als der erste Zug (nach der Referenz auf die Haltedatensatztabellen z_1 und z_2 und der Auswahl des jeweils ersten Bahnhofs („ $[...]halt_nr = 1$ “) erfolgt die Prüfung mit „ $z_2.ab > z_1.ab$ “) und der Abstand höchstens zwei Stunden beträgt („ $TIMEDIFF(z_2.ab, z_1.ab) \leq '02:00:00'$ “). Es wird außerdem noch geprüft, dass der zweite Bahnhof der Zugläufe derselbe ist (mit „ $z_3.bf_id = z_4.bf_id$ “); das ist besonders für Linien, die im Ringbetrieb betrieben werden, wichtig, um im Ergebnis nicht zu viele doppelte Zugkorrelationen zu erzeugen.

Bei der Erzeugung dieser Datenbanktabelle paralleler Züge wurde aber nur darauf geachtet, dass die ersten beiden Halte sowie der Zielbahnhof übereinstimmen. Es wäre also denkbar, dass beispielsweise Expresszüge, die Halte auslassen, keine parallelen Züge im Sinne der oben aufgeworfenen Definition sind. Denn dann könnte ein nachfolgender Zug Halte aufweisen, die dem Expresszug fehlen: Mithin wären die Züge im Sinne der Definition nicht streng parallel, sondern überlagerten sich nur auf Teilen. Entsprechend ist eine Logik zu entwickeln, die Zugpaare, bei denen der nachfolgende Zug Halte aufweist, die dem voranfahrenden Zug fehlen, aus der Tabelle `parallele_zuege` entfernt. Diese Logik ist im folgenden Algorithmus 8 dargestellt:

Algorithmus 8 Entfernung nicht streng paralleler Zugpaare aus der Datenbanktabelle `parallele_zuege`

```
DELETE p1
FROM parallele_zuege p1
INNER JOIN bp_fp_zeiten_mem z2
  ON z2.zug_nr = p1.zug2
INNER JOIN bp_fp_zeiten_mem z3
  ON z3.zug_nr = p1.zug1
LEFT JOIN bp_fp_zeiten_mem z4
  ON z2.zug_nr = z4.zug_nr
  AND z4.bf_id = z3.bf_id
WHERE
  z4.id IS NULL
```

Wörtlich beschrieben, wird in Algorithmus 8 für jeden Datensatz paralleler Züge (p1) ein Datensatz aus der Haltezeitendatenbanktabelle `bp_fp_zeiten` `z4` gesucht, der zum später parallel verkehrenden Zug gehört („ON `z2.zug_nr = z4.zug_nr`“), und einen Bahnhof des früher verkehrenden Zuges `z3` nicht aufweist. Diese Negation wird mit der `WHERE`-Bedingung umgesetzt: „WHERE `z4.id IS NULL`“ bezeichnet hierbei technisch das Fehlen eines Datensatzes, der die Bedingungen (Zugnummer des späteren Zuges, Haltebahnhof des früheren Zuges) aufweist. Wenn dieser Datensatz fehlt, kann die aufgeworfene Bedingung nicht erfüllt sein – mithin wird also mindestens ein Haltebahnhof des früher verkehrenden Zuges nicht vom später verkehrenden Zug bedient. Alle Datensätze, die die negierte Bedingung nicht erfüllen, verweisen nicht auf streng parallele Züge, entsprechend sind sie aus der Datenbanktabelle `parallele_zuege` zu löschen.

Die jetzt noch in der Datenbanktabelle `parallele_zuege` enthaltenen Zugpaare sind solche Zugpaare, bei denen zwei Züge einander folgen, dieselben Verkehrsstationen befahren und bei denen der erste Zug mindestens die Verkehrstage des zweiten Zuges aufweist; der zweite Zug mithin sich als anschlussaufnehmender Zug genau dann nie optimal sein kann, wenn auch der erste Zug bereits als anschlussaufnehmender Zug zur Wahl stand. Entsprechend können jetzt alle Umstiege aus der Datenbanktabelle `umstiege_zeiten` gelöscht werden, bei denen ein Zeitendatensatz sowohl auf den ersten als auch auf den zweiten Zug eines Datensatzes aus `parallele_zuege` verweist, denn in diesem Fall kann der zweite anschlussaufnehmende Zug keine optimale Wahl sein und soll daher gelöscht werden:

Algorithmus 9 Löschen von Umsteigealternativen aus der Tabelle `umstiege_zeiten`

```
DELETE u2
FROM umstiege_zeiten u1
INNER JOIN umstiege_zeiten u2
    ON u1.id_1 = u2.id_1
INNER JOIN bp_fp_zeiten z1
    ON z1.id = u1.id_2
INNER JOIN bp_fp_zeiten z2
    ON z2.id = u2.id_2
INNER JOIN parallele_zuege p1
    ON p1.zug1 = z1.zug_nr
    AND p1.zug2 = z2.zug_nr
```

Bei dem in Algorithmus 9 dargestellten Code wird die zuvor bereits in Worten beschriebene Aktion ausgeführt: Eine Anschlussverbindung wird dann aus der Tabelle

`umstiege_zeiten` `u2` gelöscht, wenn zum anschlussaufnehmenden Zug (in diesem Fall mit `z2` referenziert) ein paralleler, früher verkehrender, Zug derselben Linie mit mindestens denselben Wochentagen (`z1`) verkehrt (die Referenzierung erfolgt über eine Parallelinstanz der Tabelle `umstiege_zeiten`, in Algorithmus 9 abgekürzt mit dem Kürzel `u1`).

6.3.3 Ausschluss versetzt parallel verkehrender Züge mit unterschiedlichen Zielen

Auf Basis ähnlicher Erwägungen wie im vorangegangenen Kapitel wird jetzt die Datenbanktabelle `parallele_zuege` mittels des Befehls `TRUNCATE parallele_zuege` gelöscht und dann für ein verändertes Szenario, bei dem alle Züge mit derselben Startstation, aber unterschiedlichen Endhalten berücksichtigt werden, neu erzeugt. Dabei ist zu beachten, dass die Verkehrstageprüfung aus technischen Gründen umgekehrt zu der in Algorithmus 7 dargestellten ursprünglichen Erzeugung erfolgt. Im Folgenden ist der nun entsprechend veränderte Algorithmus 7 als Algorithmus 10 dargestellt:

Algorithmus 10 Erzeugung der Datenbanktabelle `parallele_zuege`

```
INSERT INTO parallele_zuege
  (zug1, zug2)
SELECT
  g1.zugnr,
  g2.zugnr
FROM bp_fp_zug g1
INNER JOIN bp_fp_zug g2
  ON g1.start = g2.start
  AND g1.ziel <> g2.ziel
  AND g1.strecke_id = g2.strecke_id
  AND g1.id <> g2.id
  AND g1.Mo <= g2.Mo
  AND g1.Di <= g2.Di
  AND g1.Mi <= g2.Mi
  AND g1.Do <= g2.Do
  AND g1.Fr <= g2.Fr
  AND g1.Sa <= g2.Sa
  AND g1.So <= g2.So
INNER JOIN bp_fp_zeiten z1
  ON z1.zug_nr = g1.zugnr
  AND z1.halt_nr = 1
INNER JOIN bp_fp_zeiten z2
  ON z2.zug_nr = g2.zugnr
  AND z2.halt_nr = 1
  AND z2.ab > z1.ab
  AND TIMEDIFF(z2.ab, z1.ab) < '02:00:00'
INNER JOIN bp_fp_zeiten z3 ON
  z3.zug_nr = g1.zugnr
  AND z3.halt_nr = 2
INNER JOIN bp_fp_zeiten z4
  ON z4.zug_nr = g2.zugnr
  AND z4.halt_nr = 2
  AND z3.bf_id = z4.bf_id
```

In der mit der Logik aus Algorithmus 10 mit neuen Werten gefüllten Datenbanktabelle `parallele_zuege` kann nun eine zugweise Prüfung auf parallele Fahrten erfolgen, wenn die Endbahnhöfe nicht gleich sind. Gesucht wird dabei nach Zügen, bei denen ein Zuglauf den anderen ab einem bestimmten Anschlusspunkt komplett überdeckt, also alle Verkehrshalte des anderen Zuges mitbedient, und gleichzeitig insgesamt vorteilhaft ist, also dieselben Randbedingungen der Fahrt aufweist (bezogen auf Preis und Komfort des Verkehrsmittels) sowie zeitlich attraktiver ist. Das Kriterium der zeitlichen Vorteilhaftigkeit muss sich nicht zwingend bei der Abfahrt an einer anschlussaufnehmenden

Verkehrshalten darstellen, sondern kann sich durch eine Überholung auch an späteren Verkehrshalten auswirken und ist daher zwingend halteweise durchzuprüfen.

Es werden dann aus der Datenbanktabelle `parallele_zuege` dann mit dem folgenden Algorithmus 11 alle Züge entfernt, die keinen parallelen Vorgängerzug besitzen, der mindestens alle Halte des zweiten Zuges abdeckt:

Algorithmus 11 Löschen von Datensätzen überlappender Züge aus der Datenbanktabelle

`parallele_zuege`

```
DELETE p1
FROM parallele_zuege p1
INNER JOIN bp_fp_zeiten_mem z2
  ON z2.zug_nr = p1.zug1
INNER JOIN bp_fp_zeiten_mem z3
  ON z3.zug_nr = p1.zug2
LEFT JOIN bp_fp_zeiten_mem z4
  ON z2.zug_nr = z4.zug_nr
  AND z4.bf_id = z3.bf_id
WHERE
  z4.id IS NULL
  AND p1.ID = :id
```

Die Vorgehensweise aus Algorithmus 11, die Datenbanktabelle `parallele_zuege` auf einen parallelen Vorgängerzug „abzutasten“, dient der effizienteren Umsetzung der bereits vorher erläuterten Erwägung, doppelte Zugläufe durch Prüfung aller Verkehrshalte gegeneinander ausschließen zu können. In diesem Fall wird mit dem `LEFT JOIN` auf `z4` darauf geprüft, dass der später verkehrende Zug keinen zusätzlichen Halt hat. Erfüllt er diese Bedingung, kann der Zug in einem späteren Schritt gelöscht werden, da er bedingungsgemäß alle Halte später erreicht als der erste Zug und keinen weiteren Vorteil (niedrigeren Fahrpreis, höheren Komfort oder kürzere Gesamtfahrzeit) bietet. Züge, die hingegen unter denselben Bedingungen zusätzliche Halte aufweisen, aber später verkehren, müssen im System erhalten bleiben und werden dabei aus dieser Datenbanktabelle gelöscht, da die Züge dieser Datenbanktabelle die Zugpaare der zu löschenden Einträge der Datenbanktabelle `umstiege_zeiten` darstellen.

Dass kein weiterer Vorteil im Sinne von Fahrpreis oder Komfort existiert, ist den Bedingungen der SQL-Ausdrücke in Algorithmus 10 und Algorithmus 11 nicht unmittelbar zu entnehmen, sondern ergibt sich aus dem Aufbau des Datenmodells und der Prüfung auf `g1.strecke_id = g2.strecke_id` in Algorithmus 10, da innerhalb derselben Strecke nur Züge desselben Fahrpreises und im Regelfall auch derselben Zuggattung verkehren. Gattungsbezogene Ausnahmen existieren hierbei nur im Regional- und

Nahverkehr, wobei dort der Unterschied im Komfort nur in Ausnahmefällen das Warten auf einen späteren Zug rechtfertigen würde und daher hier nicht separat berücksichtigt wird (vgl. dazu die Ausführungen in Kapitel 4.1).

Die jetzt in der Datenbanktabelle `parallele_zuege` übrig gebliebenen und damit nicht vorteilhaften Züge können aus der Tabelle `umstiege_zeiten` entfernt werden, sofern derselbe anschlussgebende Zug als Referenz genutzt wird; dazu dient der folgende Algorithmus 12:

Algorithmus 12 Löschen von nicht relevanten Datensätzen aus der Datenbanktabelle `umstiege_zeiten`

```
DELETE u2
FROM umstiege_zeiten u1
INNER JOIN umstiege_zeiten u2
    ON u1.id_1 = u2.id_1
INNER JOIN bp_fp_zeiten z1
    ON z1.id = u1.id_2
INNER JOIN bp_fp_zeiten z2
    ON z2.id = u2.id_2
INNER JOIN parallele_zuege p1
    ON p1.zug1 = z1.zug_nr
    AND p1.zug2 = z2.zug_nr
```

6.3.4 Ausschluss versetzt parallel verkehrender Züge mit unterschiedlichen Startbahnhöfen

Der bereits vorgestellte Ansatz wird jetzt so verändert, dass Züge derselben Zielstation, aber mit unterschiedlichen Startbahnhöfen berücksichtigt werden. Dazu wird in einem ersten Schritt die Ausgangslage wieder hergestellt und die Tabelle `parallele_zuege` mittels des Befehls `TRUNCATE parallele_zuege` gelöscht, bevor dann die Erzeugung der veränderten Daten mit dem im Folgenden dargestellten veränderten Algorithmus 13 vorgenommen wird:

Algorithmus 13 Erzeugung der Datenbanktabelle `parallele_zuege`

```
INSERT INTO parallele_zuege
  (zug1, zug2)
SELECT
  g1.zugnr,
  g2.zugnr
FROM bp_fp_zug g1
INNER JOIN bp_fp_zug g2
  ON g1.start <> g2.start
  AND g1.ziel = g2.ziel
  AND g1.strecke_id = g2.strecke_id
  AND g1.id <> g2.id
  AND g1.Mo >= g2.Mo
  AND g1.Di >= g2.Di
  AND g1.Mi >= g2.Mi
  AND g1.Do >= g2.Do
  AND g1.Fr >= g2.Fr
  AND g1.Sa >= g2.Sa
  AND g1.So >= g2.So
INNER JOIN bp_fp_zeiten z1
  ON z1.zug_nr = g1.zugnr
  AND z1.bf_id = g1.ziel
INNER JOIN bp_fp_zeiten z2
  ON z2.zug_nr = g2.zugnr
  AND z2.bf_id = g2.ziel
  AND z2.an > z1.an
  AND TIMEDIFF(z2.an, z1.an) < '02:00:00'
INNER JOIN bp_fp_zeiten z3
  ON z3.zug_nr = g1.zugnr
  AND z3.id = z1.vorh_id
INNER JOIN bp_fp_zeiten z4
  ON z4.zug_nr = g2.zugnr
  AND z4.id = z2.vorh_id
  AND z3.bf_id = z4.bf_id
```

Die Logik in Algorithmus 13 war gegenüber der Logik aus Algorithmus 10 dahingehend anzupassen, dass nun zum einen auf die Ankunft der Zeitdatensätze der beiden jeweils zu vergleichenden Züge abgestellt wird (und damit andere Datenbankfelder Berücksichtigung finden) und zum anderen aufgrund der Berücksichtigung der Zielstation des Zuglaufs die Referenzierung der Zeitdatensätze nicht mehr auf den ersten Halt des Zuges erfolgen kann, sondern neu die Zielstation des Zuglaufs als maßgeblicher Haltebahnhof des Zeitdatensatzes verwendet werden muss („z1.bf_id = g1.ziel“). Das reduziert die Effizienz des Algorithmus insoweit, als eine fixe Zuordnung einer Variablen (aus Algorithmus 10: „z1.halt_nr = 1“) durch eine weitere Referenzierung auf

eine Tabelle ersetzt werden muss. Entsprechend muss auch die Zuordnung der Tabellen z3 und z4 verändert werden und wird jetzt über eine Referenz der Tabellen z1 und z2 vorgenommen.

Der zweite Schritt besteht nun wieder darin, alle Anschlussrelationen zwischen zwei Zügen unangetastet zu lassen, die es zu behalten gilt, und dafür diejenigen Kombinationen ausfindig zu machen, die es zu eliminieren gilt. Dazu wird mit dem folgenden Algorithmus 14 die Datenbanktabelle `parallele_zuege` wieder um die zu behaltenden Anschlusszüge bereinigt, um die übrig bleibenden Züge weiter behandeln zu können:

Algorithmus 14 Löschen von Datensätzen überlappender Züge aus der Datenbanktabelle `parallele_zuege`

```
DELETE p1
FROM parallele_zuege p1
INNER JOIN bp_fp_zeiten_mem z2
  ON z2.zug_nr = p1.zug2
INNER JOIN bp_fp_zeiten_mem z3
  ON z3.zug_nr = p1.zug1
LEFT JOIN bp_fp_zeiten_mem z4
  ON z2.zug_nr = z4.zug_nr
  AND z4.bf_id = z3.bf_id
WHERE
  z4.id IS NULL
  AND p1.ID = :id
```

Im Unterschied zu Algorithmus 11 werden im hier dargestellten Algorithmus 14 der anschlussgebende und der anschlussaufnehmende Zug durch die Referenzen von z2 und z3 vertauscht; dies ist notwendig, um die „gedrehte“ Verwendung des Algorithmus bei der Nutzung der Ankunftsstationen anstelle der Startstationen korrekt abzubilden. Entsprechend werden auch wieder im nächsten Schritt die in der Datenbanktabelle `parallele_zuege` übrig gebliebenen und damit nicht vorteilhaften Züge aus der Tabelle `umstiege_zeiten` entfernt. Hierbei muss allerdings wieder die Veränderung von Start- und Zielstation beachtet werden, weswegen der hier dargestellte Algorithmus 15 vom im vorangegangenen Kapitel vorgestellten Algorithmus 12, der eine analoge Funktion erfüllt, unterscheidet:

Algorithmus 15 Löschen von nicht relevanten Datensätzen aus der Datenbanktabelle

umstiege_zeiten

```
DELETE u1
FROM umstiege_zeiten u1
INNER JOIN umstiege_zeiten u2
    ON u1.id_2 = u2.id_2
INNER JOIN bp_fp_zeiten z1
    ON z1.id = u1.id_1
INNER JOIN bp_fp_zeiten z2
    ON z2.id = u2.id_1
INNER JOIN parallele_zuege p1
    ON p1.zug1 = z1.zug_nr
    AND p1.zug2 = z2.zug_nr
```

Im Ergebnis bleiben nach vier Durchgängen der gestaffelten Aussortierung in der Datenbanktabelle `umstiege_zeiten` lediglich diejenigen Umstiege erhalten, für die innerhalb der jeweiligen Strecke kein besser passender Anschluss gefunden wurde, der denselben Bedingungen genügt und sich als zeitlich vorteilhafte Umsteigeoption darstellt.

6.3.5 Verbindungserzeugung des ersten Umstiegs

Die Erzeugung der Verbindungen des ersten Umstiegs kann jetzt trivial aus den vorangegangenen gefilterten Umsteigebeziehungen und den zugrundeliegenden Daten des Modells mit folgendem Algorithmus 16 erfolgen:

Algorithmus 16 Erzeugen der Verbindungen des ersten Umstiegs

```
INSERT INTO umstiege_dv_mem
    (dv1, dv2, bf_von, bf_nach, dauer, virtPreis,
     Mo, Di, Mi, Do, Fr, Sa, So, ab, an, Vergleich)
SELECT
    d1.id,
    d2.id,
    d1.bf_von,
    d2.bf_nach,
    d2.an-d1.ab,
    (d1.virtPreis+d2.virtPreis+5+
     (d2.an-d1.ab-d2.Dauer_Min-d1.Dauer_Min)*0.1),
    d1.Mo * d2.Mo,
    d1.Di * d2.Di,
    d1.Mi * d2.Mi,
    d1.Do * d2.Do,
    d1.Fr * d2.Fr,
    d1.Sa * d2.Sa,
    d1.So * d2.So,
    d1.ab,
    d2.an,
    CONCAT(d2.bf_nach, '.', ROUND((d1.ab+d2.an)/60))
FROM umstiege_zeiten_mem u1
INNER JOIN direktverbindungen_mem d1
    ON d1.zeile_zwei = u1.id_1
INNER JOIN direktverbindungen_mem d2
    ON d2.zeile_eins = u1.id_2
LEFT JOIN verbindungen_mem u3
    ON u3.bf_nach = d2.bf_nach
    AND u3.ab >= d1.ab
    AND u3.an <= d2.an
    AND u3.virtPreis < (d1.virtPreis + d2.virtPreis + 5 +
        (d2.an-d1.ab-d2.Dauer_Min-d1.Dauer_Min)*0.1)
    AND u3.Mo >= (d1.Mo*d2.Mo)
    AND u3.Di >= (d1.Di*d2.Di)
    AND u3.Mi >= (d1.Mi*d2.Mi)
    AND u3.Do >= (d1.Do*d2.Do)
    AND u3.Fr >= (d1.Fr*d2.Fr)
    AND u3.Sa >= (d1.Sa*d2.Sa)
    AND u3.So >= (d1.So*d2.So)
WHERE
    d1.bf_von = :bf
    AND d2.an > d1.ab
    AND u3.ID IS NULL
```

In Algorithmus 16 wird dabei bahnhofswise jedes mögliche Tupel aus zwei Verbindungen gebildet, die sich in den Umsteigebetrachtungen (festgehalten in der Datenbanktabel-

le `umstiege_zeiten`) als vorteilhaft (genauer: als nicht nachteilig) herausgestellt haben und die nicht bereits bei ihrer Erstellung von einer anderen optimalen Verbindung einer niedrigeren Umsteigezahl verdrängt werden (dazu dient die Prüfung auf „WHERE [...] `u3.ID IS NULL`“). In `verbindungen_mem u3` sind zur Laufzeit dieses Programmteils aus Algorithmus 16 sowieso nur Direktverbindungen des passenden Startbahnhofs enthalten; die Referenz auf die Datenbanktabelle `verbindungen_mem` statt auf die Datenbanktabelle `direktverbindungen_mem` (in der ebenfalls alle Direktverbindungen mit ihren Eigenschaften hinterlegt sind) hat programmtechnische Gründe: In der Datenbanktabelle `direktverbindungen_mem` sind die Eigenschaften in anderen Datenformaten und Genauigkeiten hinterlegt, als dies zur Laufzeit dieses Algorithmus benötigt wird (vergleiche dazu die Ausführungen im Abschnitt 5.2.2).

Zur weiteren Verbindungserzeugung wird das Vorgehen analog zur Verbindungserzeugung mehrerer Umstiege angewandt, dazu werden die Algorithmen 2 und 3 zur Selektion von entbehrlichen Verbindungen innerhalb der Datenbanktabelle `umstiege_mem` und damit innerhalb der Menge der zur Laufzeit durch Algorithmus 16 erzeugten Verbindungen angewandt. Als weiterer Baustein wird die Logik von Algorithmus 5 angewandt, um den Ausschluss von Umsteigeverbindungen bei lediglich partieller Überdeckung durch Direktverbindungen zu behandeln – der Ausschluss der Umsteigeverbindungen bei kompletter Überdeckung durch eine vorteilhafte Direktverbindung passiert bereits während der Verbindungserzeugung durch die Prüfung auf Nicht-Vorhandensein einer vorteilhaften Direktverbindung mit der Routine `LEFT JOIN` auf die Tabelle `verbindungen_mem u3` sowie die Prüfung auf „WHERE [...] `u3.ID IS NULL`“ in Algorithmus 16.

Zusätzlich zu den vorgenannten Sortierungsmaßnahmen sollten im Rahmen der Erzeugung von Verbindungen mit einem Umstieg weitere Prüfungen erfolgen, die die Anzahl der zu berechnenden Umsteigeverbindungen (mit mehr als einem Umstieg) reduzieren. Zuerst wird dazu eine Prüfung auf doppelt genutzte Züge durchgeführt, die dieselbe Start- und Zielstation besitzen, jedoch einen unterschiedlichen Umsteigepunkt aufweisen. Bedingt durch die vorangegangenen Prüfungen steht zu vermuten, dass zur Laufzeit des folgenden Algorithmus lediglich solche quasi-doppelten Verbindungen enthalten sind, bei denen Preis und Komfort identisch sind; es könnte jedoch in seltenen Fällen darauf hinauslaufen, dass sich der Preis- und Komfortunterschied der Verbindungen innerhalb der Verrechnung des virtuellen Preises nivellieren, sodass im Folgenden auch

schlicht über den vergebenen Primäridentifikator – und damit den Erzeugungszeitpunkt – entschieden wird:

Algorithmus 17 Löschen von quasi-doppelten Verbindungen derselben Relation

```
DELETE u1
FROM umstiege_dv_mem u1
INNER JOIN umstiege_dv_mem u2
    ON u1.Vergleich = u2.Vergleich
INNER JOIN direktverbindungen_mem d1
    ON d1.id = u1.dv1
INNER JOIN direktverbindungen_mem d2
    ON d2.id = u1.dv2
INNER JOIN direktverbindungen_mem d3
    ON d3.id = u2.dv1
    AND d3.Zug = d1.Zug
INNER JOIN direktverbindungen_mem d4
    ON d4.id = u2.dv2
    AND d4.Zug = d2.Zug
WHERE
    u1.id <> u2.id
    AND u1.an = u2.an
    AND u1.ab = u2.ab
    AND (
        u1.virtPreis >= u2.virtPreis
        AND u1.id > u2.id
        OR u1.virtPreis > u2.virtPreis
    )
```

In der Logik von Algorithmus 17 wird zur Reduktion der zu vergleichenden Verbindungspaare zuerst der Parameter `Vergleich` ausgewählt, der bedingungsgemäß für beide Verbindungen identisch sein muss, wenn Start- und Zielstation übereinstimmen sowie die Abfahrts- und Ankunftszeit gleich sind. Im zweiten Schritt wird geprüft, dass die beteiligten Züge identisch sind, obwohl sich die Direktverbindungs-Identifikatoren unterscheiden, denn durch die Wahl eines unterschiedlichen Umsteigepunkts zwischen beiden Verbindungen ergeben sich unterschiedliche Direktverbindungen der beteiligten Teil-Verbindungen.

Die Annahme, dass sich die Direktverbindungs-Identifikatoren unterscheiden, folgt dabei unmittelbar aus der Tupelbildung in der Tabelle `umstiege_zeiten`. Dort ist jeder Zeitdatensatz nur einmal für jede Umsteigekombination enthalten, entsprechend entstehen beim einfachen `INNER JOIN` als Verknüpfung der Direktverbindungen die resultierenden Verbindungen auch nur einmal – zwei unterschiedliche Umsteigeverbin-

dungen zwischen identischem Start und Ziel, gebildet aus denselben Zügen, müssen daher unterschiedliche Direktverbindungen als Basis haben.

Schließlich müssen als letzte Bedingung für die Anwendung der Logik aus Algorithmus 17 noch die Start- und Zielzeit identisch sein, und entweder eine der beiden Verbindungen – im Übrigen im dargestellten Algorithmus grundsätzlich die in der Datenbanktabelle `u1` notierte – einen höheren virtuellen Preis aufweisen oder bei identischem virtuellen Preis den niedrigeren Identifikator besitzen (gelöscht wird aus der Datenbanktabelle `u1`).

Dass der Identifikator der Umsteigeverbindung als Aussortierungskriterium genutzt wird, folgt hierbei aus der für den Fahrgast ambivalenten Entscheidung zwischen den Verbindungen, wobei keine einen Vorteil gegenüber der anderen aufweisen würde, sodass es nicht zielführend erscheint, beide anzuzeigen, sondern quasi-zufällig auszuwählen – hier über den Erzeugungszeitpunkt, der als Referenz hinterlegt ist, und keine Verbindungen systematisch benachteiligt.

Im nächsten Schritt werden nun noch alle Verbindungen, die auf einer gemeinsamen Direktverbindung basieren und deren Anschlussverbindungen selbst parallel zueinander verkehren, gelöscht:

Algorithmus 18 Löschen von Verbindungen, die teilparallel sind

```
DELETE u2
FROM parallelverbindungen_mem p1
INNER JOIN umstiege_dv_mem u1
  ON p1.v1 = u1.dv2
INNER JOIN umstiege_dv_mem u2
  ON u1.dv1 = u2.dv1
  AND p1.v2 = u2.dv2
```

Algorithmus 18 verwendet als zugrundeliegende Bedingung, dass bei teilparallelen Verbindungen mit derselben Startverbindung („ON `u1.dv1 = u2.dv1`“), sofern die zweite Verbindung streng parallel zur ersten verläuft (also nach der ersten unter denselben oder schlechteren Bedingungen verkehrt), die zweite – mithin unvorteilhaftere – Verbindung (im Algorithmus in der Datenbanktabelle `umstiege_dv_mem u2` enthalten) direkt gelöscht werden kann. Die Verbindungen sind dabei in der Datenbanktabelle `parallelverbindungen` dergestalt notiert, dass die im Feld `v1` enthaltene Verbindung grundsätzlich die vorteilhaftere darstellt, weil sie vor der im Feld `v2` gespeicherten Verbindung verkehrt und dieselben oder günstigere Nebenbedingungen (Preis, Komfort) bietet.

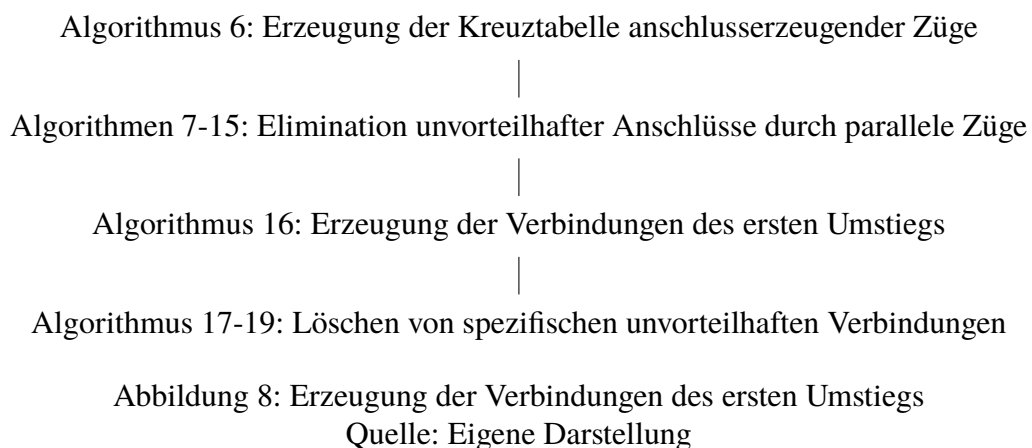
Außerdem werden in einem weiteren Schritt noch alle Verbindungen gelöscht, die den Verkehrstageswechsel (im Modell um 3 Uhr; die Zeiten zwischen 0 und 3 Uhr wurden entsprechend korrigiert und lauten auf 24–27 Uhr) überschreiten; diese Logik ist im folgenden Algorithmus 19 dargestellt:

Algorithmus 19 Löschen von Verbindungen über den Verkehrstageswechsel

```
DELETE u1
FROM umstiege_dv_mem u1
WHERE
    an < ab
```

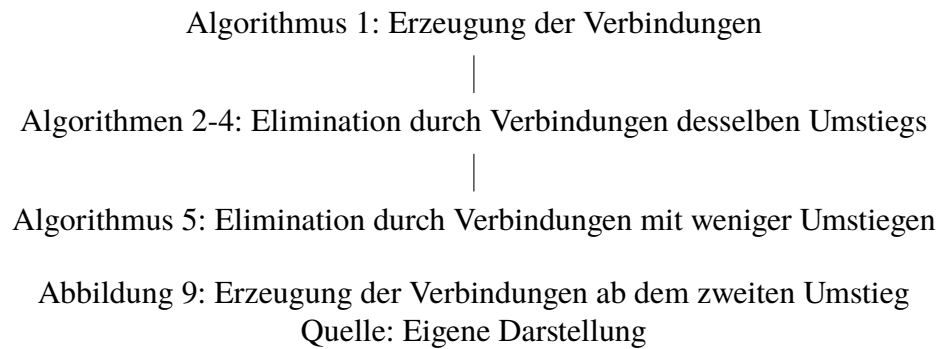
6.4 Zusammenfassung der Algorithmen

Die Nutzung der Algorithmen ist zuerst in das Szenario aufzuteilen: Nach der Erzeugung aller Direktverbindungen, die direkt zu Beginn der Laufzeit erfolgt, muss zur Vorbereitung der Verbindungen des ersten Umstiegs nach der Erstellung der Kreuztabelle anschluss erzeugender Züge eine Ermittlung aller parallelen Züge erfolgen, wobei parallele Züge hier Züge meint, die haltidentisch und zeitverschieden verkehren und dieselben Nebenbedingungen wie Preis und Komfort aufweisen. Dazu werden verschiedene Schritte ausgeführt, um die Menge der Verbindungen der auf einem Teilstück parallel verkehrenden Züge an spezifischen Punkten weiterhin zu behalten, beispielsweise am ersten und letzten möglichen Umsteigepunkt zweier Verbindungen. Die Abfolge ist in der folgenden Abbildung 8 dargestellt:



Nach der Erzeugung der Verbindungen des ersten Umstiegs werden alle Verbindungen mit mehreren Umstieg erzeugt und danach direkt mit verschiedenen weiteren Algorithmen

men geglättet, um lediglich die in den Auswertungen schlussendlich zu berücksichtigenden, weil vorteilhaften Verbindungen zu erhalten. Diese Prozessfolge ist in der folgenden Abbildung 9 dargestellt:



7 Verbindungsauswahl und Auswirkungen des Tarifs

Nach der Erzeugung aller Verbindungen muss nun eine zielgruppenspezifische Auswahl getroffen werden. Bedingungsgemäß wurde dabei bisher nur ein Vergleichskriterium – der virtuelle Preis – zur (Aus-)Sortierung von Verbindungen genutzt. In einem weiteren Schritt ist dann die konkrete Auswirkung der Anwendung verschiedener Tarifsysteme auf die Anzahl der erzeugten Verbindungen zu prüfen und zu bewerten.

7.1 Auswahl der optimalen Verbindungen

Aus den erzeugten sinnvollen Verbindungen sind zielgruppenspezifisch die Verbindungen auszuwählen, die für jede Zielgruppe zur Reise infrage kommen. Entsprechend der vorangegangenen Erwägungen gilt eine Reise dann als vorteilhaft, wenn sich keine vorteilhaftere Alternative findet, bezogen auf den Reisewiderstand und die zeitliche Verfügbarkeit; das schließt sich überholende Verbindungen prinzipbedingt aus. Die Verrechnung des Reisewiderstands erfolgt dabei, wie in Kapitel 4.1 dargestellt, monetarisiert. Zuerst müssen daher alle Verbindungen für jede Zielgruppe mit den monetarisierten Reisewiderstandswerten versehen werden, bevor dann diejenigen Verbindungen ohne Vorteil ausgeschlossen werden können. Das geschieht jeweils nach Startstationen und Zielgruppen getrennt in der Tabelle `verbindungen_zielgruppe` mit dem folgenden Algorithmus 20:

Algorithmus 20 Auswahl von Verbindungen für eine Startstation und Zielgruppe

```
INSERT INTO verbindungen_zielgruppe
  (ID, bf_von, bf_nach, an, ab, virtPreis,
   Mo, Di, Mi, Do, Fr, Sa, So, Vergleich)
SELECT
  ID,
  bf_von,
  bf_nach,
  an,
  ab,
  (Preis+Dauer/2+Umstiege*10),
  Mo,
  Di,
  Mi,
  Do,
  Fr,
  Sa,
  So,
  bf_nach
FROM verbindungen
WHERE
  bf_von = :bf
```

Bei der Anwendung dieses Algorithmus ist die Berechnung des in der Tabelle `verbindungen_zielgruppe` als `virtPreis` bezeichneten monetarisierten Reisewiderstands für die Zielgruppe vorzunehmen; in diesem Beispiel der Einfachheit halber nur aus dem Preis, der Dauer (in Minuten gerechnet und halbiert, einem Stundensatz von 30 € entsprechend) und der Anzahl der Umstiege mit 10 € Umsteigewiderstand pro Umstieg berechnet. Die Spalte `vergleich` wird hierbei jeweils mit dem Wert des Zielbahnhofs belegt; diese Information ist ausreichend, da bedingungsgemäß nur Verbindungen einer Startstation in der Tabelle notiert wurden (die Tabelle ist auch dafür vorbereitet, Verbindungen mehrere Startstationen aufzunehmen; entsprechend ist dann der Wert der Spalte `vergleich` auf eine Verknüpfung der Werte aus Start- und Zielstation zu setzen).

Die übertragenen Verbindungen müssen jetzt noch dergestalt ausgewählt werden, dass nur für die Zielgruppe vorteilhafte Verbindungen erhalten bleiben und alle anderen Verbindungen aus der Tabelle `verbindungen_zielgruppe` entfernt werden (die nicht vorteilhaften und daher entfernten Verbindungen bleiben aber in der Tabelle `verbindungen` erhalten und können daher von anderen Zielgruppen noch genutzt werden). Dazu dienen die folgenden Algorithmen, beginnend mit dem Algorithmus 21

zur Berücksichtigung aller Verbindungen mit vollständiger Überdeckung der Verkehrstage mit einer vorteilhaften Verbindung:

Algorithmus 21 Aussortieren von Verbindungen für eine Startstation und Zielgruppe mit vollständiger Überdeckung der Verkehrstage

```
DELETE u1
FROM verbindungen_zielgruppe u1
INNER JOIN verbindungen_zielgruppe u2
  ON u1.Vergleich = u2.Vergleich
  AND u1.ab <= u2.ab
  AND u1.an >= u2.an
  AND u1.virtPreis >= u2.virtPreis
  AND u1.Mo <= u2.Mo
  AND u1.Di <= u2.Di
  AND u1.Mi <= u2.Mi
  AND u1.Do <= u2.Do
  AND u1.Fr <= u2.Fr
  AND u1.Sa <= u2.Sa
  AND u1.So <= u2.So
WHERE
  u1.ID > u2.ID
```

Die Auswahl von Verbindungen u_1 , für die eine vollständig überdeckende vorteilhafte Verbindung u_2 existiert, erfolgt dabei auf Basis des Vergleichswerts (Start- und Zielstation müssen übereinstimmen), einer stattfindenden Überholung oder einer gleichzeitigen Abfahrt und Ankunft, einem zielgruppenspezifischen geringeren Reisewiderstands (definiert über die Spalte `virtPreis`) sowie der Überdeckung aller Verkehrstage. Die Verbindung u_1 wird dabei nur aussortiert, wenn eine andere Verbindung u_2 alle vorgenannten Kriterien erfüllt, also auf derselben Relation mindestens gleich schnell ist, keinen höheren Reisewiderstand bietet und an allen Verkehrstagen der nicht vorteilhaften Verbindung angeboten wird.

Die Bedingung der `WHERE`-Klausel dient im Übrigen dazu, dass bei zwei ansonsten identischen Verbindungen, die zum Beispiel aufgrund von Flügelzügen entstehen können, die aus Sicht des Systems doppelt vorhanden, aber getrennt buchbar sind, nicht beide Verbindungen wegfallen (weil zum Zeitpunkt der Durchführung des Algorithmus beide Verbindungen noch jeweils eine mindestens gleich gute Alternative besitzen, nämlich die jeweils andere; das Löschen der Verbindungen erfolgt im Algorithmus erst nach der Prüfung der `JOIN`-Verbindungen), sodass eine der Verbindungen behalten wird (hier die mit der niedrigeren `ID`).

Alle sich jetzt als vorteilhaft herausstellenden Verbindungen werden nun gewertet und zählen in die Verbindungsanzahl der Auswertung für unterschiedliche Tarifsysteme.

7.2 Vorstellung der Tarifsysteme

Um die Auswirkungen der Tarifsystematik auf die Routenauswahl und entsprechend die Anzahl angebotener Verbindungen in Routensuchsystemen zu verifizieren, werden folgend in einem ersten Schritt die zu vergleichenden Tarifsysteme beschrieben. Bestandteil der Ausarbeitung sind jedoch nur Tarifsysteme, die in ihrem Grundtypus von Entfernung und Relation abhängig sind, nicht jedoch von weiteren Instrumenten des Yield Managements (wie Kaufzeitpunkt des Tickets oder prognostizierte Auslastung der Reisekette) abhängige Tarifsysteme wie beispielsweise die Sparpreise der Deutschen Bahn.[Hof08] Auch Tarifsysteme mit Flatrate-Charakter sind nicht Bestandteil dieser Arbeit: Hier werden für einen Festpreis Fahrtberechtigungen innerhalb eines definierten Gebiets für einen festgelegten Zeitraum ohne Beschränkung der Fahrtenzahl oder des Fahrtwegs erlaubt.

7.2.1 Entfernungsabhängige Tarifierung

Die entfernungsabhängige Tarifierung (auch „Kilometertarif“ genannt) basiert dabei auf einem typischerweise stetigen Modell zur Umrechnung der zurückgelegten Tarifentfernung in den zu zahlenden Fahrpreis. Die gegebenenfalls vorhandene Abweichung zwischen Tarif- und Entfernungskilometern ist separat zu beachten, sie wurde bereits in Kapitel 5.2.1 diskutiert. Dabei sind verschiedene Modelle denkbar, die entfernungsabhängige Tarifierung kann dabei linear, regressiv oder progressiv angewandt werden:[Pfu08]

Modell	Auswirkung auf den Grenzpreis bei zunehmender Entfernung
linear	bleibt gleich
regressiv	sinkt
progressiv	steigt an

Tabelle 38: Verschiedene Modelle der entfernungsabhängigen Tarifierung
Quelle: Eigene Darstellung auf Basis von [Bös69, S.446]

Die Auswirkungen des gewählten Modells der entfernungsabhängigen Tarifierung auf den Grenzpreis – den zu zahlenden Tarif für einen weiteren Entfernungskilometer – sind

in Tabelle 38 dargestellt. Für Reisende mit kurzen Distanzen ist dabei ein progressives Modell tendenziell günstiger, für Reisende über längere Distanzen ist ein regressives Modell vorzuziehen. Der tatsächlich zu zahlende Betrag wird von dieser Modellauswahl im Übrigen nicht direkt berührt, er ergibt sich aus dem gewählten Tarif und nur mittelbar aus der Wahl des Modells. Es ist außerdem mit dem Modell vereinbar, den Preis tages- oder tageszeitabhängig anzupassen; der entfernungsabhängige Tarif bezieht sich nur auf die tatsächliche Abrechnung nach zurückgelegter Tarifentfernung, nicht jedoch auf einen zu allen Zeiten und Verkehrstagen einheitlichen Tarif.

Für die entfernungsabhängige Tarifierung ist außerdem kennzeichnend, dass Aufschläge für genutzte besondere Verkehrsmittel, die beispielsweise besondere Qualitätsansprüche versprechen wie der Trans Europ Express (TEE),²¹ und für Verkehrsmittel, die besonders schnell sind, entweder pauschal oder entfernungsabhängig erhoben werden. Zuschläge für besonders schnelle Verkehrsmittel fallen daher nur in dem Umfang an, wie diese besonderen Verkehrsmittel genutzt werden – oder werden pauschal erhoben, und sind dann für alle Nutzer gleich und von der Dauer der Benutzung unabhängig. Sie korrespondieren aber im Regelfall nicht mit der Relation, auf der sie genutzt werden; Ausnahmen bestehen hier bei Verkehrsdiensten, die lediglich auf einer Relation angeboten werden, wobei dann die Zurechnung zur Relation oder zum Verkehrsdienst wahlweise erfolgen kann.

Ein Beispiel für das beschriebene Tarifmodell der entfernungsabhängigen Tarifierung ist der Tarif der Österreichischen Bundesbahnen von 1968 (Tarif zitiert nach [Bös69]). Neben einem regressiven entfernungsabhängigen Tarif (wobei die Entfernung in Tarifkilometern gemessen wird und daher streng genommen nicht direkt entfernungsabhängig ist) für die 1. Klasse und die 2. Klasse bildet ein weiterer zusätzlicher Bestandteil den Tarif: Fahrten schnellerer und besonders komfortabler Verkehrsmittel (wie beispielsweise der TEE) werden dabei mit Zuschlägen versehen.[Bös69, S.449ff.]

Lenkungsinstrumente sind in solchen Tarifförmern durch separate Fahrkartentypen vorhanden, bei denen zu Pauschalpreisen Fahrausweise mit bestimmten Leistungszeiträumen oder -strecken unter speziellen Bedingungen vergünstigt abgegeben werden. Der Tarif der Deutschen Bundesbahn von 1991 kennt dabei neben den relationsgebundenen Zuschlägen für ICE-Züge auch zwei besondere Festpreise: Den „ICE-Sparpreis“ (zu 220,-DM in der 2. Klasse) und den „ICE-Supersparpreis“ (zu 180,-DM in der 2.

²¹Bei der Einführung des TEE wurden die Fahrten aufgrund des höheren Komforts mit höheren Zuschlägen bepreist als die ebenfalls herausgehoben schnell verkehrenden Fernschnellzüge (F-Züge) und Schnellzüge (D-Züge).[Nor12, S.268f.]

Klasse), beide als Festpreise ohne Längen- oder Abschnittsbegrenzung ausgeführt und lediglich der ICE-Supersparpreis mit der zeitlichen Bedingung, dass keine Fahrten damit an Freitagen und Sonntagen durchgeführt werden dürfen, versehen. Das Preissystem der Deutschen Bundesbahn hatte mit dem ICE-Sparpreis effektiv also ein 'price cap' für Fahrten in der 2. Klasse.[Kra91]

Das im Rahmen dieser Arbeit entwickelte Routenfindungsmodell verwendet die entfernungsabhängige Tarifierung ohne weitere Lenkungswirkung, um die durch die andere Tarifierung bedingten Lenkungswirkungen anschaulich darstellen zu können. Es haben daher auch keine weiteren Tarife Eingang in das Modell gefunden, auch nicht parallele Tarife, die im selben Verkehrsmittel gültig sind (und gegebenenfalls durch geschickte Kombination weitere Routenoptionen mit günstigen Preisen ergeben könnten).

7.2.2 Relationstarif

Der Relationstarif ist im Gegensatz zum entfernungsabhängigen Tarif nicht oder nicht unmittelbar linear von der zurückgelegten Distanz abhängig, sondern hängt von der befahrenen Start-Ziel-Relation ab. Diese Tarifart ermöglicht daher, bestimmte Verkehrsbeziehungen besonders günstig oder teuer zu bepreisen, und stellt somit eine Möglichkeit dar, den Tarif an die erwartete Nachfrage anzupassen. Gleichzeitig können mit einem Relationstarif auch bestimmte Gebiete oder Strecken stärker bepreist werden als andere; je nach konkreter Auslegung kann daher auch der Relationstarif entfernungsabhängig sein.[Pfu08]

Die höhere Bepreisung desselben Produkts wird insbesondere für besonders schnelle Services angewandt, zum Beispiel bei Fahrten über Schnellfahrstrecken. Stellvertretend hierfür berichtet die Frankfurter Allgemeine Zeitung zur Aufnahme des Verkehrs auf der Schnellfahrstrecke Köln – Rhein/Main:

„Für die Zeitersparnis müssen Bahnkunden jedoch tief in die Tasche greifen: Während sich die Gleisstrecke zwischen den Metropolen an Rhein und Main von 222 auf 177 Kilometer verkürzt, steigt der Preis fürs Ticket um 33 Prozent. Kostet die einfache IC-Fahrt ohne Bahncard bisher 35 Euro oder 39 Euro im ICE, muss der Kunde künftig 53,60 Euro berappen. Die Bahn erklärt dies mit zwölf Euro Tempo- und Komfort-Zuschlag sowie den Kosten für die Pflicht-Reservierung.“

— aus der Frankfurter Allgemeinen Zeitung vom 01.08.2002[Fra02]

Der im Zitat der Frankfurter Allgemeinen Zeitung genannte Aufschlag von 14,60 € ist dabei nicht direkt entfernungsabhängig, da die Entfernung gegenüber der zuvor befahrenen Strecke um 35 Kilometer abnimmt; er scheint Ausdruck der durch die Zeitersparnis gestiegenen Zahlungsbereitschaft zu sein. Der Relationstarif ist dabei auf der Strecke Frankfurt – Köln zusätzlich zur Relation auch noch streckenabhängig, da für dieselbe Relation abhängig vom Fahrtweg (und dem genutzten Zugprodukt) verschiedene Preise verlangt werden.

In Kombination mit einer weiteren Tarifoption wird der Relationstarif dabei allerdings noch bedeutend flexibler als der entfernungsabhängige Tarif: Wird der Relationstarif mit einer Raumbegrenzung verknüpft, so werden alle Fahrten zwischen zwei Stationen in einem zugelassenen Verkehrsraum für denselben Preis verkauft. Dies ermöglicht insbesondere bei Fahrpreisen ohne zugeordneten Zug (bei der Deutschen Bahn als „Zugbindung“ bezeichnet [DB 16a]) eine weitgehende Flexibilität und mehr Reiseoptionen (ohne zusätzliche Kosten und den entstehenden Aufwand), als das der Fall wäre, wenn lediglich eine definierte Strecke zwischen Start und Ziel zur Fahrt zulässig wäre.²²[DB 16b]

Bei der Diskussion einer Raumbegrenzung kann dabei argumentiert werden, dass die Anwendung einer solchen Raumbegrenzung zu einer gewissen Leistungsungerechtigkeit führt. Auch wer eine kürzere Route zwischen den zwei Stationen wählt, zahlt so viel wie die Fahrgäste auf einer längeren Fahrtstrecke (die entsprechende umgekehrte Argumentation ist natürlich auch zulässig). Die Bepreisung erfolgt bei der Anwendung

²²Natürlich kann auch bei entfernungsabhängigen Tarifen eine solche Option für die Fahrgäste ermöglicht werden; dann ist aber besonderes Augenmerk auf die Verhinderung von Missbrauch zu legen, damit von den Fahrgästen nicht grundsätzlich das günstigere Ticket gekauft wird, das auch zur Fahrt über die eigentlich teurere Strecke berechtigt.

eines Relationstarifs also nach dem verkauften Produkt („Fahrt von A nach B“) und nicht mehr nach dem verursachten Aufwand („Fahrt mit 281 Kilometern Länge“).

7.2.3 Kombinierte Tarifsysteme

Tarifsysteme bestehen üblicherweise aus einem Mix verschiedener Grundsysteme, wie auch bereits anhand der Mischung mit hier nicht vorgestellten Systemen beschrieben.

Beispielhaft für diese These soll an dieser Stelle der Bundesbahn-Tarif aus dem Jahr 1951 (vorliegend als Nachdruck mit Stand 01. September 1959) vorgestellt werden, der grundsätzlich die entfernungsabhängige Tarifierung (bezogen auf Tarifkilometer) verwendet, jedoch eine Raumbegrenzungstafel als Anhang enthält, sodass für bestimmte Relationen Raumbegrenzungen mit einheitlicher Tarifierung bezogen auf die Relation angewandt werden. So ist die Fahrt Dortmund Hbf – Hannover Hbf über den durch die Verkehrsstationen Hamm (Westf.) Pbf einerseits sowie Soest und Hameln andererseits aufgespannten Raum begrenzt, was eine Vielzahl von Fahrtwegen und -optionen innerhalb des zugelassenen Verkehrsraums zu einem einheitlichen Preis zulässt.[Deu59, S.109]

Außerdem werden im Bundesbahn-Tarifwerk aus 1959 für verschiedene Städte Alternativstationen hinterlegt. So berechtigt eine Fahrkarte von/nach Dortmund Hbf aus Richtung Hagen kommend auch zur Fahrt von/nach Dortmund Süd.[Deu59, S.301] Eine solche Tarifierung ist in einem entfernungsabhängigen Tarif naturgemäß nicht vorgesehen, da sie der Grundregel „Tarifierung nur nach der zurückgelegten Tarifentfernung“ durch die Nivellierung von Entfernungen mehrerer Betriebsstellen auf einen quasi-virtuellen Tarifpunkt widerspricht. Gleichzeitig jedoch ist der vorgestellte Bundesbahn-Tarif außerhalb der dargestellten Besonderheiten ein entfernungsabhängiger Tarif mit einem entfernungsabhängigen Zuschlag (der TEE-Zuschlag, der nach Entfernungen gestaffelt verkauft wurde) und zwei nicht entfernungsabhängigen Zuschlägen (dem Zuschlag für D-Züge und dem für F-Züge).[Wie06]

Die Kombination verschiedener Grundtarife in einem Tarifsystem existiert darüber hinaus in einer Vielzahl von Tarifsystemen schon deswegen, weil neben Einzelfahrkarten auch Fahrkarten mit zeitlicher Beschränkung angeboten werden: Tages- oder Monatsfahrkarten sind dabei typischerweise Fahrkarten ohne eine Einschränkung der Anzahl der Inanspruchnahme und unterscheiden sich von den lediglich begrenzt nutzbaren Einzel- und Mehrfachfahrtausweisen.

7.3 Auswirkungen der Tarifsysteme auf die Routenauswahl

Auswirkungen von Tarifsystemen auf die Routenauswahl sind dann zu erwarten, wenn der Preis für verschiedene Verbindungen ein (mit-)entscheidendes Kriterium darstellt. Bei der Bildung der Relationspreise wird im hier vorgestellten Modell der virtuelle Preis beeinflusst, welcher das über alle Zielgruppen einheitliche Entscheidungskriterium der Verbindungsauswahl zur Laufzeit der Verbindungserzeugung darstellt. Durch die Nivellierung vorhandener Unterschiede sind insbesondere für eher unkomfortable Verbindungen, die in der bisher angewandten entfernungsabhängigen Tarifierung einen Kostenvorteil besaßen, Auswirkungen zu erwarten. Zu erwarten ist auch, dass langsamere Verbindungen grundsätzlich nur noch selten berücksichtigt werden, da für parallel verkehrende schnellere Verbindungen jetzt kein Preisaufschlag mehr fällig wird.

Bei der Betrachtung der Auswirkungen bezogen auf einzelne Bahnhöfe steht zu vermuten, dass Bahnhöfe mit einer Vielzahl von Direktverbindungen weniger stark betroffen sind als Bahnhöfe, die nur wenige abgehende Verbindungen besitzen und auf günstig gelegene Umsteigebahnhöfe angewiesen sind. Im zweiten Fall ist zu erwarten, dass an den Umsteigebahnhöfen jeweils die schnellste weitere Verbindung genutzt werden wird. Die bisherige Aufteilung an dieser Stelle in eher langsame und dafür günstige und eher schnelle, aber dafür teure, Fahrten entfällt durch die Bildung des Relationspreissystems.

Insbesondere spielt die Taktung der anschließenden Verkehrsmittel eine entscheidende Rolle: Je mehr zeitlich unabhängige Verbindungen erstellt werden können, desto mehr Verbindungen werden auch behalten, weil die ansonsten vorhandene Gleichzeitigkeit ein Aussortierungskriterium darstellt (annahmegemäß wird bei gleichzeitigen Ankünften oder Abfahrten lediglich die vorteilhafteste Verbindung behalten; siehe dazu auch die Verbindungserzeugung in Kapitel 6.2).

7.3.1 Veränderungen im Algorithmus

Die Veränderungen im Algorithmus sind gering: Da für den reinen zielgruppenunabhängigen Vergleich die absolute Höhe des Preises keine weitere Rolle mehr spielt, werden für alle Relationen jeweils einheitliche Preise basierend auf dem minimalen angebotenen Preis festgelegt. Der virtuelle Preis wird dann entsprechend dieses Preisniveaus berechnet, sodass die im ursprünglichen Modell bestehenden Preisunterschiede innerhalb des virtuellen Preises von Fahrten einer Relationen nivelliert werden.

Am algorithmischen Design selbst werden daher keine Änderungen vorgenommen; lediglich die Bildung der Variablen „Preis“ und „virtueller Preis“ wird den veränderten Gegebenheiten wie geschildert angepasst und die Datenbankergebnisse danach mit den geänderten Parametern erneut komplett berechnet. Die Beibehaltung des algorithmischen Designs dient dabei auch der Vergleichbarkeit der Ergebnisse: Würden am Grundprinzip des Algorithmus oder des Datenmodells Veränderungen herbeigeführt, so wäre eine Vergleichbarkeit der Daten mit den zuvor ermittelten Ergebnissen der entfernungsabhängigen Tarifierung nicht zwingend gegeben.

7.3.2 Beispielhafte Betrachtung von Stationen und Verbindungen

In der folgenden Tabelle 39 ist nun für sechs beispielhaft ausgewählte Startstationen die Anzahl der ohne Umstieg erreichbaren Zielstationen sowie die Anzahl der Verbindungen in Abhängigkeit von der Wahl des Tarifsystems (entfernungsabhängige Tarifierung oder Relationspreissystem) und deren Verhältnis dargestellt:

Startstation	ohne Umstieg erreichbare Stationen	Anzahl der Verbindungen (entfernungsabhängige Tarifierung)	Anzahl der Verbindungen (Relationspreissystem)	Quote
Berlin Hbf	173	67.541	50.690	75,05 %
Schwerin Hbf	43	60.844	47.081	77,38 %
Hannover Hbf	157	65.354	52.689	80,62 %
Lehrte	30	82.911	69.459	83,78 %
Norddeich Mole	16	21.784	19.008	87,26 %
Westerland (Sylt)	12	33.147	30.567	92,22 %

Tabelle 39: Abhängigkeit zwischen der Anzahl direkt verbundener Ziele und den Auswirkungen eines Tarifsysteemwechsels

Quelle: Eigene Darstellung

Zur genaueren Betrachtung sind die Stationen zuerst nach ihrer Funktion sowie der Lage im Verkehrsnetz zu gliedern. Die Stationen Hannover Hbf und Berlin Hbf dienen als zentrale Umsteigestationen im Netz, die Stationen Lehrte und Schwerin Hbf bilden mittelgroße Stationen mit überschaubarem Anschluss- und Verkehrspotential ab, und die Stationen Norddeich Mole und Westerland (Sylt) agieren beide als Grenzen des Verkehrsraums (der in Norddeich Mole anstoßende Seeverkehr ist bedingungsgemäß kein Bestandteil des Datenmodells dieser Arbeit).

Entsprechend erscheinen die Werte nach einer ersten qualitativen Analyse plausibel: Stationen, von denen viele andere Verkehrshalte ohne Umstieg erreicht werden können, sind tendenziell auch in hoher Frequenz an das Netz angebunden. Mithin ergibt sich eine Vielzahl unabhängiger Abfahrten und Ankünfte von Zügen, aber eine im Verhältnis zu in geringerer Frequenz angebundene Stationen niedrigere Anzahl an folgenden Umsteigeverbindungen.

7.3.3 Gesamthafte Betrachtung von Abhängigkeiten

Nach den vorangegangenen Ausführungen ist damit für Bahnhöfe mit einer höheren Anzahl an Abfahrten tendenziell eine niedrigere Quote an zu berücksichtigenden (weil unabhängig voneinander vorteilhaften) Verbindungen zu erwarten, wenn der Wechsel von einem entfernungsabhängigen Tarifsystems auf eine Relationspreis-Tarifierung betrachtet wird.

Gleichzeitig erscheinen die Werte aus Tabelle 39 nicht stringent: die Stationen Westerland (Sylt) und Norddeich Mole sind die Verkehrsstationen mit der höchsten Quote an weiter bestehenden Verbindungen; die Reihenfolge bei den Stationen Hannover Hbf und Schwerin Hbf wäre jedoch der Logik der aufgestellten These folgend nicht zu erwarten gewesen. Es erscheint daher notwendig, die These zu verallgemeinern und dann die Betrachtung im größeren Rahmen anzustellen.

Zu erwägen sind mögliche Indikatoren zur Betrachtung gegeneinander, um Abhängigkeiten feststellen zu können. Aus der vorangegangenen Analyse ist festzuhalten, dass es einen Zusammenhang zwischen der Anzahl der direkt angebundenen Stationen und der Quote der Verbindungen nach einem Tarifsysteemwechsel sowie zwischen der Anzahl der abgehenden Direktverbindungen ab einer Station und der Quote der Verbindungen nach einem Tarifsysteemwechsel gibt. Da beide Zusammenhänge aus der Datenlage nicht direkt ersichtlich waren (was auch der Stichprobe geschuldet sein mag), bietet es sich an, auch das Verhältnis zwischen den erreichbaren Umsteigeverbindungen und den Direktverbindungen ab einer Station zu berücksichtigen, um eine mögliche Relevanz der generellen Verbindungsdichte bezogen auf die folgenden Umstiege im Bezug auf die Quote der Verbindungen nach einem Tarifsysteemwechsel feststellen zu können.

Bei der Berücksichtigung der etwa 250 größten Stationen im Modell (bezogen auf die zugeordnete Einwohnerzahl) zeigt sich ein deutlicher Zusammenhang zwischen der

Quote der zu behaltenden noch vorteilhaften Verbindungen und des Verhältnisses an zu bildenden Umsteigeverbindungen, dargestellt in Abbildung 10:

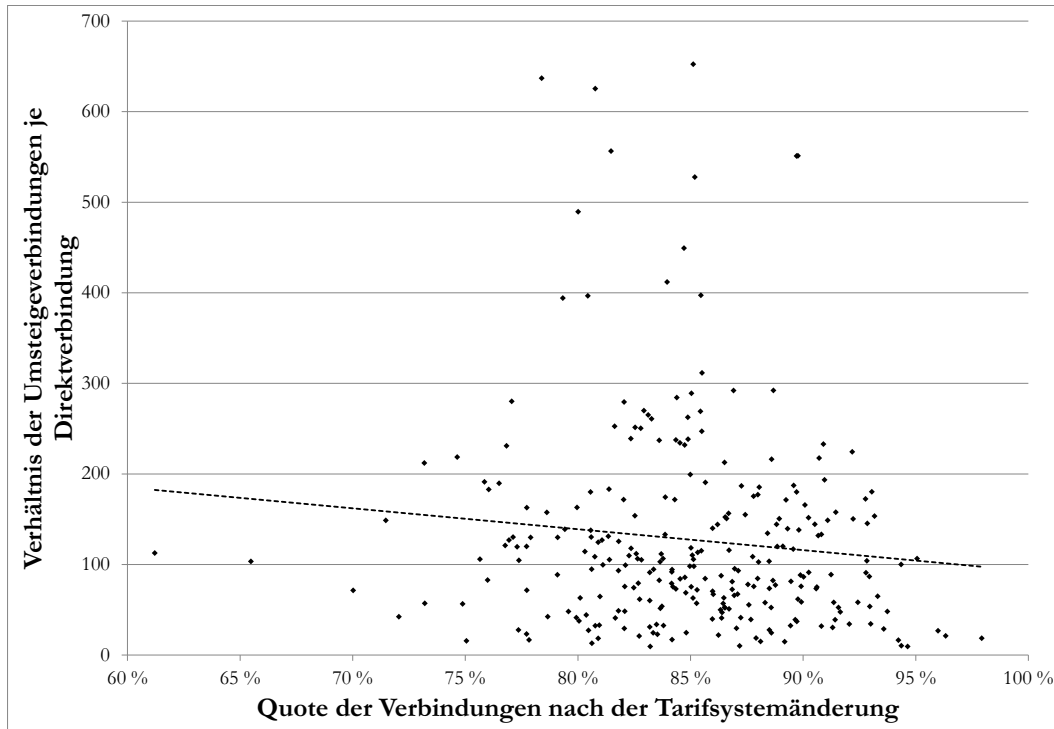


Abbildung 10: Darstellung der Verbindungsquote nach Tarifsyst m nderung gegen das Verh ltnis der Umsteigeverbindungen je Direktverbindung

Quelle: Eigene Darstellung

Dabei ist in Abbildung 10 auf der Ordinatenachse das Verh ltnis der zu bildenden Umsteigeverbindungen je Direktverbindung angegeben. Bei dem Beispiel der Station „Westerland (Sylt)“ mit den Verbindungsdaten aus Tabelle 39 und der Anzahl von 219 Direktverbindungen ergibt sich dabei ein Verh ltnis von $\frac{(30567-219)}{219} = 138,57$ Verbindungen je Direktverbindung (bedingungsgem   sind die Direktverbindungen auch in der Anzahl der gesamten Verbindungen enthalten und werden daher abgezogen, um die Anzahl der Umsteigeverbindungen zu ermitteln). Die Werte der Abszissenachse in Abbildung 10 beziehen sich auf die bereits in Tabelle 39 dargestellte Quote der noch zu behaltenden Verbindungen bei der Umstellung des Tarifsystems von entfernungsabh ngiger Tarifierung auf ein Relationspreissystem.

In Abbildung 10 ist au erdem eine Regressionsgerade eingezeichnet (als gestrichelte Linie), die den Trend der Zahlenwerte widerspiegelt: Die Quote der noch zu behaltenden Verbindungen bei dem Tarifsyst mwechsel nimmt zu, je niedriger das Verh ltnis der

Umsteigeverbindungen je Direktverbindung ist. Umgekehrt lässt sich festhalten: Je mehr Umsteigeverbindungen für jede Direktverbindung einer Abgangsstation im alten Preissystem angeboten werden, desto niedriger wird dieses Verhältnis nach dem Wechsel des Tarifsystems sein (da die Quote der Verbindungen dann entsprechend sinkt).

Auch bei der Darstellung der Anzahl der Direktverbindungen ab einer Station und der Quote der zu behaltenden, weil vorteilhaften, Verbindungen bei Änderung des Preissystems zeigt sich ein deutlicher Zusammenhang:

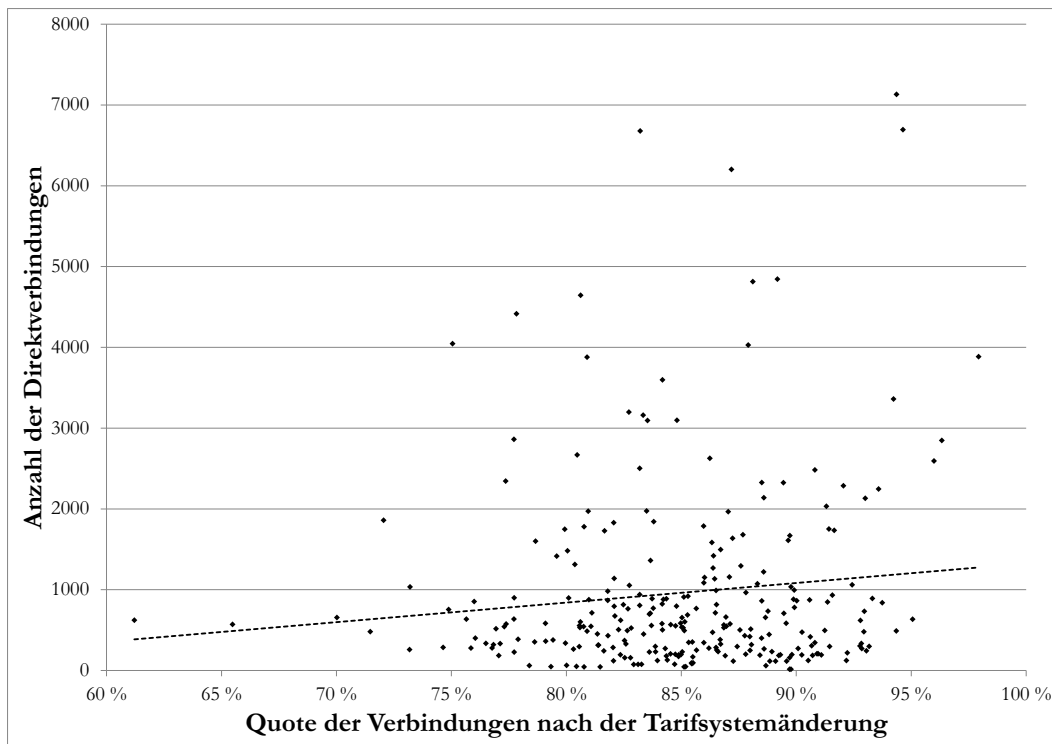


Abbildung 11: Darstellung der Verbindungsquote nach Tarifsystemänderung gegen die Anzahl der Direktverbindungen

Quelle: Eigene Darstellung

Der in Abbildung 11 dargestellte Zusammenhang zwischen der Anzahl der Direktverbindungen ab einer Station und der Quote der noch zu behaltenden gesamten Verbindungen bei Änderung des Preissystems ist dabei in der gestrichelt dargestellten Regressionsgeraden am besten sichtbar. Es zeigt sich, dass mit zunehmender Anzahl an Direktverbindungen auch das Verhältnis der insgesamt zu behaltenden Verbindungen beim Wechsel des Tarifsystems ansteigt. Argumentativ kann diese Darstellung auch untermauert werden, wenn man annimmt, dass Direktverbindungen und an Direktverbindungen anschließende

Verbindungen tendenziell werthaltigere Verbindungen in einem Relationspreissystem darstellen als Verbindungen mit vielen Umstiegen.

Beim Abtrag der Grafik mit Darstellung des Zusammenhangs zwischen der Anzahl der direkt erreichbaren Stationen und der Quote der nach Änderung des Preissystems zu behaltenden Verbindungen ergibt sich jedoch kein eindeutiges Bild:

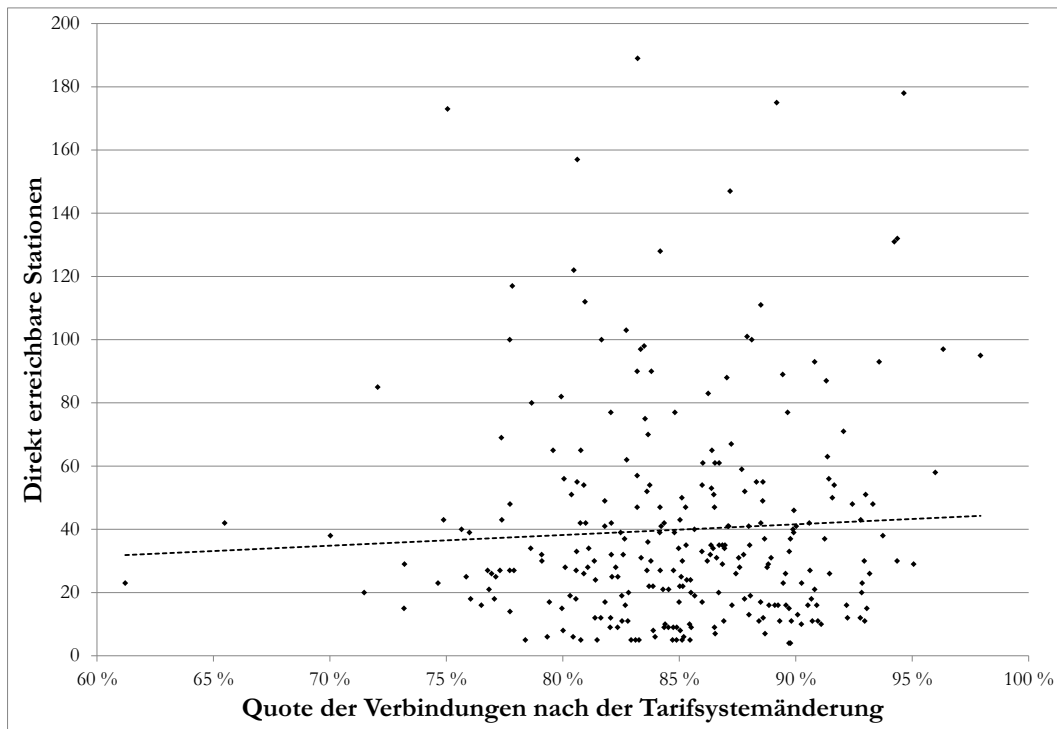


Abbildung 12: Darstellung der Verbindungsquote nach Tarifsystemänderung gegen die Anzahl der direkt erreichbaren Stationen

Quelle: Eigene Darstellung

Es zeigt sich dabei in Abbildung 12, dass die Anzahl der direkt erreichbaren Stationen einen Einfluss auf die Quote der zu behaltenden Verbindungen nach der Änderung des Tarifsystems haben könnte; der Zusammenhang scheint jedoch eher schwach zu sein, was sich auch in der verhältnismäßig geringen Steigung der wiederum gestrichelt dargestellten Regressionsgeraden manifestiert, insbesondere im Vergleich mit den Regressionsgeraden aus den Abbildungen 10 und 11. Argumentativ ist die Grafik jedoch auch in dieser Form erwartbar, da das Verhältnis zwischen den direkt erreichbaren Bahnhöfen und der Anzahl der Direktverbindungen nur indirekt zusammenzuhängen scheint, wie auch in Tabelle 39 beispielhaft dargestellt.

7.3.4 Wertung der Ergebnisse

Beim Vergleich der Anzahl der erstellten Verbindungen fällt auf, dass sowohl die absolute Anzahl der beim Wechsel des Tarifsystems obsoleten Verbindungen wie auch die Quote der Verbindungen stark schwankt. Mithilfe der Regressionsgeraden ist der Zusammenhang besonders stark bei der Anzahl der Direktverbindungen ab einer Station im Verhältnis zur Quote der zu behaltenden Verbindungen nach dem Preissystemwechsel nachweisbar (dargestellt in Abbildung 11). Ein solcher Zusammenhang war zu erwarten: Da die Anzahl der Direktverbindungen auch die Anzahl der weiteren Verbindungen mittelbar beeinflusst, indem zu vielen Direktverbindungen auch Anschlussverbindungen bereitstehen, war auch zu erwarten, dass mit der Menge der Direktverbindungen die Menge der Anschlussverbindungen unterproportional ansteigt – und damit bei einem Wechsel des Tarifsystems weg von einer entfernungsabhängigen Tarifierung hin zu einem Relationspreissystem damit auch die Quote der zu behaltenden Verbindungen höher ist (was in Abbildung 10 auch dargestellt ist).

Der eher verhaltene Zusammenhang zwischen der Anzahl der ab einer Startstation direkt erreichbaren weiteren Stationen und der Quote der Verbindungen bei einem Wechsel des Tarifsystems war nach den anderen Grafiken auch so zu erwarten: Nachdem die Menge der Direktverbindungen einen sehr starken Einfluss auf die Höhe der Quote der zu behaltenden Verbindungen beim Tarifsystemwechsel besitzt und sich das Verhältnis der Umsteigeverbindungen je Direktverbindung bei steigender Verhältniszahl negativ auswirkt, war für Stationen mit eher wenigen Direktverbindungen, die jedoch an vielen Punkten Anschluss besitzen und damit eine große Menge an Umsteigeverbindungen je Direktverbindung erzeugen, davon auszugehen, dass die Quote der Verbindungen sich nicht im selben Maße wie bei den Direktverbindungen entwickelt, sondern in gewissem Maße kompensiert wird.

Der über alle Graphen eher deutliche Verlauf der Regressionsgeraden deutet darüber hinaus darauf hin, dass der Zusammenhang in ausreichendem Maße besteht, um in weiteren Analysen quantitativ untersucht und bestätigt zu werden.

8 Fazit und Handlungsempfehlung

Im Rahmen dieser Arbeit wurde eine Modellumgebung zur Darstellung von Routenfindungsproblemen und der Auswirkung möglicher Änderungen an den Systemparametern, insbesondere am Preissystem, gebildet. Nach der Darstellung der modellbildenden Primärdaten wurde dabei auf eine Zielgruppendefinition abgestellt, die im Ergebnis die Unterteilung der in der Literatur üblicherweise aufgeworfenen Zielgruppen Privatreisende und Geschäftsreisende in insgesamt vier Zielgruppen mit verschiedenen Teilzielgruppen und zu beachtenden Aspekten dieser Teilzielgruppen zur Folge hatte.

Neben der Bildung der Zielgruppen und der Beschreibung der Präferenzen ist die Monetarisierung dieser Zielgruppenpräferenzen, jeweils bezogen auf die im Nutzenmodell einbezogenen Parameter, zentraler Bestandteil des Auswahlmodells von Verbindungen. Dabei war neben der Unterteilung der Nutzenwerte in verschiedene Teilnutzen – Zeit, Preis und Komfort – eine weitere Unterteilung speziellen Berücksichtigung der Belange verschiedener Nutzergruppen notwendig: So wird der Umsteigewiderstand in einen generellen Umsteigewiderstand aufgrund der Notwendigkeit des Zugwechsels und einen Umsteige-Zeitwiderstand aufgrund der mit Unannehmlichkeiten verbundenen Wartezeit am Umsteigebahnhof aufgeteilt. Die Bildung der Variable virtueller Preis mit Verrechnung der wesentlichen Eigenschaften auf Basisniveau zu erleichterten Aussortierung unnützer Verbindungen hat sich dabei als hilfreich erwiesen.

Die Modellbildung erfolgt auf Basis eines SQL-Datenbankmodells, das zu Beginn im Grundsatz erläutert und dann in der detaillierten Zusammensetzung vorgestellt wird. Die darauf zu bildenden Verbindungen bestehen ihrerseits aus Direktverbindungen zwischen zwei Haltezeitdatensätzen eines Zuges und werden mit spezifischen Eigenschaften wie Preis und Verkehrstagen versehen, bevor die Bildung von Umsteigeverbindungen erfolgt. Die Bildung der Umsteigeverbindungen erfolgt hierbei zweigeteilt: Neben der Vorstellung des generellen Ansatzes (Bildung von Umsteigeverbindungen als Tupel auf Basis optimaler Primärverbindungen geringerer Umstiegszahl) wird die Umsteigebildung von Verbindungen des ersten Umstiegs vorgestellt, der auf Basis der noch nicht optimierten Umsteigesituation erfolgen muss und daher weitere Hilfskonstrukte wie die Bildung einer speziellen Datenbanktabelle zur Berücksichtigung paralleler Züge beinhaltet. Die Routenauswahl wird schließlich zielgruppenweise vorgenommen, um jetzt für jede Zielgruppe ein ideales Verbindungsbild zu erhalten und bewerten zu können – auf Basis dieses Verbindungsbilds sollen sich Zugehörige der Zielgruppe dann für eine Verbindung entscheiden können.

Die Trennung der Algorithmen für den ersten und alle Umstiege einerseits sowie zur Routenfindung andererseits war für die Performance von großer Relevanz; gleichzeitig wäre zu diskutieren gewesen, ob ein etwas kleineres Modell nicht ähnliche Analysen ermöglicht hätte und einfacher handzuhaben gewesen wäre.

In der anschließenden Analyse der Auswirkungen verschiedener Tarifsysteme auf die Anzahl der anzubietenden Routen ergibt sich, dass der Wechsel von einem entfernungsabhängigen Tarif zu einem Tarifsystem mit Relationspreisen zu einer massiven Reduktion vorteilhafter (und damit zur Auswahl vorzuschlagender) Verbindungen führt, weil der bisher als Entscheidungskriterium dienende Preis nicht mehr dieselbe Funktion erfüllen kann: Alle Verbindungen einer Relation unterliegen fortan bedingungsgemäß demselben Preis. Diese Abhängigkeit zeigt sich nicht direkt bezogen auf die Anzahl der Verbindungen, aber sowohl auf die Anzahl der Direktverbindungen ab einer Station wie auch auf die durchschnittliche Anzahl der Umsteigeverbindungen für jede Direktverbindung ab einer Startstation. Das Bild beim Vergleich der Anzahl der erreichbaren Stationen ab einer Startstation im Vergleich zur Quote der zu behaltenden Verbindungen nach einem Wechsel des Tarifsystems ist dabei ambivalent, erscheint aber im Zusammenhang mit den anderen festgestellten Indikatoren ebenfalls eindeutig genug als Beweis der am Anfang getätigten These, dass die Routenauswahl elementar von Eigenschaften der Tarifierung beeinflusst wird.

Insbesondere bei der Routenauswahl kann diese Arbeit vertieft werden, indem ein Passagiererzeugungsmodell auf die bereits ermittelten Routen aufgesetzt wird. Dabei ist in einem ersten Schritt eine Bewertung der Routen nach ihrer tatsächlichen Zielgruppenrelevanz vorzunehmen und in einem weiteren Schritt nach der Bestimmung der Quelle-Ziel-Potenziale die Verteilung der Passagiere zwischen zwei Stationen auf die ermittelten Routen vorzunehmen. Auch eine weitere Vertiefung der Auswirkungen der Auswahl des Tarifsystems auf die Anzahl und Art der erzeugten Routen kann auf Basis dieser Arbeit vorgenommen werden, um die Forschung dieses Arbeitsfelds weiter voranzutreiben und das Gesamtsystem Eisenbahn mit seinen tariflichen Zusammenhängen besser verstehen und erklären zu können – und damit schlussendlich zu stärken.

Literatur

- [AEG16] AEG. *Allgemeines Eisenbahngesetz. Allgemeines Eisenbahngesetz vom 27. Dezember 1993 (BGBl. I S. 2378, 2396; 1994 I S. 2439), das durch Artikel 2 des Gesetzes vom 29. August 2016 (BGBl. I S. 2082) geändert worden ist.* 2016.
- [Air16] Air Berlin PLC & Co. Luftverkehrs KG. *Super Sunday - jeden Sonntag zwischen 13 - 18h Reiseschnäppchen sichern - airberlin.com.* 2016. URL: <https://flights.airberlin.com/de-DE/super-sunday>.
- [Bat08] Richard Batley. "On Ordinal Utility, Cardinal Utility and Random Utility". In: *Theory and Decision* 64 (1 2008), S. 37–63.
- [Bel11] Polly Trottenberg; Peter Belenky. *Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis.* Forschungsber. U.S. Department of Transportation, 28. Sep. 2011.
- [Bel14] Ilya Beliaev. *MariaDB vs. MySQL.* Hrsg. von php-dev.info. 10. Apr. 2014. URL: <https://blog.php-dev.info/2014/04/mariadb-vs-mysql/>.
- [Bod05] Karl-Dieter Bodack. *InterRegio. Die abenteuerliche Geschichte eines beliebten Zugsystems.* EK-Verlag, 2005.
- [Boe14] Julia Boek. "Schnell ans Ziel. Oder lieber günstig? Metasuchmaschinen wie Qixxit oder Waymate sollen die individuelle Reiseplanung mit Bahn, Bus und Flugzeug erleichtern. Doch halten sie auch, was sie versprechen?" In: *Zeit Online* (14. Feb. 2014). URL: <http://www.zeit.de/reisen/2014-02/kurzreisen-reisen-planen-online>.
- [Bös69] Dieter Bös. "Die Formeltarife der Österreichischen Bundesbahnen". In: *Zeitschrift für Nationalökonomie* 29 (1969), S. 445–464.
- [Cle96] Reinhard Clever. "Schnelligkeit oder Häufigkeit: Überlegungen zur Einführung des Integralen Taktfahrplans im Fernverkehr der Eisenbahn. Sonderdruck". In: *Zeitschrift für Verkehrswissenschaft* Heft 2 (67. Jahrgang, 1996). Hrsg. von Rainer Willeke; Herbert Baum, S. 137–181.
- [Col05] Stuart Cole. *Applied Transport Economics.* 3. Aufl. Kogan Page, 2005.

- [DB 15a] DB Netz AG. *Kursbuch der Deutschen Bahn AG 2016, Kursbuchtafel 770 Karlsruhe - Pforzheim - Mühlacker - Vaihingen (Enz) - Stuttgart*. Hrsg. von DB Netz AG. Frankfurt am Main, 24. November 2015. URL: http://kursbuch.bahn.de/hafas/kbview.exe/dn/KB770_H_Taeglich_G24112015.pdf?filename=KB770_H_Taeglich_G24112015.pdf.
- [DB 15b] DB Netz AG. *Kursbuch der Deutschen Bahn AG 2016, Kursbuchtafel 940 München - Mühldorf*. Hrsg. von DB Netz AG. 2015. URL: http://kursbuch.bahn.de/hafas/kbview.exe/dn/KB940_H_TGL_F_G24112015.pdf?filename=KB940_H_TGL_F_G24112015.pdf.
- [DB 15c] DB Netz AG, I.NMK 1 (Z). *TPS-Preisauskunft-Software 2016 (Stand: April 2015)*. V7.1d. DB Netz AG. Frankfurt am Main, 2015. URL: http://fahrweg.dbnetze.com/fahrweg-de/produkte/trassen/trassenpreise/trassenpreisauskunft_tpis.html.
- [DB 16a] DB Vertrieb GmbH. *Günstiges Bahnticket: Der Sparpreis der Bahn!* 2016. URL: <https://www.bahn.de/p/view/angebot/sparpreis/spartickets.shtml>.
- [DB 16b] DB Vertrieb GmbH. *Mit dem Flexpreis flexibel reisen*. 2016. URL: <https://www.bahn.de/p/view/angebot/flexpreis.shtml>.
- [DB 16c] DB Vertrieb GmbH. *Reiseauskunft der Deutschen Bahn AG*. 2016. URL: <https://reiseauskunft.bahn.de/bin/query.exe>.
- [Deu59] Deutsche Bundesbahn. *Anhang zum Entfernungszeiger für den Personen- und Gepäckverkehr - Teil I -. Raumbegrenzungsstafel*. Essen, 1959.
- [EBO16] EBO. *Eisenbahn-Bau- und Betriebsordnung. Eisenbahn-Bau- und Betriebsordnung vom 8. Mai 1967 (BGBl. 1967 II S. 1563), die zuletzt durch Artikel 1 der Verordnung vom 10. Oktober 2016 (BGBl. I S. 2242) geändert worden ist*. 2016.
- [exc16a] exchange-rates.org. *Wechselkurse des Britischen Pfundes(GBP) gegenüber dem Euro(EUR)*. 2016. URL: <http://de.exchange-rates.org/history/EUR/GBP/T>.
- [exc16b] exchange-rates.org. *Wechselkurse des US-Dollar(USD) gegenüber dem Euro(EUR)*. 2016. URL: <http://de.exchange-rates.org/history/EUR/USD/T>.

- [Fra02] Frankfurter Allgemeine Zeitung. “Mit Tempo 300 zwischen Frankfurt und Köln”. In: *Frankfurter Allgemeine Zeitung* (2002). URL: <http://www.faz.net/aktuell/gesellschaft/ice-neubaustrecke-mit-tempo-300-zwischen-frankfurt-und-koeln-171134.html>.
- [Gol76] Bruce Golden. “Technical Note - Shortest-Path Algorithms: A Comparison”. In: *Operations Research* 24.6 (1976), S. 1164–1168.
- [Gre07] Michaela Wänke; Rainer Greifeneder. *Mehr ist mehr? Die psychologische Wirkung von Angebotsvielfalt und Markenbreite*. In: Arnd Florack; Martin Scarabis; Ernst Primosch. *Psychologie der Markenführung*. 1. Aufl. Franz Wahlen, 2007, S. 149–158.
- [Höd06] Sonja Hödl. *Der europäische Markt für Nachtreisezugverkehre. Eine empirische Analyse der Nachfragedeterminanten. Schriftenreihe des Instituts für Transportwirtschaft und Logistik - Verkehr, 02/2006*. Institut für Transportwirtschaft und Logistik, WU Vienna University of Economics und Business, Wien, 2006.
- [Hof08] Marcel Rolf Hoffmann. *Sparpreise aus der Lotterie*. Hrsg. von NEON. 2008. URL: <http://www.neon.de/artikel/kaufen/produkte/sparpreise-aus-der-lotterie/666456>.
- [Hol05] Glenn Lyons; Juliet Jain; David Holley. *The use of travel time by rail passengers in Great Britain*. 2005. URL: http://sand14.com/archive/traveltimeuse/downloads/LyonsJainHolley_UseofTravelTime.pdf.
- [II03] William Barnett II. “The modern theory of consumer behavior: ordinal or cardinal?” In: *The quarterly journal of austrian economics* 6.1 (2003), S. 41–65.
- [Ins06] TRAMP - Traffic and Mobility Planning GmbH; Deutsches Institut für Urbanistik; Institut für Wirtschaftsforschung Halle. *Szenarien der Mobilitätsentwicklung unter Berücksichtigung von Siedlungsstrukturen bis 2050. Abschlussbericht*. Hrsg. von Bundesministerium für Verkehr, Bau und Stadtentwicklung. Magdeburg, 2006.
- [Jan95] Piotr Jankowski. “Integrating geographical information systems and multiple criteria decision-making methods”. In: *International Journal of Geographical Information Systems* 9 (3 1995), S. 251–273.

- [Jip09] Zhang Fuhao; Liu Jiping. “An algorithm of shortest path based on Dijkstra for huge data”. In: *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery. Hrsg. von Chinese Academy of Surveying Research Center of Government Geographic Information System und Mapping. 2009, S. 244–247.
- [Kei12] K. T. Geurs; G. H. Haarsman; B. De Keizer. *Interchanges in timetable design of railways: A closer look at customer resistance to interchange between trains*. Hrsg. von Association for European Transport und Contributors. 2012.
- [Kir16] Alexander Kirk. *RAM*. 2016. URL: <http://www.computerlexikon.com/was-ist-ram>.
- [Kra91] Günther Krauß. “<http://www.zeit.de/1991/19/2-juni-bahn-frei-fuer-die-schnellen-ice-zuege>”. In: *ZEIT Online* (3. Mai 1991). URL: <http://www.zeit.de/1991/19/2-juni-bahn-frei-fuer-die-schnellen-ice-zuege>.
- [Mar16a] MariaDB Corporation Ab. *MariaDB versus MySQL - Features*. 2016. URL: <https://mariadb.com/kb/en/mariadb/mariadb-vs-mysql-features/>.
- [Mar16b] MariaDB Corporation Ab. *MariaDB versus MySQL - Kompatibilität*. 2016. URL: <https://mariadb.com/kb/de/mariadb-vs-mysql-compatibility/>.
- [Mar16c] MariaDB Corporation Ab. *MEMORY Storage Engine*. 2016. URL: <https://mariadb.com/kb/en/mariadb/memory-storage-engine/>.
- [Mar16d] Daniel Markgraf. *Produktdifferenzierung*. *Gabler Wirtschaftslexikon*. Hrsg. von Springer Gabler Verlag. 2016. URL: <http://wirtschaftslexikon.gabler.de/Definition/produktdifferenzierung.html>.
- [Mat15] Charlotte Mathieson. *Mobility in the Victorian Novel*. 1. Aufl. Palgrave Macmillan UK, 2015.
- [Mie00] Hans De Neve; Piet Van Mieghem. “TAMCRA: a tunable accuracy multiple constraints routing algorithm”. In: *Computer Communications* 23 (7 2000), S. 667–679.

- [myS16a] MySQL AB. *12.2.1 Integer Types (Exact Value) - INTEGER, INT, SMALLINT, TINYINT, MEDIUMINT, BIGINT*. Hrsg. von Oracle Corporation. 2016. URL: <http://dev.mysql.com/doc/refman/5.7/en/integer-types.html>.
- [myS16b] MySQL AB. *Download MySQL Community Server*. Hrsg. von Oracle Corporation. 2016. URL: <http://dev.mysql.com/downloads/mysql/>.
- [Nas86] A. S. Fowkes; P. Marks; C. A. Nash. *The Value of Business Travel Time Savings. Working Paper 214*. Leeds: University of Leeds, 1986.
- [Nor12] Manuel Norta. *Ein Verkehrsmittelwahlmodell für den Personenfernverkehr auf der Basis von Verkehrswiderständen*. 2012.
- [Off14] Office of Rail Regulation. *Rail passenger experience report*. 2014. URL: http://orr.gov.uk/__data/assets/pdf_file/0003/11748/rail-passenger-experience-report.pdf.
- [Pel16] Jonas Pelkmann. *Historische Entwicklung der Inflation in Deutschland*. 2016. URL: <http://www.inflation-deutschland.de/inflation-historisch.html>.
- [Pel99] Peter Gulutzan; Trudy Pelzer. *SQL-99 Complete, Really. An Example-Based Reference Manual of the New Standard*. Lawrence, Kansas, 1999.
- [Pfu08] Carlo Pfund. *Angebot und Preise der SBB und der Nachbarbahnen SNCF, DB, FS/Trenitalia und ÖBB*. 2008.
- [Phi87] A. D. Owen; G. D. A. Phillips. "The characteristics of railway passenger demand". In: *Journal of Transport Economics and Policy* 21.3 (Sep. 1987), S. 231–253.
- [Pie16] Dirk Piekenbrock. *Nutzentheorie. Gabler Wirtschaftslexikon*. Hrsg. von Springer Gabler Verlag. 2016. URL: <http://wirtschaftslexikon.gabler.de/Definition/nutzentheorie.html>.
- [Rüg05] Bernhard Rüter. "Bereitschaftselastizität - Empirische Ermittlung zum Verkehrsmittelwahlverhalten". In: *20. Verkehrswissenschaftliche Tage*. Dresden, 2005.
- [Sch07] Silke Maria Schmitz. *Nachfrageanalyse, Economies of Scale und Preisdifferenzierung im deutschen Lebensmitteleinzelhandel*. Christian-Albrechts-Universität zu Kiel, 2007.

- [Ste14] René Steiner. *Grundkurs Relationale Datenbanken. Einführung in die Praxis der Datenbankentwicklung für Ausbildung, Studium und IT-Beruf*. 8. Aufl. Springer Vieweg, 2014.
- [Str01] Mark Wardman; Julian Hine; Stephen Stradling. *Interchange and Travel Choice Volume 1*. Edinburgh: Scottish Executive Central Research Unit, 2001.
- [Tho15] Thomson Reuters Deutschland GmbH. “Dienstreisen mit Firmenwagen gehen zurück”. In: *Handelsblatt* (27. Feb. 2015).
- [Val97] Klaus Walther; Klaus Oetting; Dirk Vallee. *Simultane Modellstruktur für die Personenverkehrsplanung - Auf der Basis eines neuen Verkehrswiderstands*. Schriftenreihe des Verkehrswissenschaftlichen Institutes der RWTH Aachen. Bd. 52. RWTH Aachen (Eigenverlag), 1997.
- [Wal08] Thomas Walter. *Kompendium der Web-Programmierung. Dynamische Websites*. Springer-Verlag, 2008.
- [Web05] Wolfgang Küchlin; Andreas Weber. *Einführung in die Informatik*. 3. Aufl. Springer, 2005.
- [Wie06] R. Engel; H. Jansen; E. Lauterbach; M. Wiegner. *Fahrgastthemen: Entwicklung der Zuschläge bei der Bahn*. Hrsg. von PRO BAHN e.V. 2006. URL: <http://www.pro-bahn.de/fakten/zuschlag.htm>.
- [Zar02] Frank Schulz; Dorothea Wagner; Christos Zaroliagis. *Using Multi-level Graphs for Timetable Information in Railway Systems*. In: *Algorithm Engineering and Experiments*. Hrsg. von David M. Mount; Clifford Stein. 2002, S. 43–59.
- [Zha97] F. Benjamin Zhan. “Three Fastest Shortest Path Algorithms on Real Road Networks: Data Structures and Procedures”. In: *Journal of Geographic Information and Decision Analysis* 1.1 (1997), S. 70–82.