

Reglerentwurf zur dezentralen Online-Steuerung von Lichtsignalanlagen in Straßennetzwerken

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät
Verkehrswissenschaften „Friedrich List“
an der Technischen Universität Dresden

von
Dipl.-Ing. Stefan Lämmer
geboren am 15. Mai 1977 in Annaberg-Buchholz

eingereicht am 10. April 2007

Kurzfassung

Die vorliegende Dissertationsschrift widmet sich einer systemtheoretischen Untersuchung zur verkehrsabhängigen Steuerung von Lichtsignalanlagen in Straßennetzwerken. Aus einem mathematischen Modell für den Verkehrsablauf auf Knotenzufahrten wird ein Verfahren abgeleitet, mit dem sich Umschaltzeitpunkte und Phasenwechsel flexibel an das tatsächliche Verkehrsgeschehen anpassen lassen. Der Ansatzpunkt ist, die einzelnen Knotenpunkte des Netzwerks lokal zu optimieren. Eine „Grüne Welle“ soll sich von selbst einstellen, und zwar genau dann, wenn dadurch lokal Wartezeiten eingespart werden. Indem die lokale Optimierung in ein lokales Stabilisierungsverfahren eingebettet wird, können Instabilitäten aufgrund netzwerkweiter Rückkopplungen ausgeschlossen werden. Das vorgestellte Verfahren setzt sich aus drei Teilen zusammen:

- (i) einem lokalen Prognoseverfahren zur Bewertung von Schaltzuständen und Phasenübergängen bezüglich zukünftig entstehender Wartezeiten,
- (ii) einem lokalen Optimierungsverfahren, das jeder Phase einen dynamischen Prioritätsindex zuweist und die Phase mit höchster Priorität zur Bedienung auswählt und
- (iii) einem lokalen Stabilisierungsverfahren, das zum Einhalten einer mittleren und einer maximalen Bedienperiode korrigierend in die lokale Optimierung eingreift.

Indem die Knotenpunkte ausschließlich über die Verkehrsströme gekoppelt sind, ergeben sich die Umschaltzeitpunkte unmittelbar aus den Ankunftszeitpunkten der Fahrzeuge selbst. Die Phasenwechsel stellen sich somit von selbst bedarfsgerecht ein. Simulationsergebnisse machen deutlich, dass sich aufgrund der höheren Flexibilität sowohl die Wartezeiten als auch der Kraftstoffverbrauch senken lassen.

Danksagung

Mein ganz besonderer Dank geht an Karsten Peters, der mich an seiner umfassenden Erfahrung mit hybriden dynamischen Systemen teilhaben ließ, Erjen Lefeber, mit dem ich unermüdlich nach optimalen Trajektorien für geschaltete Warteschlangennetzwerke suchen konnte, Dieter Armbruster, der mir den Anstoß gab, komplexe Systeme durch gezielte Beeinflussung ihrer internen Wechselwirkungen sich selbst organisieren zu lassen, und Jochen Sprickerhof, mit dem ich die ersten Ideen zu dieser Arbeit diskutieren und in Computersimulationen erproben konnte.

Als studentische Hilfskräfte haben mich Winnie Pohl, Kristin Meier, Christine Jost und Carsten Knoll begleitet. Ihnen besten Dank für die engagierte und anregende Unterstützung! Gunter Thiele danke ich für die empirische Tagesganglinie aus Abb. 7.11.

Der Studienstiftung des Deutschen Volkes verdanke ich ein zweijähriges Promotionsstipendium. Vielen Dank auch der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung durch das Forschungsprojekt He 2789/5-1 (2003). Bedanken möchte ich mich auch beim Benson-Ford-Research-Center (Dearborn, USA) für die zuvorkommende Mitwirkung bei der Recherche in den Archiven des Henry-Ford-Museums.

Am herzlichsten jedoch danke ich meinem Doktorvater, Professor Dirk Helbing. Er hat mich angespornt, mit unkonventionellen Ideen gänzlich neue Zugänge zu bislang ungelösten Problemen zu finden. Seine hohen Erwartungen waren immer mit unnachgiebigem Vertrauen verbunden. Zu verdanken habe ich ihm auch die Teilnahme an zahlreichen internationalen Konferenzen und Arbeitskreisen.

Inhaltsverzeichnis

Kurzfassung	ii
Danksagung	iii
Abbildungsverzeichnis	viii
1. Einleitung	1
1.1. Einordnung	2
1.1.1. Ausgangssituation	2
1.1.2. Entwicklungstrend	2
1.1.3. Besondere Probleme dezentraler Steuerungsansätze	3
1.2. Beitrag dieser Dissertationsschrift	5
1.2.1. Ansatzpunkt	5
1.2.2. Lösungsschritte	5
1.2.3. Annahmen	6
1.2.4. Wesentliche Ergebnisse	7
1.3. Kapitelübersicht und verwendete Methoden	8
2. Stand der Forschung	11
2.1. Entwicklung der Lichtsignalsteuerung	12
2.1.1. Erste Lichtsignalanlagen	12
2.1.2. Steuerungsverfahren der ersten Generation	14
2.1.3. Verkehrsadaptive Steuerungsverfahren	16
2.2. Geschaltete Warteschlangennetzwerke	20
2.2.1. Steuerungsverfahren	21
2.2.2. Stabilität und Steuerbarkeit	23
2.3. Umlaufbasierte Steuerungen	24
2.3.1. Randbedingungen	24
2.3.2. Entwurfsparameter	26
2.3.3. Warteschlangen und Wartezeiten	30
2.3.4. Versatzzeitenanpassung	32
2.3.5. Schwierigkeiten netzwerkweiter Koordinierung	35
2.4. Zusammenfassung und Kritik	37
3. Modellierung und Prognose von Bedienvorgängen	39

3.1.	Verkehrsfusserfassung	40
3.1.1.	Zeitreihe der erwarteten Ankünfte	40
3.1.2.	Anzahl abgefertigter Fahrzeuge	44
3.2.	Warteschlangenmodell	45
3.2.1.	Warteschlangenlänge	45
3.2.2.	Wartezeiten	46
3.2.3.	Kraftstoffverbrauch	49
3.3.	Prognose von Bedienvorgängen	50
3.3.1.	Erforderliche Freigabezeit zum Leeren der Warteschlange	51
3.3.2.	Eigenschaften der prognostizierten Kenngrößen	53
3.3.3.	Prognostizierte Gesamtwartezeit	58
3.3.4.	Abbruch eines Bedienvorgangs	62
3.4.	Zusammenfassung	64
4.	Lokale Optimierung	67
4.1.	Prioritätsbasierte Verfahren	68
4.1.1.	Existenz optimaler Prioritätsindizes	68
4.1.2.	Herleitung der μc -Regel	69
4.2.	Dynamische Priorisierung	73
4.2.1.	Erweiterung um Kurzzeitprognose	73
4.2.2.	Priorisierung bereits geleerter Warteschlangen	77
4.2.3.	Verallgemeinerte dynamische Priorisierung	79
4.3.	Spezifikation des Optimierungsverfahrens	82
4.3.1.	Schaltprogramm und Zusammenfassung	83
4.3.2.	Schaltverhalten	84
4.3.3.	Instabiles Verhalten	87
5.	Netzwerkweite Stabilität	90
5.1.	Stabilisierungsprinzip	91
5.1.1.	Lokales Effizienzkriterium	91
5.1.2.	Stabilisierender Regler	92
5.2.	Wunsch-Bedienperiode	94
5.2.1.	Wahl des Schwellenwerts	94
5.2.2.	Hinreichende Stabilitätsbedingung	95
5.3.	Spezifikation des Stabilisierungsverfahrens	96
5.3.1.	Priorisierungsregeln	96
5.3.2.	Zuweisung der Freigabezeiten	97
5.3.3.	Heuristischer Ansatz	99
5.4.	Maximale Bedienperiode	100
5.4.1.	Virtueller Fahrzeugstrom	100
5.4.2.	Zeitabhängiger Schwellenwert	102
5.5.	Schaltverhalten und Zusammenfassung	104

5.5.1. Schaltverhalten	104
5.5.2. Zusammenfassung	108
6. Kombiniertes Verfahren	110
6.1. Schaltprogramm	111
6.1.1. Grundidee	111
6.1.2. Programmablauf	111
6.2. Schaltverhalten	113
6.2.1. Eingriff des stabilisierenden Regimes	113
6.2.2. Grenzyklus bei konstanter Ankunftsrate	113
6.2.3. Kumar-Seidman-Netzwerk	115
6.2.4. Zusammenfassung	115
7. Simulationsergebnisse	117
7.1. Simulationsdurchführung	118
7.1.1. Verwendete Parameter	118
7.1.2. Verkehrsflussmodell	118
7.2. Isolierter Knotenpunkt	119
7.2.1. Konstante Zuflussraten	120
7.2.2. Kraftstoffverbrauch	125
7.2.3. Stochastische Zufüsse	127
7.3. Arterie	129
7.3.1. Variation der Richtungsströme	129
7.3.2. Freigabezeitbänder	132
7.3.3. Variation der Knotenanzahl	135
7.4. Komplexes Netzwerk	137
7.4.1. Tagesganglinie	137
7.4.2. Mittlere Warteschlangenlänge	139
7.4.3. Kraftstoffverbrauch	141
8. Ausblick und Zusammenfassung	143
8.1. Ausblick	144
8.1.1. Heterogene Verkehrsströme	144
8.1.2. Stauraumbegrenzung	145
8.1.3. Mehrphasenbetrieb	147
8.1.4. Weitere Vorschläge	148
8.2. Zusammenfassung	149
A. Herleitungen	152
A.1. Instabilität der Clear-Largest-Buffer-Policy	153
A.1.1. Kumar-Seidman-Netzwerk	153
A.1.2. Zeitliche Entwicklung der Warteschlangen	154

A.1.3. Diskussion	156
A.2. Optimale Platzierung des Freigabezeitfensters	156
A.2.1. Gesamtwartezeiten an einer Zufahrt	157
A.2.2. Konstruktion der Wartezeitenkurve	158
A.3. Konvergenz zum optimalen Grenzyklus	159
A.3.1. Zeitliche Entwicklung der Systemgrößen	159
A.3.2. Konvergenz der Iterationsvorschrift	164
A.4. Theoretische Verteilung der Bedienperiode	165
A.4.1. Definition	165
A.4.2. Theoretische Verteilung mit $T^{\max} \rightarrow \infty$	166
A.4.3. Theoretische Verteilung mit endlichem T^{\max}	167
Symbolverzeichnis	168
Literaturverzeichnis	171
Stichwortverzeichnis	186

Abbildungsverzeichnis

1.1. Gegenüberstellung von Steuerung und Regelung	4
1.2. Prinzipielle Anordnung	7
2.1. Erste Lichtsignalanlage der Welt	13
2.2. US-Patente zur Lichtsignalsteuerung	13
2.3. Erste Lichtsignalanlage Deutschlands	15
2.4. MAXBAND und SCOOT	17
2.5. Gleitende Prognose und Hierarchisch-Dezentrale Dekomposition . .	19
2.6. Kumar-Seidman-Netzwerk	22
2.7. Signalbildabfolge und effektiver Freigabezustand	27
2.8. Durchlassfähigkeit eines lichtsignalgesteuerten Knotenpunktes . . .	29
2.9. Mittlere Gesamtwartezeiten in Abhängigkeit der Umlaufzeit	31
2.10. Zeitscheibe einer umlaufbasierten Steuerung	32
2.11. Optimale Anpassung der Lage des Freigabezeitfensters	33
3.1. Zeitreihe der erwarteten Ankünfte	41
3.2. Stückweise lineare Warteschlangendynamik	47
3.3. Warteschlangenlänge und Wartezeit	48
3.4. Berechnung der benötigten Freigabezeit \hat{g}_i	52
3.5. Zustandsabhängiger Verlauf von $\tau_i(t)$	56
3.6. Wirkungsbereich	57
3.7. Prognostizierte Gesamtwartezeit	59
3.8. Wartezeitenprognose bei Abbruch eines Bedienvorgangs	64
3.9. Funktionsbaustein zur Zustandsbeobachtung und -prognose	65
4.1. Optimale Bedienreihenfolge bei zwei Warteschlangen (Szenario 4.i) .	71
4.2. Maximale Zeitlücken bei Freigabezeitverlängerung (Szenario 4.iii) .	79
4.3. Drei Problemstellungen	84
4.4. Grenzyklus des Optimierungsverfahrens (Szenario 4.vi)	86
4.5. Instabilität des Optimierungsverfahrens (Szenario 4.vii)	88
5.1. Lineares Optimierungsproblem	98
5.2. Zeitabhängiger Schwellenwert	103
5.3. Verteilung der tatsächlichen Bedienperiode (Szenario 5.ii)	106
5.4. Grenzyklus des Stabilisierungsverfahren	107

5.5. Versatzzeitenanpassung (Szenario 5.iv)	108
6.1. Programmablaufgraph	112
6.2. Grenzyklus des kombinierten Verfahrens (Szenario 6.i)	114
6.3. Stabilisiertes Kumar-Seidman-Netzwerk (Szenario 6.ii)	116
7.1. Isolierter vierarmiger Knotenpunkt (Szenario 7.i)	121
7.2. Mittlere Warteschlangenlängen (Szenario 7.i)	122
7.3. Schaltmuster des kombinierten Verfahrens (Szenario 7.i)	125
7.4. Mittlerer Kraftstoffverbrauch je Fahrzeug (Szenario 7.i)	126
7.5. Box-Whisker-Plot (Szenario 7.ii)	128
7.6. Arterie mit irregulären Knotenpunktabständen (Szenario 7.iii)	130
7.7. Variation der Richtungsströme (Szenario 7.iii)	131
7.8. Freigabezeitbänder (Szenario 7.iii)	133
7.9. Variation der Knotenanzahl (Szenario 7.iv)	136
7.10. Komplexes Netzwerk (Szenario 7.v)	138
7.11. Tagesganglinie (Szenario 7.v)	138
7.12. Mittlere Warteschlangen (Szenario 7.v)	140
7.13. Kraftstoffverbrauch (Szenario 7.v)	142
8.1. Begrenzter Stauraum	146
8.2. Lokale Recheneinheit	151
A.1. Kumar-Seidman-Netzwerk	154
A.2. Instabiles Verhalten der Clear-Largest-Buffer-Policy	155
A.3. Knotenpunkt mit konstanten Ankunftsraten (Szenario 4.vi)	160

Kapitel 1.

Einleitung

1.1. Einordnung

1.1.1. Ausgangssituation

Funktionierende Städte setzen funktionierende Straßennetzwerke voraus. Dafür zu sorgen, dass der Verkehr auf den Straßen fließt und nicht still steht, zahlt sich in vielerlei Hinsicht aus: Die Verkehrsteilnehmer erreichen schneller ihr Ziel, der Wirtschaft gehen weniger Arbeitsstunden verloren, und die Belastung der Umwelt durch Schadstoffe und Lärm wird gesenkt.¹ Damit der Verkehr fließen kann, können beispielsweise Fahrbahnen und Knotenpunkte um zusätzliche Fahrspuren erweitert werden. Das ist aufwändig, kostenintensiv und beansprucht Raum, der den Fußgängern, Radfahrern und überhaupt der nutzbaren Stadtfläche verloren geht. Flüssigeren Verkehr kann man aber auch durch intelligentere Lichtsignalsteuerungen erreichen.

Ganz vermeiden lassen sich Rückstaus und Wartezeiten nicht, denn um bestimmten Verkehrsströmen mit Grün freie Fahrt zu geben, müssen andere mit Rot gestoppt werden. Das ist das Prinzip der Lichtsignalsteuerung. Längere Grünzeiten für die einen heißt automatisch längere Rotzeiten für die anderen. Die Steuerung der Lichtsignalanlagen muss darum abwägen, wann welche Verkehrsströme wie lange Grün bekommen. Damit die Grünzeiten weder zu früh noch zu spät enden, kommt es vor allem darauf an, die Umschaltzeitpunkte präzise an die tatsächlichen Fahrzeugankünfte anzupassen.

1.1.2. Entwicklungstrend

Der klassische Ansatz ist, dem Verkehrsablauf durch wiederkehrende Schaltfolgen ein festes periodisches Muster aufzuprägen. Die Muster werden dann so aufeinander abgestimmt, dass beispielsweise Grüne Wellen entstehen. In Dresden gibt es mehr als 30 davon [28]. Werden diese Schaltfolgen einmal berechnet und dann für lange Zeit auf den Knotenpunkten angewendet, spricht man von koordinierten Festzeitsteuerungen. Diese können allerdings nur dann wie gewünscht funktionieren, wenn die Verkehrsbedingungen nicht von denen abweichen, die der Planung zugrunde gelegt wurden [26].

Modernere verkehrsadaptive Verfahren passen die Schaltfolgen daher flexibel an das aktuelle Verkehrsaufkommen an. Hierzu wird das Verkehrsgeschehen mit Verkehrsdetektoren erfasst und an Verkehrszentralen übermittelt. Doch selbst Hochleistungsrechner sind mit der gewaltigen Datenflut überfordert. Auch mit vereinfachten Berechnungsverfahren können die Schaltprogramme für das gesamte Straßennetz nur für Zeiträume von mehreren Minuten reoptimiert werden. So wurde

¹Verkehrsstaus verursachen allein in den USA jährlich 3.7 Milliarden Stunden Wartezeit und 8.7 Milliarden Liter Kraftstoffverbrauch. Der Gesamtschaden wird auf 63.1 Milliarden US-Dollar geschätzt [169].

beispielsweise im Zuge des aktuellen Forschungsprojektes „Travolution“ gegenüber der Presse geäußert, dass die angestrebten fünf Minuten Berechnungszeit für 46 Lichtsignalanlagen eventuell nicht ausreichen werden [67]. Eine flexible Anpassung der Umschaltzeitpunkte an einzelne Fahrzeugankünfte erfordert allerdings eine sekundliche Aktualisierung. Es bleibt deshalb nichts anderes übrig, als die Detektordaten vor Ort zu verarbeiten: dort wo sie erhoben und gleichzeitig benötigt werden.

Aus diesem Grund lagern modernste verkehrsabhängige Verfahren einen Teil der Rechenleistung auf dezentrale Steuergeräte aus. Die Verkehrszentrale gibt in diesem Fall nur noch Rahmensignalpläne vor [181]. Darin sind für die Phasenwechsel nicht mehr Zeitpunkte sondern Zeitfenster vorgeschrieben. Die Steuergeräte vor Ort können dann entscheiden, wo innerhalb dieser Zeitfenster umgeschaltet werden soll. Der höhere Grad an Dezentralisierung erlaubt somit einen höheren Grad an Flexibilität.

Diesen Entwicklungstrend fortzuschreiben ist das Ziel der vorliegenden Dissertationsschrift. Der Verkehrsablauf soll nicht das Ergebnis vorher berechneter Programmparameter sein, im Gegenteil: Phasenwechsel und Umschaltzeitpunkte sollen sich unmittelbar aus dem aktuellen tatsächlichen Verkehrszustand heraus ableiten. Die Lichtsignalsteuerung ist vielmehr als Teil eines Regelkreises zu verstehen, wie ihn Abb. 1.1(b) zeigt. Das setzt allerdings die Verwendung dezentraler Online-Steuerung voraus. Jedoch wurden die besonderen Schwierigkeiten, die mit der Entwicklung und dem Einsatz dezentraler Online-Steuerungen verbunden sind, bislang noch nicht vollständig untersucht und verstanden.

1.1.3. Besondere Probleme dezentraler Steuerungsansätze

So stellt Papageorgiou, Herausgeber der Enzyklopädie „Concise Encyclopedia of traffic & transportation systems“ [140], in einem aktuelleren Artikel [141] fest:

„The properties of a completely decentralized operation (e.g., independent algorithm application at each intersection) are currently not fully analyzed or understood. This kind of thorough analysis may be useful [...] [for reducing] the possibly negative impact of decentralization.“

Die möglichen negativen Auswirkungen liegen darin begründet, dass jeder lokale Schalteingriff netzwerkweite Auswirkungen haben kann, und diese Auswirkungen auf lokaler Ebene unmöglich vorhersehbar sind.

Jede Masche im Straßennetz, das heißt jeder Straßenabschnitt mit zwei Richtungsfahrbahnen, wirkt als eine Rückkopplungsschleife. Da sich bei dezentralen Verfahren die Schaltprogramme aus den lokalen Fahrzeugankünften ableiten, und diese Fahrzeugankünfte wiederum das Resultat der umliegenden Schaltprogramme sind, wirkt ein lokaler Schalteingriff zeitversetzt auf den selben Knotenpunkt

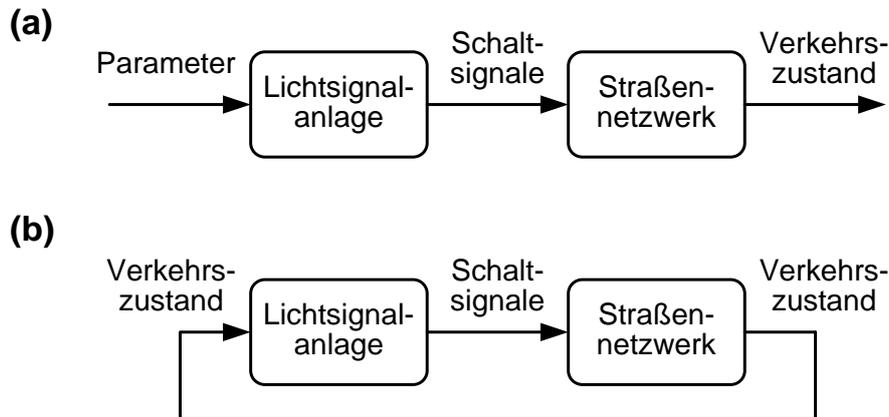


Abbildung 1.1.: (a) Bei den herkömmlichen Steuerungsverfahren handelt es sich auch dem Wortsinn nach um Steuerungen. Der Verkehrsablauf das direkte Ergebnis der vorher berechneten Programmparameter. (b) Das in dieser Dissertationsschrift vorgestellte Verfahren hat die Struktur eines Reglers. Aufgrund der Zustandsrückführung führen die Fahrzeuge ihre Umschaltzeitpunkte selbst herbei.

zurück. Diese Rückkopplungen können Instabilitäten² hervorrufen. Kumar und Seidman demonstrieren in Referenz [98], dass dies beispielsweise für die einfache Schaltregel „Längste Schlange zuerst“ der Fall ist. Mit dieser Schaltregel wachsen die Warteschlangen in einem Netzwerk aus nur zwei Knotenpunkten unaufhörlich an, obwohl sich beide Knotenpunkte weit unterhalb der Auslastungsgrenze befinden.

Wenn die Knotenpunkte nur noch über die Verkehrsströme selbst gekoppelt sind, kann eine Netzkoordinierung nicht mehr auf dem Grundprinzip der Grünen Welle beruhen. Zwar schließt das nicht aus, dass herannahende Fahrzeugkolonnen freie Fahrt erhalten, aber ob dies gelingt, hängt vom konkret verwendeten Verfahren ab. In einem Modellversuch in Münster ist das anscheinend nicht gelungen. Dort wurden die dezentralen Online-Steuerungen am Albersloher Weg wieder durch eine umlaufbasierte Netzsteuerung ersetzt [134].

Ein weiteres Problem dezentraler Verfahren ist, dass die Prognosehorizonte beschränkt sind. Mit umlaufbasierten Steuerungen, bei denen sich die Schaltprogramme nur geringfügig in der Zeit ändern, kann der Verkehrsablauf über mehrere Minuten prognostiziert werden. Das ist übrigens eine Grundannahme herkömmlicher Optimierungsverfahren. Werden die Knotenpunkte dagegen mit dezentralen verkehrabhängigen Verfahren betrieben, ist der Prognosehorizont kaum größer als die Reisezeiten zwischen den Knotenpunkten. Sogar mit zusätzlichen Kommunikationsmechanismen kann der Prognosehorizont nur schwerlich größer als das

²Vgl. Abschnitt 2.2.2

Doppelte dieser Reisezeit sein, denn ab diesem Moment beziehen sich die Prognosen wieder auf sich selbst.

Weichenmeier behauptet [67]: „*Verkehrsabhängige Knotenpunktsteuerungen stoßen bei hohem Verkehrsaufkommen an ihre Grenzen, da sie nur lokale Optimierungen vornehmen können*“. Er hat damit Recht, dass nur lokale Optimierungen vorgenommen werden können. Denn der Optimierungshorizont ist aufgrund der dezentralen Architektur zum einen räumlich, und aufgrund des beschränkten Prognosehorizonts zum anderen auch zeitlich begrenzt. Problematisch wird es, wenn sich bestimmte Schalteingriffe zu einem Zeitpunkt oder an einem Ort auszahlen, der außerhalb des Optimierungshorizonts liegt. Aber die Behauptung, dass dezentrale verkehrsabhängige Steuerungen deshalb bei großen Verkehrslasten versagen müssen, soll durch diese Dissertationsschrift widerlegt werden.

1.2. Beitrag dieser Dissertationsschrift

In der vorliegenden Dissertationsschrift werden Ansätze zur dezentralen Online-Steuerung vernetzter Verkehrsströme entwickelt. Die Ansätze werden systematisch hergeleitet, mathematisch begründet und am Ende zu einem in sich geschlossenen Verfahren zusammengefasst.

1.2.1. Ansatzpunkt

Der Ansatzpunkt ist, die einzelnen Knotenpunkte des Netzwerks lokal und unabhängig voneinander zu optimieren. Dies soll im Ergebnis zu einer flexiblen Anpassung der Umschaltzeitpunkte an die tatsächlichen Fahrzeugankünfte führen. Eine „Grüne Welle“ soll sich von selbst einstellen, und zwar genau dann, wenn dadurch lokal Wartezeiten eingespart werden. Damit dieser Ansatz überhaupt funktionieren kann und nicht etwa zu Instabilitäten führt, wird das lokale Optimierungsverfahren in ein lokales Stabilisierungsverfahren eingebettet.

Im Unterschied zu den herkömmlichen Ansätzen wird keine Kommunikation zwischen benachbarten Knotenpunkten oder zu einer Verkehrszentrale vorausgesetzt. Das heißt, die Knotenpunkte sind lediglich über die Verkehrsströme selbst gekoppelt. Es müssen weder die Bedienreihenfolge noch die Umschaltzeitpunkte vorher festgelegt werden, da sie sich unmittelbar aus den Ankunftszeitpunkten der Fahrzeuge ergeben. Ob ein Bedienvorgang fortgesetzt oder unterbrochen werden soll, ob dem einen oder dem anderen Verkehrsstrom Vorrang zu geben ist, das wird zu jedem Zeitpunkt aufs Neue entschieden.

1.2.2. Lösungsschritte

Die Lösung wird in vier Schritten vom Kleinen zum Großen entwickelt.

Im ersten Schritt wird eine einzelne Knotenzufahrt betrachtet und für diese ein Prognoseverfahren zur Abschätzung der zur Warteschlangenleerung erforderlichen Freigabezeit entwickelt. Der Prognosehorizont braucht hierfür nicht größer als die Dauer eines Bedienvorgangs zu sein, also nicht größer als etwa 30 Sekunden.

Der zweite Schritt untersucht, wie die Zufahrten eines Knotenpunktes optimal zu bedienen sind. Die lokale Minimierung der Gesamtwartezeiten wird hierbei auf eine dynamische Priorisierungsregel zurückgeführt. Die Lösungsvorschrift lautet lediglich: „Bediene die Zufahrt mit höchster Priorität.“

Im dritten Schritt wird die hinreichende Bedingung für netzwerkweite Stabilität auf ein lokales Kriterium für die Effizienz der Bedienvorgänge reduziert. Dieser Ansatz garantiert jedem Verkehrsstrom die minimal benötigte Abfertigungskapazität. Damit soll sichergestellt werden, dass die Fahrzeuge nirgends im Netzwerk systematisch aufgehalten werden und die Warteschlangen auch im Falle eventueller Rückkopplungseffekte nicht beliebig anwachsen können.

Im vierten und letzten Schritt werden Optimierungs- und Stabilisierungsverfahren miteinander kombiniert. Das Stabilisierungsverfahren braucht nicht einzugreifen, wenn die lokale Optimierung von selbst zum gewünschten Schaltverhalten führt, d.h. wenn jeder Verkehrsstrom mindestens einmal innerhalb einer gegebenen Zeitspanne ausreichend lange bedient wird.

1.2.3. Annahmen

Aus regelungstechnischer Sicht stellt das Verkehrsgeschehen an den Zufahrten eines Knotenpunktes die „Regelstrecke“ und die Lichtsignalanlage selbst den „Regler“ dar. Die Schnittstelle zwischen Regelstrecke und Regler wird einerseits durch die Verkehrsdetektoren und andererseits durch die Signalgeber realisiert. Eine Skizze des prinzipiellen Aufbaus zeigt Abb. 1.2. Dem analytischen Reglerentwurf werden die folgenden vereinfachenden Annahmen zugrunde gelegt:

- A1: Die Fahrzeuge einer Wartekolonne fahren während der Freigabe entsprechend dem Abflussgesetz [166] mit einer bekannten Rate aus der Knotenzufahrt heraus. Der Zufluss in nachfolgende Streckenabschnitte sei ungehindert möglich.
- A2: Die Prognose der erwarteten Fahrzeugankünfte kann fehlerbehaftet sein. Als sicher bekannt wird dagegen vorausgesetzt, ob sich auf einer Knotenzufahrt eine Warteschlange befindet oder nicht.
- A3: In jeder Phase wird nur genau ein Verkehrsstrom bedient. Das ist zwar gleichbedeutend mit der Annahme, dass alle Verkehrsströme zueinander unverträglich sind und daher nicht gleichzeitig bedient werden dürfen. Aber diese Annahme ist notwendig, um bei der mathematischen Herleitung des Verfahrens

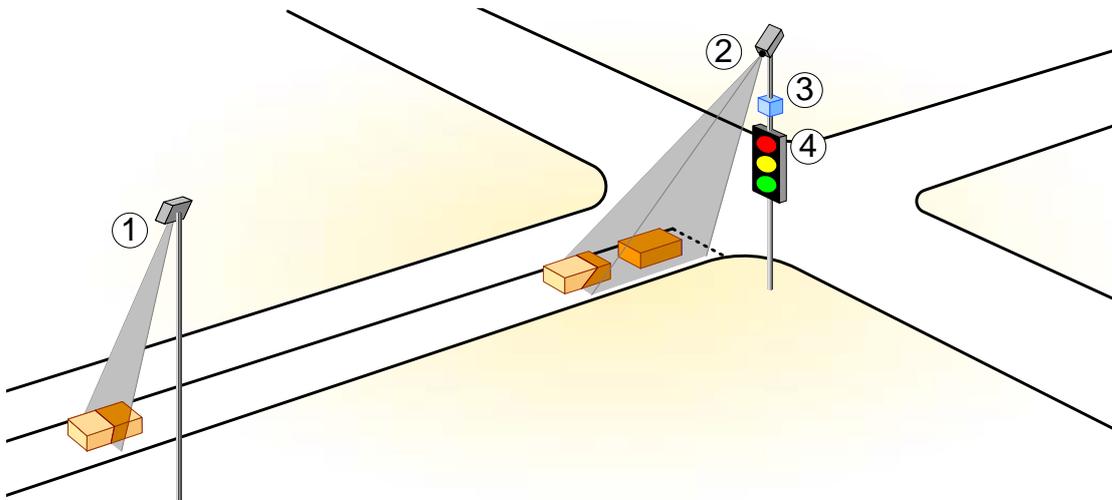


Abbildung 1.2.: Verkehrsdetektoren erfassen das aktuelle Verkehrsgeschehen auf den Zufahrten eines Knotenpunktes. Es genügt, einen Detektor ① hinreichend weit stromaufwärts der Haltelinien zu platzieren, und einen weiteren ② unmittelbar vor den Haltelinien. Aus den Detektordaten leitet die lokale Recheneinheit ③ Schaltsignale für die Signalgeber ④ ab. Eine Kommunikation mit benachbarten Knotenpunkten oder einer Verkehrszentrale ist nicht erforderlich.

nicht explizit die kombinatorische Vielfalt möglicher Phasenzusammenstellungen berücksichtigen zu müssen. Entsprechende Verallgemeinerungen werden im Ausblick 8.1 diskutiert.

A4: Das Umschalten ist mit einer Zwischenzeit verbunden, während der alle Knotenzufahrten gesperrt sind. Wie groß die Zwischenzeit jeweils ist, soll nur von der zu bedienenden Knotenzufahrt abhängen.

Auf diese Annahmen und ihre Implikationen wird in den Abschnitten 2.3.1, 2.3.2 und 3.1 genauer Bezug genommen.

1.2.4. Wesentliche Ergebnisse

Die flexible Anpassung der Umschaltzeitpunkte an die tatsächlichen Fahrzeugankünfte weist gegenüber herkömmlichen Verfahren deutliche Vorteile auf.

Zum Ersten erlaubt das vorgestellte Prognoseverfahren, Wartezeiten und Haltevorgänge zu vermeiden anstatt sie lediglich zu minimieren. Um einer herannahenden Kolonne freie Fahrt zu gewähren, beginnt der Bedienvorgang gerade rechtzeitig, um vorher den Umschaltvorgang abzuschließen und die bereits auf der Zufahrt wartenden bzw. noch vor der Kolonne eintreffenden Fahrzeuge abzufertigen.

Zum Zweiten werden die Grünzeiten bedarfsgerecht zugewiesen. Schwankungen in der Verkehrsstärke werden somit unmittelbar kompensiert und führen nicht zwangsläufig zu ebenso großen Schwankungen in den Wartezeiten. Der Fall, dass auf einer Zufahrt die Warteschlange nicht geleert werden kann und gleichzeitig eine andere Zufahrt unnötig lange Grün bekommt, ist damit von vornherein ausgeschlossen.

Zum Dritten haben die Knotenpunkte des Netzwerks mit der dezentralen Steuerung exakt die gleiche Durchlassfähigkeit wie mit einem umlaufbasierten Verfahren.

Zum Vierten werden die nicht unbedingt benötigten Abfertigungskapazitäten für flexible Umschaltvorgänge oder Freigabezeitverlängerungen genutzt. Das heißt, immer wenn ein Knotenpunkt nicht maximal ausgelastet ist, werden die Zufahrten häufiger oder länger bedient. Dies führt in der Folge zu deutlich geringeren Wartezeiten.

1.3. Kapitelübersicht und verwendete Methoden

Der Aufbau dieser Dissertationsschrift folgt den vier Lösungsschritten aus Abschnitt 1.2.2. Diese werden nacheinander in den Kapiteln 3 bis 6 behandelt. Dem vorangestellt ist ein Überblick über den aktuellen Stand der Forschung in Kapitel 2. Den Abschluss bilden die Simulationsergebnisse in Kapitel 7 und der Ausblick in Kapitel 8.

Kapitel 2 stellt den gegenwärtigen Stand der Forschung vor und geht auf aktuelle Forschungstrends sowie Lösungsansätze aus verwandten Problemfeldern ein. Die wesentlichen Randbedingungen für die Steuerung von Lichtsignalanlagen werden am Beispiel einer umlaufbasierten Steuerung erläutert. In diesem Zusammenhang werden auch die Möglichkeiten zur Netzkoordinierung diskutiert.

Kapitel 3 bereitet den analytischen Zugang für die nachfolgenden Schritte vor. Dies geschieht, indem eine Knotenzufahrt als hybrides dynamisches System modelliert wird. Ein hybrides dynamisches System zeichnet sich dadurch aus, dass seine Zustandsgrößen sowohl kontinuierlich als auch diskret sind. Die Potentiale dieses Modellierungsansatzes werden von Savkin und Evans in ihrer Monographie „Hybrid Dynamical Systems“ [162] wie folgt beschrieben:

„The study of hybrid dynamical systems represents a difficult and exciting challenge to control engineers and is referred to as “the control theory of tomorrow”³ because of its future potential for solving problems.“

Dieser Modellierungsansatz erfordert, zunächst die diskreten Zustände und möglichen Zustandsübergänge zu identifizieren und anschließend das kontinuierliche Systemverhalten jedes Zustandes durch eine Entwicklungsgleichung zu spezifizieren. Daraus wird dann ein Prognoseverfahren abgeleitet. Damit lässt sich zum einen prognostizieren, welche Freigabezeit zur Leerung einer Warteschlange erforderlich ist, und zum anderen, um welchen Betrag die Gesamtwartezeiten bis zur Warteschlangenleerung anwachsen, falls die Bedienung nicht sofort, sondern erst kurze Zeit später beginnt.

In **Kapitel 4** wird ein heuristisches Optimierungsverfahren zur lokalen Minimierung der Gesamtwartezeiten entwickelt. Die Lösung des nichtlinearen Optimierungsproblems wird auf die Spezifikation eines dynamischen Prioritätsindex zurückgeführt. Wie dieser zu berechnen ist, leitet sich aus der Analyse folgender drei Problemstellungen ab: (i) In welcher Reihenfolge sind Warteschlangen ohne Zuflüsse zu bedienen? (ii) Wann ist ein Bedienvorgang abzuberechnen um einen anderen vorzuziehen? (iii) Welche Zeitlücken im Verkehrsstrom können zur Verlängerung der Freigabezeit maximal akzeptiert werden? Durch Verallgemeinerung der jeweiligen Optimalitätskriterien erhält man als Ergebnis eine Formel für den Prioritätsindex. Mit diesem löst die Schaltregel „Bediene die Zufahrt mit größter Priorität“ jede der drei Problemstellungen optimal.

Kapitel 5 führt den Begriff der Bedieneffizienz ein. Darunter wird der Teil der Bedienzeit einschließlich des Umschaltens verstanden, der für die Abfertigung der Fahrzeuge aufgebracht werden kann. Aus einem hinreichenden Kriterium für netzwerkweite Stabilität wird folgende lokale Schaltregel abgeleitet: „Bediene eine Zufahrt erst dann, wenn die Effizienz des Bedienvorgangs größer als ein kritischer Schwellenwert ist.“ Die Schwellenwerte lassen sich so wählen, dass jede Zufahrt im Mittel einmal innerhalb einer vorgegebenen Wunschperiode bedient wird. Eine Maximalperiode wird eingehalten, indem man einen Teil der tatsächlich detektierten Fahrzeuge durch virtuelle Fahrzeuge ersetzt und gleichzeitig fordert, dass auch die virtuellen allein den Schwellenwert nach einer bestimmten Zeit erreichen. Die Wunschperiode und die Maximalperiode sind die beiden einzigen Parameter des gesamten Steuerungsverfahrens.

In **Kapitel 6** werden die beiden grundsätzlich verschiedenen Ansätze aus den Kapiteln 4 und 5 miteinander kombiniert. Dabei wird das Optimierungsverfahren in das Stabilisierungsverfahren eingebettet. Insbesondere sollen damit zwei Ziele erreicht werden: Einerseits soll jeder Verkehrsstrom mindestens die minimal benö-

³aus Referenz [72]

tigte Abfertigungskapazität erhalten und andererseits soll die verbleibende Kapazität zur lokalen Optimierung genutzt werden, d.h. für häufigere Umschaltvorgänge und für Verlängerungen der Freigabezeiten. Mit diesem Ansatz können die oben erwähnten Rückkopplungen im Netzwerk toleriert werden. Denn das übergeordnete Stabilisierungsverfahren greift korrigierend ein noch bevor die Rückkopplungen zu instabilem Verhalten führen können.

Kapitel 7 unterzieht das dezentrale Steuerungsverfahren in Simulationsstudien einem Vergleich mit einer Festzeitsteuerung. Zunächst wird ein einzelner Knotenpunkt unter regelmäßigen sowie irregulären Verkehrslasten betrachtet. Anschließend werden die Verfahren auf einer Arterie mit unregelmäßigen Knotenpunktabständen untersucht. Hierbei wird die Festzeitsteuerung im Sinne einer Grüne Welle koordiniert, während die dezentralen Online-Steuerungen lediglich lose über die Verkehrsströme selbst gekoppelt sind. Die Betrachtung der Freigabezeitbänder veranschaulicht die wesentlichen Unterschiede der beiden Verfahren. Abschließend wird ein komplexes Netzwerk über einen gesamten Tag simuliert, wobei die Verkehrsnachfrage einer empirischen Tagesganglinie folgt. Darin wird deutlich, dass der Ansatz, auf die Netzkoordinierung zu verzichten und stattdessen die Umschaltzeitpunkte flexibel an das tatsächliche Verkehrsgeschehen anzupassen, zum Erfolg führt: Es können sowohl die Wartezeiten als auch der Kraftstoffverbrauch erheblich gesenkt werden.

Kapitel 2.

Stand der Forschung

Kurzfassung

Lichtsignalanlagen regeln seit beinahe einhundert Jahren das Verkehrsgeschehen an Knotenpunkten von Straßennetzwerken. Der erste Abschnitt dieses Kapitels fasst die Epochen ihrer Entwicklung chronologisch zusammen und gibt einen Überblick über den aktuellen Forschungsstand.

Nahezu alle derzeit eingesetzten Verfahren sind umlaufbasiert. Am Beispiel einer umlaufbasierten Steuerung werden die wichtigsten Randbedingungen erläutert. Anschließend werden sie mit den relevanten Steuerungsparametern, beispielsweise der Freigabe- und Umlaufzeit, quantitativ in Beziehung gesetzt.

Durch Koordinierung der Knotenpunkte im Netzwerk können die Gesamtwarthezeiten aller Fahrzeuge verringert werden, im Idealfall mit einer Grünen Welle. Es wird ein einfaches Modell zur Abschätzung der Gesamtwarthezeiten in Abhängigkeit von der Lage der Freigabezeitfenster vorgestellt. Dieses erlaubt zum einen, die grundlegenden Zusammenhänge herauszustellen, und zum anderen, die prinzipiellen Schwierigkeiten deutlich zu machen.

Der aktuelle Stand der Forschung wird abschließend kritisch diskutiert.

2.1. Entwicklung der Lichtsignalsteuerung

2.1.1. Erste Lichtsignalanlagen

England

Die erste Lichtsignalanlage der Welt ging am 10. Dezember 1868 in Betrieb [130]. In London, an der Kreuzung George und Bridge Street, befestigte der Erfinder Knight an einem Mast zwei Eisenbahn-Signalflügel. Verkehrspolizisten konnten diese über eine mechanische Steuerung entweder waagrecht (Halt!) oder schräg abwärts (Vorsicht!) stellen, wie in Abb. 2.1(a) zu sehen. Das Signal galt nur den Pferdefuhrwerken und Fahrzeugen, es sollte vor Allem den Fußgängern ein sicheres Überqueren der Straßen ermöglichen. Trotz der mechanischen Signalflügel kann diese Konstruktion bereits Lichtsignalanlage genannt werden, da sich zusätzlich auf der Mastspitze eine Gaslaterne mit Farbfilter befand. Die Gaslaterne könnte übrigens auch der Grund dafür gewesen sein, warum die Anlage 1872 nach nur vier Betriebsjahren wieder demontiert wurde. Denn durch ihre Explosion wurde ein diensthabender Verkehrspolizist schwer verletzt. In London wurde die nächste Lichtsignalanlage erst im Jahre 1926 am Piccadilly Circus errichtet.

USA

In seiner Ansprache zum National Defense Transportation Day am 10. Mai 1997 kam der damalige US-Präsident Bill Clinton [38] auch auf die Erfindung der Lichtsignalanlage zu sprechen: „*An African American, Garrett Morgan invented the traffic signal and is recognized as the father of our safe transportation technology . . .*“. Ebenso vermerkt Brown in einem Buch über große amerikanische Erfinder [27]: „*In 1922, after Morgan witnessed a horrible traffic accident, he turned his inventive mind to automobile safety. A year later, he was awarded the first patent for a traffic signal.*“ Tatsächlich ließ Morgan im Jahre 1923 eine Lichtsignalanlage patentieren [128], aber technisch übertraf sie die englische Erfindung aus dem Jahre 1868 nur wenig: Um die Signalflügel zu bewegen, musste, wie im Patent in Abb. 2.1(b) zu sehen, per Hand der Kurbelarm Nr. 28 gedreht werden. Außerdem wurden bereits fast zwei Jahre vor Morgans Erfindung die Kreuzungen von Manhattan zentral und elektrisch gesteuert. Dazu war am 5. Januar 1922 in der New York Times [2] zu lesen: „*[T]raffic in most of Manhattan will be simultaneously stopped and started by red, green and yellow lights all operated by a single switch in Times Square.*“ Morgans Erfindung war also nicht die erste ihrer Art. So befindet sich beispielsweise im Henry Ford Museum in Dearborn (Michigan) [15] eine Erfindung von Inspektor Potts aus dem Jahre 1920. Sie ist unternitelt mit: „*The world’s first three color, four direction traffic signal.*“ Anders als ihr Preis, sie kostete damals 37\$, hat sich das Design bis heute kaum verändert: ein frei hängender schwarzer Kasten mit den drei übereinander angeordneten farbigen

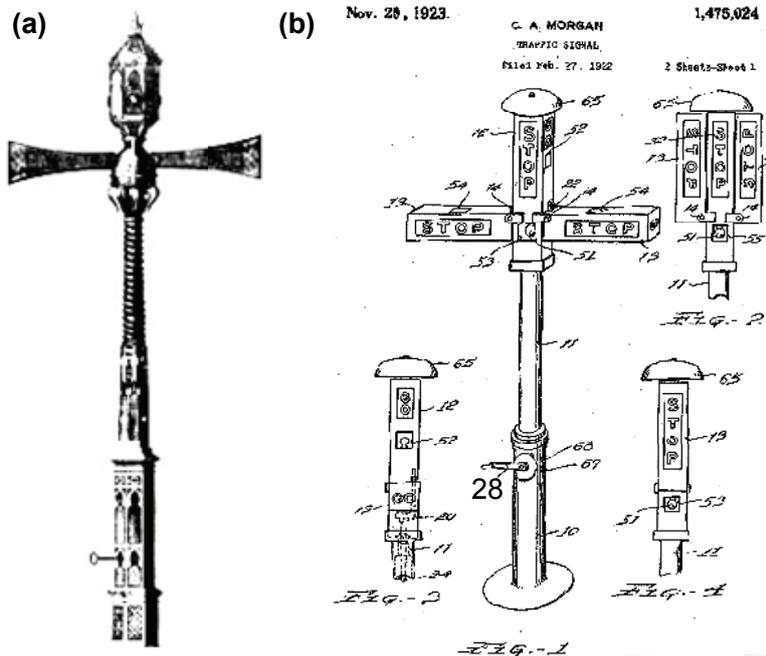


Abbildung 2.1.: (a) Die erste Lichtsignalanlage der Welt wurde 1868 in England installiert. (b) Morgans US-Patent [128] aus dem Jahre 1923 übertrifft die englische Konstruktion nur wenig.

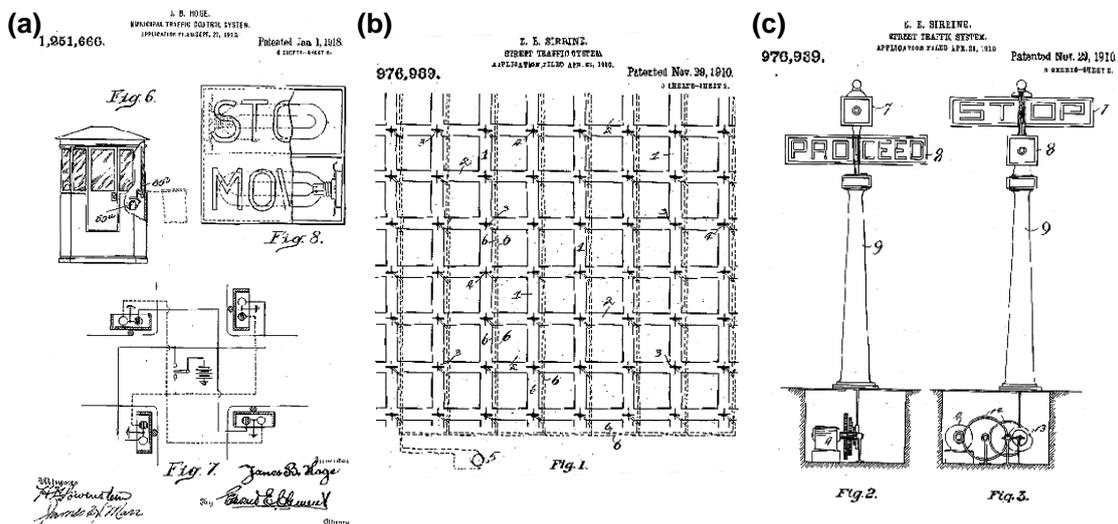


Abbildung 2.2.: (a) US-Patent des Erfinders Hoge [84] aus dem Jahre 1913. (b) und (c) Das vermutlich erste US-Patent zur Lichtsignalsteuerung stammt vom Erfinder Serrine [179] aus dem Jahre 1910.

Leuchtfeldern und Blendschuten. Aber das Design war wohl das einzig Neue an Potts' Erfindung, denn bereits im Jahre 1914 hatte die American Traffic Signal Company eine Lichtsignalanlage in Cleveland (Kreuzung 105th Street und Euclid Avenue) in Betrieb genommen. Hierzu schrieb die Monatszeitschrift „Motorist“ [1] in der August-Ausgabe von 1914: *„This system is, perhaps, destined to revolutionize the handling of traffic in congested city streets and should be seriously considered by traffic committees for general adoption.“* Dem lag eine Erfindung von Hoge [84] aus dem Jahre 1913 zugrunde. Er entwarf eine wettergeschützte Kabine, wie sie Abb. 2.2(a) zeigt, in der ein Verkehrspolizist per Knopfdruck zwischen zwei Phasen umschalten konnte. Da die Erfindung keine Übergangszeiten (z.B. Rot-Gelb) vorsah, wurden Phasenwechsel mit einem Klingelzeichen eingeleitet. Zwar wird auch Hoges Erfindung oft für die erste Lichtsignalanlage gehalten [166], aber in einem Übersichtsartikel [115] wird eine noch frühere erwähnt: die Signalanlage von Wire. Sie ging 1912 in Salt Lake City an der Kreuzung 2nd South Street und Main Street in Betrieb und hatte zwei farbige Leuchtfelder für jede der vier Richtungen. Und eine Patentrecherche ergibt: 13 Jahre bzw. über 60 US-Lichtsignal-Patente vor Morgans Signalflügel wurde im Jahre 1910 das erste automatische „Street Traffic System“ [179] zur Steuerung eines Straßenverkehrsnetzes patentiert. Ob es in Betrieb ging, ist nicht bekannt. Abbildungen 2.2(b) und (c) zeigen die ersten beiden Seiten des Patents.

Deutschland

In Deutschland ging die erste Lichtsignalanlage 1922 am Stephansplatz in Hamburg in Betrieb, siehe Abb. 2.3(a). Fälschlicherweise trägt jedoch ein Berliner Denkmal die Aufschrift *„Verkehrsturm vom Potsdamer Platz aus dem Jahre 1924 – Nachbildung der ersten Lichtsignalanlage Deutschlands.“* Von diesem Verkehrsturm aus, der ursprünglich aus New York stammt [92], konnte ein Verkehrspolizist in drei Metern Höhe den gesamten Platz überschauen und steuern, siehe Abb. 2.3(b). Vor der Installation des Turmes waren dafür elf Verkehrspolizisten nötig.

Die Technik entwickelte sich stetig weiter: Mit einer Schalttafel und zwei Schaltwalzen ausgestattet, ging 1926 im Polizeipräsidium am Alexanderplatz die erste Verkehrszentrale Deutschlands in Betrieb. Die erste adaptive Steuerung folgte im Jahre 1934 mit in die Straße eingelassenen Kontaktschwellen.

2.1.2. Steuerungsverfahren der ersten Generation

Gegenüber dem technischen Fortschritt blieb das theoretische Verständnis zunächst noch zurück. 1961 leitete Miller einen Aufsatz [124] mit den Worten ein: *„In Britain today, the majority of traffic signals are of the vehicle-actuated kind. That is, there are pneumatic pads on each approach which transmit a signal to*

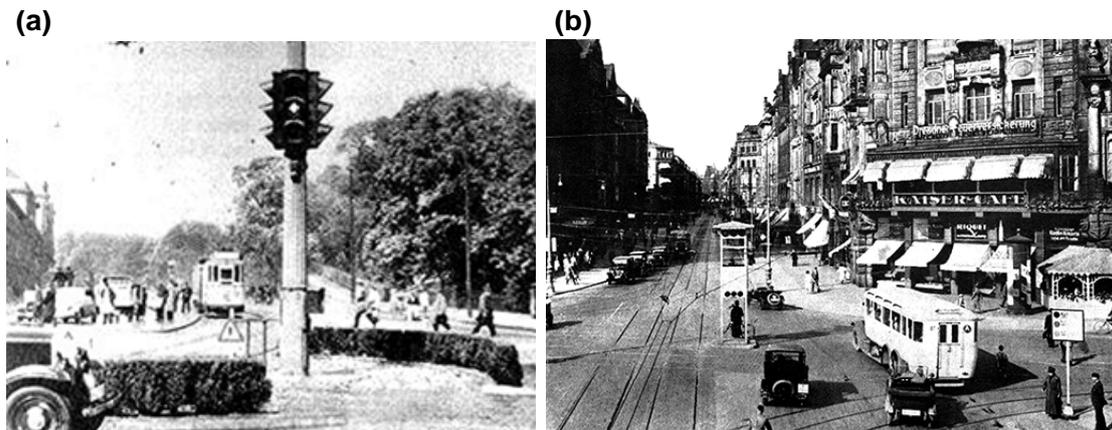


Abbildung 2.3.: (a) Die erste Lichtsignalanlage Deutschlands am Stephansplatz in Hamburg aus dem Jahre 1922. (b) In Dresden wurde 1927 am Wiener Platz (Blick in die Prager Straße) ein Verkehrsturm installiert [189], wie er noch heute als Denkmal am Potsdamer Platz in Berlin zu besichtigen ist.

the controller when each new vehicle arrives. [...] There is, however, still a need to know how to set fixed-cycle traffic signals as, when there is heavy traffic, or when the detectors are not functioning correctly, vehicle-actuated signals operate as fixed-time signals.“ Obwohl in den 50er und 60er Jahren bereits die ersten elektrischen Steuergeräte mit Verkehrsdetektoren ausgestattet waren, war noch immer unklar, wie die Wartezeiten der Fahrzeuge von den Parametern einer Festzeitsteuerung abhängen. Pionierarbeit in der Beantwortung dieser Frage leisteten Webster, Miller und Little, der Begründer des Satzes von Little. Ihre Arbeiten legten die Grundsteine für alle später entwickelten Verfahren. Einige daraus werden im Folgenden vorgestellt.

Grafische Lösungsverfahren

Das Grundprinzip der Grünen Welle , d.h. „...dass bei der Einfahrt eines Wagens in eine Hauptstraße dieser Wagen die Hauptstraße ohne Fahrunterbrechung durchfährt ...“, wurde bereits 1925 von Dr.-Ing. Johannes Adolph zum Patent angemeldet [8]. Eine umfassende systemtheoretische Untersuchung des Problems der Linienkoordinierung wurde jedoch erst viel später, im Jahre 1964, von Morgan und Little [129] durchgeführt. Sie entwickelten eine grafische Berechnungsvorschrift für die Schaltprogramme an den Kreuzungen entlang einer Arterie. Den Verkehrsströmen soll dabei ein maximales Freigabezeitfenster bzw. eine maximale Bandbreite zur Verfügung stehen, siehe Abb. 2.4(a). Die Berechnungsvorschrift trägt deshalb den Namen MAXBAND. Kurz darauf formulierte Little das zugehörige Optimierungsproblem als ein gemischt-ganzzahliges lineares Problem [112].

Aber erst 15 Jahre später, nach dem Einzug der digitalen Rechentechnik, konnte er einen FORTRAN-Code präsentieren, der das Optimierungsproblem mit dem Branch-and-Bound-Verfahren löst [111].

Offline-Optimierung

Mit Etablierung der digitalen Rechentechnik konnten überhaupt erstmals komplexere Verkehrsnetze offline, d.h. auf historischen Messdaten beruhend, behandelt werden [182]. Eins der ersten Verfahren – und wohl selbst heute noch eins der am weitesten verbreiteten [57] – ist TRANSYT (**T**raffic **N**etwork **S**tud**Y** **T**ool). Es wurde 1969 von Robertson [155] entwickelt und wird seitdem kontinuierlich ausgebaut [119, 37]. Das Programm TRANSYT sucht nach dem optimalen Schaltprogramm für Festzeitsteuerungen, indem es das Verkehrsgeschehen modellbasiert simuliert und mit einer Gütefunktion aus Gesamtwarzeiten und Anzahl der Stopps bewertet [156]. Das Schaltprogramm des besten Simulationslaufs wird übernommen. Weil sich das tatsächliche Verkehrsaufkommen im Gegensatz zu dem, das der Offline-Optimierung zugrunde gelegt wurde, im Laufe von Monaten und Jahren ändert, nehmen die mittleren Wartezeiten pro Jahr um zirka 3-4% zu [14]. Als gegen Ende der 70er Jahre die ersten Computer die Schaltprogramme während des Betriebs inkrementell nachjustieren konnten, weil sie mit aktuellen Messdaten gespeist wurden, war die Methode der verkehrsadaptiven Verfahren geboren.

2.1.3. Verkehrsadaptive Steuerungsverfahren

Inkrementelle Parametervariation

SCOOT (**S**plit, **C**ycle, and **O**ffset **O**ptimisation **T**echnique) ist die verkehrsadaptive Variante von TRANSYT. Das Verfahren geht zum Teil auf den gleichen Entwickler zurück, beruht auf dem gleichen Verkehrsmodell und maximiert die gleiche Gütefunktion [88, 87, 156]. Die wichtigste Erweiterung ist jedoch der zentrale SCOOT-Rechner, der sekundlich die Messdaten aller Sensoren im SCOOT-Netzwerk erhält [119]. Abbildung 2.4(b) zeigt den prinzipiellen Aufbau. Auf dem Zentralrechner passt ein Optimierungsmodul die Freigabezeiten langsam und inkrementell der aktuellen Verkehrsnachfrage an. Die Philosophie von SCOOT ist es, nur kleine inkrementelle Änderungen vorzunehmen [156]: „*SCOOT uses an ,elastic‘ coordination plan that can be stretched or shrunk to match the latest situation.*“ Die Freigabezeiten können zwar kurzzeitig um bis zu ± 4 s von der Bemessungsfreigabezeit abweichen, aber die Bemessungsfreigabezeit selbst wird nur um höchstens ± 1 s je Umlauf variiert. Die Umlaufzeit ist konstant und für alle Knotenpunkte des SCOOT-Netzwerks gleich. Sie wird anhand der am meis-

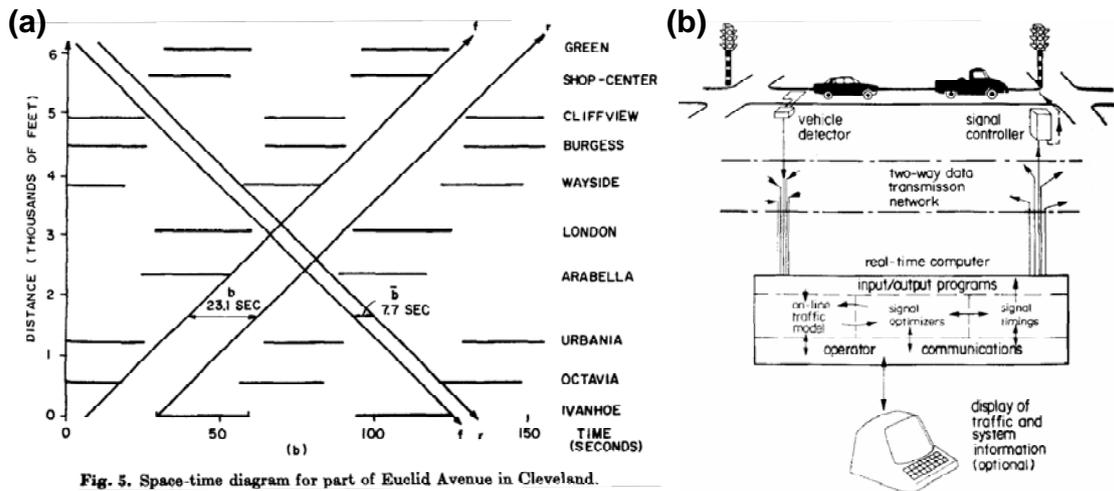


Fig. 5. Space-time diagram for part of Euclid Avenue in Cleveland.

Abbildung 2.4.: (a) Das grafische Optimierungsverfahren MAXBAND [129] maximiert Freigabezeitbänder im Raum-Zeit-Diagramm. Heute ist das Verfahren unter „Arterial Progression“ [60] oder Bandbreitenmaximierung [166] bekannt. (b) Die Systemarchitektur von SCOOT [88, 119]. Verschiedene Weiterentwicklungen dieses Verfahrens sind heute im Einsatz.

ten ausgelasteten Kreuzung bemessen: ihr Sättigungsgrad⁴ darf höchstens 90% betragen. Aufgrund der gemeinsamen Umlaufzeit kann SCOOT mit zyklischen Flussprofilen arbeiten: Das bildet die Grundlage der Koordinierung. Später wurde das Verfahren um das „Stage-Skipping“-Konzept zum Phasentausch für die ÖPNV-Priorisierung erweitert [24].

Gleitende Prognose

Zeitgleich mit SCOOT wurden andere Verfahren entwickelt, die im Unterschied zu diesem keine gemeinsame Umlaufzeit an allen Knotenpunkten voraussetzen. Weil die Verfahren folglich auch nicht auf immer wiederkehrende Verkehrszustände zurückgreifen können, müssen sie das Verkehrsgeschehen prognostizieren. Zu den am meisten verbreiteten prognosebasierten Verfahren gehören beispielsweise OPAC (**O**ptimization **P**olicies for **A**daptive **C**ontrol) [59, 35], PRO-DYN (**D**ynamic **P**rogramming) [80, 81] und CRONOS (**C**ontrol of **N**etworks by **O**ptimization of **S**witchovers) [21, 22]. Sie optimieren die Schaltfolge über einen gewissen Prognosehorizont von üblicherweise 60 Sekunden. Weil die gefundene Lösung fortwährend, beispielsweise alle 4 Sekunden, reoptimiert wird, nennt sich das Vorgehen auch „Rolling Horizon Dynamic Programming“. Abbildung 2.5(a) zeigt

⁴Der Sättigungsgrad ist das Verhältnis zwischen tatsächlicher und maximaler Auslastung (vgl. Abschnitt 2.3.2).

das Funktionsprinzip. Der Unterschied zwischen den Verfahren besteht darin, wie das Optimierungsproblem gelöst wird. OPAC diskretisiert die Zeit und enumeriert den Lösungsraum vollständig. PRODYN hingegen verwendet ein heuristisches Lösungsverfahren, den so genannten Forward-Dynamic-Programming-Algorithmus. Beide Ansätze weisen eine exponentielle Komplexität auf und sind aus diesem Grund in der Praxis nur für isolierte Knotenpunkte anwendbar [141]. Dagegen verwendet CRONOS zur Abschätzung der Freigabezeit und des Versatzes den Algorithmus von Box mit polynomialer Komplexität zweiter Ordnung. Dennoch kann das Verfahren in Echtzeit nur auf eine kleine Anzahl von Knotenpunkten angewendet werden.

Ein Steuerungsverfahren der gleichen Kategorie wird derzeit im Forschungsprojekt „Travolution“ (**Traffic Evolution**) entwickelt [67, 183, 68]. Das Projekt wurde im April 2006 von mehreren Unternehmen, u.a. Audi AG, und der TU München ins Leben gerufen und verfolgt den innovativen Ansatz, die Fahrzeuge mit den Lichtsignalanlagen in eine aktive Kommunikationsbeziehung über WLAN treten zu lassen. Die Steuerung soll für die Fahrzeuge „mitdenken“. Der Denkprozess besteht dabei in der Anwendung eines Genetischen Algorithmus [12]. Auf einem Zentralrechner werden dafür zufallsbasiert die Schaltfolgen sämtlicher Lichtsignalanlagen miteinander kombiniert und über einen Prognosehorizont von fünf Minuten simuliert. Die beste Schaltfolge wird ausgewählt und an die Lichtsignalanlagen gesendet. Das Verfahren wird derzeit in Ingolstadt an 46 Knotenpunkten erprobt. Weil aufgrund der großen Datenmengen fünf Minuten Berechnungszeit eventuell nicht ausreichen [67], wird bereits eine Untergliederung in mehrere kleine Teilnetze erwogen.

Hierarchisch-dezentrale Dekomposition

Andere Steuerungsverfahren bewältigen den hohen Komplexitätsgrad, indem sie das Gesamtproblem hierarchisch in isolierte Teilprobleme zerlegen. Meist unterscheiden sie zwischen strategischer, taktischer und operationaler Ebene. Auf jeder dieser Ebenen werden verschiedene Zeitskalen betrachtet, beispielsweise Stunden, Minuten und Sekunden [182, 57]. Auf den unteren Hierarchieebenen wird die Rechenlast dezentral auf die lokalen Steuergeräte an den Knotenpunkten verteilt. Abbildung 2.5(b) zeigt eine typische Systemarchitektur. Weil die dezentralen Steuerungen nicht von selbst in der Lage sind, sich zu koordinieren [117], wird ihr Aktionsspielraum über Referenzpläne und Rahmenvorgaben der übergeordneten Ebenen eingeschränkt. Die oberste Ebene befindet sich immer auf einem Zentralrechner [57].

Eins der ersten Verfahren dieser Art ist UTOPIA (**Urban Traffic Optimization by Integrated Automation**) [53, 118, 117]. Es besteht aus zwei Hierarchie-Ebenen namens „Area Level“ und „Intersection Level“. Die Schaltfolge wird mittels semi-heuristischer Optimierung berechnet. Ein anderes Verfahren, RHODES (**Real-**

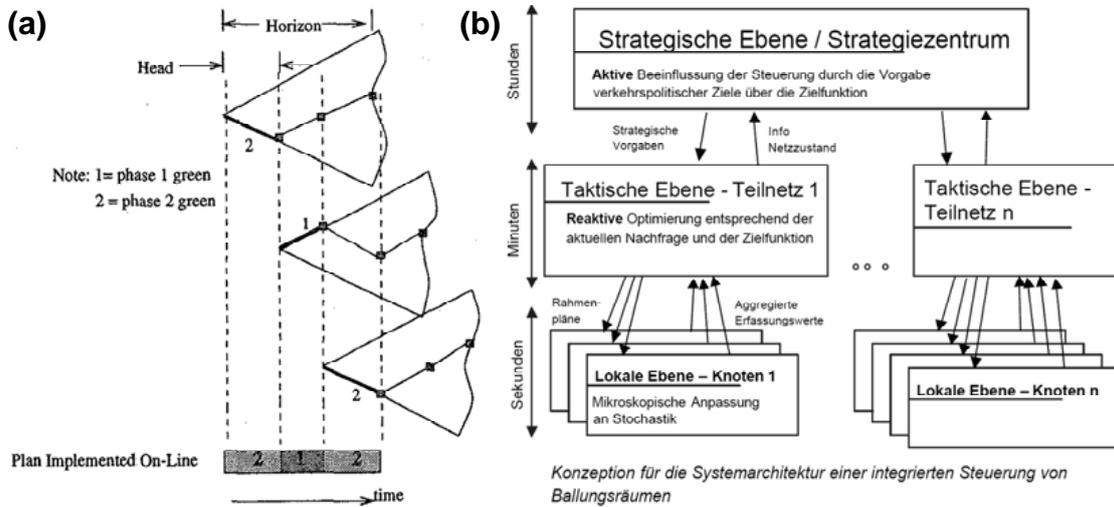


Abbildung 2.5.: (a) Prinzip des gleitenden Prognosehorizonts (aus [146]). In jedem Zeitschritt werden die zukünftigen Umschaltzeitpunkte neu optimiert. (b) Mit einer hierarchischen Systemarchitektur wird ein Teil der Funktionalität auf dezentrale Recheneinheiten ausgelagert (aus [58]).

Time **H**ierarchical **O**ptimized **D**istributed **E**ffective **S**ystem) [125, 126], verwendet zur Kurzzeitprognose ein Bayes-basiertes Vorhersagemodell namens PREDICT und zur Berechnung der Phasenwechsel den COP-Algorithmus (Controlled Optimization of Phases) [176].

Daneben gibt es das Verfahren **B**ALANCE (**B**alancing **A**daptive **N**etwork **C**ontrol **M**ethod) [57, 58, 185], zu dessen Entwicklung Friedrich 1997 in seiner Dissertation [57] maßgeblich beigetragen hat. Auf drei Hierarchieebenen werden dabei die klassischen Parameter Phaseneinteilung, Umlaufzeit, Freigabezeit und Versatz optimiert. Die Zielfunktion berücksichtigt für jede Verkehrsart die Wartezeiten, die Anzahl der Halte sowie die Menge an Schadstoffemission.

Zu erwähnen ist auch **M**OTION (**M**ethod for the **O**ptimization of **T**raffic **S**ignals in **O**nline **C**ontrolled **N**etworks) [30, 17, 31, 181]. Auf der taktischen Ebene hat MOTION zwei Besonderheiten: Das „Incident-Detection“-Modul erkennt Störungen auf Streckenabschnitten und reagiert darauf mit vordefinierten Schaltplänen. Das „Stage-Sequence-Selection“-Modul hingegen erkennt, ob Phasen ausgelassen oder vertauscht werden können. MOTION ist ein Teil des Verkehrsmanagementsystems „Sittraffic Concert“ der Siemens AG und kam beispielsweise zur Fußballweltmeisterschaft 2006 rund um die Schalker Veltins-Arena zum Einsatz.

Store-and-forward Ansatz

Einen anderen Ansatz verfolgt die Gruppe um Papageorgiou mit dem Steuerungsverfahren TUC (**T**raffic-responsive **U**rban **C**ontrol) [18, 47, 48, 50, 49]. Das Verfahren ist durch die Regelung von Füllständen in Kanalisations-Netzwerken inspiriert, bei denen anstelle binärer Variablen wie „Sperrung“ und „Freigabe“ kontinuierliche Ventilstellungen verwendet werden. Mit dem so genannten „store-and-forward“-Ansatz leitet TUC aus einer linear-quadratischen Problemformulierung einen multivariablen Reglerentwurf ab.

TUC nimmt für die Geschwindigkeit, mit der die maximale Rückstaulänge auf einer Knotenzufahrt von Umlauf zu Umlauf anwächst, eine lineare Funktion der Freigabe- und Sperrzeiten an. Die maximale Rückstaulänge ist somit das Integral dieser Funktion. TUC nutzt aus, dass sich ein lineares System mit einem solchen Verhalten durch einen Proportionalregler stabilisieren lässt. Eine offline ermittelte Steuermatrix bildet die Grundlage einer linearen Berechnungsvorschrift, mit der online die Freigabezeiten kontinuierlich variiert werden. Dabei berücksichtigt TUC ein so genanntes „gating feature“, d.h. die Freigabezeiten werden reduziert, wenn sich stromabwärts ein Rückstau bildet [48].

In der Steuermatrix sind die Wechselwirkungen aller Knotenpunkte innerhalb des Netzwerks abgebildet. Wenn sich diese ändern, muss die Steuermatrix neu berechnet werden. Die besondere Stärke von TUC besteht darin, die Echtzeit-Berechnungen auf ein linear-quadratisches Optimierungsproblem mit polynomialer Komplexität zu reduzieren. Deswegen ist es für den Einsatz in sehr großen Netzwerken geeignet. Derzeit wird es in europäischen Großstädten wie München, Glasgow (Schottland), Chania (Griechenland) und Southampton (England) erprobt, wo es nach eigenen Angaben gute Ergebnisse auch bei gesättigten Verkehrsbedingungen liefert. In den letzten Jahren wurde der Kern des Verfahrens um eine Anpassung der Umlauf- und Versatzzeiten sowie um eine ÖV-Priorisierung erweitert [18, 50].

2.2. Geschaltete Warteschlangennetzwerke

In zu Straßennetzwerken verwandten Bereichen gibt es andere innovative Lösungsansätze. Straßennetzwerke lassen sich, ebenso wie Produktions- oder Telekommunikationssysteme, unter bestimmten Voraussetzungen als „geschaltete Warteschlangennetzwerke“ modellieren. Die lichtsignalgesteuerten Knotenpunkte nehmen die Rolle von Bedienstationen ein, der Rückstaubereich auf den Knotenzufahrten die Rolle von Puffern. Diese Betrachtungsweise erlaubt es, einen bedeutenden Teil der für die Steuerung anderer technischer Systeme gewonnenen Erkenntnisse und Methoden auf Straßennetzwerke zu übertragen. Aus dem Literaturfundus zu „geschalteten Warteschlangennetzwerken“ sind für unsere Unter-

suchungen insbesondere die Beiträge zur Optimalität und Stabilität bestimmter Reglerstrukturen von Interesse. Im Folgenden werden einige Nennenswerte daraus vorgestellt.

2.2.1. Steuerungsverfahren

Clearing Policies

Ein intuitiver und häufig diskutierter Ansatz zur dezentralen Steuerung von Warteschlangennetzwerken sind sogenannte Clearing-Policies [142, 29, 150, 104]. Besonders bekannt ist die „Clear-Largest-Buffer-Policy“. Sie lautet: „Bediene die Zufahrt mit der längsten Warteschlange und leere sie vollständig.“ Während der Leerung einer Warteschlange wird der Bedienvorgang nicht unterbrochen. Deshalb charakterisiert man die Schaltregel als „erschöpfend“ (engl.: „exhaustive“). Stabilität⁵ ist für Clearing-Policies nur in maschenfreien⁶ Netzwerken garantiert [98, 86]. In nicht-maschenfreien Netzwerken können Clearing-Policies selbst bei geringem Verkehrsaufkommen instabiles Verhalten zeigen.

Ein nicht-maschenfreies Netzwerk ist beispielsweise in Abb. 2.6 dargestellt. Dieses Netzwerk wurde von Kumar und Seidman in [98] analytisch untersucht. Der für diese Dissertationsschrift relevante Teil ihrer Arbeit ist im Anhang A.1 zusammengefasst. Aus ihrer Arbeit wird ersichtlich, dass jede Masche eine potentielle Rückkopplungsschleife darstellt. Die Kopplung entsteht dabei durch die Verkehrsströme selbst, denn die Ankünfte auf den Knotenzufahrten sind nicht unabhängig vom Schaltprogramm, sondern eine Folge dessen. Die Rückkopplungsschleife schließt sich indem die Schaltprogramme auf diese Ankünfte reagieren. Falls die Rückkopplung positiv ist, das heißt falls die Wirkung ihre Ursache verstärkt, entsteht instabiles Verhalten.

Die Möglichkeiten der Stabilisierung sind vielfältig [98, 86, 127, 159, 55, 44]. Ein Ansatz ist, die instabile Clearing-Policy um eine übergeordnete, stabilisierende Schaltregel zu ergänzen. Dieser Ansatz wird in Kapitel 5 verfolgt.

Erweitertes lineares Komplementaritätsproblem

In den Referenzen [173, 174, 171, 172] betrachten de Schutter und de Moor einen isolierten Knotenpunkt mit konstanten Ankunftsrate und fest vorgegebener Phasenfolge.

⁵Vgl. Abschnitt 2.2.2

⁶Maschenfrei ist ein Netzwerk genau dann, wenn sich kein geschlossener Pfad finden lässt, der von einem beliebigen Knoten über andere zum selben Knoten zurück führt. Da in Straßennetzwerken aber jeder Knotenpunkt von jedem anderen aus erreichbar ist, sind diese niemals maschenfrei. Jede Straße mit zwei Richtungsfahrbahnen stellt eine Masche dar.

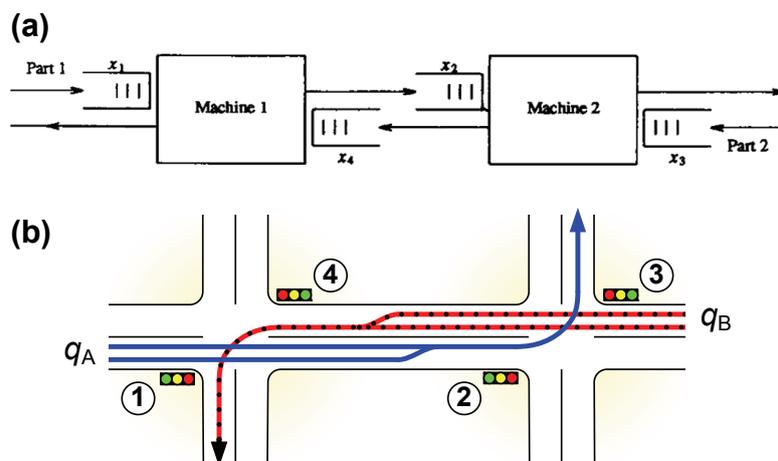


Abbildung 2.6.: Das Prinzip, die längste Schlange zuerst zu bedienen, kann selbst bei geringem Verkehrsaufkommen zu instabilem Verhalten führen. (a) Kumar und Seidman [98] beobachten dieses Phänomen in einem Produktionsnetzwerk mit nur zwei Bedienstationen. (b) Das gleiche Netzwerk mit zwei verschränkten Verkehrsströmen A und B . Die analytische Berechnung ist in Anhang A.1 aufgeführt. Dieses Beispiel demonstriert eindrucksvoll die Schwierigkeit beim Entwurf dezentraler Steuerungsverfahren.

Im Schaltzustand Freigabe ist die Ausflussrate aus den Knotenzufahrten entweder maximal oder gleich der konstanten Ankunftsrate.⁷ Durch Aufstellen bestimmter nichtlinearer Nebenbedingungen können diese beiden Ausflussraten zusammengefasst werden. Damit formulieren die Autoren das Optimierungsproblem als ein erweitertes lineares Komplementaritätsproblem [39]. Wird anstatt der Gesamtwarteziten die mittlere Warteschlangenlänge zu den Zeitpunkten des Umschaltens auf Freigabe minimiert, so können die nichtlinearen Nebenbedingungen in lineare Nebenbedingungen überführt werden. Die Arbeiten von de Schutter und de Moor liefern ein eindrucksvolles Beispiel dafür, wie aufwändig die Berechnung der optimalen Umschaltzeitpunkte ist – trotz dem die Zuflussraten konstant sind.

Vorgeben einer Wunschtrajektorie

Anstatt einen Regler vorzugeben und die resultierende Dynamik zu untersuchen, schlagen Lefeber, Rooda und van Eekelen in den Referenzen [105, 56] den umgekehrten Weg vor: Sie gehen von einem gewünschten Systemverhalten aus und leiten dafür den zugehörigen Regler ab. Das gewünschte Systemverhalten ist dabei in Form einer periodischen Schaltfolge und einer zugehörigen Trajektorie für die

⁷Vgl. Abschnitt 3.1.2

Warteschlangenlängen im Gleichgewichtszustand vorzugeben. Für den geschlossenen Regelkreis bildet die Wunschtrajektorie einen globalen Attraktor.

Der Reglerentwurf spezifiziert die Bedingungen für die Übergänge von einer Schaltkombination zur nächsten. Der Regler hat die Struktur eines endlichen Zustandsautomaten [85]. Durch schrittweise Minimierung einer Lyapunov-Funktion [95] werden die Übergangsbedingungen berechnet. Die tatsächlichen Umschaltzeitpunkte ergeben sich dann dynamisch aus den aktuellen Systemzuständen.

Prioritätsbasierte Verfahren

Bei den prioritätsbasierten Verfahren, auch „Index-Policies“, wird jeder Warteschlange i ein Prioritätsindex π_i zugewiesen. Dieser kann als ein Maß für die Dringlichkeit der Bedienung verstanden werden. Die Warteschlange mit höchster Dringlichkeit wird zur Bedienung ausgewählt. Die Schaltregel lautet allgemein:

$$\sigma = \arg \max_i \pi_i. \quad (2.1)$$

Hierbei nimmt die Entscheidungsvariable σ das Argument, d.h. den Index i , der zu bedienenden Warteschlange an.

Der Prioritätsindex π_i kann sich als Funktion der Warteschlangenlänge, der Wartezeit und vielem mehr berechnen. Hinsichtlich der Spezifikation von π_i wird eine Vielzahl von prioritätsbasierten Verfahren unterschieden. Panwalkar und Iskander geben in Referenz [139] einen Eindruck über das breite Spektrum indem sie 113 verschiedene Spezifikationen auflisten. Inhaltlich mehr in die Tiefe gehen die Referenzen [40, 191, 137, 36, 54, 150, 135, 114]. In Kapitel 4 wird ein heuristisches Optimierungsverfahren auf der Basis eines dynamischen Prioritätsindex entwickelt.

2.2.2. Stabilität und Steuerbarkeit

Stabilität

Die Stabilität geschalteter Warteschlangennetzwerke ist nach Perkins und Kumar [142] wie folgt definiert: Ein Schaltprogramm ist dann stabil, wenn die Warteschlangenlänge⁸ $n_i(t)$ sämtlicher Knotenzufahrten i unter der Annahme unbegrenzter Rückstaukapazität zu jeder Zeit t kleiner als eine endliche Konstante n^{\max} ist, d.h.

$$\text{es existiert } n^{\max} > 0, \text{ so dass } n_i(t) < n^{\max} \text{ für alle } i, t. \quad (2.2)$$

⁸Vgl. Abschnitt 3.2.1

Notwendige Stabilitätsbedingung

Ein beliebiges Schaltprogramm kann nur dann stabil sein, wenn folgende notwendige Bedingung erfüllt ist: An jedem Knotenpunkt muss die Summe der Auslastungen λ_i der Knotenzufahrten i kleiner als 100% sein, d.h.

$$\sum_i \lambda_i < 1 \quad \text{für jeden Knotenpunkt im Netzwerk.} \quad (2.3)$$

Unter der Auslastung einer Knotenzufahrt versteht man dabei das Verhältnis aus mittlerer Ankunfts- und maximaler Ausflussrate. Falls mehrere Verkehrsströme in einer Phase gemeinsam bedient werden können, so zählt in jeder Phase nur der Verkehrsstrom mit der größten Auslastung.⁹

Steuerbarkeit

Savkin [160] überträgt das Konzept der Steuerbarkeit, welches Kalman [93] für lineare Regelsysteme aufgestellt hat, auf geschaltete Warteschlangennetzwerke. Savkin nennt ein geschaltetes Warteschlangennetzwerk mit konstanten Randbedingungen steuerbar, falls ein Regler existiert, mit dem alle Zustandsgrößen des geschlossenen Regelkreises beschränkt sind und ihre Trajektorien gegen einen Grenzyklus mit gegebener Periode streben. Er zeigt, dass für einen solchen Regler beispielsweise eine „Festzeitsteuerung“ in Betracht kommt. In einer anderen Arbeit [161] beweist Savkin, dass die Periode des Grenzyklus nicht beliebig klein sein kann.

2.3. Umlaufbasierte Steuerungen

Unter den Verfahren zur Steuerung von Lichtsignalanlagen dominieren die umlaufbasierten Verfahren. Ihr Ansatz ist es, das Verkehrsgeschehen nach einem periodischen Muster ablaufen zu lassen. Dazu werden die Schaltzustände der Lichtsignalanlagen auf einer Zeitscheibe (Abb. 2.10) abgebildet. Die Koordinierung von Knotenpunkten kommt dadurch zustande, dass diese Zeitscheibe an allen Knotenpunkten mit der gleichen Periode, der sogenannten Umlaufzeit T^{cyc} , durchlaufen wird. Verkehrsadaptive Verfahren passen diese Zeitscheibe im Tagesverlauf an das aktuelle Verkehrsaufkommen an [182, 57, 177]. Die wesentlichen Nebenbedingungen und die freien Parameter, die beim Entwurf einer umlaufbasierten Steuerung auftreten, werden im Folgenden vorgestellt.

2.3.1. Randbedingungen

⁹Vgl. Abschnitt 2.3.2, insbesondere S. 28

Unverträgliche Verkehrsströme

Zunächst muss beachtet werden, dass nur paarweise verträgliche Verkehrsströme, d.h. nur solche ohne gemeinsamen Konfliktpunkt im Kreuzungsbereich [4], gemeinsam in einer Phase bedient werden dürfen. Nicht verträgliche Ströme werden in verschiedenen Phasen bedient. Eine Ausnahme bilden die bedingt verträglichen oder signaltechnisch ungesicherten Verkehrsströme nach §9 (3,4) der Straßenverkehrsordnung (StVO).

Die Menge aller zulässigen Phasen lässt sich mit Hilfe eines sogenannten Verträglichkeitsgraphen [90, 57, 70, 71] bestimmen. In diesem Graph stellen die Verkehrsströme die Knoten dar. Sind zwei Ströme verträglich, werden die zugehörigen Knoten durch eine Kante verbunden. Jeder Teilgraph, in dem sämtliche Knotenpaare miteinander durch Kanten verbundenen sind, entspricht einer Phase mit paarweise zueinander verträglichen Verkehrsströmen. Welche Phasen im Schaltprogramm zu verwenden und wie sie dort anzuordnen sind, ist in erster Linie vom periodischen Ankunftsprofil der Fahrzeuge auf den Zufahrten abhängig und kann mit Methoden der kombinatorischen Optimierung entschieden werden. Üblicherweise wird dabei jede Zufahrt pro Umlauf genau ein Mal bedient. Moderne Verfahren können jedoch außerdem im laufenden Betrieb Phasen vertauschen oder überspringen.

Entsprechend Annahme A3 wird in den nachfolgenden Kapiteln davon ausgegangen, dass alle Verkehrsströme eines Knotenpunktes zueinander unverträglich sind. Jede Phase wird deshalb nur durch genau einen Verkehrsstrom repräsentiert. Diese Annahme erlaubt es beim analytischen Reglerentwurf, auf die Betrachtung der unterschiedlichen Schaltkombinationen zu verzichten. Wie diese Annahme aufgehoben werden kann, diskutiert Abschnitt 8.1.3.

Zwischenzeiten

Bevor die Freigabe einer Phase p beginnen kann, müssen die anderen Phasen, deren Verkehrsströme zu denen von p unverträglich sind, rechtzeitig gesperrt werden [166]. Die Zeitspanne zwischen dem Ende der Freigabe der anderen Phasen und dem Beginn der Freigabe von p wird Zwischenzeit τ_p^0 genannt. Die Größe der Zwischenzeiten hängt wesentlich von der Kreuzungsgeometrie und der Art der Verkehrsteilnehmer ab. Typische Zwischenzeiten liegen zwischen 4 und 6 Sekunden, bei kreuzenden Fußgängern oder Straßenbahnen sind auch 10 Sekunden üblich [4, 186]. Auch Annahme A4, dass sich die Zwischenzeiten ausschließlich an der zu bedienenden Phase bemessen, wird zugunsten eines einfacheren analytischen Zugangs getroffen.

Im Folgenden wird unter Zwischen- bzw. Umschaltzeit der effektive Zeitverlust eines Phasenwechsels verstanden. Sie erstreckt sich von dem Zeitpunkt, zu dem der Phasenwechsel beginnen soll, bis zu dem Zeitpunkt, zu dem das erste Fahrzeug

der neuen Phase effektiv in den Kreuzungsbereich einfährt. Nach dieser Definition sind in der Zwischenzeit sowohl die Anfahrverluste als auch die Überfahrzeiten mit enthalten.

Anfahrverluste und Überfahrzeiten

Die Zeitpunkte, zu denen das erste Fahrzeug frühestens bzw. das letzte Fahrzeug spätestens über die Haltelinie fährt, fallen aufgrund von Reaktionszeiten und Verzögerungen beim Anfahren und Abbremsen nicht exakt mit den Signalbildwechseln zusammen. Die Anfahrverluste bei „Rot-Gelb“ und die Überfahrzeit bei „Gelb“ lassen sich entweder empirisch bestimmen oder mit den Formeln aus [4, 166, 196] berechnen. Long [113] ermittelte in einer empirischen Studie Anfahrverluste von 2,0 Sekunden. Überfahrzeiten sind laut Schabel und Lose [166] für Geradeausfahrer mit drei Sekunden und für Abbieger mit zwei Sekunden anzusetzen. Abbildungen 2.7(a) und (b) stellen Signalbildwechsel und Fahrzeugtrajektorien gegenüber.

Der Zustand „Freigabe“ soll sich auf die Zeitspanne beziehen, während der die Fahrzeuge potenziell in den Kreuzungsbereich einfahren können. Dieser Freigabezustand wird mit $\gamma_i(t) \in \{0, 1\}$ bezeichnet (1 für Freigabe, 0 für Sperrung). Abb. 2.7(c) zeigt seinen Verlauf. Um auf Annahme A4 verzichten zu können, müsste man in den Modellen, die in den nachfolgenden Kapiteln entwickelt werden, eine zusätzliche Zeitspanne zwischen den berechneten und den tatsächlichen Umschaltzeitpunkten berücksichtigen.

2.3.2. Entwurfsparameter

Aus den genannten Randbedingungen folgen bestimmte Regeln, nach denen die Parameter einer umlaufbasierten Steuerung zu bemessen sind.

Freigabezeiten

Entsprechend Annahme A1 fahren die auf einer Knotenzufahrt i aufgestauten Fahrzeuge während der Freigabe mit der maximalen Rate q_i^{\max} heraus [166, 79]. Diese maximale Ausflussrate wird auch Sättigungsverkehrsstärke genannt. Sie kann als konstant angenommen werden, wenn der Zufluss in nachfolgende Streckenabschnitte nicht durch Rückstaueffekte gemindert oder blockiert ist.

Um eine Regel für die Bemessung der Freigabezeit für eine Knotenzufahrt i abzuleiten, wird zunächst die Bilanz der zu- und abfließenden Fahrzeuge während einer Umlaufzeit T^{cyc} betrachtet. Innerhalb der Umlaufzeit T^{cyc} können maximal $\Delta n_i^{\text{out}} = q_i^{\max} \int_{T^{\text{cyc}}} \gamma_i(t) dt$ Fahrzeuge aus der Zufahrt heraus fahren. In der gleichen Zeitspanne kommen $\Delta n_i^{\text{exp}} = \int_{T^{\text{cyc}}} q_i^{\text{exp}}(t) dt$ neue Fahrzeuge hinzu. Hierbei bezeichnet $q_i^{\text{exp}}(t)$ die Ankunftsrate der Fahrzeuge.¹⁰ Der zufließende Verkehrs-

¹⁰Vgl. Abschnitt 3.1.1

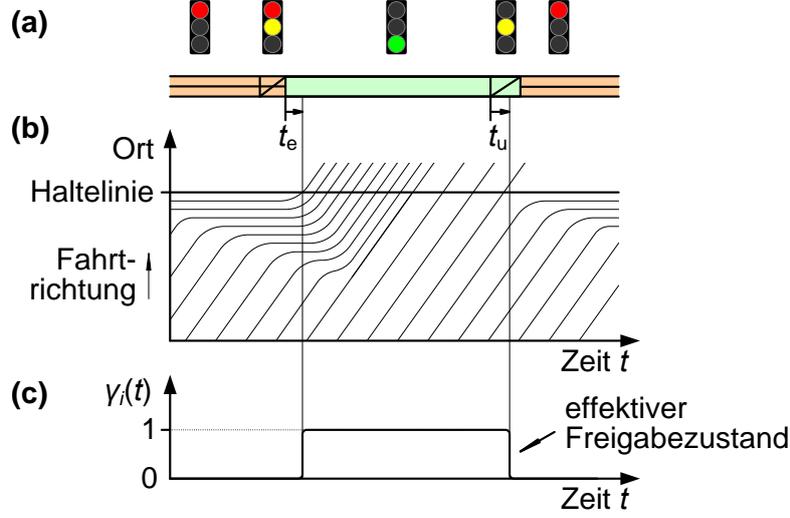


Abbildung 2.7.: (a) Signalbildabfolge nach §37 (2) der StVO und zugehörige Signalzeitendarstellung nach RiLSA [4]. (b) Fahrzeugtrajektorien auf einer Knotenzufahrt bei konstanter Zufussrate. Die Einfahrzeit t_e und die Überfahrzeit t_u entsprechen der Zeitspanne zwischen dem jeweiligen Signalbildwechsel und den frühesten bzw. spätesten Überfahrzeitpunkten der Haltelinie. (c) Verlauf des effektiven Freigabezustands $\gamma_i(t) \in \{0, 1\}$.

trom kann nur unter der Bedingung $\Delta n_i^{\text{out}} > \Delta n_i^{\text{exp}}$ abgefertigt werden. Diese ist erfüllt, wenn

$$\bar{\gamma}_i > \bar{\lambda}_i \quad \text{mit} \quad \bar{\lambda}_i = \bar{q}_i^{\text{exp}} / q_i^{\text{max}} . \quad (2.4)$$

Hierbei bezeichnet $\bar{\gamma}_i$ den mittleren Freigabezeitanteil und $\bar{\lambda}_i$ die mittlere Auslastung der Knotenzufahrt i . Die Bedingung (2.4) gibt also an, welcher Anteil der Umlaufzeit mindestens für die Freigabe von Zufahrt i aufgewendet werden muss. Die Größe des Freigabezeitfensters g_i^0 für Zufahrt i muss folglich größer als eine Mindestfreigabezeit g_i^{min} sein, d.h.

$$g_i^0 > g_i^{\text{min}} = \bar{\lambda}_i T^{\text{cyc}} . \quad (2.5)$$

In einer Phase p werden im Allgemeinen mehrere Zufahrten gleichzeitig bedient. Die Mindestfreigabezeit g_p^{min} einer Phase p muss dann so groß sein wie größte Mindestfreigabezeit ihrer Zufahrten:

$$g_p^{\text{min}} = \max_{i \in p} \bar{\lambda}_i T^{\text{cyc}} . \quad (2.6)$$

Hierfür sei angenommen, dass jede Zufahrt in genau einer Phase bedient wird. Der Maximum-Operator in Gleichung (2.6) macht deutlich, dass für die Bemessung

der erforderlichen Freigabezeit jeder Phase nur eine einzige Zufahrt entscheidend ist: die mit der größten Auslastung. Webster drückt dies in seiner Arbeit „Traffic Signal Settings“ [192] so aus:

„Each phase can be represented by one approach only – the one with the highest ratio of flow to saturation flow.“

Die Verkehrsströme, die bei einer gegebenen Phaseneinteilung die jeweils größte mittlere Auslastung $\bar{\lambda}_i$ haben und folglich die größte Freigabezeit erfordern, werden im Weiteren die „kritischen Verkehrsströme“ genannt.

Umlaufzeit

Die benötigte Zeit für den Bedienvorgang einer Phase p setzt sich aus der Zwischenzeit τ_p^0 und der Freigabezeit g_p^0 zusammen. Zum Durchlaufen einer gegebenen Phasenfolge wird die Umlaufzeit T^{cyc} benötigt, für die gelten muss:

$$T^{\text{cyc}} > \sum_p (\tau_p^0 + g_p^0) . \quad (2.7)$$

Während die Zwischenzeiten konstant sind, hängt die Mindestfreigabezeiten g_p^{min} entsprechend Gleichung (2.6) von der Umlaufzeit selbst ab. Durch Einsetzen und Umstellen erhält man eine Bedingung für die Mindestumlaufzeit $T^{\text{cyc,min}}$:

$$T^{\text{cyc}} > T^{\text{cyc,min}} = \frac{\sum_p \tau_p^0}{1 - \Lambda} . \quad (2.8)$$

Hierbei bezeichnet Λ die Auslastung des Knotenpunktes. Sie ist die Summe der maximalen Auslastungen jeder einzelnen Phase p :

$$\Lambda = \sum_p \max_{i \in p} \bar{\lambda}_i . \quad (2.9)$$

Der Ausdruck $1 - \Lambda$ im Nenner von Gleichung (2.8) macht deutlich, dass die ankommenden Verkehrsströme nur bedient werden können, wenn die Knotenpunktauslastung Λ kleiner als eins ist. Das ist konsistent mit der notwendigen Stabilitätsbedingung (2.3).

Je größer die Knotenpunktauslastung ist, desto größer ist die benötigte Umlaufzeit. Der sichere Betrieb von Lichtsignalanlagen setzt voraus, dass maximale Sperr- und damit auch maximale Umlaufzeiten nicht überschritten werden [4]. Für eine gegebene Umlaufzeit T^{cyc} muss also die Knotenpunktauslastung Λ , wie durch Umstellen von Ungleichung (2.8) folgt, kleiner als eine Maximalauslastung Λ^{max} sein. Für diese gilt:

$$\Lambda < \Lambda^{\text{max}} = 1 - \sum_p \tau_p^0 / T^{\text{cyc}} . \quad (2.10)$$

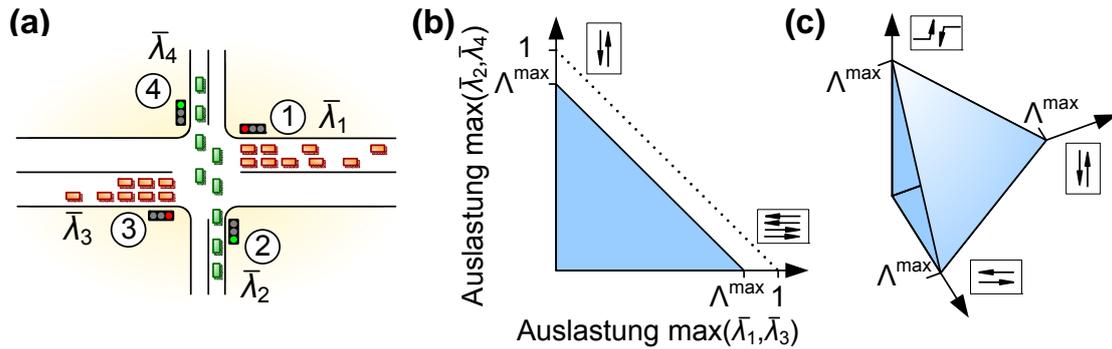


Abbildung 2.8.: (a) Ein Knotenpunkt mit mehreren Verkehrsströmen wird umlaufbasiert geschaltet. (b) Zweiphasenbetrieb ohne Linksabbieger. Das Verkehrsaufkommen kann genau dann bewältigt werden, wenn die mittleren Auslastungen $\bar{\lambda}_i$ der Zufahrten i innerhalb des schattierten Dreiecks liegen. Die gepunktete Linie markiert die Kapazitätsgrenze für den theoretischen Fall ohne Zwischenzeiten. (c) Im Dreiphasenbetrieb sind die zulässigen Auslastungen $\bar{\lambda}_i$ durch eine Ebene begrenzt.

Für ein umlaufbasiertes Verfahren mit gegebener Umlaufzeit T^{cyc} und nach (2.5) bemessenen Freigabezeiten g_i^0 stellt die Ungleichung (2.10) eine hinreichende Stabilitätsbedingung dar.

Abbildung 2.8 zeigt für einen vierarmigen Knotenpunkt im Zwei- bzw. Dreiphasenbetrieb, welche Kombinationen der mittleren Auslastungen $\bar{\lambda}_i$ der Zufahrten i die Bedingung (2.10) erfüllen. Die Menge der zulässigen Kombinationen stellt die Durchlassfähigkeit des Knotenpunktes dar [166].

Berücksichtigung stochastischer Einflüsse

Weil die Verkehrsnachfrage zufälligen Schwankungen unterliegt, sind in manchen Umläufen weniger und in anderen dagegen mehr Fahrzeuge abzufertigen. Stochastische Warteschlangenmodelle erlauben es, die Verteilung der Ankunftszeitpunkte analytisch zu begründen. Aus den Quantilen dieser Verteilungen können die Freigabezeiten dann so bemessen werden, dass sie beispielsweise in 90 oder 95% aller Fälle ausreichend groß sind [5].

Stochastische Warteschlangenmodelle beschreiben das Verkehrsgeschehen üblicherweise im stationären Gleichgewicht, das heißt unter der Annahme, dass sich die statistischen Verteilungen der Verkehrskenngrößen im Betrachtungszeitraum nicht ändern. Diese Modelle finden in der Literatur große Beachtung, beispielsweise in den Referenzen [7, 110, 73, 197, 158, 94]. Einen umfassenden Überblick geben Troutbeck und Brilon in [187, 25] und van Woensel und Vandaale in [195]. Welches Modell im konkreten Fall zu verwenden ist, hängt vom erforderlichen Detaillierungsgrad ab. Beispielsweise können Bündelungseffekte durch zusammengesetzte,

sogenannte dichotomisierte, Verteilungen berücksichtigt werden [184, 42]. Im einfachsten Fall werden die Ankünfte als stochastisch unabhängig angenommen und als Poisson-Prozess modelliert. Insbesondere an isolierten Knotenpunkten bei geringer mittlerer Auslastung ist diese Annahme sehr gut erfüllt [166, 26]. Die Bruttozeitlücken zwischen je zwei ankommenden Fahrzeugen sind dann exponential- und die Summe der Bruttozeitlücken zwischen mehreren ankommenden Fahrzeugen Erlang-verteilt. Die Anzahl der Fahrzeuge, die innerhalb einer gegebenen Zeitspanne ankommen, ist binomialverteilt.

2.3.3. Warteschlangen und Wartezeiten

Mit Hilfe der obigen Betrachtungen können Freigabe- und Umlaufzeiten so bemessen werden, dass das Verkehrsaufkommen abgefertigt werden kann und in der Folge die Warteschlangenlängen beschränkt bleiben. Wie groß die Warteschlangen jedoch tatsächlich werden und welche Gesamtwartezeiten entstehen, hängt außerdem vom Zeitverlauf der Fahrzeugankünfte ab. Zunächst wird von in der Zeit gleichmäßig verteilten Ankunftszeitpunkten ausgegangen. Anschließend, im nächsten Abschnitt 2.3.4, werden beliebige periodische Ankunftsprofile behandelt.

Konstante Ankunftsrate

Bei konstanten Ankunftsraten lässt sich die kumulierte Anzahl der auf einer Zufahrt i angekommenen bzw. abgefertigten Fahrzeuge $N_i^{\text{exp}}(t)$ bzw. $N_i^{\text{out}}(t)$ durch Geradenstücke approximieren. Abbildung 2.9(a) zeigt ihren Verlauf. Die dreieckige Fläche zwischen den beiden Kurven entspricht den Gesamtwartezeiten W_i^{cyc} aller Fahrzeuge während eines Umlaufs. Sie lässt sich unter Ausnutzung der geometrischen Beziehungen mit folgender Formel berechnen [89]:

$$W_i^{\text{cyc}} = \frac{q_i^{\text{exp}}}{2} \frac{(T^{\text{cyc}} - g_i^0)^2}{1 - \lambda_i}. \quad (2.11)$$

Die Gleichung macht deutlich, dass sich die Gesamtwartezeiten proportional zum Quadrat der Differenz aus Umlauf- und Freigabezeit, also zum Quadrat der Sperrzeit, verhalten. Summiert man über alle Zufahrten auf, erhält man die Gesamtwartezeiten $W^{\text{cyc}} = \sum_i W_i^{\text{cyc}}$ aller Fahrzeuge an der gesamten Kreuzung während eines Umlaufs.

Stochastisch unabhängige Ankunftszeitpunkte

Die Abschätzung (2.11) ist nur für konstante Ankunftsraten gültig. Unter stochastischen Bedingungen ist laut Webster [192] der Gesamtwartezeit aus (2.11) eine Funktion überlagert, die gegen unendlich strebt je näher die Umlaufzeit an der Mindestumlaufzeit liegt. In seiner Arbeit führt er aus:

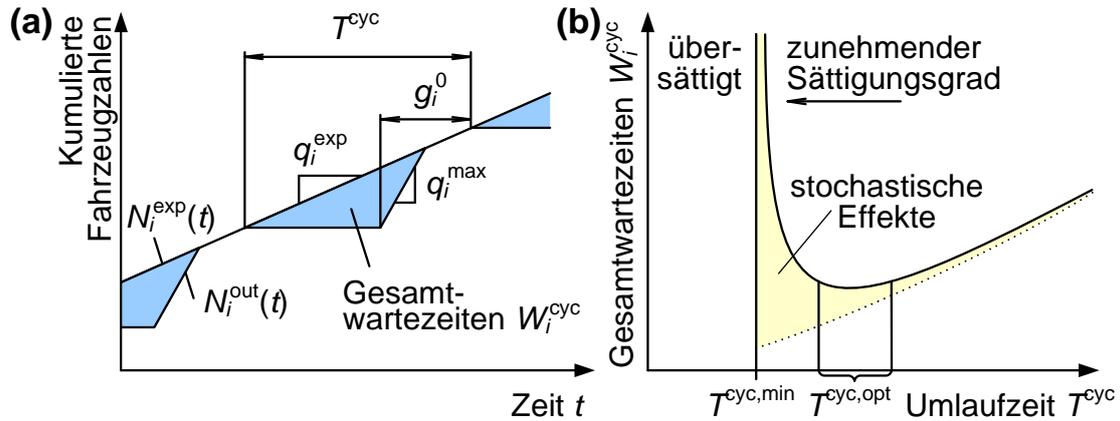


Abbildung 2.9.: (a) Kumulierte Anzahl der auf einer Zufahrt i mit konstanter Zuflussrate ankommenden und abgefertigten Fahrzeuge $N_i^{\text{exp}}(t)$ bzw. $N_i^{\text{out}}(t)$. Die Gesamtwartezeit W_i^{cyc} innerhalb eines Umlaufs T^{cyc} entspricht der Fläche des schattierten Dreiecks. (b) Gesamtwartezeiten W_i^{cyc} in Abhängigkeit der Umlaufzeit T^{cyc} bei stochastischen Ankünften nach Webster [192]. Je stärker die Kreuzung ausgelastet ist, d.h. je näher die Umlaufzeit T^{cyc} an der Mindestumlaufzeit $T^{\text{cyc},\text{min}}$ aus Gleichung (2.8) liegt, desto größer ist der Einfluss der stochastischen Effekte. Ohne diese würde W_i^{cyc} entlang der gepunkteten Kurve verlaufen.

*„When traffic is of a truly random character
the minimum cycle is associated with infinite delay.“*

Abbildung 2.9(b) zeigt Websters Kurve unter der Annahme Poisson-verteilter Fahrzeugankünfte. Aus seinen empirischen Beobachtungen und analytischen Interpretationen schätzt er das Intervall der wartezeitoptimalen Umlaufzeit $T^{\text{cyc},\text{opt}}$ wie folgt ab:

$$T^{\text{cyc},\text{opt}} \in [0, 75; 1, 5] \cdot \frac{1.5 \cdot \sum_p \tau_p + 5 \text{ s}}{1 - \Lambda}. \quad (2.12)$$

Hierbei steht Λ für die Knotenpunktauslastung aus Gleichung (2.9). Der Vorfaktor $[0, 75; 1, 5]$ spiegelt die stochastischen Schwankungen in den Ankünften wider. Bei gleichmäßigen Verkehrsflüssen sind kleinere, bei stark schwankenden Verkehrsflüssen eher größere Vorfaktoren zu wählen [192, 166]. Auch wenn man den zufälligen Ankunftsprozess weiter spezifiziert [124, 10], bleibt der prinzipielle Zusammenhang zwischen Gesamtwartezeiten und Umlaufzeit der gleiche. Die Freigabezeiten müssen größer gewählt werden, als es im statistischen Mittel nötig wäre.

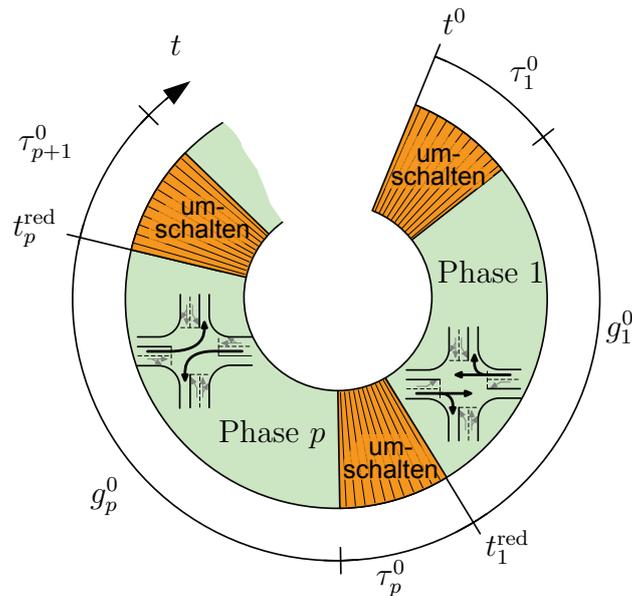


Abbildung 2.10.: Zeitscheibe einer umlaufbasierten Steuerung. Durch Verschiebung der Umschaltzeitpunkte t_p^{red} oder durch Drehung der gesamten Zeitscheibe können die Freigabezeitfenster an die periodischen Ankunftsprofile angepasst und somit die Gesamtwartezeiten am Knotenpunkt lokal minimiert werden.

2.3.4. Versatzzeitenanpassung

„Die Grundvoraussetzung für jede Koordinierung ist die einheitliche Umlaufzeit an allen angeschlossenen Lichtsignalanlagen“ [188]. Der periodische Zeitverlauf der Verkehrskenngrößen bildet somit die Zeitbasis der Koordinierung [60, 61, 5, 181]. Durch die Bedienprozesse an den lichtsignalgesteuerten Knotenpunkten werden die Verkehrsströme gebündelt und zu Kolonnen geformt [196]. Folglich werden die Wartezeiten in erster Linie von der zeitlichen Anordnung der Freigabezeitfenster in Bezug auf das periodische Ankunftsprofil bestimmt. Die Berechnung dessen, wie die Schaltmuster benachbarter Knotenpunkte zeitlich versetzt zueinander anzuordnen sind, nennt man „Versatzzeitenanpassung“.

Um die Gesamtwartezeiten aller Fahrzeuge an einem Knotenpunkt lokal zu minimieren, soll im Folgenden untersucht werden, wie die Lage der Zeitscheibe (siehe Abb. 2.10) für beliebige periodische Ankunftsprofile zu wählen ist. Die Anzahl der Haltevorgänge¹¹ bleibt zunächst unberücksichtigt.

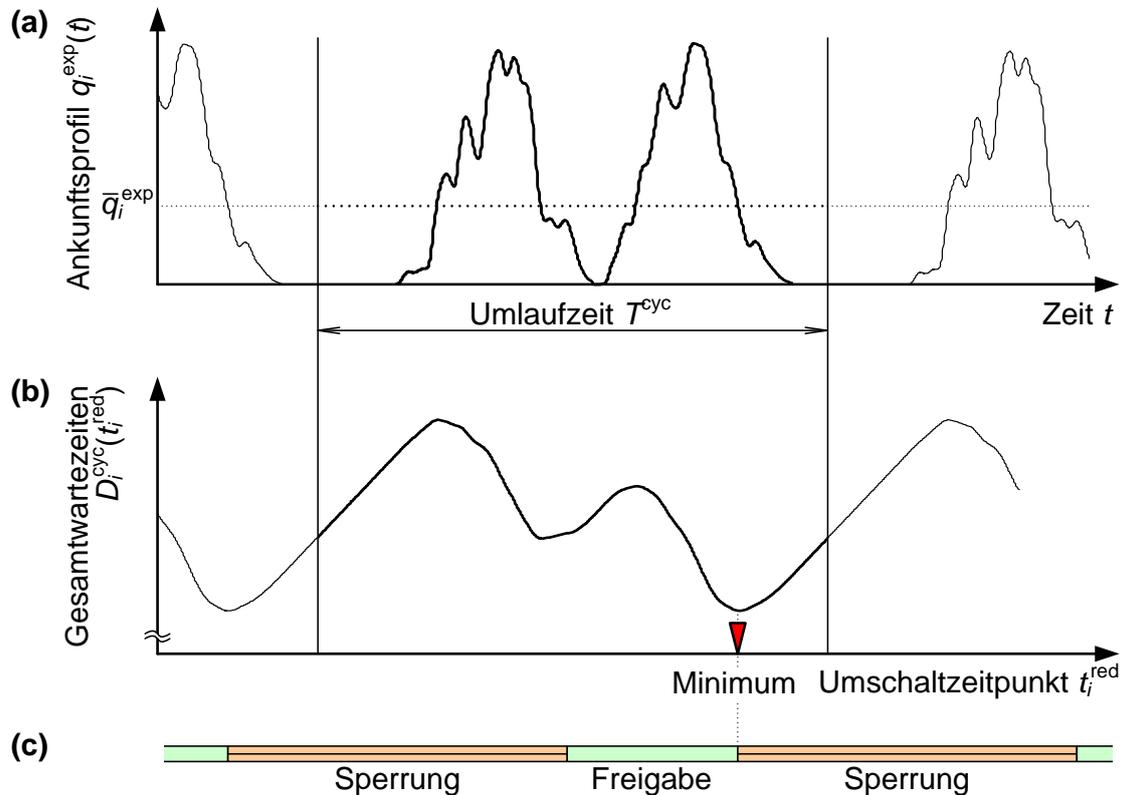


Abbildung 2.11.: (a) Periodisches Ankunftsprofil $q_i^{\text{exp}}(t)$ einer Zufahrt i mit dem Mittelwert \bar{q}_i^{exp} . (b) Kurve der Gesamtwartezeiten $W_i^{\text{cyc}}(t_i^{\text{red}})$ aller Fahrzeuge innerhalb der Umlaufzeit T^{cyc} in Abhängigkeit des Zeitpunktes t_i^{red} , zu dem von Freigabe auf Sperrung umgeschaltet wird. Der optimale Umschaltzeitpunkt lässt sich am Minimum der Kurve ablesen. (c) Optimale Lage des Freigabezeitfensters innerhalb des Umlaufs.

Wartezeitenkurve

Die Wartezeiten W_i^{cyc} , die während eines gesamten Umlaufs auf einer Knotenzufahrt i anfallen, lassen sich als eine Funktion in Abhängigkeit der Lage des Freigabezeitfensters angeben. Den Verlauf dieser Kurve für ein gegebenes Ankunftsprofil zeigt Abb 2.11. Die optimale Lage des Freigabezeitfensters ist dort, wo die Kurve W_i^{cyc} ihr Minimum hat.

Ein Verfahren zur Bestimmung der Wartezeitenkurve wurde vom Autor dieser Dissertationsschrift zusammen mit Peters, Kori und Helbing in Referenz [102] veröffentlicht. Schönemann [168] hat das Verfahren in einer Seminararbeit auf empirische Verkehrsdaten von der Kreuzung Strehleener Straße und Fritz-Löffler-Straße in Dresden erfolgreich angewendet.

¹¹Vgl. Abschnitt 3.2.3

Die Herleitung des Verfahrens ist in Anhang A.2 aufgeführt. Es beruht auf der Annahme, dass sich die Ankunftszeitpunkte der Fahrzeuge innerhalb der Umlaufzeit linear auf die Ausfahrzeitpunkte innerhalb des Freigabezeitfensters abbilden lassen. Je später ein Fahrzeug innerhalb des Umlaufs ankommt, desto später fährt es innerhalb des Freigabezeitfensters heraus. Damit lässt sich die Konstruktion der Wartezeitenkurve mit Gleichung (A.17) auf eine Integration des Ankunftsprofils zurückführen. Die Schlussfolgerungen, die sich aus dem Modell ableiten lassen, werden im Folgenden aufgelistet.

Eigenschaften der optimalen Lösung

In Referenz [60] untersuchen Gartner et al. die gleiche Fragestellung, wobei sie jedoch anstatt eines beliebigen Ankunftsprofils die Ankunft einer einzigen Kolonne mit konstanter Flussrate in ihrem Inneren annehmen. Damit konnten sie die Wartezeitenkurve mit Geraden und Parabeln stückweise annähern. Unsere allgemeinen Beobachtungen sind mit ihrem Spezialfall vollständig konsistent:

- (i) Die Platzierung des Freigabezeitfensters ist umso entscheidender für die Gesamtwartezeit W_i^{cyc} pro Umlauf, je stärker das Ankunftsprofil ausgeprägt ist, d.h. je mehr die Ankunftsrate $q_i^{\text{exp}}(t)$ zeitweilig von der mittleren Ankunftsrate \bar{q}_i^{exp} abweicht. Bei konstanten Ankunftsrate $q_i^{\text{exp}}(t) = \bar{q}_i^{\text{exp}}$ verschwindet die rechte Seite von Gleichung (A.14) und die Gesamtwartezeiten sind vom Umschaltzeitpunkt t_i^{red} unabhängig, nicht jedoch von der Freigabezeit g_i^0 .
- (ii) Als optimale Zeitpunkte t_i^{red} zum Umschalten auf Sperrung kommen nur Zeitpunkte in Frage, zu denen die Ankunftsrate $q_i^{\text{exp}}(t)$ unter den Mittelwert \bar{q}_i^{exp} abfällt. Ist die Flussrate innerhalb einer Kolonne größer als \bar{q}_i^{exp} , so ist die Abfertigung dieser Kolonne nicht zu unterbrechen.
- (iii) Gibt es nur eine einzige zusammenhängende Kolonne, dann wird Optimalität erreicht, wenn das Ende der Freigabezeit mit der Ankunft des letzten Fahrzeugs zusammenfällt. Gegebenenfalls ist das erste Fahrzeug zu verzögern, um lang gezogene Kolonnen zu stauchen.
- (iv) Kommt innerhalb eines Umlaufs nur eine einzige Kolonne mit konstanter Flussrate q_i^{exp} an und wird Δt Sekunden ($\Delta t \ll T^{\text{cyc}}$) später als optimal umgeschaltet, dann erhöhen sich die Gesamtwartezeiten um $\Delta W_i^{\text{cyc}} = (T^{\text{cyc}} - g_i^0)\bar{q}_i^{\text{exp}}\Delta t$ Sekunden. Wird dagegen Δt Sekunden früher als optimal umgeschaltet, dann erhöhen sich die Gesamtwartezeiten um $\Delta W_i^{\text{cyc}} = (T^{\text{cyc}} - g_i^0)(q_i^{\text{exp}} - \bar{q}_i^{\text{exp}})\Delta t$ Sekunden. Falls $q_i^{\text{exp}} > 2\bar{q}_i^{\text{exp}}$ gilt, d.h. falls sich die Ankunft der Kolonne über weniger als die halbe Umlaufzeit erstreckt, ist das um Δt spätere Umschalten günstiger, anderenfalls das frühere.

- (v) Auch bei der Ankunft mehrerer Kolonnen zeichnet sich die optimale Lösung dadurch aus, dass das Ende der Freigabezeit mit der Ankunft des letzten Fahrzeugs einer Kolonne zusammenfällt.
- (vi) Kommen innerhalb eines Umlaufs mehrere Kolonnen gleicher Größe an, dann ist es optimal, wenn das Ende der Freigabezeit mit der Ankunft des letzten Fahrzeugs derjenigen Kolonne zusammenfällt, die von der größten Zeitlücke bis zur Ankunft der nächsten Kolonne gefolgt wird.
- (vii) Falls innerhalb eines Umlaufs mehrere Kolonnen unterschiedlicher Größe ankommen, dann ist es niemals optimal, diejenige Kolonne mit einer Grünen Welle zu bedienen, deren Quotient aus Kolonnengröße und zeitlichem Abstand zwischen dem jeweils letzten Fahrzeug der vorherigen und der betrachteten Kolonne größer ist als die mittlere Ankunftsrate \bar{q}_i^{exp} .

2.3.5. Schwierigkeiten netzwerkweiter Koordinierung

Zur umlaufbasierten Netzkoordinierung reicht es nicht aus, lediglich die Freigabezeitfenster der Knotenpunkte individuell und unabhängig an die lokalen Fahrzeugankünfte anzupassen. Da die Verkehrsströme vernetzt sind, wirkt sich jede lokale Parametervariation an anderen Stellen im Netzwerk aus. Welche Schwierigkeiten das mit sich bringt, soll im Folgenden genauer untersucht werden.

Lokale und globale Optimalität

Ist für jeden einzelnen Knotenpunkt das individuelle Optimum gefunden, heißt das jedoch noch nicht, dass die Steuerung für das gesamte Straßennetzwerk optimal ist. Sobald ein einzelner Knotenpunkt sein Schaltprogramm verändert, um es lokal zu optimieren, verändert sich gleichzeitig das Ankunftsprofil an den benachbarten Knotenpunkten. Unter Umständen nehmen dort die Wartezeiten um einen größeren Betrag zu, als am ersten Knotenpunkt lokal eingespart werden konnte. Wenn die benachbarten Knotenpunkte daraufhin ihre Schaltprogramme reoptimieren, beeinflusst das wiederum das Ankunftsprofil am ersten Knotenpunkt. Diese Wechselwirkungen zerstören die Optimalität des ursprünglichen Schaltprogramms. Das liegt daran, dass die Ankünfte an einem Knotenpunkt von dessen eigenem Schaltprogramm abhängig sind: Das lokale Schaltprogramm eines Knotenpunktes bestimmt nicht nur das Profil der Ausflüsse, sondern über netzwerkweite Rückkopplungen auch das Profil der Ankünfte. Ein anschauliches Beispiel hierfür wird im Anhang A.1 vorgestellt.

Bedingung für globale Optimalität

Gartner et al. formulieren in [60, 61], unter welcher Bedingung das lokale Optimum an jedem einzelnen Knotenpunkt auch dem globalen Optimum des gesamten Straßennetzwerks entspricht:

„An important physical constraint on the system is that the sum of the offsets [...] around any loop of the network equals an integer multiple of the [common] cycle time.“

Das heißt, die Reisezeit entlang jedes geschlossenen Pfades bzw. entlang jeder Masche muss bei freiem Verkehr ein ganzzahliges Vielfaches der kollektiven Umlaufzeit sein. Die optimale Lösung zeichnet sich dadurch aus, dass ein einzelnes Fahrzeug mit der zulässigen Höchstgeschwindigkeit auf beliebigen Wegen durch das Straßennetz fahren kann, ohne mehr als ein einziges Mal (das erste Mal) zu halten.

Unter Umständen lässt sich diese Bedingung in Straßennetzwerken mit regulärer Struktur erfüllen: Die kollektive Umlaufzeit müsste dann – falls es die Verkehrsnachfrage überhaupt zulässt – an die Reisegeschwindigkeiten und Streckenlängen angepasst werden. Aber derartige Netzwerkstrukturen sind in den größeren deutschen Städten eigentlich nie zu finden [101].

Dominanzverfahren

Alternativ schlagen Gartner et al. [60] vor, nur einen bestimmten Teil des Straßennetzwerks zu betrachten, in dem keine geschlossenen Pfade enthalten sind. Der gleiche Ansatz wurde 1970 von Inose et al. [91] in den USA als Patent angemeldet. Um einen solchen maschenfreien Teilgraphen zu erhalten, müssen aus dem gerichteten Graphen des Straßennetzwerks mit K Knotenpunkten und I gerichteten Kanten (Richtungsfahrbahnen) mindestens $I - K + 1$ Kanten eliminiert werden. Das betrifft, wenn das Straßennetz, wie üblich, nicht von Einbahnstraßen dominiert wird, weit über die Hälfte aller Kanten. Der resultierende Graph ist ein Baum [97]. Nur die darin verbliebenen Kanten bzw. die zugehörigen Richtungsfahrbahnen dürfen in der Zielfunktion der Optimierung berücksichtigt werden. Steierwald, Künne und Vogt drücken dies in Referenz [181] so aus: *„Netzkoordinationen werden durch Verknüpfung von Linienkoordinierungen gebildet.“* Zwar können diese Kanten oder „Linien“ entsprechend der größten Verkehrsstärke beispielsweise mit Hilfe des Dominanzverfahrens [166, 11] ausgewählt werden. Aber ob es in einem weiteren Schritt gelingt, durch die Variation von Umlauf- und Freigabezeiten beispielsweise mittels genetischer Algorithmen [12] zusätzliche Kanten mit einzubeziehen, hängt von der konkreten Netzwerkgeometrie ab.

2.4. Zusammenfassung und Kritik

Mit dem Ansatz, die Schaltfolge an jedem Knotenpunkt fortwährend zu wiederholen, werden die Freiheitsgrade des ursprünglichen Optimierungsproblem drastisch reduziert. Schaltfolgen, mit denen die Verkehrsströme flexibel und vielleicht irregulär, aber dennoch hinreichend häufig und hinreichend lange bedient werden, sind von vornherein ausgeschlossen. Die verbleibenden Freiheitsgrade sind im Wesentlichen die im Abschnitt 2.3.2 diskutierten Parameter Phaseneinteilung, Freigabezeit, Umlaufzeit und Versatz.

Die Schranken, in denen diese Parameter liegen müssen, lassen sich allein aus der Kenntnis der mittleren Verkehrsstärke ableiten. Das erlaubt den Einsatz einfacher Sensortechnologie und eine Abschwächung der Echtzeitanforderung für die Übertragung zur Verkehrszentrale. Die Parameter können mit Quantilen von bekannten theoretischen Verteilungen so bemessen werden, dass eine gewünschte Durchlassfähigkeit trotz gewisser Abweichungen zwischen tatsächlicher und angenommener Verkehrsnachfrage gewährt bleibt. In dieser Hinsicht sind umlaufbasierte Verfahren robust.

Diese Robustheit wird jedoch mit höheren Gesamtwartezeiten erkaufte. Die Freigabezeiten werden an der stochastischen Schwankungsbreite der Verkehrsnachfrage bemessen und sind deshalb in den meisten Fällen – je nach Bemessungsgrundlage in nahezu 90 bzw. 95% aller Fälle – länger als notwendig. Mit den Freigabezeiten verlängern sich jedoch auch die Sperrzeiten. Zwar können moderne Verfahren bedarfsgerecht Phasen vertauschen oder Freigabezeiten verkürzen. Aber dies geschieht nur in einem gewissen Rahmen und, falls die Knotenpunkte in eine Netzkoordinierung eingebunden sind, ausschließlich unter der Maßgabe, die kollektive Umlaufzeit einzuhalten.

Die Grundlage der Netzkoordinierung ist das periodische Muster des Verkehrsablaufs. Das Ziel ist dabei meist eine Grüne Welle für den größten Verkehrsstrom. Aber sind Grüne Wellen immer die beste Wahl? In einem Forschungsbericht [26] des Bundesministeriums für Verkehr und der Forschungsgesellschaft für Straßen- und Verkehrswesen e.V. leiten die Autoren Brilon, Großmann und Blanke ihre Kritikpunkte mit folgenden an Deutlichkeit kaum zu übertreffenden Worten ein:

„Die Grüne Welle ist ein pragmatisches Koordinierungsverfahren, deren Ursprünge auf die Zeiten zurückgehen, in denen noch keine Computer zur Verfügung standen.“

Zwar ist die Idee der Grünen Welle aufs Erste überzeugend: Die Fahrzeuge des ausgewählten Verkehrsstroms erhalten freie Fahrt und brauchen nicht anzuhalten. Aber dem werden alle anderen Verkehrsströme untergeordnet. Besonders negativ wirkt sich dies bei ungünstigen Knotenpunktständen aus. Bei hohen Auslastungen kann die Grüne Welle sogar *„zu einer systematisch besonders schlechten*

Lösung mit langen Wartezeiten und vielfachen Haltevorgängen führen“ [26]. Außerdem wird die gemeinsame Umlaufzeit durch den am meisten ausgelasteten Knotenpunkt des Netzwerks vorgegeben. Folglich müssen alle anderen Knotenpunkte langsamer schalten als sie könnten, wodurch unnötig lange Rotzeiten entstehen.

Die kollektive Umlaufzeit wird als einzige Möglichkeit zur Koordinierung angesehen. Dabei kann auf diese Weise nur ein maschenfreier Teilgraph des Straßennetzwerks systematisch optimiert werden. Der Ansatz, die Lage der Freigabezeitfenster optimal an das Profil der Ankünfte anzupassen, lässt sich nur an weniger als jeder zweiten Zufahrt umsetzen.

Ein weiterer Kritikpunkt betrifft die Qualität der optimalen Lösungen. Die Optimierungsverfahren werden nicht auf das ursprüngliche, sondern auf das reduzierte Problem angewendet. Zwar erlaubt das Aufstellen zusätzlicher Nebenbedingungen, beispielsweise das Einhalten einer gegebenen Umlaufzeit oder durch das Außer-Acht-Lassen bestimmter Streckenabschnitte, effiziente Optimierungsverfahren einzusetzen. Aber die gefundenen Lösungen können nur ein Optimum für das reduzierte Problem darstellen.

Die Literatur zur Steuerung von Warteschlangennetzwerken bietet eine solide theoretische Grundlage, mit der sich die prinzipiellen Zusammenhänge und Probleme bei der Steuerung vernetzter Verkehrsströme verstehen lassen. Dennoch ist kein Verfahren bekannt, das in der vorliegenden Weise für die Steuerung von Lichtsignalanlagen geeignet wäre. Zum einen werden unverzichtbare Voraussetzungen verletzt, wie zum Beispiel das Einhalten maximaler Sperrzeiten. Zum anderen basieren sie auf unrealistischen Annahmen, wie zum Beispiel, dass die Ankunftsraten konstant sein müssen.

Kapitel 3.

Modellierung und Prognose von Bedienvorgängen

Kurzfassung

Dieses Kapitel soll erläutern, welche Zustandsgrößen des Verkehrsgeschehens zu erfassen und zur Verwendung in einem Prognoseverfahren aufzubereiten sind.

Erfasst werden muss lediglich die Zeitreihe der erwarteten Ankünfte und die Anzahl der tatsächlich abgefertigten Fahrzeuge. Beide Größen beziehen sich auf den Fahrbahnquerschnitt der Haltelinie. Wann die Fahrzeuge die Haltelinie frühestens erreichen können, lässt sich durch Verkehrsflussmessungen stromaufwärts der Haltelinie vorhersagen.

Mit einem deterministischen Warteschlangenmodell können aus diesen Messgrößen die Warteschlangenlänge, die Gesamtwartezeiten und der Kraftstoffverbrauch berechnet werden. Aus diesem Modell wird ein Prognoseverfahren abgeleitet, mit dem sich die zur Warteschlangenleerung benötigte Freigabezeit eines bevorstehenden Bedienvorgangs abschätzen lässt. Die während des Umschaltens und während der Freigabe hinzukommenden Fahrzeuge sind dabei mit erfasst. Neben der benötigten Freigabezeit lassen sich auch die Gesamtwartezeiten aller Fahrzeuge während des Bedienvorgangs prognostizieren. Es gelingt anzugeben, welche Wartezeiten infolge der Verzögerung oder Unterbrechung eines Bedienvorgangs zusätzlich entstehen.

Beide Größen, die prognostizierte Freigabe- und die prognostizierte Wartezeit, bilden die Grundlage für den Reglerentwurf in den nächsten Kapiteln.

3.1. Verkehrsflusserfassung

Zur Regelung des Verkehrsgeschehens ist die Erfassung seines aktuellen Zustands erforderlich. Dieser Abschnitt führt den aktuellen Verkehrszustand einer Knotenzufahrt auf zwei Größen zurück, die Zeitreihe der erwarteten Ankünfte und die Anzahl der abgefertigten Fahrzeuge. Unter Knotenzufahrten werden dabei die Abbiegespuren des betrachteten Knotenpunktes verstanden. Abbiegespuren mit gleicher Abbiegerichtung werden zu einer Zufahrt zusammengefasst.

3.1.1. Zeitreihe der erwarteten Ankünfte

Die erste der beiden Größen, die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$, soll Aufschluss über die zeitliche Verteilung der Fahrzeugankünfte an der Haltelinie einer Knotenzufahrt i geben. Sie gibt an, wie viele Fahrzeuge die Haltelinie bis zum Zeitpunkt t unter Annahme freien Verkehrs potenziell erreicht haben können. Das ist gleichbedeutend mit der Aussage, dass auf Knotenzufahrt i bis zum Zeitpunkt t höchstens $N_i^{\text{exp}}(t)$ Fahrzeuge abgefertigt werden können. Rückstaubedingte Verzögerungen bis zur tatsächlichen Ankunft an der Haltelinie sind hierbei explizit nicht mit zu erfassen. Damit lässt sich $N_i^{\text{exp}}(t)$ beispielsweise durch Verkehrsflussmessungen an Fahrbahnquerschnitten stromaufwärts der Haltelinie über einen gewissen Zeitraum prognostizieren. Abbildung 3.1 stellt das Prinzip grafisch dar.

In dieser Arbeit wird davon ausgegangen, dass diese Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ bekannt ist. Es wird ein Prognosehorizont benötigt, der so groß ist wie die Dauer der Bedienvorgänge, also der in der Größenordnung von etwa 30 Sekunden liegt.

Definition

Definiert sei die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ über folgendes Integral:

$$N_i^{\text{exp}}(t) = N_i^{\text{exp}}(t_0) + \int_{t_0}^t q_i^{\text{exp}}(t') dt'. \quad (3.1)$$

Hierbei bezeichnet $q_i^{\text{exp}}(t)$ die erwartete Ankunftsrate über alle Fahrspuren der betrachteten Knotenzufahrt. Je nachdem, ob der Verkehrsfluss mit mikroskopischen, makroskopischen, probabilistischen oder anderen Modellen beschrieben wird, kann der Verlauf von $N_i^{\text{exp}}(t)$ sowohl stetig als auch stufenförmig sein. Angenommen wird jedoch, dass die Werte der Ankunftsrate $q_i^{\text{exp}}(t)$ nicht größer als die Sättigungsverkehrsstärke q_i^{max} sind, d.h.

$$0 \leq q_i^{\text{exp}}(t) \leq q_i^{\text{max}}. \quad (3.2)$$

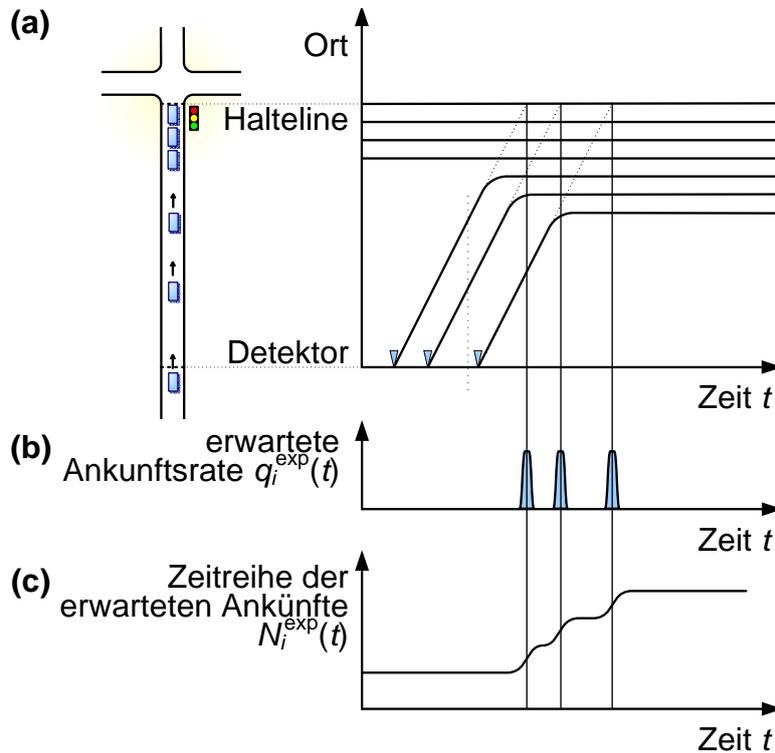


Abbildung 3.1.: (a) Fahrzeugtrajektorien auf Knotenzufahrt i . (b) Erwartete Ankunftsrate $q_i^{\text{exp}}(t)$ an der Haltelinie unter Annahme freien Verkehrs. (c) Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$. Sie gibt an, wie viele Fahrzeuge bis zum Zeitpunkt t potenziell abgefertigt werden können.

Damit hat die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ einen stetigen aber nicht notwendigerweise differenzierbaren Verlauf. Weil $q_i^{\text{exp}}(t)$ niemals negativ ist, wächst $N_i^{\text{exp}}(t)$ monoton in der Zeit. Weiterhin sei die mittlere Ankunftsrate \bar{q}_i^{exp} bekannt. Es wird angenommen, dass sich diese nur langsam ändert und auf der Zeitskala der Bedienvorgänge konstant ist.

Prognose der Ankünfte

Zur Verkehrsflusserfassung steht eine Vielzahl unterschiedlicher Technologien zur Verfügung. Während Induktivschleifen-, Ultraschall-, Radar- oder Infrarot-Detektoren [106, 3, 16] die Überfahrzeitpunkte an bestimmten Fahrbahnquerschnitten messen, lassen sich mit Videoerfassungssystemen [51] sogar Informationen zum räumlichen Verkehrsgeschehen gewinnen. Die Erfassung von Fahrzeugtrajektorien, sogenannter „Floating-Car-Daten“, auf Basis der aktiven Kommuni-

kation zwischen Fahrzeugen und Lichtsignalanlagen über WLAN wird derzeit im Rahmen des Projektes „Travolution“¹² in Ingolstadt erprobt [67, 183, 68].

Aus aktuellen Messdaten stromaufwärts der Haltelinie lassen sich unter Verwendung modellgestützter Prognoseverfahren zukünftige Werte der erwarteten Ankunftsrate $q_i^{\text{exp}}(t + \Delta t)$ abschätzen. Nimmt man im einfachsten Fall an, dass die Geschwindigkeit der Fahrzeuge im freien Verkehr der zulässigen Höchstgeschwindigkeit V_i^0 entspricht, ist die Reisezeit vom Messquerschnitt bis zur Haltelinie konstant. Die Prognose kann dann auf eine Zeitverschiebung zurückgeführt werden. Aus dem Verkehrsfluss $q_i^x(t)$ an einem Fahrbahnquerschnitt x Meter stromaufwärts der Haltelinie von Knotenzufahrt i erhält man als Prognose für die erwartete Ankunftsrate:

$$q_i^{\text{exp}}(t + \Delta t) = q_i^x(t) \quad \text{mit} \quad \Delta t = x/V_i^0. \quad (3.3)$$

Auf langen Streckenabschnitten können jedoch kleine Geschwindigkeitsdifferenzen zu einer Streuung der Reisezeiten führen, wie beispielsweise in den Referenzen [69, 196] durch empirische Untersuchungen festgestellt wurde. Dies lässt sich berücksichtigen, indem man die zukünftig erwartete Ankunftsrate $q_i^{\text{exp}}(t + \Delta t)$ durch eine Faltung der stromaufwärts gemessenen Datenreihe $q_i^x(t)$ mit der empirisch ermittelten Verteilung $p_i^x(D)$ der Reisezeiten D berechnet. Das unbestimmte Faltungsintegral lautet dann:

$$q_i^{\text{exp}}(t + \Delta t) = \int q_i^x(t') p_i^x(t + \Delta t - t') dt'. \quad (3.4)$$

Für $p_i^x(D)$ nimmt Pacey [138] eine Normalverteilung an. Das Berechnungsverfahren TRANSYT [154] verwendet dagegen eine geometrische Verteilung. Noch einen anderen Ansatz verfolgen Geroliminis und Skabardonis [66]. Sie prognostizieren das Ankunftsprofil mit Hilfe eines Markov-Prozesses. Trotz dieser Unterschiede treffen alle diese Arbeiten die Annahme, dass die Geschwindigkeit jedes einzelnen Fahrzeugs unabhängig von den anderen Fahrzeugen ist und folglich, dass sich die Fahrzeuge ungehindert überholen können. Bei fehlenden Überholmöglichkeiten zerfällt der Verkehrsstrom allerdings in mehrere Kolonnen, die jeweils von den langsameren Fahrzeugen geführt werden [131, 123, 119]. Um solche und auch andere Phänomene zu berücksichtigen, bei denen die Interaktion der Fahrzeuge eine wichtige Rolle spielt, ist die zeitliche Entwicklung des Verkehrsgeschehens modellgestützt zu prognostizieren. Ein makroskopisches Verkehrsflussmodell, das sich insbesondere zur Beschreibung des Verkehrsgeschehens um sogenannte Inhomogenitätsstellen wie Fahrbahnverengungen oder Einfädelpunkten eignet, stellen Helbing, der Autor dieser Dissertationsschrift und Lebacque in Referenz [79] vor.

¹²Vgl. Abschnitt 2.1.3.

Abbiegeanteile

Zur Prognose der erwarteten Ankunftsrate $q_i^{\text{exp}}(t + \Delta t)$ muss weiterhin bekannt sein, zu welchen Teilen sich der Verkehrsstrom einer Richtungsfahrbahn auf die nachfolgenden Knotenzufahrten verteilt. Hierzu ist eine Schätzung der Abbiegeanteile notwendig. Die Abbiegeanteile resultieren aus dem Routenwahlverhalten der Verkehrsteilnehmer [167, 163, 194, 109], welches insbesondere auch von den mittleren Wartezeiten an den Knotenpunkten abhängig ist. Unter Verwendung verkehrsadaptiver Steuerungsverfahren kann dadurch eine dynamische Wechselbeziehung zwischen Lichtsignalsteuerung und Routenwahlverhalten entstehen [122, 57]. Einen modellbasierten Ansatz zur dynamischen Schätzung der Abbiegeanteile stellen Helbing et al. in Referenz [79] vor. Andere Ansätze basieren auf rekurrenten neuronalen Netzen [116, 198] oder auf nichtlinearen Kalman-Filtern [103].

Prognosegenauigkeit

Im Folgenden wird die Genauigkeit verschiedener Prognosehorizonte untersucht. Wir betrachten die Knotenzufahrt einer Richtungsfahrbahn mit 500 Metern Länge und einer mittleren Reisegeschwindigkeit von 60 km/h.

- **5 Sekunden:** Innerhalb dieser Zeitspanne legen die Fahrzeuge etwa 80 Meter zurück. Es ist anzunehmen, dass sich die Fahrzeuge bereits auf der richtigen Abbiegespur eingefädelt haben. Die Streuung der Reisezeiten kann vernachlässigt werden. Für diesen Prognosehorizont ist die höchste Genauigkeit zu erwarten.
- **30 Sekunden:** Innerhalb dieser Zeitspanne legen die Fahrzeuge etwa 500 Meter zurück. Befindet sich der Messquerschnitt unmittelbar hinter dem vorgelagerten Knotenpunkt, ist zusätzlich zu berücksichtigen, dass die Fahrzeuge unter Umständen noch nicht auf ihre Wunschgeschwindigkeit beschleunigt haben. Außerdem ist es möglich, dass ein Teil des Verkehrsstroms in eventuell vorhandene Nebenstraßen abzweigt oder andere Verkehrsströme einmünden.
- **60 Sekunden:** Für diesen Prognosehorizont sind zusätzlich die Bedienvorgänge am vorgelagerten Knotenpunkt mit zu erfassen. Weil auch der vorgelagerte Knotenpunkt die Ankünfte an seinen Haltelinien prognostiziert und daraus eine Prognose der Bedienvorgänge¹³ ableiten kann, lässt sich der Zufluss in die betrachtete Richtungsfahrbahn um weitere 30 Sekunden prognostizieren. Dies setzt eine Kommunikation zwischen den Knotenpunkten voraus. Dieser erweiterte Prognosehorizont ist mit zusätzlichen Unsicherheiten verbunden.

¹³Vgl. Abschnitt 3.3

- **Mehr als 60 Sekunden:** Für größere Prognosehorizonte müsste der Nachbarknoten die Prognose des lokal betrachteten Knotenpunktes mit einbeziehen. Insofern nicht ausgeschlossen werden kann, dass diese wechselseitige Abhängigkeit die Prognose systematisch verfälscht, ist die erwartete Ankunftsrate mit ihrem Mittelwert \bar{q}_i^{exp} abzuschätzen.

Beim Reglerentwurf wird das Vorhandensein unsicherer Informationen explizit berücksichtigt. In Abschnitt 5.4.1 wird gezeigt, dass zur Garantie sowohl netzwerkweiter Stabilität als auch maximaler Sperrzeiten die Zuverlässigkeit folgender beiden Informationen ausreicht: (i) ob sich auf der Zufahrt eine Warteschlange befindet oder nicht und (ii) der Wert der mittleren Ankunftsrate \bar{q}_i^{exp} . Diese wesentlichen Aspekte des Reglerverhaltens, also Stabilität und maximale Sperrzeiten, werden von den Ungenauigkeiten der Prognose unabhängig sein.

3.1.2. Anzahl abgefertigter Fahrzeuge

Definition

Bei der zweiten Größe handelt es sich um die Anzahl der Fahrzeuge, die auf einer Knotenzufahrt i bis zum Zeitpunkt t tatsächlich abgefertigt worden sind. Sie wird mit $N_i^{\text{out}}(t)$ bezeichnet und über folgendes Integral definiert:

$$N_i^{\text{out}}(t) = N_i^{\text{out}}(t_0) + \int_{t_0}^t q_i^{\text{out}}(t') dt'. \quad (3.5)$$

Hierbei bezeichnet $q_i^{\text{out}}(t)$ die Ausflussrate zum Zeitpunkt t . Sie habe den gleichen Wertebereich wie die erwartete Ankunftsrate $q_i^{\text{exp}}(t)$ aus Gleichung (3.2):

$$0 \leq q_i^{\text{out}}(t) \leq q_i^{\text{max}}. \quad (3.6)$$

Damit ist auch die Zeitreihe $N_i^{\text{out}}(t)$ stetig und monoton wachsend.

Zustandsabhängigkeit der Ausflussrate

Der Wert der Ausflussrate $q_i^{\text{out}}(t)$ hängt vom aktuellen Freigabezustand $\gamma_i(t) \in \{0, 1\}$ (0 für Sperrung und 1 für Freigabe) ab. Es sei daran erinnert, dass der Freigabezustand einschließlich der Überfahrzeiten und ausschließlich der Anfahrverluste zu verstehen ist¹⁴. Es gilt:

$$q_i^{\text{out}}(t) = 0, \quad \text{falls } \gamma_i(t) = 0. \quad (3.7)$$

¹⁴Vgl. Abschnitt 2.3.1

Während der Freigabe ist weiterhin zu unterscheiden, ob auf Knotenzufahrt i eine Warteschlange vorhanden ist ($n_i(t) > 0$) oder nicht ($n_i(t) = 0$).¹⁵ Haben sich Fahrzeuge aufgestaut, so wird angenommen, dass diese bei Freigabe mit der konstanten maximalen Ausflussrate q_i^{\max} aus der Knotenzufahrt herausfahren [5, 4, 166], d.h.

$$q_i^{\text{out}}(t) = q_i^{\max}, \quad \text{falls } \gamma_i(t) = 1 \text{ und } n_i(t) > 0. \quad (3.8)$$

Diese Annahme ist ohne Weiteres erfüllt, wenn es sich bei den Fahrzeugen der Wartekolonnen um PKWs handelt. Anderenfalls ist jedes Fahrzeug mit einem Faktor entsprechend seiner Fahrzeugklasse zu gewichten.¹⁶ Zunächst wird davon ausgegangen, dass die Verkehrsströme im Wesentlichen aus PKWs bestehen und Gleichung (3.8) gültig ist. Außerdem sei angenommen, dass der Zufluss in die nachfolgenden Streckenabschnitte nicht durch Rückstaus oder Ähnliches behindert wird.

Nachdem die Warteschlange vollständig geleert ist, haben die während der verbleibenden Freigabezeit ankommenden Fahrzeuge freie Fahrt. Sie fahren zu exakt den Zeitpunkten aus der Knotenzufahrt heraus, zu denen ihre Ankunft an der Haltelinie prognostiziert wurde. Die Ausflussrate $q_i^{\text{out}}(t)$ ist somit identisch zur erwarteten Ankunftsrate $q_i^{\text{exp}}(t)$:

$$q_i^{\text{out}}(t) = q_i^{\text{exp}}(t), \quad \text{falls } \gamma_i(t) = 1 \text{ und } n_i(t) = 0. \quad (3.9)$$

In diesem Zustand ist die Zahl der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ gleich der Anzahl der abgefertigten Fahrzeuge $N_i^{\text{out}}(t)$. Das erlaubt einen regelmäßigen Abgleich der Integrationskonstanten in (3.1) und (3.5).

3.2. Warteschlangenmodell

Das im Folgenden vorgestellte deterministische Warteschlangenmodell erlaubt es, aus den bekannten Größen $N_i^{\text{exp}}(t)$ und $N_i^{\text{out}}(t)$ wichtige Kenngrößen des aktuellen Verkehrszustands abzuleiten. Diese sind beispielsweise die Anzahl der verzögerten Fahrzeuge, die Gesamtwartezeiten und der Kraftstoffverbrauch. Weil das verwendete Modell nur den zeitlichen Aspekt des Verkehrsgeschehens beschreibt, zählt es zu den sogenannten „Point-Queue-Models“ [132, 23, 133, 43].

3.2.1. Warteschlangenlänge

Anzahl der verzögerten Fahrzeuge

Bis zum Zeitpunkt t haben unter der Annahme freien Verkehrs $N_i^{\text{exp}}(t)$ Fahrzeuge die Haltelinie erreichen können. Von diesen sind $N_i^{\text{out}}(t)$ Fahrzeuge tatsächlich ab-

¹⁵Die Warteschlangenlänge $n_i(t)$ wird im Abschnitt 3.2.1 eingeführt.

¹⁶Die Gewichtung (und Priorisierung) einzelner Fahrzeuge wird in Ausblick 8.1.1 diskutiert.

gefertigt worden. Die Abfertigung der anderen Fahrzeuge wurde folglich verzögert. Die Anzahl der zum Zeitpunkt t auf Knotenzufahrt i verzögerten Fahrzeuge wird als Warteschlangenlänge $n_i(t)$ bezeichnet. Für sie soll Folgendes gelten:

$$\boxed{n_i(t) = N_i^{\text{exp}}(t) - N_i^{\text{out}}(t).} \quad (3.10)$$

Bei freiem Verkehr sind $N_i^{\text{exp}}(t)$ und $N_i^{\text{out}}(t)$ identisch und es gilt $n_i(t) = 0$.

Hybride Modellgleichung

Die zeitliche Entwicklung der Warteschlangenlänge $n_i(t)$ lässt sich in Verbindung mit den Gleichungen (3.7) bis (3.9) durch folgende zustandsabhängige Modellgleichung ausdrücken:

$$\boxed{\frac{dn_i}{dt} = \begin{cases} q_i^{\text{exp}}(t), & \text{falls } \gamma_i(t) = 0, & \textcircled{1} \\ q_i^{\text{exp}}(t) - q_i^{\text{max}}, & \text{falls } \gamma_i(t) = 1 \text{ und } n_i(t) > 0, & \textcircled{2} \\ 0, & \text{falls } \gamma_i(t) = 1 \text{ und } n_i(t) = 0. & \textcircled{3} \end{cases}} \quad (3.11)$$

Die Warteschlangenlänge $n_i(t)$ folgt in jedem der drei Zustände $\textcircled{1}$, $\textcircled{2}$ und $\textcircled{3}$ einer linearen Differenzialgleichung erster Ordnung. Weil die Linearität jedoch nur stückweise gegeben ist, ist das resultierende Systemverhalten nichtlinear. Aufgrund der Tatsache, dass sowohl kontinuierliche als auch diskrete Zustandsgrößen vorkommen, handelt es sich um ein hybrides dynamisches System [162, 99, 171, 107, 19]. Abbildung 3.2 zeigt den stückweise linearen Verlauf der Warteschlangenlänge $n_i(t)$ bei konstanten Ankunftsrate.

Bis auf einen Zustandsübergang können alle anderen von der Lichtsignalsteuerung durch den Freigabezustand $\gamma_i(t)$ direkt beeinflusst werden. Die Ausnahme bildet der Zustandsübergang von $\textcircled{2}$ nach $\textcircled{3}$, der bei Leerung der Warteschlange auftritt. Abschnitt 3.3.1 wird zeigen, wie sich der Zeitpunkt, zu dem die Warteschlange erstmals vollständig geleert ist, aus dem Verlauf der Zeitreihe $N_i^{\text{exp}}(t)$ prognostizieren lässt.

3.2.2. Wartezeiten

Gesamtwartezeiten

Die Gesamtwartezeiten $W_i(t)$, die auf der Knotenzufahrt i für alle Fahrzeuge bis zur Zeit t angefallen ist, wächst in jeder Zeiteinheit dt um die Anzahl der momentan verzögerten Fahrzeuge $n_i(t)$ an. Sie berechnet sich folglich über das Integral

$$\boxed{W_i(t) = W_i(t_0) + \int_{t_0}^t n_i(t') dt'.} \quad (3.12)$$

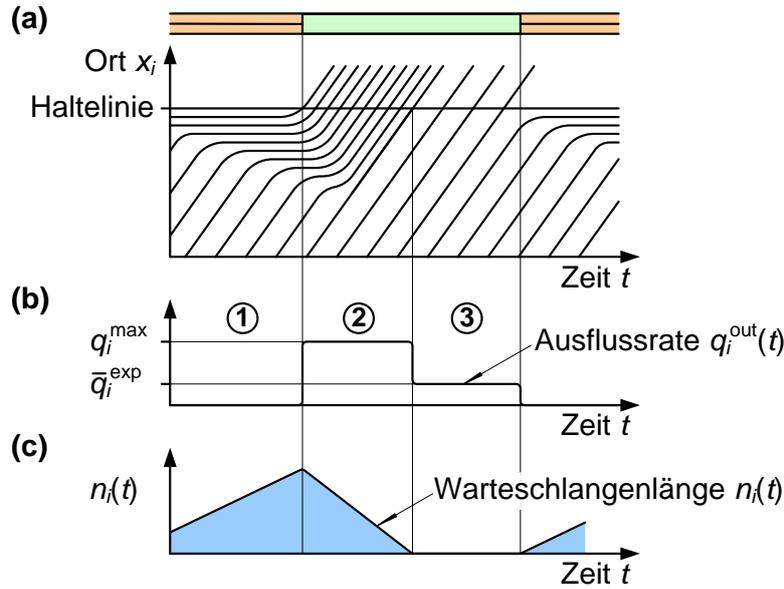


Abbildung 3.2.: (a) Fahrzeugtrajektorien bei konstanter Ankunftsrate. (b) Die Ausflussrate $q_i^{\text{out}}(t)$ ist zustandsabhängig. Im Zustand ‚Sperrung‘ ① ist sie null. Im Zustand ‚Freigabe‘ ist sie während des Abfließens der Wartekolonne ② maximal und anschließend ③ gleich der Ankunftsrate. (c) In jedem dieser drei Zustände lässt sich die Warteschlangenlänge durch eine lineare Differenzialgleichung erster Ordnung beschreiben. Weil die Linearität aber nur stückweise gegeben ist, handelt es sich um ein hybrides dynamisches System.

Individuelle Wartezeit

Die individuelle Wartezeit $w_i(t)$ des zum Zeitpunkt t an der Haltelinie befindlichen Fahrzeugs lässt sich durch einen Vergleich der beiden Zeitreihen $N_i^{\text{exp}}(t)$ und $N_i^{\text{out}}(t)$ ermitteln. Für $w_i(t)$ gilt die implizite Beziehung

$$N_i^{\text{exp}}(t - w_i(t)) = N_i^{\text{out}}(t). \quad (3.13)$$

Diskussion

Abbildung 3.3 stellt die in den Gleichungen (3.10), (3.12) und (3.13) formulierten Zusammenhänge grafisch in Beziehung zu den Fahrzeugtrajektorien. Es ist zu beachten, dass ein Fahrzeug nicht schon während des Abbremsvorgangs zur Warteschlangenlänge $n_i(t)$ beiträgt, sondern erst nachdem es unter der Annahme freien Verkehrs die Haltelinie erreicht. Der dadurch entstehende Fehler wird jedoch ausgeglichen, weil dasselbe Fahrzeug erst zum Zeitpunkt des Herausfahrens von der Warteschlangenlänge subtrahiert wird, und nicht schon zum Zeitpunkt des

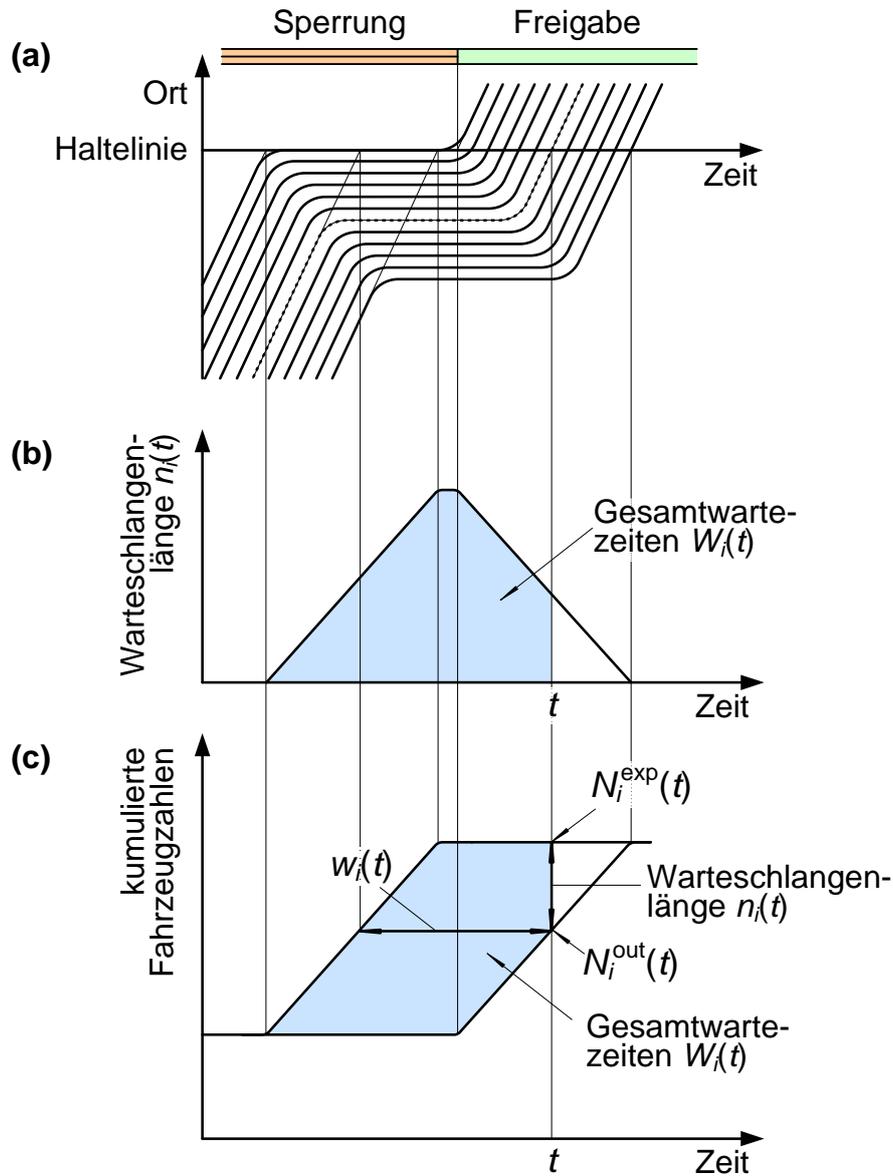


Abbildung 3.3.: (a) Fahrzeugtrajektorien auf einer Knotenzufahrt. (b) Als Warteschlangenlänge $n_i(t)$ wird nicht die Rückstaulänge, sondern die Anzahl verzögerten Fahrzeuge bezeichnet. Sie entspricht der Fahrzeuganzahl, die zwar bei freiem Verkehrs bereits die Haltelinie erreicht haben könnten, aber noch nicht abgefertigt worden sind. (c) Die Zeitreihen der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ und der abgefertigten Fahrzeuganzahl $N_i^{\text{out}}(t)$. Aus diesen lassen sich die Warteschlangenlänge $n_i(t)$, die Gesamtwartezeiten aller Fahrzeuge $W_i(t)$ und die individuelle Wartezeit des ersten Fahrzeugs der Warteschlange $w_i(t)$ berechnen.

Beschleunigen. Für die Gesamtwartezeiten der Fahrzeuge ist nicht entscheidend, wo oder wann sie abbremsen. Entscheidend sind allein die Zeitpunkte, zu denen sie frühestens abgefertigt werden können und zu denen sie tatsächlich abgefertigt werden [41].

Aspekte der Wartezeitenminimierung

Die Minimierung der Gesamtwartezeiten $W_i(t)$ ist das Ziel des Optimierungsverfahrens, das in Kapitel 4 vorgestellt wird. Dass bei gegebenen Ankunftsdaten $q_i^{\text{exp}}(t)$ der Zeitverlauf der Ausflussrate $q_i^{\text{out}}(t)$ eine wesentliche Rolle spielt, lässt sich durch folgende Betrachtung bereits an dieser Stelle erkennen. Den prinzipiellen Zusammenhang zwischen Ausflussrate und Gesamtwartezeiten findet man durch das Einsetzen der Definitionsgleichungen (3.1) und (3.5) in die Gleichungen (3.10) und (3.12). In der Schreibweise als unbestimmtes Integral erhält man:

$$W_i(t) = \iint q_i^{\text{exp}}(t) dt dt - \iint q_i^{\text{out}}(t) dt dt. \quad (3.14)$$

Die doppelte Integration deutet auf den besonderen Einfluss des Zeitverlaufs von $q_i^{\text{out}}(t)$ hin: Es kommt darauf an, die Ausflussrate so früh wie möglich zu erhöhen. Die Schwierigkeiten bei der Lösung des Optimierungsproblems liegen darin, dass die Ausflussrate den Abhängigkeiten (3.7) bis (3.9) unterliegt und während der Bedienung einer Zufahrt die anderen gesperrt sind. Bevor dieses Problem in Kapitel 4 erneut aufgegriffen werden kann, sind noch einige Vorbetrachtungen notwendig.

3.2.3. Kraftstoffverbrauch

Anzahl der Haltevorgänge

Eine zentrale Größe zur Abschätzung des Kraftstoffverbrauchs, aber auch zur Bewertung der Lärmemissionen und des Fahrkomforts, ist die Anzahl der Anhalte- bzw. Beschleunigungsvorgänge [165, 4, 82]. Für die Anzahl der Haltevorgänge $C_i^{\text{stop}}(t)$, die bis zur Zeit t auf dem Streckenabschnitt i stattfinden, lässt sich folgende Formel angeben:

$$C_i^{\text{stop}}(t) = C_i^{\text{stop}}(t_0) + \int_{t_0}^t c_1(t') + c_2(t') dt'. \quad (3.15)$$

Der erste Integrand $c_1(t)$ erfasst dabei die Fahrzeuge, die während des Herausfahrens aus der Knotenzufahrt beschleunigen müssen. Das sind all jene Fahrzeuge, die verzögert wurden und folglich unter der Bedingung $n_i(t) > 0$ herausfahren:

$$c_1(t) = \begin{cases} q_i^{\text{max}}, & \text{falls } \gamma_i(t) = 1 \text{ und } n_i(t) > 0 \\ 0, & \text{sonst.} \end{cases} \quad (3.16)$$

Der zweite Integrand $c_2(t)$ soll die Fälle berücksichtigen, in denen die Warteschlange während der Freigabe nicht vollständig geleert wird und aufgrund des Nachrückens der Wartekolonnen zusätzliche Anfahr- und Beschleunigungsmanöver hinzukommen. Hiervon sind die Fahrzeuge betroffen, die zu den Zeitpunkten t^{red} des Umschaltens von Freigabe auf Sperrung verzögert werden. Die Anzahl der Haltevorgänge C_i^{stop} erhöht sich dabei um den Betrag der momentanen Warteschlangenlänge $n_i(t)$. Dieser Sprung lässt sich im Integral aus Gleichung (3.15) mit

$$c_2(t) = \delta(t - t^{\text{red}}) n_i(t) \quad (3.17)$$

berücksichtigen, wobei $\delta(t)$ für den Dirac-Puls steht.

Gesamtkraftstoffverbrauch

Der Kraftstoffverbrauch kann in erster Näherung mit einer Linearkombination der Reiseweite, der Wartezeit und der Anzahl der Haltevorgänge berechnet werden. Da nur die beiden letzten Größen von der Lichtsignalsteuerung direkt beeinflusst werden können, wird die folgende Betrachtung auf diese beiden beschränkt. Ein durchschnittlicher Mittelklasse-Pkw hat einen Leerlaufverbrauch von einem Liter je Stunde und einen Mehrverbrauch von 0,02 Litern je Beschleunigungsvorgang [164, 166, 165]. Damit lässt sich der Gesamtkraftstoffverbrauch $C_i^{\text{fuel}}(t)$ aller Fahrzeuge auf Zufahrt i bis zur Zeit t , gemessen in Litern, wie folgt abschätzen:

$$C_i^{\text{fuel}}(t) = a_1 W_i(t) + a_2 C_i^{\text{stop}}(t). \quad (3.18)$$

Entsprechend der Verbrauchsangaben sind die Koeffizienten $a_1 = 28 \cdot 10^{-5}$ l/s und $a_2 = 0,02$ l zu wählen. Die Zahlenwerte ergeben, dass für einen Beschleunigungsvorgang genau so viel Kraftstoff verbraucht wird, wie für $a_2/a_1 \approx 70$ Sekunden Warten im Leerlauf.¹⁷

3.3. Prognose von Bedienvorgängen

Ein Bedienvorgang umfasst den Umschaltvorgang für die Dauer der Zwischenzeit τ_i^0 und die anschließende Freigabe. Das Verkehrsgeschehen eines zum Zeitpunkt t beginnenden bzw. fortgesetzten Bedienvorgangs zu prognostizieren ist das Ziel dieses Abschnittes. Begonnen wird mit der Berechnung der zur Warteschlangenleerung benötigten Freigabezeit \hat{g}_i und der dabei abzufertigende Fahrzeuganzahl \hat{n}_i . Diese beiden Größen werden anschließend in Abschnitt 3.3.2 auf ihre Zeit- und Zustandsabhängigkeit hin untersucht. In Abschnitt 3.3.3 werden die prognostizierten Gesamtwartezeiten \hat{W}_i genauer untersucht. Das Ergebnis ist, dass im Zustand

¹⁷Für die Bewertung des Fahrkomforts nimmt man für das Verhältnis a_2/a_1 den Wert 30 Sekunden an [26].

„keine Bedienung“ die Beziehung $d\hat{W}_i/dt = \hat{n}_i$ gilt. Ansonsten ist $d\hat{W}_i/dt = 0$. Auf diese grundlegenden Zusammenhänge wird im nächsten Kapitel ein heuristisches Optimierungsverfahren aufgebaut.

3.3.1. Erforderliche Freigabezeit zum Leeren der Warteschlange

Es soll nun berechnet werden, welche Freigabezeit \hat{g}_i aufzubringen ist, um die Warteschlange einer Knotenzufahrt i vollständig zu leeren. Dabei ist insbesondere zu berücksichtigen, dass während des Umschaltens und während der Abfertigung der bereits vorhandenen Warteschlange weitere Fahrzeuge hinzukommen können. Jedes zusätzliche Fahrzeug verlängert die benötigte Freigabezeit \hat{g}_i und erlaubt dadurch potenziell weiteren Fahrzeugen das Erreichen der Haltelinie.

Unmittelbar mit der benötigten Freigabezeit \hat{g}_i verbunden ist die Anzahl der abzufertigenden Fahrzeuge \hat{n}_i . Weil die Ausflussrate während der Abfertigung einer Warteschlange entsprechend Gleichung (3.8) als konstant angenommen wird, gilt der proportionale Zusammenhang

$$\hat{n}_i = \hat{g}_i q_i^{\max}. \quad (3.19)$$

Schnittpunktproblem

Um die Warteschlange auf Knotenzufahrt i vollständig zu leeren, muss sich der Bedienvorgang mindestens über das Zeitintervall $[t, t + \tau_i + \hat{g}_i]$ erstrecken. Hierbei bezeichnet $\tau_i \in [0, \tau_i^0]$ die noch verbleibende Umschaltzeit bis zum Beginn der Freigabe.

Zum Ende des minimal erforderlichen Bedienvorgangs, zum Zeitpunkt $t' = t + \tau_i + \hat{g}_i$, sind auf der betrachteten Knotenzufahrt $N_i^{\text{out}}(t) + \hat{g}_i q_i^{\max}$ Fahrzeuge abgefertigt worden. Wie viele Fahrzeuge bis dahin die Haltelinie erreicht haben, lässt sich direkt aus der Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t')$ ablesen. Zu dem Zeitpunkt t' , zu dem gleich viele Fahrzeuge abgefertigt worden wie angekommen sind, ist die Warteschlange vollständig geleert. Der gesuchte Wert für die erforderliche Freigabezeit \hat{g}_i muss demnach folgende Gleichung erfüllen:

$$\boxed{N_i^{\text{out}}(t) + \hat{g}_i q_i^{\max} = N_i^{\text{exp}}(t + \tau_i + \hat{g}_i)}. \quad (3.20)$$

Die linke Seite der Gleichung stellt eine Gerade, die rechte Seite eine als bekannt angenommene Zeitreihe bezüglich \hat{g}_i dar. Die Lösung der Gleichung lässt sich damit, wie in Abb. 3.4(a) zu sehen, auf die Bestimmung des Schnittpunkts der beiden Kurven zurückführen. Aufgrund der Eigenschaft¹⁸ $q_i^{\text{exp}}(t) \leq q_i^{\max}$ ist der Anstieg der Zeitreihe niemals größer als der Anstieg der Geraden. Folglich kann es nur

¹⁸Vgl. Gleichung (3.2)

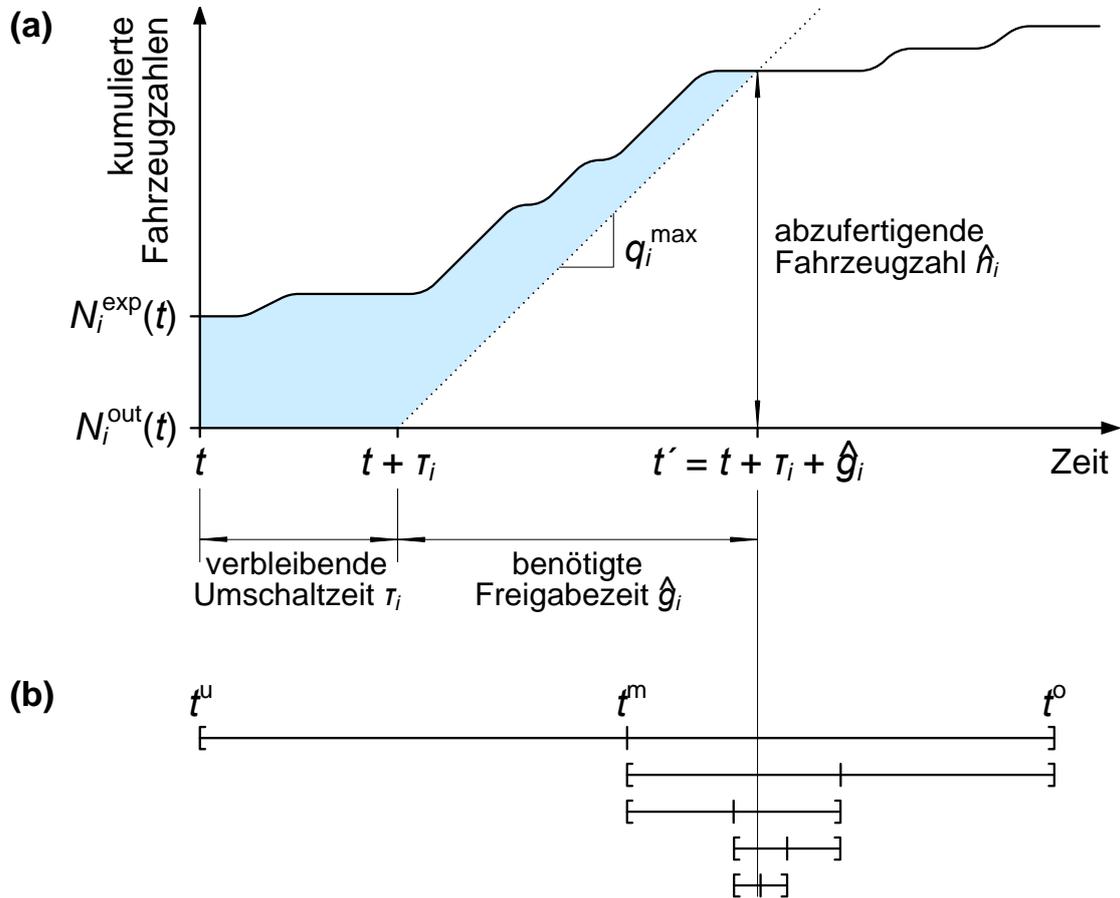


Abbildung 3.4.: Um die Warteschlange auf Knotenzufahrt i zu leeren, wird die Freigabezeit \hat{g}_i benötigt. (a) Die Berechnung von \hat{g}_i lässt sich auf die Lösung eines Schnittpunktproblems zurückführen. Nach Ablauf der verbleibenden Umschaltzeit τ_i werden die Fahrzeuge mit maximaler Rate $q_i^{\text{out}}(t) = q_i^{\text{max}}$ abgefertigt (gepunktete Gerade). Wurden genauso viele Fahrzeuge abgefertigt wie angekommen sind (Schnittpunkt mit durchgezogener Kurve), ist die Warteschlange vollständig geleert. (b) Das Schnittpunktproblem lässt sich numerisch beispielsweise durch iterative Halbierung des Suchintervalls $[t^u, t^o]$ lösen.

einen einzigen Schnittpunkt bzw. einen einzigen zusammenhängenden Schnittbereich geben. Ein Schnittbereich tritt dann auf, wenn zum Zeitpunkt t' der Warteschlangenleerung die Zuflussrate maximal ist. Es wird festgelegt, dass in diesem Fall für \hat{g}_i die obere Grenze des Schnittbereichs zu wählen ist. Durch Eingrenzungsverfahren, beispielsweise durch Anwendung der Newton-Raphsonschen Methode [46] oder durch Bisektion [175] lässt sich \hat{g}_i numerisch effizient bestimmen.

Numerisches Lösungsverfahren

Am Beispiel der iterativen Halbierung des Suchintervalls soll kurz ein Bisektionsverfahren zur numerischen Lösung von Gleichung (3.20) vorgestellt werden. Abbildung 3.4(b) skizziert den Lösungsweg. Das Verfahren besteht aus drei Schritten:

1. Man beginnt mit einem hinreichend großen Suchintervall $[t^u, t^o]$, das den Zeitpunkt t' der Warteschlangenleerung mit Sicherheit enthält:

$$t^u \leq t' \leq t^o. \quad (3.21)$$

2. Die Mitte des Suchintervalls liegt bei $t^m = (t^u + t^o)/2$. Ob t^m größer oder kleiner als t' ist, lässt sich durch Einsetzen von t^m in die beiden Seiten von Gleichung (3.20), wie gleich gezeigt, feststellen. Falls $t^m \leq t'$, wird die untere Intervallgrenze auf t^m gesetzt, ansonsten die obere. Die Rechenvorschrift lautet:

$$t^u := t^m, \quad \text{falls } N_i^{\text{out}}(t) + (t^m - t - \tau_i) q_i^{\text{max}} \leq N_i^{\text{exp}}(t^m) \quad \text{und} \quad (3.22)$$

$$t^o := t^m, \quad \text{falls } N_i^{\text{out}}(t) + (t^m - t - \tau_i) q_i^{\text{max}} > N_i^{\text{exp}}(t^m). \quad (3.23)$$

3. Der zweite Schritt ist solange zu wiederholen, bis die Breite $t^o - t^u$ des Suchintervalls eine vorher festgelegte Fehlerschranke $\epsilon > 0$ unterschreitet. Den gesuchten Wert für \hat{g}_i erhält man schließlich mit

$$\hat{g}_i = t^m - t - \tau_i. \quad (3.24)$$

Da sich das Suchintervall in jedem Iterationsschritt halbiert, nimmt seine Breite mit jedem Iterationsschritt exponentiell ab.

3.3.2. Eigenschaften der prognostizierten Kenngrößen

Im vorherigen Abschnitt wurden die benötigte Freigabezeit \hat{g}_i und die abzufertigende Fahrzeuganzahl \hat{n}_i bis zur vollständigen Leerung der Warteschlange eingeführt und ein numerisches Verfahren zu ihrer Berechnung vorgestellt. Nun soll

untersucht werden, wie sich diese Größen im Laufe der Zeit und unter verschiedenen Schaltzuständen verhalten. Dazu werden die beiden Größen

$$\hat{g}_i(t) \quad \text{und} \quad \hat{n}_i(t) \quad (3.25)$$

von nun an als Funktionen des Zeitpunktes t betrachtet, zu dem sie berechnet werden. Das Vorgehen ist zweigeteilt: Zuerst wird ein allgemeiner Ausdruck für ihre Zeitableitungen bestimmt. Anschließend wird dieser für die möglichen Schaltzustände spezifiziert und diskutiert.

Zeitabhängigkeit

Aufgrund der direkten Proportionalität $\hat{n}_i(t) = \hat{g}_i(t)q_i^{\max}$ nach Gleichung (3.19) kann die folgende Untersuchung auf die Betrachtung von $\hat{g}_i(t)$ beschränkt werden. Zur Bestimmung der Zeitableitung $d\hat{g}_i/dt$ wird zunächst die Bestimmungsgleichung (3.20) durch Einsetzen der Definitionsgleichungen (3.1) und (3.5) für N_i^{exp} und N_i^{out} in folgende Form überführt:

$$N_i^{\text{out}}(t_0) + \int_{t_0}^t q_i^{\text{out}}(t') dt' + \hat{g}_i q_i^{\max} = N_i^{\text{exp}}(t_0) + \int_{t_0}^{t+\tau_i(t)+\hat{g}_i(t)} q_i^{\text{exp}}(t') dt'. \quad (3.26)$$

Beide Seiten dieser Gleichung lassen sich nun nach t ableiten, wobei für die Ableitung der Integrale folgende Regel anzuwenden ist:

$$\frac{d}{dt} \int_{t_0}^{b(t)} f(t') dt' = f(b(t)) \frac{db}{dt}. \quad (3.27)$$

Beim Integral der linken Seite von (3.26) ist $b(t)$ durch t und $f(t)$ durch $q_i^{\text{out}}(t)$ zu substituieren, beim Integral der rechten Seite $b(t)$ durch $t + \tau_i(t) + \hat{g}_i(t)$ und $f(t)$ durch $q_i^{\text{exp}}(t)$. Man erhält zunächst die Gleichung:

$$q_i^{\text{out}}(t) + \frac{d\hat{g}_i}{dt} q_i^{\max} = q_i^{\text{exp}}(t + \tau_i(t) + \hat{g}_i(t)) \left(1 + \frac{d\tau_i}{dt} + \frac{d\hat{g}_i}{dt} \right). \quad (3.28)$$

Diese lässt sich anschließend mit den Ersetzungen $t' = t + \tau_i(t) + \hat{g}_i(t)$ und $\dot{\tau}_i(t) = d\tau_i/dt$ nach $d\hat{g}_i/dt$ auflösen:

$$\boxed{\frac{d\hat{g}_i}{dt} = \frac{(1 + \dot{\tau}_i(t))q_i^{\text{exp}}(t') - q_i^{\text{out}}(t)}{q_i^{\max} - q_i^{\text{exp}}(t')}}. \quad (3.29)$$

Die Eigenschaften dieser nichtlinearen Entwicklungsgleichung werden nun gesondert für die vier Zustände „keine Bedienung“, „Umschalten“, „Freigabe mit Warteschlange“ und „Freigabe ohne Warteschlange“ untersucht.

Zustandsabhängigkeit

Der zustandsabhängige Wert der Ausflussrate $q_i^{\text{out}}(t)$ ist durch die Gleichungen (3.7) bis (3.9) gegeben. Weiterhin ist auch der zustandsabhängige Verlauf der noch verbleibenden Umschaltzeit $\tau_i(t)$ bekannt: Solange die Bedienung noch nicht begonnen hat, entspricht ihr Wert der Zwischenzeit τ_i^0 . Während des Umschaltens fällt $\tau_i(t)$ mit $\dot{\tau}_i(t) = -1$ kontinuierlich auf den Wert null ab, den sie während der Freigabe behält. Am Ende des Bedienvorgangs springt $\tau_i(t)$ wieder auf τ_i^0 zurück. Ihren Verlauf zeigt Abb. 3.5.

Damit sind $q_i^{\text{out}}(t)$, $\tau_i(t)$ und $\dot{\tau}_i(t)$ in jedem der vier Zustände bekannt. Betrachtet wird nun, was daraus für die Zustandsabhängigkeit von \hat{g}_i folgt:

- (i) **Keine Bedienung:** Mit $q_i^{\text{out}}(t) = 0$, $\tau_i(t) = \tau_i^0$ und $\dot{\tau}_i(t) = 0$ vereinfacht sich Gleichung (3.29) zu:

$$\frac{d\hat{g}_i}{dt} = \frac{q_i^{\text{exp}}(t')}{q_i^{\text{max}} - q_i^{\text{exp}}(t')}. \quad (3.30)$$

Die zum Leeren der Warteschlange benötigte Freigabezeit ist wegen $q_i^{\text{exp}} \geq 0$ monoton wachsend. Kommt zum prognostizierten Zeitpunkt der Leerung t' eine Fahrzeugkolonne mit maximaler Rate $q_i^{\text{exp}}(t') = q_i^{\text{max}}$ an der Haltelinie an, springt $\hat{g}_i(t)$ zum Zeitpunkt t auf einen größeren Wert. Wie aus den Betrachtungen des Abschnitts 3.3.1 folgt, ist der Sprung im Verlauf von $\hat{g}_i(t)$ so groß, wie die zusätzlich benötigte Freigabezeit zur Abfertigung dieser Fahrzeugkolonne. Ebenso entspricht die Höhe des Sprungs im Verlauf von $\hat{n}_i(t)$ genau der Fahrzeuganzahl der Kolonne.

Die Zeitspanne $t' - t$ zwischen dem Sprung von $\hat{g}_i(t)$ und der Ankunft der Kolonne ist so groß, wie zum Umschalten und zur Abfertigung der bereits wartenden und der noch vor der Kolonne ankommenden Fahrzeuge nötig ist. Diesen Sachverhalt verdeutlicht Abb. 3.6 anhand von Fahrzeugtrajektorien im Weg-Zeit-Diagramm.

- (ii) **Umschalten:** Mit $q_i^{\text{out}}(t) = 0$ und $\dot{\tau}_i(t) = -1$ verschwinden beide Summanden des Zählers von Gleichung (3.29). Es gilt:

$$\frac{d\hat{g}_i}{dt} = 0. \quad (3.31)$$

Die vor dem Umschalten prognostizierte Freigabezeit $\hat{g}_i(t)$ ändert sich also während des Umschaltvorgangs nicht. Hierin drückt sich aus, dass der

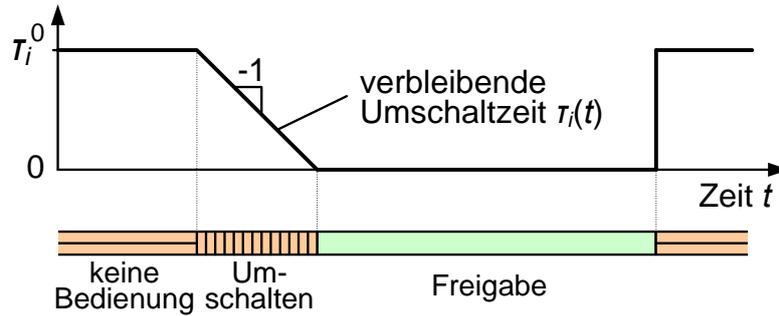


Abbildung 3.5.: Zustandsabhängiger Verlauf der verbleibenden Umschaltzeit $\tau_i(t)$. Während des Umschaltens fällt sie kontinuierlich auf null ab und springt am Ende des Bedienvorgangs wieder auf τ_i^0 zurück.

zum Freischalten notwendige Umschaltvorgang von der Prognose vollständig erfasst ist.

- (iii) **Freigabe mit Warteschlange:** Mit $q_i^{\text{out}}(t) = q_i^{\text{max}}$ und $\dot{\tau}_i(t) = 0$ vereinfacht sich Gleichung (3.29) zu

$$\frac{d\hat{g}_i}{dt} = \frac{q_i^{\text{exp}}(t') - q_i^{\text{out}}(t)}{q_i^{\text{max}} - q_i^{\text{exp}}(t')} = -1. \quad (3.32)$$

In Worten: Mit jeder Sekunde, in der die Warteschlange abgebaut wird, nimmt die verbleibende Zeit bis zur Leerung um eine Sekunde ab. Deshalb ändert sich der Zeitpunkt t' , der noch vor Beginn des Bedienvorgangs prognostiziert wurde, auch während der Abfertigung der Warteschlange nicht. Die in dieser Zeit neu hinzukommenden Fahrzeuge wurden von der Prognose vollständig erfasst.

- (iv) **Freigabe ohne Warteschlange:** Mit $t' = t$, $q_i^{\text{out}}(t) = q_i^{\text{exp}}(t)$ und $d\tau_i/dt = 0$ erhält man:

$$\frac{d\hat{g}_i}{dt} = \frac{0}{q_i^{\text{max}} - q_i^{\text{exp}}(t)}. \quad (3.33)$$

Der Nenner verschwindet sobald eine Fahrzeugkolonne (oder ein einzelnes Fahrzeug) mit maximaler Flussrate ankommt. In diesem Fall ist die Zeitableitung $d\hat{g}_i/dt$ unbestimmt und die Bestimmungsgleichung (3.20) hat mehrere Lösungen. Durch die Festlegung,¹⁹ dass für $\hat{g}_i(t)$ der größtmögliche Wert zu wählen ist, erfährt $\hat{g}_i(t)$ in diesem Fall einen Sprung. Der Zustand maximaler Ausflussrate $q_i^{\text{out}}(t) = q_i^{\text{max}}$ ist äquivalent zum Zustand (iii) der Warteschlangenleerung. Dieser Fall ist in Abb. 3.6 rechts unten zu sehen.

¹⁹Vgl. Abschnitt 3.3.1, insbesondere S. 53

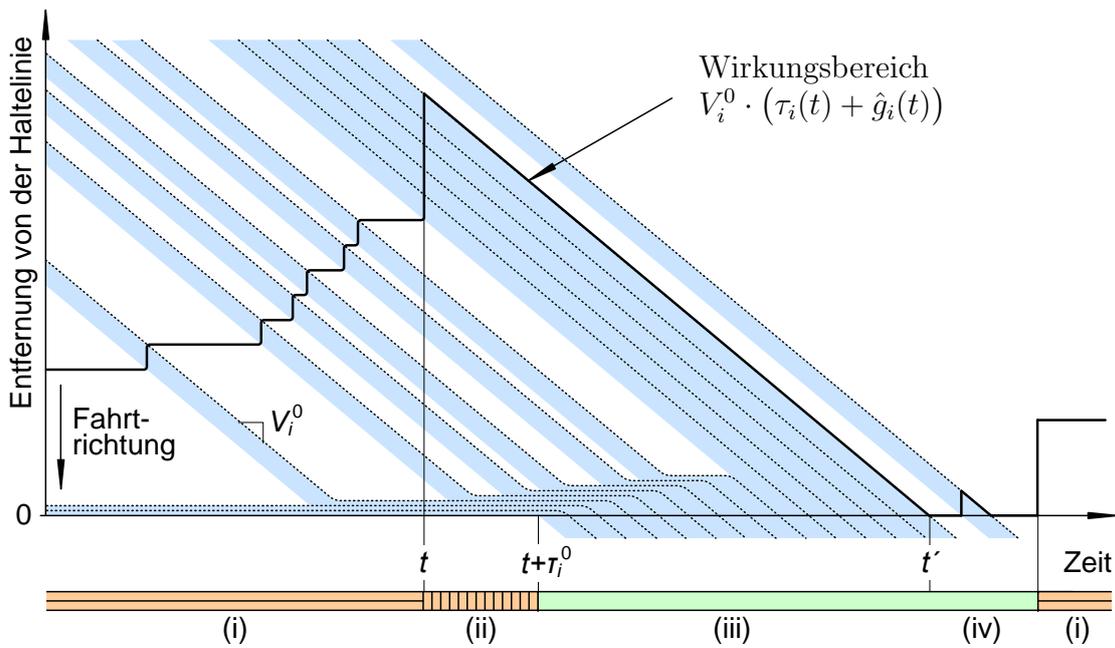


Abbildung 3.6.: Zustandsabhängiger Wirkungsbereich. Die in diesem Bereich erfassten Fahrzeuge können im nächsten Bedienvorgang mit maximaler Rate abgefertigt werden. Seine stromaufwärtige Grenze (dicke Linie) ist die mit der Reisegeschwindigkeit V_i^0 multiplizierte Zeitspanne $\tau_i(t) + \hat{g}_i(t)$ bis zur Warteschlangenleerung. Mit jedem Fahrzeug, das diese Grenze überfährt, rückt die Bereichsgrenze weiter stromaufwärts. Bei einer Kolonne mit maximaler Flussrate erfährt die Bereichsgrenze einen größeren Sprung. Die Zeitspanne $t' - t$ zwischen dem Sprung und der Ankunft der Kolonne an der Haltelinie so groß ist, wie zum Umschalten und zur Abfertigung der bereits wartenden und der noch vor der Kolonne ankommenden Fahrzeuge benötigt wird. Die Fahrzeuge, die zum Beginn des Bedienvorgangs die Bereichsgrenze überfahren, erhalten freie Fahrt.

Beim Abbruch des Bedienvorgangs springt $\tau_i(t)$ zurück auf τ_i^0 . Aufgrund dieser Unstetigkeit ist die Ableitung $d\hat{g}_i/dt$ beim Umschalten auf Sperrung nicht definiert. Dieser Zustandsübergang wird in Abschnitt 3.3.4 genauer untersucht.

3.3.3. Prognostizierte Gesamtwartezeit

Der in den beiden vorherigen Abschnitten vorgestellte Ansatz erlaubt es, den Verkehrsablauf während des gesamten Bedienvorgangs zu prognostizieren. Deshalb können, noch bevor der Bedienvorgang für eine Zufahrt i begonnen hat, die zur Warteschlangenleerung benötigte Freigabezeit $\hat{g}_i(t)$ und die Anzahl der abzufertigenden Fahrzeuge $\hat{n}_i(t)$ berechnet werden. Als dritte und letzte Kenngröße sollen nun die Gesamtwartezeiten $\hat{W}_i(t)$, die bis zum Ende des Bedienvorgangs anfallen, prognostiziert werden.

Dies soll in vier Schritten geschehen. Zuerst wird eine Definitionsgleichung für $\hat{W}_i(t)$ hergeleitet, die anschließend in eine zeitliche Entwicklungsgleichung $d\hat{W}_i/dt$ überführt wird. Im dritten Schritt wird diese wieder in verschiedenen Schaltzuständen untersucht. Zuletzt wird die Unstetigkeit von $\hat{W}_i(t)$ beim Abbruch des Bedienvorgangs betrachtet.

Definition

Als prognostizierte Gesamtwartezeiten $\hat{W}_i(t)$ werden die Gesamtwartezeiten aller Fahrzeuge auf Knotenzufahrt i bis zum Ende eines zum Zeitpunkt t beginnenden bzw. fortgesetzten Bedienvorgangs verstanden. Es sei zunächst angenommen, dass dabei die Warteschlange vollständig geleert wird. Der vorzeitige Abbruch des Bedienvorgangs wird weiter unten genauer betrachtet. Nachdem die Warteschlange vollständig geleert wurde, können keine zusätzlichen Wartezeiten entstehen. Deshalb interessiert im Folgenden nur die Zeitspanne bis zur Leerung.

In Abb. 3.7 sind die prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ als Flächeninhalt des Bereichs zwischen den beiden bekannten Zeitreihen $N_i^{\text{exp}}(t)$ und $N_i^{\text{out}}(t)$ dargestellt. Dieser Flächeninhalt wird zur Definition von $\hat{W}_i(t)$ in drei Teile zerlegt:

$$\hat{W}_i(t) = A(t) + B(t) + C(t). \quad (3.34)$$

Der erste Summand $A(t)$ sei die bis zum Zeitpunkt t angefallene Gesamtwartezeit $W_i(t)$. Entsprechend den Gleichungen (3.10) und (3.12) ist diese definiert als

$$A(t) = W_i(t) = W_i(t_0) + \int_{t_0}^t N_i^{\text{exp}}(t') - N_i^{\text{out}}(t') dt'. \quad (3.35)$$

Der zweite Summand $B(t)$ umfasst die Gesamtwartezeiten während des Umschaltens bis zum Beginn der Freigabe. Sie entspricht dem Integral der Warteschlan-

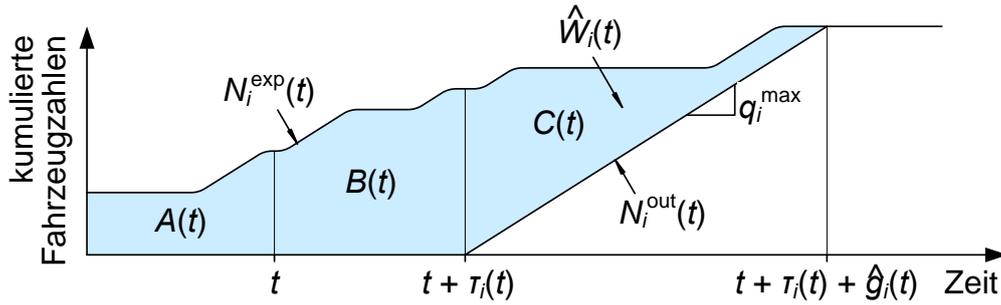


Abbildung 3.7.: Die prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ sind die Wartezeiten aller Fahrzeuge bis zum Ende eines zum Zeitpunkt t beginnenden bzw. fortgesetzten Bedienvorgangs.

genlänge $n_i(t)$ im Zeitraum $[t, t + \tau_i(t)]$. Es ist zu beachten, dass sich die Anzahl der abgefertigten Fahrzeuge $N_i^{out}(t)$ in diesem Zeitraum nicht ändert. Damit gilt:

$$B(t) = \int_t^{t+\tau_i(t)} N_i^{exp}(t') - N_i^{out}(t) dt'. \quad (3.36)$$

Der dritte Summand $C(t)$ umfasst die Gesamtwartezeiten während der Freigabe. Während die Warteschlange im Zeitraum $[t + \tau_i, t + \tau_i + \hat{g}_i]$ geleert wird, wächst die Anzahl der abgefertigten Fahrzeuge $N_i^{out}(t')$ mit der konstanten Rate q_i^{max} an. Sie lässt sich deshalb durch eine Geradengleichung (eckige Klammer) beschreiben:

$$C(t) = \int_{t+\tau_i(t)}^{t+\tau_i(t)+\hat{g}_i(t)} N_i^{exp}(t') - \left[N_i^{out}(t) + (t' - (t + \tau_i(t))) q_i^{max} \right] dt'. \quad (3.37)$$

Mit den drei Summanden $A(t)$, $B(t)$ und $C(t)$ ist die Gleichung (3.34) für die prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ vollständig bekannt. Diese soll nun bezüglich ihrer Zeit- und Zustandsabhängigkeit ausgewertet werden.

Zeitabhängigkeit

Die Zeitableitung der prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ ist die Summe der Zeitableitungen der drei Summanden:

$$\frac{d\hat{W}_i(t)}{dt} = \frac{dA(t)}{dt} + \frac{dB(t)}{dt} + \frac{dC(t)}{dt}. \quad (3.38)$$

Bei der Zeitableitung d/dt der Integrale ist zu beachten, dass t sowohl in den Integrationsgrenzen als auch im Integranden vorkommen kann. Deshalb verwenden wir folgende Ableitungsregel:

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(t, t') dt' = f(t, b(t)) \frac{db}{dt} - f(t, a(t)) \frac{da}{dt} + \int_{a(t)}^{b(t)} \frac{\partial f(t, t')}{\partial t} dt'. \quad (3.39)$$

Es soll zunächst die Ableitung des ersten Summanden $A(t)$ aus Gleichung (3.35) betrachtet werden. Mit den Substitutionen $a(t) = t_0$, $b(t) = t$ und $f(t, t') = N_i^{\text{exp}}(t') - N_i^{\text{out}}(t')$ verschwinden die Zeitableitungen da/dt und $\partial f(t, t')/\partial t$. Man erhält:

$$\frac{dA(t)}{dt} = \left[\underbrace{N_i^{\text{exp}}(t) - N_i^{\text{out}}(t)}_{= n_i(t) \text{ nach (3.10)}} \right] \underbrace{\frac{dt}{dt}}_{=1}. \quad (3.40)$$

Für die Ableitung von $B(t)$ aus Gleichung (3.36) sind die Substitutionen $a(t) = t$, $b(t) = t + \tau_i(t)$ und $f(t, t') = N_i^{\text{exp}}(t') - N_i^{\text{out}}(t')$ vorzunehmen. Man erhält:

$$\begin{aligned} \frac{dB(t)}{dt} &= \left[N_i^{\text{exp}}(t + \tau_i(t)) - N_i^{\text{out}}(t) \right] \underbrace{\frac{d(t + \tau_i(t))}{dt}}_{= 1 + \dot{\tau}_i(t)} \\ &\quad - \left[\underbrace{N_i^{\text{exp}}(t) - N_i^{\text{out}}(t)}_{= n_i(t) \text{ nach (3.10)}} \right] \underbrace{\frac{dt}{dt}}_{=1} \\ &\quad + \underbrace{\int_t^{t+\tau_i(t)} -q_i^{\text{out}}(t) dt'}_{= -\tau_i(t) q_i^{\text{out}}(t)}. \end{aligned} \quad (3.41)$$

Nach dem gleichen Schema erhält man für die Ableitung von $C(t)$:

$$\begin{aligned} \frac{dC(t)}{dt} &= \left[\underbrace{N_i^{\text{exp}}(t + \tau_i(t) + \hat{g}_i(t)) - N_i^{\text{out}}(t)}_{= \hat{n}_i(t) \text{ nach (3.20)}} - \underbrace{\hat{g}_i(t) q_i^{\text{max}}}_{= \hat{n}_i(t) \text{ nach (3.19)}} \right] \frac{d(t + \tau_i(t) + \hat{g}_i(t))}{dt} \\ &\quad - \left[N_i^{\text{exp}}(t + \tau_i(t)) - N_i^{\text{out}}(t) \right] \underbrace{\frac{d(t + \tau_i(t))}{dt}}_{= 1 + \dot{\tau}_i(t)} \\ &\quad + \underbrace{\int_{t+\tau_i(t)}^{t+\tau_i(t)+\hat{g}_i(t)} -q_i^{\text{out}}(t) + q_i^{\text{max}} \frac{d(t + \tau_i(t))}{dt} dt'}_{= \hat{n}_i(t) (1 + \dot{\tau}_i(t)) - \hat{g}_i(t) q_i^{\text{out}}(t)}. \end{aligned} \quad (3.42)$$

Hierbei ist der erste Summand gleich null.

Für die Berechnung von $d\hat{W}_i/dt$ sind die drei Ableitungen zu addieren, wobei vier weitere Terme wegfallen. Die rechte Seite von (3.40) kürzt sich mit dem zweiten Summanden aus (3.41), der erste Summand aus (3.41) kürzt sich mit dem zweiten aus (3.42). Übrig bleiben die jeweils dritten Summanden aus (3.41) und (3.42):

$$\boxed{\frac{d\hat{W}_i}{dt} = \hat{n}_i(t) (1 + \dot{\tau}_i(t)) - q_i^{\text{out}}(t) (\tau_i(t) + \hat{g}_i(t))}. \quad (3.43)$$

Damit ist es gelungen, eine Gleichung für die zeitliche Entwicklung der prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ herzuleiten. Die besondere Bedeutung dieser Gleichung wird klar, wenn im Folgenden die Schaltzustände mit einbezogen werden.

Zustandsabhängigkeit

Für jeden Schaltzustand ist der Verlauf der Ausflussrate $q_i^{\text{out}}(t)$ und der verbleibenden Zwischenzeit $\tau_i(t)$ bekannt.²⁰ Darauf aufbauend kann für die einzelnen Zustände Folgendes festgestellt werden:

- (i) **Keine Bedienung:** Mit $q_i^{\text{out}}(t) = 0$ und $\dot{\tau}_i(t) = 0$ vereinfacht sich Gleichung (3.43) zu:

$$\frac{d\hat{W}_i}{dt} = \hat{n}_i(t). \quad (3.44)$$

Daraus kann eine wesentliche Aussage abgeleitet werden: Das Herauszögern eines Bedienvorgangs um eine kleine Zeitspanne dt wird die Wartezeiten aller Fahrzeuge um den Betrag $\hat{n}_i(t)dt$ erhöhen.

- (ii) **Umschalten:** Ist $q_i^{\text{out}}(t) = 0$ und $\dot{\tau}_i(t) = -1$, so verschwinden beide Summanden von Gleichung (3.43) und man erhält:

$$\frac{d\hat{W}_i}{dt} = 0. \quad (3.45)$$

- (iii) **Freigabe mit Warteschlange:** Mit $q_i^{\text{out}}(t) = q_i^{\text{max}}$, $\tau_i(t) = 0$ und $\dot{\tau}_i(t) = 0$ kürzen sich beide Summanden aus Gleichung (3.43) heraus:

$$\frac{d\hat{W}_i}{dt} = 0. \quad (3.46)$$

²⁰Vgl. Abschnitt 3.3.2

- (iv) **Freigabe ohne Warteschlange:** Für $\hat{g}_i(t) = 0$ und $\tau_i(t) = 0$ verschwinden wiederum beide Summanden:

$$\frac{d\hat{W}_i}{dt} = 0. \quad (3.47)$$

Aus den Punkten (ii) bis (iv) lässt sich schlussfolgern: Während des gesamten Bedienvorgangs ändert sich der Wert der prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ nicht. Sämtliche Wartezeiten wurden bereits vor Beginn des Bedienvorgangs erfasst. Dies lässt sich zusammenfassen mit:

$$\frac{d\hat{W}_i}{dt} = \begin{cases} \hat{n}_i(t), & \text{falls „keine Bedienung“,} \\ 0, & \text{während des gesamten Bedienvorgangs,} \\ \text{n.d.,} & \text{beim Abbruch des Bedienvorgangs.} \end{cases} \quad (3.48)$$

Beim Abbruch des Bedienvorgangs ist die Ableitung $d\hat{W}_i/dt$ wegen der Unstetigkeit von $\tau_i(t)$ nicht definiert (n.d.). Dieser Zustandsübergang wird nun genauer betrachtet.

3.3.4. Abbruch eines Bedienvorgangs

Vorbetrachtung

Die prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ wachsen, solange die betrachtete Zufahrt nicht bedient wird, mit der Rate $\hat{n}_i(t)$. Damit lassen sich die Wartezeiten nicht erfassen, die im hypothetischen Grenzfall „keine Bedienung für *null* Zeiteinheiten“ allein durch das erneute Abwarten der Zwischenzeit τ_i^0 zusätzlich entstehen. Um exakt den Betrag dieser zusätzlichen Wartezeiten muss $\hat{W}_i(t)$ beim Abbruch des Bedienvorgangs anwachsen.

Aus dieser Überlegung kann zunächst Folgendes abgeleitet werden: Die prognostizierten Gesamtwartezeiten $\hat{W}_i(t)$ wachsen beim Abbruch des Bedienvorgangs um den Betrag

$$\Delta W_i^0(t) = \hat{W}_i^{(2)}(t) - \hat{W}_i^{(1)}(t), \quad (3.49)$$

wobei für die beiden Summanden der rechten Seite Folgendes gilt:

- (1) Wird der aktuelle Bedienvorgang fortgesetzt, so entstehen bis zur Leerung der Warteschlange die Gesamtwartezeiten $\hat{W}_i^{(1)}(t)$.
- (2) Wird der Bedienvorgang zum Zeitpunkt t unterbrochen und sofort wieder begonnen, so entstehen bis zur Leerung der Warteschlange die Gesamtwartezeiten $\hat{W}_i^{(2)}(t)$.

Zwar lassen sich $\hat{W}_i^{(1)}(t)$ und $\hat{W}_i^{(2)}(t)$ mit der Formel (3.34) berechnen, aber für den Reglerentwurf in Kapitel 4 ist das noch nicht ausreichend. Deshalb wird nun eine Entwicklungsgleichung für $\Delta W_i^0(t)$ hergeleitet.

Entwicklungsgleichung für Zunahme der Gesamtwartezeiten

Hierzu wird erneut Gleichung (3.43) herangezogen. Weiterhin ist es erforderlich, die prognostizierte Freigabezeit und die abzufertigende Fahrzeuganzahl

$$\hat{g}_i(t, \tau_i(t)) \quad \text{und} \quad \hat{n}_i(t, \tau_i(t)) \quad (3.50)$$

zusätzlich auch als Funktionen der verbleibenden Umschaltzeit $\tau_i(t)$ zu betrachten.

Zum Zeitpunkt t werde von Freigabe auf Sperrung umgeschaltet, die Ausflussrate $q_i^{\text{out}}(t)$ sei bereits null. Damit lässt sich Gleichung (3.43) in folgende Form bringen:

$$\frac{d\hat{W}_i}{dt} = \hat{n}_i(t, \tau_i(t)) \frac{d(t + \tau_i(t))}{dt}. \quad (3.51)$$

Durch Umstellen erhält man die Ableitung von $\hat{W}_i(t)$ nach dem Zeitpunkt $(t + \tau_i(t))$, zu dem die Freigabe frühestens wieder beginnen kann:

$$\frac{d\hat{W}_i}{d(t + \tau_i(t))} = \hat{n}_i(t, \tau_i(t)). \quad (3.52)$$

Während der Zeitpunkt t fest ist, verschiebt sich der Zeitpunkt $(t + \tau_i(t))$, zu dem die Freigabe beginnen kann, von $(t + \tau_i(t))$ nach $(t + \tau_i^0)$. Die verbleibende Umschaltzeit erhöht sich also von $\tau_i(t)$ auf τ_i^0 . Folglich berechnet sich der Betrag $\Delta W_i^0(t)$, um den die prognostizierte Gesamtwartezeit $\hat{W}_i(t)$ zum Zeitpunkt t wächst, mit dem Integral:

$$\Delta \hat{W}_i^0(t) = \int_{\tau_i(t)}^{\tau_i^0} \hat{n}_i(t, \tau') d\tau'. \quad (3.53)$$

Der Betrag $\Delta W_i^0(t)$ umfasst genau die Wartezeiten, die allein durch das erneute Abwarten der Zwischenzeit τ_i^0 *zusätzlich* entstehen. Das Intervall $[\tau_i(t), \tau_i^0]$ ist umso kleiner, je weniger der Zwischenzeit τ_i^0 zum Zeitpunkt t bereits verstrichen ist. Damit wird deutlich, dass $\Delta W_i^0(t)$ tatsächlich nur die *zusätzlich* entstehenden Wartezeiten erfasst. Die Wartezeiten, die aufgrund der Sperrung im Zeitraum $[t, t + \tau_i(t)]$ ohnehin anfallen, sind bereits in der Prognose enthalten. Abb. 3.8 skizziert den zum Integral aus (3.53) zugehörigen Flächeninhalt im Verlauf der Zeitreihen $N_i^{\text{exp}}(t)$ und $N_i^{\text{out}}(t)$.

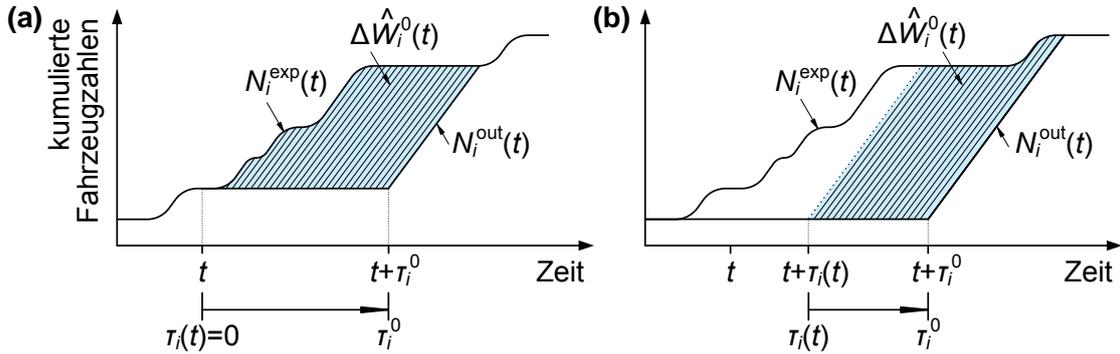


Abbildung 3.8.: Zusätzliche Wartezeiten $\Delta W_i^0(t)$, die durch eine Unterbrechung des Bedienvorgangs für die Dauer von *null* Zeiteinheiten entstehen. (a) Während der Freigabezeit wurde die Warteschlange zum Zeitpunkt t bereits vollständig geleert. (b) Die Freigabe kann erst nach Ablauf der verbleibenden Umschaltzeit $\tau_i(t)$ zum Zeitpunkt $t + \tau_i(t)$ beginnen. In beiden Szenarien (a) und (b) werde der Bedienvorgang zum Zeitpunkt t unterbrochen und sofort wieder aufgenommen. Die Freigabe beginnt dann (erneut) zum Zeitpunkt $t + \tau_i^0$. Die dadurch zusätzlich entstehenden Wartezeiten $\Delta W_i^0(t)$ entsprechen dem Inhalt der schraffierten Fläche.

3.4. Zusammenfassung

In diesem Kapitel wurde ein Verfahren zur Zustandsbeobachtung und -prognose des Verkehrsgeschehens auf einer Knotenzufahrt vorgestellt. Abbildung 3.9 zeigt, wie dieses Verfahren in die Struktur des Regelkreises eingebettet ist.

Das Verkehrsgeschehen um die Haltelinie einer Knotenzufahrt i wird auf die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ und die Anzahl der abfertigten Fahrzeuge $N_i^{\text{out}}(t)$ zurückgeführt. Die Zeitreihe $N_i^{\text{exp}}(t)$ lässt sich mit Hilfe von Verkehrsflussmessungen stromaufwärts der Haltelinie prognostizieren. Weil für das verwendete Prognoseverfahren die Rückstaueffekte nicht zu berücksichtigen sind, wird auch lediglich ein Verkehrsmodell zur Beschreibung des freien Verkehrsflusses benötigt.

Eine Knotenzufahrt lässt sich als ein hybrides dynamisches System beschreiben. Unter verschiedenen diskreten Schalt- und Systemzuständen gelten für die kontinuierlichen Systemgrößen unterschiedliche Entwicklungsgleichungen. Ein Beispiel dafür ist die Entwicklungsgleichung (3.11) der Warteschlangenlänge. Obwohl sich das Verhalten in jedem einzelnen Zustand mit einer linearen Differenzialgleichung erster Ordnung beschreiben lässt, ist die resultierende Dynamik nichtlinear. Andere Beispiele sind die Entwicklungsgleichungen der prognostizierten Freigabezeit (3.30) bis (3.33) und der prognostizierten Gesamtwartezeiten (3.48).

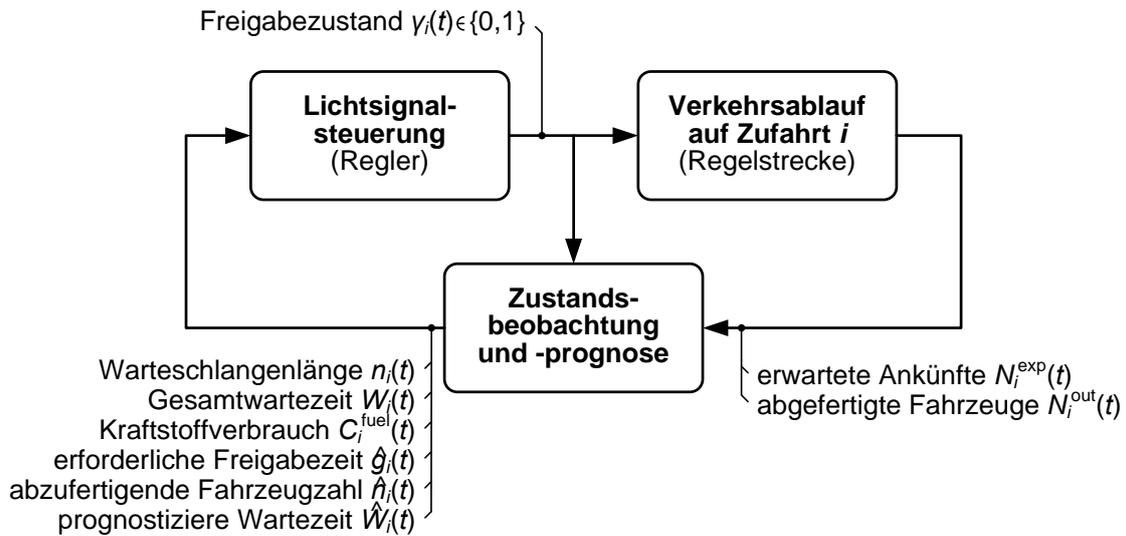


Abbildung 3.9.: Struktur des Regelkreises und Bezeichnung der relevanten Kenngrößen. Im Funktionsbaustein „Zustandsbeobachtung und -prognose“ wird die betrachtete Knotenzufahrt als hybrides dynamisches System modelliert. Das Prognoseverfahren beruht auf den beiden Annahmen, dass (i) die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ über einen Prognosehorizont der Dauer eines Bedienvorgangs bekannt ist, und dass (ii) die Fahrzeuge einer Wartekolonne mit konstanter Rate abfließen. Damit können die links unten aufgeführten Kenngrößen berechnet und beim Reglerentwurf im nächsten Kapitel verwendet werden.

Auf Grundlage des verwendeten Warteschlangenmodells lässt sich der Verkehrsablauf während eines Bedienvorgangs vollständig prognostizieren. Wie Abb. 3.6 (auf S. 57) zeigt, wird dabei ein variabler Prognosehorizont verwendet. Das Prognoseverfahren setzt voraus, dass (i) die erwarteten Ankünfte $N_i^{\text{exp}}(t)$ in diesem Zeitraum bekannt sind und (ii) die Fahrzeuge der Wartekolonne mit konstanter Rate q_i^{max} herausfahren. Die Anzahl der abgefertigten Fahrzeuge $N_i^{\text{out}}(t)$ kann dann, wie Abb. 3.4 (auf S. 52) zeigt, mit Geradenstücken beschrieben werden. Für die benötigte Freigabezeit $\hat{g}_i(t)$ bis zur Leerung der Warteschlange folgt daraus die Bestimmungsgleichung (3.20), die sich mit dem numerischen Eingrenzungsverfahren (3.21) bis (3.24) lösen lässt. Bei der Untersuchung des Zeitverhaltens von $\hat{g}_i(t)$ wurde mit Gleichung (3.30) folgende Eigenschaft deutlich: Im Verlauf von $\hat{g}_i(t)$ tritt zum Zeitpunkt t ein Sprung auf, wenn zu einem späteren Zeitpunkt t' die Ankunft einer Fahrzeugkolonne erwartet wird. Zum Ersten ist der Sprung umso größer je mehr Fahrzeuge die Kolonne umfasst, und zum Zweiten verbleibt bis zur Ankunft der Kolonne gerade die benötigte Zeit $(t' - t)$, um den Umschaltvor-

gang abzuschließen und die Warteschlange zu leeren. Dieser Zusammenhang ist in Abb. 3.6 (auf S. 57) ersichtlich.

Unmittelbar mit $\hat{g}_i(t)$ verknüpft ist die Größe $\hat{n}_i(t)$. Einerseits stellt sie die bis zur vollständigen Leerung der Warteschlange abzufertigende Fahrzeuganzahl dar. Andererseits ist $\hat{n}_i(t)$ der Betrag, um den die Gesamtwartezeiten $\hat{W}_i(t)$ eines noch nicht begonnenen Bedienvorgangs anwachsen. Wird also ein Bedienvorgang zum Zeitpunkt t um eine kleine Zeitspanne dt herausgezögert, so erhöhen sich die Gesamtwartezeiten $W_i(t')$, die bis zum Ende des Bedienvorgangs zum Zeitpunkt t' anfallen, um den Betrag $\hat{n}_i(t)dt$. Um welchen Betrag sich $\hat{W}_i(t)$ infolge des Abbruchs eines Bedienvorgangs erhöht, wird mit Gleichung (3.53) erfasst. Diese Zusammenhänge spielen eine zentrale Rolle beim Entwurf des lokalen Optimierungsverfahrens im nächsten Kapitel.

Kapitel 4.

Lokale Optimierung

Kurzfassung

In den Kapiteln 4 und 5 werden die beiden Kernbausteine des Verfahrens zur dezentralen Online-Steuerung vorgestellt. Der erste Baustein, das „Optimierungsverfahren“, dient der Minimierung der Gesamtwartezzeiten am lokal betrachteten Knotenpunkt.

Das Verfahren beruht auf einer dynamischen Priorisierung der Knotenzufahrten. Die Spezifikation der Prioritätsindizes wird aus den optimalen Lösungen dreier Testszenarien systematisch abgeleitet und anschließend verallgemeinert. Unter Verwendung des im Abschnitt 3.3 vorgestellten Prognoseverfahrens werden all jene Fahrzeuge in die Optimierung einbezogen, die sich momentan im Wirkungsbereich (siehe Abb. 3.6) befinden und im folgenden Bedienvorgang mit maximaler Rate abgefertigt werden können. Das Optimierungsverfahren wählt zu jedem Zeitpunkt den bezüglich dieser Fahrzeuge optimalen Schaltzustand.

Das Optimierungsverfahren berücksichtigt weder die Wechselwirkung mit benachbarten Knotenpunkten noch maximale Sperrzeiten einzelner Zufahrten. Diese Aspekte werden im nachfolgenden Kapitel 5 behandelt. Das dort vorgestellte Stabilisierungsverfahren wird, obwohl es ebenso lediglich auf die lokalen Fahrzeugankünfte reagiert, für netzwerkweite Stabilität sorgen und maximale Bedienperioden einhalten. In Kapitel 6 werden beide Verfahren miteinander kombiniert.

4.1. Prioritätsbasierte Verfahren

Prioritätsbasierte Verfahren wurden bereits kurz in Abschnitt 2.2.1 erwähnt. Sie weisen jeder Zufahrt i einen Prioritätsindex π_i zu. Entsprechend der allgemeinen Schaltregel

$$\sigma = \arg \max_i \pi_i \quad (4.1)$$

ist die Zufahrt mit größtem positiven Prioritätsindex zur Bedienung auszuwählen. Der entscheidende Vorteil prioritätsbasierter Verfahren ist ihre Einfachheit: Die Lösung des nichtlinearen kombinatorischen Optimierungsproblems reduziert sich auf die Berechnung eines einzigen Zahlenwerts für jede Zufahrt.

4.1.1. Existenz optimaler Prioritätsindizes

Prioritätsbasierte Verfahren können nur unter bestimmten Voraussetzungen optimal sein. Nachdem seit 1966 keine Arbeit mehr veröffentlicht wurde, die über die Entdeckung neuer optimaler Prioritätsindizes berichtet, gingen Rothkopf und Smith im Jahre 1984 der Frage nach, ob denn überhaupt noch andere als die bereits bekannten existieren können. Sie kamen zum Schluss:

*„There are no undiscovered priority index
sequencing rules for minimizing total delay costs.“*

Den Beweis für diese Feststellung veröffentlichten sie unter gleichem Titel im Referenz [157]. Ihre Argumentation soll im Folgenden genauer betrachtet werden.

Notwendige Voraussetzungen

Ein optimaler Prioritätsindex existiert nur dann, wenn die optimale Reihenfolge, in der zwei beliebig herausgegriffene Warteschlangen abzufertigen sind, sowohl

- (i) vom Zustand und Vorhandensein anderer Warteschlangen als auch
- (ii) von der bereits verstrichenen Wartezeit

unabhängig ist. Dies setzt zwingend voraus, dass die zu minimierenden Kosten mit jedem wartenden Fahrzeug entweder exponentiell oder linear in der Zeit anwachsen. In unserem Fall ist die Kostenfunktion die Gesamtwartezeit aller Fahrzeuge, und sie wächst, wie in Abschnitt 3.3.3 gezeigt, linear mit der Anzahl der wartenden Fahrzeuge. Damit ist die notwendige Voraussetzung erfüllt. Weiterhin impliziert Forderung (ii), dass im Betrachtungszeitraum zu den bereits wartenden Fahrzeugen keine weiteren hinzukommen dürfen. Nur so bleibt der Zustand einer nicht bedienten Warteschlange unverändert.

Optimale Prioritätsregel

Die optimale Priorisierungsregel zur Abfertigung von Warteschlangen ohne Zufluss geht auf die Arbeiten von Smith [180] und McNaughton [120] aus den Jahren 1956 bzw. 1959 zurück. Sie ist in der Warteschlangentheorie unter dem Namen μc -Regel bekannt. Die Symbole μ und c stehen dort für die Abfertigungs- bzw. die Kostenzuwachsrate. Die Regel besagt: „Leere die Warteschlangen in absteigender Reihenfolge des Produkts aus μ und c “. Während die ersten Arbeiten noch von verschwindend kleinen Umschaltzeiten ausgingen, wurde diese Regel später, beispielsweise von van Oyen et al. [137], auf verlustbehaftete Umschaltvorgänge verallgemeinert.

Weiteres Vorgehen

Die μc -Regel soll der Ausgangspunkt für Spezifikation eines dynamischen Prioritätsindex sein. Das Vorgehen ist viergeteilt. Im nächsten Abschnitt 4.1.2 wird zunächst die μc -Regel hergeleitet. Im darauf folgenden Abschnitt 4.2.1 wird die Priorisierungsregel um das Prognoseverfahren aus Abschnitt 3.3 erweitert. Somit können auch herannahende Fahrzeuge mit in die Optimierung einbezogen werden. Gleichzeitig wird die Frage geklärt, wann ein Bedienvorgang zu unterbrechen ist, um einer anderen Zufahrt Vorrang zu gewähren. In Abschnitt 4.2.2 wird die Verlängerung der Freigabezeit zur Abfertigung von später ankommenden Nachzüglern untersucht. Im letzten Schritt werden die bis dahin gewonnenen Ergebnisse im Abschnitt 4.2.3 verallgemeinert. Schließlich fasst Abschnitt 4.3 die Ergebnisse in geschlossener Form zusammen.

4.1.2. Herleitung der μc -Regel

Im ersten Schritt sollen die μc -Regel hergeleitet und die aus ihr folgenden Schlüsse diskutiert werden. Die μc -Regel gilt nur für den Fall, dass im betrachteten Zeitraum keine weiteren Fahrzeuge hinzukommen. Nach Rothkopf und Smith [157] ist dann die optimale Bedienreihenfolge zweier beliebig herausgegriffener Zufahrten vom Vorhandensein weiterer Zufahrten unabhängig. Deshalb brauchen lediglich zwei Zufahrten betrachtet zu werden. Denn aus dem paarweisen Vergleich, welche Zufahrt vor welcher zu bedienen ist, ergibt sich unmittelbar die optimale Bedienreihenfolge für beliebig viele Zufahrten.

Nach van Oyen et al. [137] ist es in diesem Fall für die optimale Schaltregel notwendig, jede Warteschlange in einem einzigen Bedienvorgang vollständig zu leeren. Deshalb kommen zum Umschalten nur die Zeitpunkte in Frage, zu denen eine Warteschlange leer wird. Letztlich ist allein die Bedienreihenfolge entscheidend. Schaltprogramme, die einen einmal begonnenen Bedienvorgang bis zur voll-

ständigen Leerung der Warteschlange fortsetzen, werden als „erschöpfend“ (engl.: „exhaustive“) charakterisiert.

Szenario 4.i

Es wird der Knotenpunkt aus Abb. 4.1(a) betrachtet. Auf seinen zwei Zufahrten 1 und 2 befinden sich anfangs $n_1^0 > 0$ und $n_2^0 > 0$ Fahrzeuge. Bevor mit der Freigabe einer Zufahrt – auch der zuerst ausgewählten – begonnen werden kann, ist die jeweilige Zwischenzeit τ_1^0 bzw. τ_2^0 abzuwarten.

Wartezeitenbilanz

Die optimale Bedienreihenfolge soll sich aus dem Vergleich der Gesamtwartezeiten, die in den beiden möglichen Fällen entstehen, ergeben. In Abbildung 4.1(c) und (d) sind beide Fälle gegenübergestellt. Beginnen wir mit Fall (A), in dem zuerst Zufahrt 1 bedient wird. Während ihrer Bedienung ($\sigma = 1$) fallen auf den Zufahrten 1 und 2 die Wartezeiten $W_1^{\sigma=1}$ bzw. $W_2^{\sigma=1}$ an. Diese ergeben sich beispielsweise aus den entsprechenden Flächeninhalten in Abb. 4.1(c). Man erhält somit:

$$W_1^{\sigma=1} = n_1^0 \tau_1^0 + \frac{(n_1^0)^2}{2q_1^{\max}} \quad \text{und} \quad W_2^{\sigma=1} = n_2^0 \tau_1^0 + \frac{n_1^0 n_2^0}{q_1^{\max}}. \quad (4.2)$$

Nachdem die Warteschlange auf der einen Zufahrt 1 vollständig geleert wurde, wird die andere Zufahrt 2 bedient. Dabei fallen zusätzlich noch die Wartezeiten

$$W_2^{\sigma=2} = n_2^0 \tau_2^0 + \frac{(n_2^0)^2}{2q_2^{\max}} \quad (4.3)$$

an. Nun untersuchen wir den Fall (B) mit umgekehrter Bedienreihenfolge. Wenn Zufahrt 2 zuerst bedient wird, fallen zunächst die Wartezeiten $W_1^{\sigma=2}$ und $W_2^{\sigma=2}$ an. Dabei ist $W_2^{\sigma=2}$ identisch mit dem zuvor berechneten Wert in Gleichung (4.3). Die Wartezeiten $W_1^{\sigma=2}$ berechnen sich mit:

$$W_1^{\sigma=2} = n_1^0 \tau_2^0 + \frac{n_1^0 n_2^0}{q_2^{\max}}. \quad (4.4)$$

Die bei der anschließenden Leerung von Zufahrt 1 entstehenden Wartezeiten $W_1^{\sigma=1}$ wurden bereits in Gleichung (4.2) berechnet.

Nun können die Gesamtwartezeiten beider Bedienreihenfolgen verglichen werden. Fall (A) ist dann optimal, d.h. Zufahrt 1 ist dann als erste auszuwählen, wenn gilt:

$$W_1^{\sigma=1} + W_2^{\sigma=1} + W_2^{\sigma=2} < W_1^{\sigma=2} + W_2^{\sigma=2} + W_1^{\sigma=1}. \quad (4.5)$$

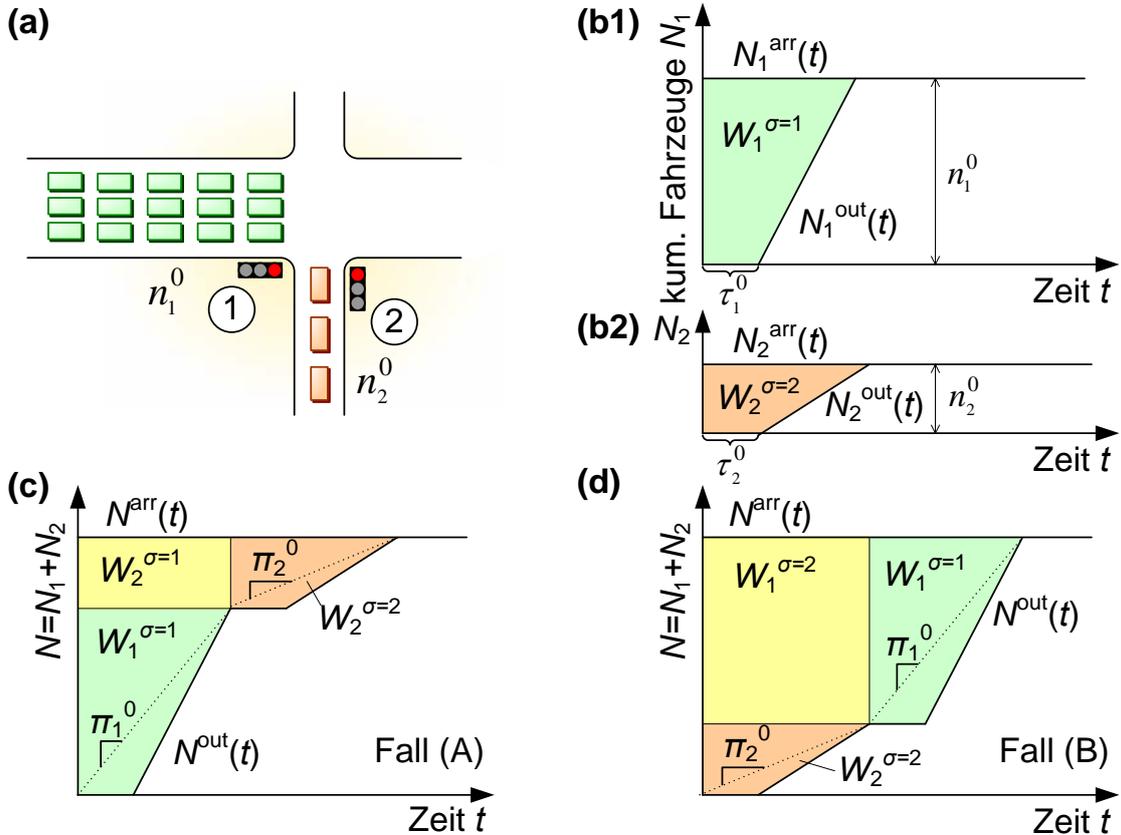


Abbildung 4.1.: (a) Knotenpunkt mit zwei Zufahrten 1 und 2. Zu den Warteschlangen der Länge n_1^0 bzw. n_2^0 kommen im Betrachtungszeitraum keine weiteren Fahrzeuge hinzu. (b1, b2) Berechnung der Wartezeiten auf Zufahrt 1 bzw. 2, die während ihrer Bedienung anfallen. (c, d) Gesamtwartezeiten für die beiden Fälle, dass Zufahrt 1 bzw. Zufahrt 2 zuerst gelehrt wird.

In dieser Bilanz fallen die Terme $W_1^{\sigma=1}$ und $W_2^{\sigma=2}$ weg. Die Wartezeiten, die auf einer Zufahrt während ihrer Leerung anfallen, sind von der Bedienreihenfolge unabhängig. Übrig bleibt die Bedingung

$$W_2^{\sigma=1} < W_1^{\sigma=2}. \quad (4.6)$$

Durch Einsetzen von (4.2) und (4.4) erhält man eine Ungleichung, deren beiden Seiten sich bezüglich der Indizes 1 und 2 separieren lassen:

$$\frac{\tau_1^0}{n_1^0} + \frac{1}{q_1^{\max}} < \frac{\tau_2^0}{n_2^0} + \frac{1}{q_2^{\max}}. \quad (4.7)$$

Jede Seite dieser Ungleichung weist den Zufahrten eine Maßzahl zu, in deren aufsteigender Reihenfolge sie wartezeitenoptimal zu bedienen sind. Die optimale Prioritätsregel ergibt sich daraus, dass man ihren Kehrwert als Prioritätsindex π_i^0 definiert. Dann ist es optimal, die Zufahrt mit größtem π_i^0 zuerst zu bedienen. Es soll also gelten:

$$\pi_i^0 = \frac{n_i^0}{\tau_i^0 + g_i^0} \quad \text{mit der Freigabezeit} \quad g_i^0 = n_i^0 / q_i^{\max}. \quad (4.8)$$

Eine bereits geleerte Zufahrt mit $n_i^0 = 0$ hat den Prioritätsindex $\pi_i^0 = 0$ und wird entsprechend der allgemeinen Schaltregel (4.1) nicht noch einmal bedient.

Die Ähnlichkeit zur oben besprochenen μc -Regel wird deutlich, wenn man den Bruch $n_i^0 / (\tau_i^0 + g_i^0)$ als effektive Abfertigungsrate „ μ “ interpretiert. Die Kostenzuwachsrate „ c “ ist in unserem Fall gleich eins, denn die Gesamtwartezeit wächst mit jedem wartenden Fahrzeug in jeder Sekunde um eine Sekunde. Damit wurde die μc -Regel hergeleitet.

Zusammenfassend lautet die Regel zur wartezeitenminimalen Abfertigung von Warteschlangen ohne Zufluss: Solange nicht alle Warteschlangen leer sind, ist die Zufahrt mit dem größtem Prioritätsindex π_i^0 auszuwählen und vollständig zu leeren. Die Werte für π_i^0 , berechnet nach Gleichung (4.8), entsprechen der effektiven Ausflussrate während des Bedienvorgangs zur Warteschlangenleerung.

Implikationen

Folgende Schlussfolgerungen können gezogen werden: Für die optimale Bedienreihenfolge ist nicht entscheidend, wie lang die Warteschlangen sind, sondern wie „effizient“²¹ sie abgefertigt werden können. Nur im Sonderfall symmetrischer Anordnungen, das heißt wenn alle Knotenzufahrten identische Zwischenzeiten und identische maximale Ausflussraten haben, können längere Warteschlangen effizienter abgefertigt werden. Hingegen spielt die Warteschlangenlänge im Sonderfall verschwindend kleiner Zwischenzeiten gar keine Rolle: Die Zufahrten sind dann in der Reihenfolge ihrer maximalen Ausflussraten zu bedienen.

Das Ergebnis (4.8) lässt sich auch geometrisch deuten. Der Inhalt der rechteckigen Flächen in den Abb. 4.1(c) und (d) entspricht den Gesamtwartezeiten während der Bedienvorgänge. Er ist dann minimal, wenn der Anstieg der Geradenstücke, die jeweils aufeinander folgende Umschaltzeitpunkte verbinden, zu Beginn am größten ist und dann immer kleiner wird. Der Prioritätsindex π_i^0 wird durch den Anstieg der Geradenstücke repräsentiert.

²¹Der Begriff „effizient“ wird hier im Sinne des Verhältnisses aus Freigabezeit zu insgesamt benötigter Bedienzeit (einschließlich der Zwischenzeit) verstanden. Vgl. Abschnitt 5.1.1.

4.2. Dynamische Priorisierung

Im zweiten Schritt wird zugelassen, dass sich die optimale Bedienreihenfolge aufgrund neu hinzukommender Fahrzeugkolonnen jederzeit ändern kann. Wird dadurch der aktuelle Bedienvorgang unterbrochen und umgeschaltet, so können die verbliebenen Fahrzeuge erst später nach einem zusätzlichen Umschaltvorgang abgefertigt werden. Aus dem optimalen Kriterium dafür, wann der aktuelle Bedienvorgang zu unterbrechen ist, um einen anderen vorzuziehen, wird die Spezifikation des dynamischen Prioritätsindex $\pi_i(t)$ abgeleitet. Er ist zu jedem Zeitpunkt t für die Zufahrt i am größten, die entsprechend der „momentan“ optimalen Bedienreihenfolge als erste zu leeren ist. Bei dieser Art der dynamischen Reoptimierung, das heißt beim Verfolgen der jeweils momentan optimalen Lösung, handelt es sich um einen heuristischen Ansatz. Dieser Ansatz ist optimal, wenn im betrachteten Zeitraum keine weiteren Fahrzeuge hinzukommen. Das Schaltprogramm ist mit dieser Erweiterung nicht mehr „exhaustive“ sondern „präemptiv“.

4.2.1. Erweiterung um Kurzzeitprognose

Zusätzlich wird für die Priorisierung das in Abschnitt 3.3 eingeführte Prognoseverfahren herangezogen. Deshalb wird die Wartezeitenbilanz nun nicht mehr für die Gesamtwartezeiten $\sum_i W_i(t)$ aufgestellt, sondern für die prognostizierten Gesamtwartezeiten $\sum_i \hat{W}_i(t)$. Die zeitliche Entwicklung von $\hat{W}_i(t)$ ist in Gleichung (3.48) zusammengefasst. Während eine Zufahrt i nicht bedient wird, wächst $\hat{W}_i(t)$ mit der Anzahl der Fahrzeuge $\hat{n}_i(t, \tau)$, die im nachfolgenden Bedienvorgang mit maximaler Rate abgefertigt werden können. Beim Umschalten auf Sperrung erhöhen sich die prognostizierten Gesamtwartezeiten um den Betrag $\Delta \hat{W}_i^0(t)$ nach Gleichung (3.53), während des Bedienvorgangs bleiben sie konstant.

Das Prognoseverfahren erfasst insbesondere auch die Fahrzeuge, die sich noch auf freier Strecke befinden und erst während des Umschaltens oder während der Freigabe die Haltelinie erreichen können.²² Dabei wird berücksichtigt, dass jedes zusätzliche Fahrzeug die benötigte Freigabezeit entsprechend verlängert und dadurch potenziell weiteren Fahrzeugen das Eintreffen vor Leerung der Warteschlange ermöglicht.

Szenario 4.ii

Die folgenden Betrachtungen werden wieder auf den Knotenpunkt aus Abb. 4.1(a) bezogen. Zu einem Zeitpunkt t werde Zufahrt 2 bedient, d.h. $\sigma(t) = 2$. Bis zur Freigabe auf Zufahrt 2 muss noch die verbleibende Zwischenzeit $\tau_2(t)$ verstreichen.²³ Auf den beiden Zufahrten können $\hat{n}_1(t)$ bzw. $\hat{n}_2(t)$ Fahrzeuge mit maximaler Rate

²²Vgl. Abb. 3.6 (auf S. 57)

²³Falls die Freigabe bereits begonnen hat, ist $\tau_i(t)$ gleich null, vgl. Abb. 3.5 (auf S. 56).

abgefertigt werden. Es sei angenommen, dass zu diesen prognostizierten Fahrzeugen keine weiteren hinzukommen.

Bei der Herleitung des statischen Prioritätsindex π_i^0 mussten nur jene Zeitpunkte betrachtet werden, zu denen eine Warteschlange vollständig geleert wurde und deshalb ein Umschaltvorgang erforderlich war. Das ist nun anders. Es wird nämlich zugelassen, dass während der Bedienung von Zufahrt 2 auf der anderen Zufahrt 1 neue Fahrzeuge hinzugekommen sind. Folglich besteht die Wahl zwischen folgenden beiden Alternativen: (A) Die Bedienung von Zufahrt 2 wird fortgesetzt und erst nach beendeter Warteschlangenleerung zur anderen Zufahrt 1 umgeschaltet. (B) Die Bedienung von Zufahrt 2 wird sofort unterbrochen, um zuerst Zufahrt 1 vollständig zu leeren. Anschließend wird zurück geschaltet und Zufahrt 2 weiter bedient.

Wartezeitenbilanz

Das optimale Entscheidungskriterium zur Wahl zwischen den beiden Alternativen wird im Folgenden aus einer Bilanz der prognostizierten Gesamtwartezeiten abgeleitet. Die Zunahme der prognostizierten Gesamtwartezeiten auf Zufahrt i bei Wahl von Alternative (A) bzw. (B) soll hierbei mit $\Delta\hat{W}_i^A$ bzw. $\Delta\hat{W}_i^B$ bezeichnet werden. Damit lautet das Entscheidungskriterium für die Wahl von Alternative (A) in allgemeiner Form:

$$\Delta\hat{W}_1^A + \Delta\hat{W}_2^A < \Delta\hat{W}_1^B + \Delta\hat{W}_2^B. \quad (4.9)$$

Beginnen wir mit Fall (A). Die während der verbleibenden Umschaltzeit $\tau_2(t)$ und während der Freigabezeit $\hat{g}_2(t) = \hat{n}_2(t)/q_2^{\max}$ auf Zufahrt 2 anfallenden Wartezeiten sind bereits in der Prognose erfasst. Deshalb bleiben die dort prognostizierten Gesamtwartezeiten $\hat{W}_2(t)$ konstant. Aufgrund der Annahme, dass keine weiteren Fahrzeuge hinzukommen, fallen auch nach der Leerung der Warteschlange keine weiteren Wartezeiten an. Folglich gilt:

$$\Delta\hat{W}_2^A = 0. \quad (4.10)$$

Zufahrt 1 wird erst bedient, nachdem Zufahrt 2 vollständig geleert wurde. Bis zur Bedienung von Zufahrt 1 entspricht die Zuwachsrate der prognostizierten Gesamtwartezeiten der Anzahl der abzufertigenden Fahrzeuge $\hat{n}_1(t)$. Diese wurde als konstant angenommen. Damit erhält man $\Delta\hat{W}_1^A$ aus der Multiplikation der Wachstumsrate $\hat{n}_1(t)$ mit der Zeitspanne $\tau_2(t) + \hat{g}_2(t)$ bis zum Beginn des Bedienvorgangs. Folglich gilt:

$$\Delta\hat{W}_1^A = \hat{n}_1(t)(\tau_2(t) + \hat{g}_2(t)). \quad (4.11)$$

Im Fall (B) wird die Bedienung der Zufahrt 1 vorgezogen. Sowohl während als auch nach der Bedienung von Zufahrt 1 bleiben die dort prognostizierten Gesamt-

wartezeiten konstant. Es gilt:

$$\Delta \hat{W}_1^B = 0. \quad (4.12)$$

Auf Zufahrt 2 wird der bereits begonnene Bedienvorgang unterbrochen. Dabei erhöhen sich die prognostizierten Gesamtwarezeiten $\Delta \hat{W}_2^B$ sowohl beim Umschalten auf Sperrung als auch während des gesamten Bedienvorgangs für Zufahrt 1. Der erste Teil, $\Delta \hat{W}_2^0(t)$, wird durch Gleichung (3.53) beschrieben und soll im Folgenden genauer untersucht werden. Der zweite Teil berechnet sich wie oben aus der Multiplikation der Wachstumsrate $\hat{n}_2(t)$ mit der Zeitspanne $\tau_1^0 + \hat{g}_1(t)$ bis zur Neuaufnahme des Bedienvorgangs. Man erhält zunächst:

$$\Delta \hat{W}_2^B = \Delta \hat{W}_2^0(t) + \hat{n}_2(t)(\tau_1^0 + \hat{g}_1(t)). \quad (4.13)$$

Der erste Summand berücksichtigt, dass sich durch den Sprung der verbleibenden Zwischenzeit $\tau_2(t)$ auf τ_2^0 auch die Wartezeitenprognose abrupt ändert. Weil angenommen wurde, dass keine weiteren Fahrzeuge hinzukommen und deshalb die Anzahl der abzufertigenden Fahrzeuge $\hat{n}_2(t)$ konstant ist, kann das Integral aus Gleichung (3.53) auf eine Multiplikation zurückgeführt werden:

$$\Delta \hat{W}_2^0(t) = \int_{\tau_2(t)}^{\tau_2^0} \hat{n}_2(t, \tau') d\tau' = \hat{n}_2(t)(\tau_2^0 - \tau_2(t)). \quad (4.14)$$

Die Differenz $\tau_2^0 - \tau_2(t)$ entspricht der bereits verstrichenen und – aufgrund der Unterbrechung – zusätzlich aufs Neue abzuwartenden Umschaltzeit. Der andere Teil der Zwischenzeit τ_1^0 , die verbleibende Zwischenzeit $\tau_2(t)$, wurde noch nicht aufgewendet und liefert deshalb keinen zusätzlichen Beitrag. Wird Gleichung (4.14) in Gleichung (4.13) eingesetzt, erhält man:

$$\Delta \hat{W}_2^B = \hat{n}_2(t)(\tau_1^0 + \tau_2^0 - \tau_2(t) + \hat{g}_1(t)). \quad (4.15)$$

Damit sind alle Terme im Entscheidungskriterium (4.9) spezifiziert. Als Kriterium, unter dem Alternative A (Fortsetzen der Bedienung) optimal ist, erhält man durch Einsetzen und Umformen folgende Ungleichung:

$$\frac{\hat{n}_2(t)}{\tau_2(t) + \hat{g}_2(t)} > \frac{\hat{n}_1(t)}{\tau_2^0 - \tau_2(t) + \tau_1^0 + \hat{g}_1(t)} \quad \text{mit} \quad \hat{g}_i(t) = \frac{\hat{n}_i(t)}{q_i^{\max}}. \quad (4.16)$$

Im Unterschied zur Herleitung des statischen Prioritätsindex gelingt es zunächst nicht, die beiden Seiten der Ungleichung (4.16) bezüglich der Indizes 1 und 2 zu separieren. Aber der folgende Ansatz hilft weiter: Der Summand $\tau_2^0 - \tau_2(t)$ im Nenner der rechten Seite bezieht sich auf die zur Zeit t ausgewählte Zufahrt $\sigma(t) = 2$. Wie weiter unten gezeigt wird, kann deshalb $\tau_2^0 - \tau_2(t)$ ohne Einschränkung durch

$\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t)$ ersetzt werden. Falls zur Zeit t keine Zufahrt bedient wird ($\sigma(t) = \emptyset$), soll $\tau_{\sigma(t)}^0 = \tau_{\sigma(t)}(t) = 0$ gelten.

Spezifikation

Aus den beiden Seiten der Ungleichung (4.16) lässt sich der dynamische Prioritätsindex $\pi_i(t)$ wie folgt spezifizieren:

$$\pi_i(t) = \begin{cases} \frac{n_i(t)}{\tau_i(t) + g_i(t)}, & \text{falls } \sigma(t) = i \\ \frac{n_i(t)}{\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t) + \tau_i^0 + g_i(t)}, & \text{sonst.} \end{cases} \quad (4.17)$$

Die Hürde, einen bereits begonnenen Bedienvorgang zu unterbrechen, ist umso größer, je mehr Zwischenzeit bereits aufgebracht wurde. In der Spezifikation (4.17) ist Folgendes zu sehen: Der Prioritätsindex einer nicht bedienten Zufahrt $i \neq \sigma(t)$ wird um den Betrag der verstrichenen Zwischenzeit $\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t)$ im Nenner „geschwächt“. Gleichzeitig wird der Prioritätsindex $\pi_i(t)$ der ausgewählten Zufahrt $i = \sigma(t)$ umso größer, je kleiner die verbleibende Zwischenzeit $\tau_{\sigma(t)}(t)$ ist, d.h. je weiter der Umschaltvorgang bereits fortgeschritten ist.²⁴ Hat die Freigabe begonnen und ist $\tau_{\sigma(t)}(t)$ auf null abgefallen, ist $\pi_i(t) = q_i^{\max}$ maximal.

Hat auf der momentan bedienten Zufahrt $i = \sigma(t)$ die Abfertigung bereits begonnen ($\tau_i(t) = 0$), so kann eine Unterbrechung nur dann optimal sein, wenn die vorzuziehende Zufahrt $j \neq \sigma(t)$ eine größere maximale Ausflussrate hat als Zufahrt i . Denn aus Gleichung (4.17) folgt für die momentan bediente Zufahrt $\pi_i(t) = q_i^{\max}$, für die nicht bediente dagegen $\pi_j(t) < q_j^{\max}$. Mit identischen maximalen Flussraten auf jeder Fahrspur kann die Abfertigung einer Zufahrt nur von einer anderen Zufahrt mit höherer Spurenanzahl unterbrochen werden.

Erste Forderung von Rothkopf und Smith

Nun soll noch allgemein gezeigt werden, dass die Ersetzung von $\tau_2^0 - \tau_2(t)$ durch $\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t)$ in (4.16) zulässig war. Nach Rothkopf und Smith [157] darf sich dadurch die optimale Bedienreihenfolge zweier beliebig herausgegriffener Zufahrten i und j mit $i, j \neq \sigma(t)$ nicht ändern. Welcher der beiden Prioritätsindizes $\pi_i(t)$ und $\pi_j(t)$ größer ist, hängt weder vom Zustand noch dem Vorhandensein der aktuell bedienten Zufahrt $\sigma(t)$ ab. Diese Aussage trifft zu, weil

$$\frac{\hat{n}_i(t)}{\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t) + \tau_i^0 + \hat{g}_i(t)} > \frac{\hat{n}_j(t)}{\tau_{\sigma(t)}^0 - \tau_{\sigma(t)}(t) + \tau_j^0 + \hat{g}_j(t)} \quad (4.18)$$

²⁴Vgl. Abb. 4.4 (auf S. 86)

genau dann erfüllt ist, wenn

$$\frac{\hat{n}_i(t)}{\tau_i^0 + \hat{g}_i(t)} > \frac{\hat{n}_j(t)}{\tau_j^0 + \hat{g}_j(t)} \quad (4.19)$$

gilt. Also auch mit der erweiterten Funktionalität des dynamischen Prioritätsindex, die Bedienung einer Zufahrt gegebenenfalls spontan zu unterbrechen, bleibt die optimale Reihenfolge, in der die anderen Zufahrten nacheinander zu leeren sind, unverändert. Außerdem bestätigt dies den Ansatz, den optimalen Prioritätsindex aus der Betrachtung von lediglich zwei Zufahrten abzuleiten.

Aufgrund der Tatsache, dass der in Gleichung (4.17) spezifizierte Prioritätsindex $\pi_i(t)$ nicht allein von der betrachteten Zufahrt i , sondern zusätzlich von der momentan bedienten Zufahrt $\sigma(t)$ abhängt, handelt es sich nicht mehr um einen Prioritätsindex im klassischen Sinne. Dass diese Erweiterung dennoch zulässig ist, wurde soeben gezeigt. Erst die zusätzliche Abhängigkeit von $\sigma(t)$ erlaubt es, die Kosten für die Unterbrechung des aktuellen Bedienvorgangs zu berücksichtigen.

Im nächsten Schritt soll eine noch nicht befriedigende Eigenschaft der Priorisierungsregel aufgehoben werden: Sobald die Warteschlange einer Zufahrt i vollständig geleert wurde und keine Abfertigung mit maximaler Rate mehr möglich ist, fällt der Prioritätsindex $\pi_i(t)$ nach Spezifikation (4.17) auf null ab. Der Bedienvorgang wird sofort abgebrochen. Kurz danach ankommende Nachzügler werden unweigerlich gestoppt. Das Problem, die Freigabezeit über den Zeitpunkt der Warteschlangenleerung hinaus zu verlängern, wird im Folgenden genauer untersucht.

4.2.2. Priorisierung bereits geleerter Warteschlangen

Das Prognoseverfahren aus Abschnitt 3.3 erfasst nur die Fahrzeuge, die während der Freigabezeit mit maximaler Rate abgefertigt werden können. Das sind diejenigen Fahrzeuge, die entweder aus einer aufgestauten Wartekolonnen herausfahren oder als kompakte Kolonne mit einer Grünen Welle bedient werden.²⁵ Erreicht eine zweite Kolonne die Haltelinie nur kurze Zeit später, so wird sie nicht mit berücksichtigt. Das führt dazu, dass die bisherige Spezifikation den Bedienvorgang abbricht. Aber im selben Moment wird die nachfolgende Kolonne dennoch erfasst: Während der anschließenden Sperrung, die sich mindestens über die Zwischenzeit τ_i^0 für das sofortige Zurückschalten erstreckt, rücken die Nachzügler auf und können dann doch mit maximaler Rate abgefertigt werden. Würde daraufhin der soeben abgebrochene Bedienvorgang sofort neu gestartet werden, entstünden unnötige Wartezeiten. Wie dieser Fall zu behandeln ist, und wie die Nachzügler in die dynamische Priorisierung einzubeziehen sind, wird nun im dritten Szenario untersucht.

²⁵Vgl. Abb. 3.6 (auf S. 57)

Szenario 4.iii

Es wird die in Abb. 4.2(a) dargestellte Situation betrachtet. Zum Zeitpunkt t wird auf Knotenzufahrt 1 das letzte Fahrzeug einer Wartekolonne abgefertigt. Kurze Zeit später, zum Zeitpunkt $t + \Delta t_1$ mit $\Delta t_1 \leq \tau_1^0$, erreicht das erste Fahrzeug einer kompakten Kolonne aus Δn_1 Fahrzeugen die Halteline. Würde Zufahrt 1 für die Dauer der Zwischenzeit τ_1^0 gesperrt werden, so könnten danach Δn_1 Fahrzeuge mit maximaler Rate abfließen. Folglich hängt die Prognose der abzufertigenden Fahrzeuganzahl $\hat{n}_1(t, \tau)$ von der angenommenen Zwischenzeit τ ab:

$$\hat{n}_1(t, \tau) = \begin{cases} 0, & \text{falls } \tau < \Delta t_1 \\ \Delta n_1, & \text{sonst.} \end{cases} \quad (4.20)$$

Diese Abhängigkeit zeigt auch Abb. 4.2(b). Auf Zufahrt 2 können $\hat{n}_2(t)$ Fahrzeuge mit maximaler Rate abgefertigt werden. Zu den genannten Fahrzeugen kommen keine weiteren hinzu.

Wartezeitenbilanz

Es stehen zwei Alternativen zur Auswahl: (A) die Verlängerung der Freigabezeit für Zufahrt 1 zur Abfertigung der Nachzügler und (B) das sofortige Umschalten zur Leerung von Zufahrt 2 mit anschließendem Zurückschalten zur Zufahrt 1. Bei (A) erhöhen sich die prognostizierten Gesamtwartezeiten nur auf Zufahrt 2, bei (B) nur auf Zufahrt 1. Deshalb ist Alternative (A) optimal, wenn gilt:

$$\Delta \hat{W}_2^A < \Delta \hat{W}_1^B. \quad (4.21)$$

Im Fall (A) beginnt die Bedienung von Zufahrt 2, nachdem alle Nachzügler abgefertigt wurden, d.h. nach $\Delta t_1 + \Delta n_1 / q_1^{\max}$ Zeiteinheiten. Bis dahin wachsen die prognostizierten Gesamtwartezeiten mit der konstanten Rate $\hat{n}_2(t)$. Damit erhält man:

$$\Delta \hat{W}_2^A = (\Delta t_1 + \Delta n_1 / q_1^{\max}) \hat{n}_2(t). \quad (4.22)$$

Dem stellen wir nun Fall (B) gegenüber, bei dem die Nachzügler gestoppt werden und während der Leerung von Zufahrt 2 warten müssen. Die Zunahme der prognostizierten Gesamtwartezeiten $\Delta \hat{W}_1^B$ setzt sich wieder aus zwei Teilen zusammen, dem ersten für die Erweiterung des Prognosehorizonts nach Gleichung (3.53) und dem zweiten für das kontinuierliche Anwachsen der prognostizierten Gesamtwartezeiten bis zum Beginn des Bedienvorgangs. Es gilt:

$$\Delta \hat{W}_1^B = \Delta \hat{W}_1^0(t) + (\tau_2^0 + \hat{g}_2(t)) \Delta n_1. \quad (4.23)$$

Der Term $\Delta \hat{W}_1^0(t)$ entspricht den Wartezeiten, die den Nachzüglern auf Zufahrt 1 beim Umschalten allein durch das erneute Abwarten der Zwischenzeit τ_i^0 ent-

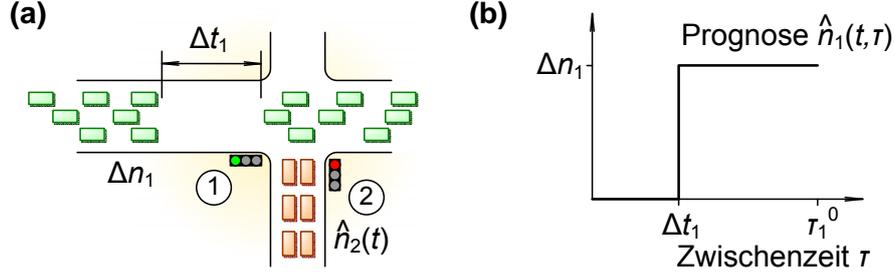


Abbildung 4.2.: (a) Auf Zufahrt 1 wurde eine Wartekolonne abgefertigt. Bis zur Ankunft des ersten von Δn_1 Nachzügler vergeht die Zeit Δt_1 . Auf Zufahrt 2 können $\hat{n}_2(t)$ Fahrzeuge mit maximaler Rate abgefertigt werden. (b) Die Prognose der abzufertigenden Fahrzeuganzahl $\hat{n}_1(t, \tau)$ hängt von der angenommenen Zwischenzeit τ ab.

stehen. Diese Wartezeiten sind umso größer, je kleiner die Zeitlücke Δt_1 bis zur Ankunft der Nachzügler ist. Keine Wartezeiten entstehen, wenn die Nachzügler erst nach einer Zeitlücke $\Delta t_1 = \tau_1^0$ ankommen. Der genaue Wert von $\Delta \hat{W}_i^0(t)$ berechnet sich nach Gleichung (3.53) und dem gegebenen Prognoseverlauf (4.20) mit der Formel:

$$\Delta \hat{W}_i^0(t) = \int_{\tau_1(t)}^{\tau_1^0} \hat{n}_1(t, \tau) d\tau = (\tau_1^0 - \Delta t_1) \Delta n_1. \quad (4.24)$$

Setzt man diesen Ausdruck in Gleichung (4.23) ein, erhält man für die rechte Seite der Wartezeitenbilanz:

$$\Delta \hat{W}_1^B = (\tau_1^0 - \Delta t_1 + \tau_2^0 + \hat{g}_2(t)) \Delta n_1. \quad (4.25)$$

Damit ist Alternative (A), die Verlängerung der Freigabezeit, unter folgender Bedingung optimal:

$$\frac{\hat{n}_2(t)}{\tau_1^0 - \Delta t_1 + \tau_2^0 + \hat{g}_2(t)} < \frac{\Delta n_1}{\Delta t_1 + \Delta g_1} \quad \text{mit} \quad \Delta g_1 = \frac{\Delta n_1}{q_1^{\max}}. \quad (4.26)$$

Dieses Kriterium soll wieder zur Spezifikation eines Prioritätsindex $\pi_i(t)$ herangezogen werden. Weil diese Spezifikation aber weder im Widerspruch zu den vorherigen Spezifikationen (4.8) und (4.17) stehen darf, noch von den verwendeten Parametern abhängen soll, sind zunächst einige Verallgemeinerungen notwendig.

4.2.3. Verallgemeinerte dynamische Priorisierung

Im vierten und letzten Schritt wird aus den Ergebnissen der drei vorherigen Abschnitte eine verallgemeinerte Priorisierungsregel abgeleitet. Diese Regel soll, wenn sie auf jede der drei betrachteten Szenarien 4.i bis 4.iii angewendet wird, das opti-

male Verhalten zeigen. Aber sie soll nicht explizit von den speziellen Parametern Δt_1 und Δn_1 abhängig sein, sondern ausschließlich von der als bekannt angenommenen Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ bzw. der daraus abgeleiteten Prognose für $\hat{n}_i(t, \tau)$.

Diese Verallgemeinerung erfolgt in zwei Schritten. Zuerst wird ein verallgemeinerter Strafterm für die Unterbrechung des aktuellen Bedienvorgangs abgeleitet, der in den beiden Szenarien 4.ii und 4.iii die Summanden $(\tau_2^0 - \tau_2(t))$ bzw. $(\tau_1^0 - \Delta t_1)$ ersetzt. Im zweiten Schritt wird dann untersucht, wie Nachzügler mit beliebig verteilten Ankunftszeitpunkten zu behandeln sind. Die Ergebnisse werden abschließend in Abschnitt 4.3 zusammengefasst.

Strafterm für die Unterbrechung von Bedienvorgängen

Die Unterbrechung des Bedienvorgangs für eine Zufahrt i erhöht die prognostizierten Gesamtwartezeiten um den sich aus Gleichung (3.53) ergebenden Wert $\Delta \hat{W}_i^0(t)$. Dieser Wert entspricht den Wartezeiten bei einer fiktiven Unterbrechung des Bedienvorgangs für null Zeiteinheiten.²⁶ Wartezeiten entstehen deshalb, weil bis zum Beginn der Freigabe erneut die Zwischenzeit τ_i^0 verstreichen muss. Die Anzahl der davon betroffenen Fahrzeuge ist die Anzahl der Fahrzeuge $\hat{n}_i(t, \tau_i^0)$, die anschließend mit maximaler Rate abgefertigt werden können. Wie lange dabei jedes einzelne dieser Fahrzeuge warten muss, hängt von der konkreten Verteilung der Ankunftszeitpunkte ab. Die mittleren Wartezeiten der betroffenen Fahrzeuge wird als Strafterm $\tau_i^{\text{pen}}(t)$ definiert. Die Bezeichnung „pen“ soll für „penalty“ stehen. Der Strafterm $\tau_i^{\text{pen}}(t)$ sei definiert als:

$$\tau_i^{\text{pen}}(t) := \frac{\Delta \hat{W}_i^0(t)}{\hat{n}_i(t, \tau_i^0)} = \frac{\int_{\tau_i(t)}^{\tau_i^0} \hat{n}_i(t, \tau) d\tau}{\hat{n}_i(t, \tau_i^0)}. \quad (4.27)$$

Für $\hat{n}_i(t, \tau_i^0) = 0$ wird $\tau_i^{\text{pen}}(t) = 0$ gesetzt. Weil $\hat{n}_i(t, \tau)$ nach Gleichung (3.29) monoton steigend in τ ist, gilt stets $\tau_i^{\text{pen}}(t) \leq \tau_i^0 - \tau_i(t)$. Diesen Maximalwert hat $\tau_i^{\text{pen}}(t)$ dann, wenn jedes der betroffenen Fahrzeuge die gesamte Zwischenzeit warten muss. Zu beachten ist, dass nur die bereits verstrichene Zwischenzeit $\tau_i^0 - \tau_i(t)$, nicht aber die noch verbleibende Zwischenzeit $\tau_i(t)$ in den Strafterm $\tau_i^{\text{pen}}(t)$ eingeht. Die „Strafe“ bezieht sich also nur auf die durch die Unterbrechung *zusätzlich* entstehende Wartezeit, nicht auf die ohnehin notwendige.

Die Definition von $\tau_i^{\text{pen}}(t)$ deckt die Ergebnisse aus den vorherigen Abschnitten vollständig ab. Im Szenario 4.ii kann der Term $(\tau_2^0 - \tau_2(t))$ aus Gleichung (4.14) uneingeschränkt durch $\tau_i^{\text{pen}}(t)$ ersetzt werden. Im Szenario 4.iii gilt das Gleiche für den Term $(\tau_1^0 - \Delta t_1)$ aus Gleichung (4.24).

²⁶Vgl. Abschnitt 3.3.4

Deshalb lässt sich der dynamische Prioritätsindex $\pi_i(t)$ für eine momentan nicht bediente Zufahrt $i \neq \sigma(t)$ verallgemeinert wie folgt angeben:

$$\pi_i(t) = \frac{\hat{n}_i(t)}{\tau_{\sigma(t)}^{\text{pen}}(t) + \tau_i^0 + \hat{g}_i(t)}, \quad \text{falls } i \neq \sigma(t). \quad (4.28)$$

Der Strafterm $\tau_{\sigma(t)}^{\text{pen}}(t)$ bezieht sich auf die momentan bediente Zufahrt $\sigma(t)$ und berechnet sich über die Formel (4.27).

Zeitlückenverteilung

Nun ist zu verallgemeinern, wie nach erfolgter Warteschlangenleerung mit später ankommenden Fahrzeugen zu verfahren ist. Betrachtet wird dazu eine Zufahrt i , die zum Zeitpunkt t bedient wird ($\sigma(t) = i$) und deren Warteschlange leer ist ($n_i(t) = 0$). Aus der Verteilung der erwarteten Ankunftszeitpunkte der später eintreffenden Fahrzeuge kann dann direkt auf den Verlauf der Prognose $\hat{n}_i(t, \tau)$ für $\tau \in [0, \tau_i^0]$ geschlossen werden, wie beispielsweise im Szenario 4.iii mit Gleichung (4.20). Allgemein lässt sich Folgendes sagen: Wird die Freigabezeit um

$$\tau + \hat{g}_i(t, \tau) \quad (4.29)$$

Zeiteinheiten verlängert, können währenddessen genau

$$\hat{n}_i(t, \tau) \quad (4.30)$$

Fahrzeuge abgefertigt werden. Die Größe τ entspricht dabei der Summe der Nettozeitlücken²⁷ zwischen den Fahrzeugen. Während der verlängerten Freigabezeit (4.29) werden also τ Zeiteinheiten lang keine Fahrzeuge abgefertigt.

Diese Überlegung soll nun auf das Szenario 4.iii übertragen werden. Dort blieb die Zeitspanne Δt_1 bis zur Ankunft der Nachzügler ungenutzt. Nach ihrer Abfertigung wurde die Freigabezeit sofort beendet. Während der Freigabezeit wurden also $\tau = \Delta t_1$ Zeiteinheiten lang keine Fahrzeuge abgefertigt. Die rechte Seite von Bedingung (4.26) soll als Kandidat für den Prioritätsindex einer bedienten Zufahrt ohne Warteschlange verwendet werden. Setzt man dort anstatt der gegebenen Zeitlücke Δt_1 die allgemeine Zeitlückensumme τ ein, erhält man

$$\frac{\hat{n}_i(t, \tau)}{\tau + \hat{g}_i(t, \tau)}. \quad (4.31)$$

Im Szenario 4.iii ist dieser Ausdruck nur für $\tau = \Delta t_1$ maximal. Für kleinere τ verschwindet \hat{n}_i und damit der Zähler,²⁸ wohingegen für größere τ bei konstantem

²⁷In einer kompakten Kolonne sind die Nettozeitlücken gleich null.

²⁸aufgrund des Prognoseverlaufs (4.20), vgl. Abb. 4.2 (auf S. 79)

\hat{n}_i bzw. \hat{g}_i der Nenner wächst. Wird also ein dynamischer Prioritätsindex $\pi_i(t)$ mit dem Ausdruck (4.31) spezifiziert, und wird für τ der Wert gewählt, der diesen Ausdruck maximiert, so führt $\pi_i(t)$ im Szenario 4.iii zum optimalen Umschaltkriterium (4.26).

Für den Fall, dass eine Zufahrt i zum Zeitpunkt t bedient wird ($\sigma(t) = i$) und ihre Warteschlange leer ist ($n_i(t) = 0$), wird ihr dynamischer Prioritätsindex $\pi_i(t)$ wie folgt spezifiziert:

$$\pi_i(t) = \max_{0 < \tau \leq \tau_i^0} \frac{\hat{n}_i(t, \tau)}{\tau + \hat{g}_i(t, \tau)}, \quad \text{falls } \sigma(t) = i \quad \text{und} \quad n_i(t) = 0. \quad (4.32)$$

Der Zähler entspricht der abzufertigenden Fahrzeuganzahl, der Nenner der dafür benötigten Freigabezeitverlängerung nach Gleichung (4.29). Damit lässt sich der Quotient wieder als die erwartete effektive Ausflussrate interpretieren. Die Summe der Nettozeitlücken τ , während der keine Abfertigung erfolgt, ist dabei so zu wählen, dass die erwartete effektive Ausflussrate maximal ist.

Gleichung (4.32) gibt den Prioritätsindex für eine momentan bediente Zufahrt ohne Warteschlange an. Den Fall mit Warteschlange beschreibt dagegen die erste Zeile aus Gleichung (4.17). Beide Gleichungen lassen sich wie folgt zusammenfassen:

$$\pi_i(t) = \max_{\tau_i(t) < \tau \leq \tau_i^0} \frac{\hat{n}_i(t, \tau)}{\tau + \hat{g}_i(t, \tau)}, \quad \text{falls } \sigma(t) = i. \quad (4.33)$$

Diese Zusammenfassung ist als folgenden Gründen zulässig: Im Fall ohne Warteschlange ist die verbleibende Umschaltzeit $\tau_i(t) = 0$, und Gleichung (4.33) geht direkt in Gleichung (4.32) über. Im anderen Fall mit Warteschlange ist der Quotient aus Gleichung (4.33) für den kleinsten Wert von $\tau \in (\tau_i(t), \tau_i^0]$ maximal. Das folgt aus der Annahme, dass im Szenario 4.ii keine weiteren Fahrzeuge hinzukommen. Damit geht auch in diesem Fall Gleichung (4.33) in Gleichung (4.17) über. Folglich deckt Gleichung (4.33) beide Fälle, mit und ohne Warteschlange, vollständig ab.

Damit sind alle notwendigen Verallgemeinerungen abgeschlossen. Die Ergebnisse werden nun zusammengefasst.

4.3. Spezifikation des Optimierungsverfahrens

In den vorherigen Abschnitten ist es gelungen, das Problem der wartezeitenminimalen Abfertigung aller im Wirkungsbereich²⁹ befindlichen Fahrzeuge durch die Spezifikation einer allgemeinen Priorisierungsregel optimal zu lösen. Der Wirkungsbereich umfasst all jene Fahrzeuge, die in einem momentan (erneut) be-

²⁹Vgl. Abb. 3.6 (auf S. 57)

ginnenden Bedienvorgang mit maximaler Rate abgefertigt werden können. Die dynamische Priorisierungsregel wird nun zusammengefasst.

4.3.1. Schaltprogramm und Zusammenfassung

Priorisierungsregel

Zu jedem Zeitpunkt t wird unter allen Zufahrten i mit $\pi_i(t) > 0$ die Zufahrt zur Bedienung ausgewählt, deren Prioritätsindex $\pi_i(t)$ am größten ist. Bezeichnet $\sigma(t)$ den Index i der zur Zeit t ausgewählten Zufahrt, so lässt sich die Schaltregel wie folgt angeben:

$$\sigma(t) = \arg \max_i \pi_i(t). \quad (4.34)$$

Falls es keine Zufahrt i mit positivem Prioritätsindex $\pi_i(t) > 0$ gibt, d.h. $\sigma(t) = \emptyset$, sind alle Zufahrten gesperrt. Die Freigabe einer ausgewählten Zufahrt $i = \sigma(t)$ beginnt erst nach Ablauf der Zwischenzeit τ_i^0 .

Der Prioritätsindex $\pi_i(t)$ ist wie folgt spezifiziert:

$$\pi_i(t) = \begin{cases} \max_{\tau_i(t) < \tau \leq \tau_i^0} \frac{\hat{n}_i(t, \tau)}{\tau + \hat{g}_i(t, \tau)}, & \text{falls } i = \sigma(t) \\ \frac{\hat{n}_i(t, \tau_i^0)}{\tau_{\sigma(t)}^{\text{pen}}(t) + \tau_i^0 + \hat{g}_i(t, \tau_i^0)}, & \text{falls } i \neq \sigma(t). \end{cases} \quad (4.35)$$

Hierbei ergeben sich die prognostizierten Freigabezeiten $\hat{g}_i(t, \tau)$ und abzufertigenden Fahrzeuganzahlen $\hat{n}_i(t, \tau)$ aus den Gleichungen (3.19) bzw. (3.20), sowie der Strafterm $\tau_{\sigma(t)}^{\text{pen}}(t)$ aus Gleichung (4.27). Die Werte des Prioritätsindex $\pi_i(t)$ liegen im Bereich $[0, q_i^{\text{max}}]$ und haben die Einheit des Flusses, d.h. Fahrzeuge je Sekunde.

Optimalität

Beim Optimierungsverfahren handelt es sich um ein heuristisches Lösungsverfahren zur Minimierung der Gesamtwartezeiten. Indem es zu jedem Zeitpunkt t die Zufahrt mit größtem Prioritätsindex $\pi_i(t)$ bedient, stellt sich von selbst die „momentan“ optimale Bedienreihenfolge ein. Kommen im betrachteten Zeitraum keine neuen Fahrzeuge hinzu, so werden die Warteschlangen mit minimalen Gesamtwartezeiten geleert. Kommen anderenfalls neue Fahrzeuge hinzu, so kann sich jederzeit eine neue „momentan“ optimale Bedienreihenfolge einstellen. Gegebenenfalls wird der aktuelle Bedienvorgang unterbrochen und den Fahrzeugen einer anderen Zufahrt Vorrang gewährt. Dass dabei zusätzliche Umschaltverluste entstehen, wird berücksichtigt.

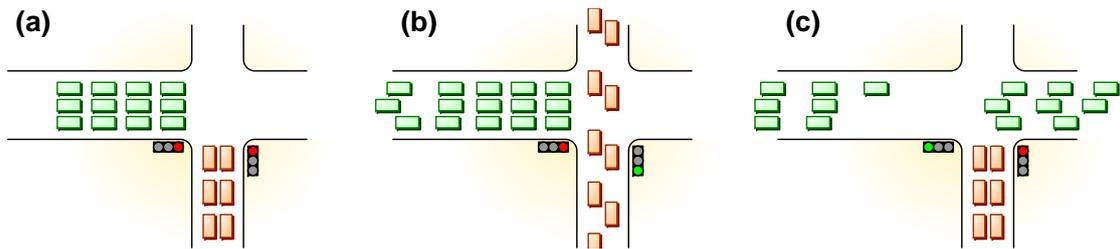


Abbildung 4.3.: Drei Problemstellungen: (a) In welcher Reihenfolge sind Warteschlangen ohne Zuflüsse abzufertigen? (b) Wann ist ein Bediensvorgang abubrechen um einen anderen vorzuziehen? (c) Welche Zeitlücken im Verkehrsstrom können zur Freigabezeitverlängerung maximal akzeptiert werden? Jede dieser drei Problemstellungen wird mit der Schaltregel (4.34), „Bediene die Zufahrt mit höchster Priorität“, wartezeitenminimal gelöst. Der Prioritätsindex $\pi_i(t)$ berechnet sich in jedem Fall mit der allgemeinen Formel (4.35).

Komplexität

In den Prioritätsindex $\pi_i(t)$ einer Zufahrt i gehen nur ihre eigenen Zustandsgrößen und, falls Zufahrt i zur Zeit t nicht bedient wird, außerdem die Zustandsgrößen der momentan bedienten Zufahrt $\sigma(t)$ ein. Aufgrund dieser Eigenschaft wächst der benötigte Rechenaufwand lediglich linear mit der Anzahl der Knotenzufahrten.

Zusammenfassung

Die Spezifikation (4.35) des Prioritätsindex $\pi_i(t)$ wurde aus den jeweils optimalen Entscheidungskriterien für die drei Problemfälle der Szenarien 4.i bis 4.iii abgeleitet und verallgemeinert. Diese sind nochmals in Abb. 4.3 zusammengefasst. Durch Anwenden der Schaltregel (4.34), das heißt „Bediene die Zufahrt mit höchster Priorität“, wird also jeder dieser drei Problemfälle optimal gelöst. Die heuristische Vermutung hierbei ist, dass sich die Priorisierungsregel in davon abweichenden Fällen ebenfalls günstig verhält. Diese Vermutung wird in den nachfolgenden Abschnitten genauer untersucht.

4.3.2. Schaltverhalten

Das Optimierungsverfahren wird in Abschnitt 7.2.1 mit numerischen Simulationen einem umlaufbasierten Verfahren gegenübergestellt. Mit welchem der beiden Verfahren geringere Gesamtwartezeiten entstehen, ist von der Knotenpunktauslastung abhängig. Bei geringem Verkehrsaufkommen sind die Gesamtwartezeiten unter Verwendung des Optimierungsverfahren am kleinsten. Dagegen tritt bei höherem Verkehrsaufkommen der Fall auf, dass einige Knotenzufahrten zu kurz oder

gar nicht bedient werden. Dieser instabile Fall wird in Abschnitt 4.3.3 genauer erläutert.

Grüne Wellen für Kolonnen (Szenario 4.iv)

Mit dem Optimierungsverfahren können herannahende kompakte Kolonnen³⁰ verzögerungsfrei als „Grüne Welle“ bedient werden. Nähert sich der Warteschlange einer momentan nicht bedienten Knotenzufahrt i eine kompakte Kolonne, so springt die Anzahl der abzufertigenden Fahrzeuge $\hat{n}_i(t)$, wie im Abschnitt 3.3.2 mit Gleichung (3.30) gezeigt, abrupt auf einen höheren Wert. Der Sprung ist umso größer, je mehr Fahrzeuge die Kolonne umfasst. Aufgrund der dynamischen Reoptimierung wird dieser Kolonne nun spontan Vorrang gegeben und die Bedienung einer anderen Zufahrt gegebenenfalls unterbrochen, wenn dadurch am lokalen Knotenpunkt geringere Wartezeiten prognostiziert werden als durch das Anhalten der Kolonne. Außerdem geschieht der Sprung von $\hat{n}_i(t)$ exakt zu dem Zeitpunkt, zu dem gerade noch so viel Zeit bis zur Ankunft der Kolonne verbleibt, um die Zwischenzeit τ_i^0 abzuwarten und um sowohl die bereits wartenden als auch die noch vor der Kolonne ankommenden Fahrzeuge abzufertigen. Die Kolonne passiert den Knotenpunkt dann verzögerungsfrei. Dieses Szenario kann in Abb. 3.6 (auf S. 57) nachvollzogen werden.

Bündelungseffekt bei Einzelankünften (Szenario 4.v)

Es wird nun der Fall betrachtet, dass die Fahrzeuge auf Zufahrt i nicht in einer kompakten Kolonne sondern einzeln ankommen. Der Bedienvorgang wird gestartet, sobald der Prioritätsindex $\pi_i(t)$ größer ist als auf den anderen Zufahrten. Das heißt, das Verhältnis aus abzufertigender Fahrzeuganzahl und dafür benötigter Bedienzeit muss einen bestimmten Wert überschreiten. Solange dies nicht der Fall ist, werden die ankommenden Fahrzeuge aufgehalten. Dies führt zu einer Bündelung des Verkehrsstroms. Das Optimierungsverfahren leitet einen Phasenwechsel erst dann ein, wenn der Umschaltverlust durch die im Anschluss höhere Abfertigungsrate kompensiert wird. Ob die Freigabezeit nach der Leerung der Warteschlange fortgesetzt wird, hängt von der Zeitlückenverteilung der danach ankommenden Fahrzeuge ab.³¹ Die Freigabezeit wird verlängert, wenn das Verhältnis aus abzufertigender Fahrzeuganzahl und dafür benötigter Bedienzeit größer ist als auf den anderen Zufahrten.

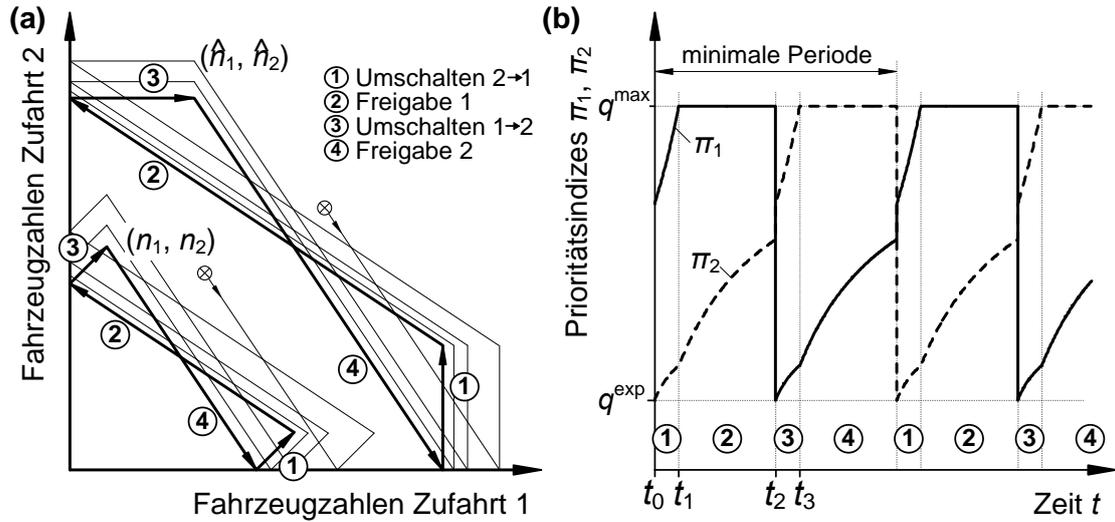


Abbildung 4.4.: (a) Optimaler Grenzyklus für einen Knotenpunkt mit zwei Zufahrten und konstanten Zuflussraten. Für den gezeigten Kurvenverlauf wurden identische Ankunftsraten q^{exp} und identische maximale Ausflussraten q^{max} auf beiden Zufahrten angenommen. Die Trajektorien der Warteschlangenlänge (n_1, n_2) und der abzufertigenden Fahrzeuganzahl (\hat{n}_1, \hat{n}_2) nähern sich exponentiell dem optimalen Grenzyklus. (b) Verlauf der Prioritätsindizes $\pi_1(t)$ und $\pi_2(t)$ im stationären Fall. Die Kurvenverläufe aus (a) und (b) sowie das Konvergenzverhalten werden in Anhang A.3 analytisch berechnet.

Optimaler Grenzyklus (Szenario 4.vi)

Nun soll untersucht werden, wie sich das Optimierungsverfahren bei konstanten Ankunftsraten q_i^{exp} verhält. In diesem Fall, so zeigen Savkin und Evans in Referenz [162], muss die stationäre Lösung eines stabilen deterministischen Schaltprogramms periodisch sein.

Für einen Knotenpunkt mit zwei Zufahrten wird der Zeitverlauf der Zustandsgrößen $n_i(t)$ und $\hat{n}_i(t)$ und des Prioritätsindex $\pi_i(t)$ in Anhang A.3 analytisch berechnet. Für Zuflussraten q_i^{exp} , unter denen das Optimierungsverfahren stabil ist, streben die Zustandsgrößen von Periode zu Periode exponentiell gegen eine stationäre periodische Lösung. Dies wird insbesondere durch die lineare Iterationsvorschrift (A.44) deutlich, die angibt, wie sich der Wert \hat{n}^{k+1} in Periode $k+1$ aus dem Wert \hat{n}^k der vorherigen Periode k ergibt:

$$\hat{n}^{k+1} = a + b \hat{n}^k. \quad (\text{A.44})$$

³⁰mit maximaler Flussrate in ihrem Inneren

³¹Vgl. Szenario 4.iii in Abschnitt 4.2.2

Hierbei ergeben sich die beiden Konstanten a und b aus den Zuflussraten und den Zwischenzeiten. Die Zustandsgröße $\hat{n}_i(t)$ konvergiert folglich mit $|b| < 1$, was bei erfüllter Stabilitätsbedingung (2.3) der Fall ist, gegen eine stationäre periodische Lösung, den sogenannten „Grenzyklus“ (engl.: „limit cycle“). Im betrachteten Szenario ist dieser Grenzyklus global stabil und optimal. Er ist in Abb. 4.4(a) dargestellt. Abbildung 4.4(b) zeigt den zugehörigen Verlauf der Prioritätsindizes $\pi_i(t)$.

4.3.3. Instabiles Verhalten

Im Beispiel des Kumar-Seidman-Netzwerks³² sind Rückkopplungseffekte zwischen benachbarten Knotenpunkten für die Instabilität eines Online-Steuerungsverfahrens verantwortlich. Aber Instabilitäten können auch an isolierten Knotenpunkten auftreten, wie beispielsweise Duenyas und van Oyen in Referenz [54] gezeigt haben.

Szenario 4.vii

Es wird die in Abb. 4.5 gezeigte Situation betrachtet. Auf Knotenzufahrt 1 sei die Zuflussrate q_1^{exp} konstant. Zum betrachteten Zeitpunkt t wird Zufahrt 1 bei geleerter Warteschlange weiter bedient, während auf der zweiten Zufahrt Fahrzeuge warten.

Zeitliche Entwicklung

Das Optimierungsverfahren fährt mit der Bedienung von Zufahrt 1 solange fort wie die Ungleichung $\pi_1(t) > \pi_2(t)$ erfüllt ist. Der Prioritätsindex $\pi_1(t)$ für Zufahrt 1 berechnet sich mit Hilfe der ersten Zeile aus Gleichung (4.35). Aufgrund der konstanten Ankunftsrate gilt für die prognostizierte Freigabezeit $\hat{g}_1(t, \tau)$ folgender

³²Vgl. Anhang A.1

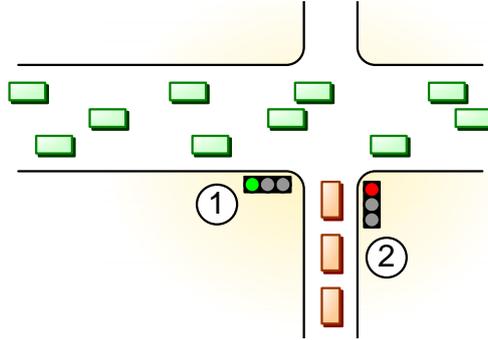


Abbildung 4.5.: Falls die konstante Ankunftsrate q_1^{exp} auf Zufahrt 1 größer ist als die maximale Ausflussrate q_2^{max} auf Zufahrt 2, wird Zufahrt 2 niemals bedient. Das Optimierungsverfahren ist in diesem Fall instabil.

linearer Zusammenhang³³ bezüglich τ :

$$\hat{g}_1(t, \tau) = \frac{\tau}{1 - q_1^{\text{exp}}/q_1^{\text{max}}}. \quad (4.38)$$

Der Quotient $\hat{n}_i(t, \tau)/(\tau + \hat{g}_i(t, \tau))$, der nach Gleichung (4.35) bezüglich τ zu maximieren ist, ist in diesem Fall unabhängig von τ . Der Prioritätsindex $\pi_1(t)$ hat damit den Wert

$$\pi_1(t) = q_1^{\text{exp}} \quad (4.39)$$

und ist damit ebenso wie die Zufussrate konstant. Ohne den Prioritätsindex $\pi_2(t)$ zu berechnen ist bekannt,³⁴ dass er seinen Maximalwert q_2^{max} niemals überschreiten wird:

$$\pi_2(t) < q_2^{\text{max}}. \quad (4.40)$$

Folglich wird das Optimierungsverfahren niemals Zufahrt 2 zur Bedienung auswählen, wenn die konstante Ankunftsrate auf Zufahrt 1 größer ist als die maximale

³³Zur Berechnung von $\hat{g}_1(t, \tau)$ wird die Bestimmungsgleichung (3.20) verwendet. Aufgrund der konstanten Ankunftsrate q_1^{exp} kann die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t + \tau + \hat{g}_1)$ durch folgende Geradengleichung ausgedrückt werden:

$$N_1^{\text{exp}}(t + \tau + \hat{g}_1) = N_1^{\text{exp}}(t) + (\tau + \hat{g}_1)q_1^{\text{exp}}. \quad (4.36)$$

Da die Warteschlange zum Zeitpunkt t leer ist, gilt weiterhin $N_1^{\text{exp}}(t) = N_1^{\text{out}}(t)$. Setzt man beides in die Bestimmungsgleichung (3.20) ein, vereinfacht sich diese zu folgendem Ausdruck:

$$\hat{g}_1 q_1^{\text{max}} = (\tau + \hat{g}_1) q_1^{\text{exp}}. \quad (4.37)$$

Dieser lässt sich nach der gesuchten Größe $\hat{g}_1(t, \tau)$ umstellen. Als Ergebnis erhält man Gleichung (4.38).

³⁴Vgl. Abschnitt 4.3.1

Ausflussrate von Zufahrt 2, also wenn gilt:

$$q_1^{\text{exp}} > q_2^{\text{max}}. \quad (4.41)$$

In diesem Fall ist das Optimierungsverfahren instabil.

Diskussion

Das instabile Verhalten kann selbst bei mittleren Knotenpunktauslastungen auftreten. Ist die Zufahrt 1 beispielsweise dreispurig und Zufahrt 2 einspurig, wird Zufahrt 2 nicht mehr bedient, wenn die Auslastung $\lambda_1 = q_1^{\text{exp}}/q_1^{\text{max}}$ von Zufahrt 1 größer als $1/3$ ist. Hierbei ist jedoch zu beachten, dass konstante Ankunftsdaten und damit gleichverteilte Ankunftszeitpunkte angenommen wurden. Zwar würde das Optimierungsverfahren zufällige Zeitlücken im Verkehrsstrom von Zufahrt 1 sofort zur Bedienung von Zufahrt 2 ausnutzen. Aber es wird ein Schaltprogramm benötigt, das sich unter allen Umständen³⁵ stabil verhält.

Das betrachtete Szenario 4.vii macht ein prinzipielles Problem lokal optimierender Verfahren deutlich. Bei einem ausschließlich lokal optimierenden Verfahren wird ein Schalteingriff nur dann in Betracht gezogen, wenn der Optimierungshorizont größer ist als die Zeitspanne, nach der sich der Schalteingriff auszahlt. Diese Zeitspanne kann jedoch beliebig groß sein. Der Optimierungshorizont ist dagegen begrenzt, beispielsweise aufgrund des endlichen Prognosehorizonts oder aufgrund der maximal nutzbaren Rechenkapazität.³⁶ Deshalb ist es erforderlich, neben der lokalen Optimierung Mechanismen zu verwenden, die stabiles Verhalten garantieren und maximale Sperrzeiten einhalten. Dem Entwurf eines solchen Mechanismus wendet sich das nächste Kapitel zu.

³⁵solange die hinreichende Stabilitätsbedingung (2.10) erfüllt ist

³⁶Vgl. Abschnitte 1.1.3 und 3.1.1

Kapitel 5.

Netzwerkweite Stabilität

Kurzfassung

Dieses Kapitel stellt den zweiten Baustein des Verfahrens zur dezentralen Online-Steuerung vor, das „Stabilisierungsverfahren“.

Die Ursachen für instabiles Verhalten dezentraler Verfahren sind vielfältig und kaum vollständig zu erfassen. Immer jedoch geht instabilem Verhalten die ineffiziente Nutzung der Abfertigungskapazität voraus. An diesem Punkt setzt das hier vorgestellte Stabilisierungsprinzip an. Zuerst wird ein hinreichendes Stabilitätskriterium für die Effizienz der Bedienvorgänge formuliert, und anschließend daraus ein stabiler Regler abgeleitet. Das Effizienzkriterium lässt sich dezentral auf die einzelnen Knotenpunkte anwenden und garantiert netzwerkweite Stabilität.

Dem Entwurf des stabilisierenden Reglers wird ein gewünschtes Schaltverhalten zugrunde gelegt: Jede Zufahrt soll im Mittel einmal innerhalb einer Wunschperiode T bedient werden. Das Messen der Zeit wird dabei auf das Zählen von Fahrzeugen zurückgeführt. Indem ein Teil des tatsächlichen Fahrzeugstroms durch virtuelle Fahrzeuge ersetzt wird, lässt sich mit einem zweiten Parameter T^{\max} die maximale Bedienperiode festlegen. Diese wird selbst dann eingehalten, wenn keine Fahrzeuge detektiert wurden. Damit ist netzwerkweite Stabilität auch bei Sensorausfällen und Prognoseungenauigkeiten sichergestellt.

Welche Wartezeiten mit diesem Verfahren entstehen, soll zunächst nur eine untergeordnete Rolle spielen. Im nachfolgenden Kapitel 6 werden beide Verfahren, das lokal optimierende und das netzwerkweit stabilisierende, miteinander kombiniert.

5.1. Stabilisierungsprinzip

Für die Stabilisierung geschalteter Warteschlangennetze sind in der Warteschlangentheorie sogenannte „local supervisory mechanisms“ bekannt [98, 54, 32, 33, 148]. Das im Folgenden vorgestellte Verfahren baut auf diesen Mechanismen auf. Gleichzeitig sollen diese Mechanismen jedoch um den Aspekt der Bedienperioden, das heißt um ein präzises Modell für die Dauer der Sperrzeiten, erweitert werden. Deshalb soll das Stabilisierungsprinzip im Folgenden neu motiviert und hergeleitet werden.

5.1.1. Lokales Effizienzkriterium

Bedieneffizienz

Ein Bedienvorgang soll dahingehend bewertet werden, wie *effizient* die Ressource Bedienzeit zur Abfertigung von Fahrzeugen genutzt wird. Im theoretischen Fall, dass während der gesamten Bedienzeit die Fahrzeuge mit maximaler Rate abgefertigt werden, soll die Bedieneffizienz 100% betragen. Weil aber der Ausfluss zu Beginn jedes Bedienvorgangs für die Dauer der Zwischenzeit gesperrt ist, wird die Bedieneffizienz immer kleiner als 100% sein. Genauer gesagt, wird im Folgenden unter der Effizienz eines Bedienvorgangs das Verhältnis aus minimal benötigter Freigabezeit und insgesamt aufgebrauchter Bedienzeit verstanden:

$$\text{Bedieneffizienz} = \frac{\text{minimal benötigte Freigabezeit}}{\text{insgesamt aufgebrauchte Bedienzeit}}. \quad (5.1)$$

Die Effizienz eines Bedienvorgangs für Knotenzufahrt i wird mit η_i bezeichnet. Zur Abfertigung von n_i Fahrzeugen mit maximaler Rate q_i^{\max} wird die Freigabezeit n_i/q_i^{\max} benötigt. Die insgesamt aufgebrauchte Bedienzeit hingegen setzt sich aus der Zwischenzeit τ_i^0 und der tatsächlichen Freigabezeit g_i zusammen. Die Bedieneffizienz η_i berechnet sich folglich mit:

$$\eta_i = \frac{n_i/q_i^{\max}}{\tau_i^0 + g_i}. \quad (5.2)$$

Zur Berechnung der mittleren Effizienz $\bar{\eta}_i$ mehrerer Bedienvorgänge ist das harmonische Mittel³⁷ anzuwenden.

³⁷Es ist das harmonische und nicht das arithmetische Mittel zu verwenden, weil die Quotienten η_i^b der einzelnen Bedienvorgänge b nicht gleichnamig sind. Es gilt $\bar{\eta}_i = 1/(\sum_b f^b/\eta_i^b)$, wobei $f^b = (\tau_i^0 + g_i^b)/\sum_{b'}(\tau_i^0 + g_i^{b'})$ dem Anteil der Bedienzeit von b an der Summe aller Bedienzeiten entspricht. Den gleichen Wert für $\bar{\eta}_i$ erhält man auch mit der Formel $\sum_b(n_i^b/q_i^{\max})/\sum_b(\tau_i^0 + g_i^b)$.

Hinreichende Stabilitätsbedingung

Die mittleren Ankunftsraten \bar{q}_i^{exp} jeder Knotenzufahrt seien bekannt. Wie aus den Betrachtungen in Abschnitt 2.3.2 und insbesondere aus Gleichung (2.4) folgt, gibt die mittlere Auslastung $\bar{\lambda}_i = \bar{q}_i^{\text{exp}}/q_i^{\text{max}}$ unmittelbar an, welcher Freigabezeitanteil zur Abfertigung des ankommenden Verkehrsstrom mindestens benötigt wird. Nur wenn auf der Zufahrt i der Anteil der Freigabezeit größer als $\bar{\lambda}_i$ ist, bleiben die Warteschlangen beschränkt.

Nun hat aber die minimal benötigte Freigabezeit an der für den Bedienvorgang insgesamt benötigten Zeit nur den Anteil η_i . Folglich erfordert der Mindest-Freigabezeitanteil $\bar{\lambda}_i$ einen Mindest-Bedienzeitanteil $\bar{\lambda}_i/\bar{\eta}_i$. Das heißt, Stabilität ist nur dann gegeben, wenn das Schaltprogramm mindestens den Teil $\bar{\lambda}_i/\bar{\eta}_i$ einer betrachteten Zeitspanne für die Bedienung von Zufahrt i aufbringt. Mit der Annahme, dass die Zufahrten eines Knotenpunktes einzeln bedient werden, muss die Summe der Mindest-Bedienzeitanteile kleiner als 100% sein:

$$\boxed{\sum_i \bar{\lambda}_i/\bar{\eta}_i < 1.} \quad (5.3)$$

Durch diese Stabilitätsbedingung drückt sich nicht nur eine Forderung an das Verhältnis aus Verkehrsnachfrage und Abfertigungskapazität eines Knotenpunktes aus, sondern auch an die Effizienz seines Schaltprogramms.

Netzwerkweite Stabilität

Die besondere Stärke des lokalen Effizienzkriteriums (5.3) liegt darin, dass es sich unabhängig auf jeden einzelnen Knotenpunkt im Netzwerk anwenden lässt. Wenn jeder einzelne Knotenpunkt im Netzwerk stabil ist, so sind die mittleren Ankunfts-raten \bar{q}_i^{exp} unabhängig von den an den Knotenpunkten verwendeten Schaltprogrammen. Die Schaltprogramme haben zwar einen Einfluss darauf, ob die Fahrzeuge an den anderen Knotenzufahrten früher oder später ankommen. Aber sie haben keinen Einfluss auf die mittlere Rate, in der die Fahrzeuge dort ankommen.³⁸ Netzwerkweite Stabilität ist garantiert, wenn jeder Knotenpunkt für sich das Kriterium (5.3) lokal erfüllt.

5.1.2. Stabilisierender Regler

Schaltregel

Das lokale Effizienzkriterium (5.3) lässt sich für den Entwurf eines stabilen Reglers verwenden. Das Schaltprogramm soll zunächst wie folgt lauten: Im Grundzustand

³⁸Hierbei sei angenommen, dass sich das Routenwahlverhalten infolge veränderter Schaltprogramme langsamer ändert als die Erfassung der mittleren Ankunftsrate.

sind alle Zufahrten gesperrt und der Bedienvorgang einer Zufahrt i beginnt erst dann, wenn dieser mit mindestens einer kritischen Bedieneffizienz η_i^{crit} durchgeführt werden kann. Damit ist die mittlere Bedieneffizienz $\bar{\eta}_i$ mindestens so groß wie der Schwellenwert η_i^{crit} :

$$\bar{\eta}_i \geq \eta_i^{\text{crit}}. \quad (5.4)$$

Es ist folglich hinreichend für die Stabilität des Reglers, wenn die Schwellenwerte η_i^{crit} das lokale Effizienzkriterium (5.3) erfüllen.³⁹ Wie die kritischen Schwellenwerte zu berechnen sind, wird in Abschnitt 5.2.1 behandelt.

Prognostizierte Bedieneffizienz

Zunächst ist noch zu klären, wie sich die Effizienz eines noch nicht begonnenen Bedienvorgangs prognostizieren lässt. Hierzu soll das Prognoseverfahren aus Abschnitt 3.3 herangezogen werden. Außerdem sei zunächst davon auszugehen, dass der Bedienvorgang nach vollständiger Leerung der Warteschlange sofort beendet wird. Für einen zum Zeitpunkt t beginnenden Bedienvorgang ist bekannt, dass während der prognostizierten Freigabezeit $\hat{g}_i(t)$ $\hat{n}_i(t)$ Fahrzeuge mit maximaler Rate abgefertigt werden können. Aus Gleichung (5.2) kann mit der Ersetzung $\hat{n}_i/q_i^{\text{max}} = \hat{g}_i$ die prognostizierte Bedieneffizienz $\hat{\eta}_i(t)$ also wie folgt definiert werden:

$$\hat{\eta}_i(t) = \frac{\hat{g}_i(t)}{\tau_i^0 + \hat{g}_i(t)}. \quad (5.5)$$

Weiterhin ist es wichtig zu zeigen, dass $\hat{\eta}_i(t)$ während des Zustands „keine Bedienung“ monoton in der Zeit wächst. An der Gleichung (5.5) ist sofort zu sehen, dass $\hat{\eta}_i$ wegen $d\hat{\eta}_i/d\hat{g}_i > 0$ für $\tau_i^0 > 0$ streng monoton in \hat{g}_i ist. Die Monotonie von \hat{g}_i bezüglich t wurde im Abschnitt 3.3.2 anhand von Gleichung (3.30) gezeigt. Daraus folgt die Beziehung

$$\frac{d\hat{\eta}_i}{dt} = \underbrace{\frac{d\hat{\eta}_i}{d\hat{g}_i}}_{>0} \cdot \underbrace{\frac{d\hat{g}_i}{dt}}_{\geq 0} \geq 0 \quad \text{im Zustand „keine Bedienung“,} \quad (5.6)$$

die zeigt, dass die prognostizierte Bedieneffizienz $\hat{\eta}_i(t)$ eines noch nicht begonnenen Bedienvorgangs monoton in der Zeit wächst. Damit ist gezeigt, dass $\hat{\eta}_i(t)$ den Schwellenwert η_i^{crit} nach einer gewissen Zeit erreichen wird. Eine obere Schranke für diese Zeitspanne wird in Abschnitt 5.4 eingeführt.

³⁹Duenyas und van Oyen [54] hatten in Simulationsexperimenten beobachtet, dass ein solcher Regler beispielsweise mit $\eta_i^{\text{crit}} = \eta^{\text{crit}} > \sum_i \bar{\lambda}_i 1$ stabiles Verhalten zeigt. Die Allgemeingültigkeit ihrer Beobachtung kann durch Einsetzen in (5.3) bestätigt werden.

5.2. Wunsch-Bedienperiode

Für gegebene Auslastungen $\bar{\lambda}_i$ mit $\sum_i \bar{\lambda}_i < 1$ sollen nun die kritischen Schwellenwerte gefunden werden, die das lokale Effizienzkriterium (5.3) erfüllen. Als zusätzliche Bedingung soll folgendes gewünschte Schaltverhalten erreicht werden: Jede Zufahrt i eines Knotenpunktes wird im Mittel einmal innerhalb einer vorgegebenen Wunsch-Bedienperiode T bedient.

5.2.1. Wahl des Schwellenwerts

Kritische Fahrzeuganzahl

Der kritische Schwellenwert lässt sich anstatt für die prognostizierte Bedieneffizienz $\hat{\eta}_i$ ebenso bezüglich der abzufertigenden Fahrzeuganzahl \hat{n}_i formulieren. Das Schaltprogramm des stabilen Reglers lautet dann wie folgt: Im Grundzustand sind alle Zufahrten gesperrt und der Bedienvorgang einer Zufahrt i beginnt erst, wenn dort die prognostizierte Anzahl abzufertigender Fahrzeuge $\hat{n}_i(t)$ mindestens so groß ist wie die kritische Fahrzeuganzahl n_i^{crit} , wenn also gilt:

$$\hat{n}_i(t) \geq n_i^{\text{crit}}. \quad (5.7)$$

Aus dem lokalen Effizienzkriterium (5.3) und der Definition (5.5) der Bedieneffizienz folgt dann als hinreichende Stabilitätsbedingung:

$$\sum_i \bar{\lambda}_i \left(\underbrace{\frac{n_i^{\text{crit}}/q_i^{\text{max}}}{\tau_i^0 + n_i^{\text{crit}}/q_i^{\text{max}}}}_{=n_i^{\text{crit}}} \right)^{-1} < 1. \quad (5.8)$$

Als Schwellenwert wird im Weiteren ausschließlich die kritische Fahrzeuganzahl n_i^{crit} verwendet.

Gewünschtes Schaltverhalten

Ein Bedienvorgang soll nicht dann beginnen, wenn die Warteschlange eine bestimmte Länge erreicht hat, sondern dann, wenn eine bestimmte Anzahl an Fahrzeugen mit maximaler Rate abgefertigt werden kann. Für genau diese Unterscheidung eignet sich das Prognoseverfahren aus Abschnitt 3.3. Dieses berücksichtigt nämlich auch jene Fahrzeuge, die erst während des Umschaltens oder während der Warteschlangenleerung eintreffen und mit abzufertigen sind. Abbildung 3.6 (auf S. 57) macht diesen Sachverhalt deutlich.

Es werden nun die Konsequenzen für den neuen Schwellenwert n_i^{crit} betrachtet, die sich aus dem gewünschten Schaltverhalten ergeben. Um jede Zufahrt im Mittel einmal innerhalb T zu bedienen, muss mit dem betrachteten Schaltprogramm ein

Bedienvorgang genau dann starten, wenn so viele Fahrzeuge mit maximaler Rate abgefertigt werden können, wie innerhalb der Zeitspanne T im Mittel ankommen. Für den Schwellenwert n_i^{crit} muss demnach gelten:

$$\boxed{n_i^{\text{crit}} = T \bar{q}_i^{\text{exp}}.} \quad (5.9)$$

Die lokale Stabilitätsbedingung (5.8) wird damit auf einen einzigen freien Parameter T reduziert. Welche Werte für T zulässig sind, soll nun genauer untersucht werden.

5.2.2. Hinreichende Stabilitätsbedingung

Setzt man Gleichung (5.9) in die hinreichende Stabilitätsbedingung (5.8) ein, so erhält man mit der Ersetzung $\bar{\lambda}_i = \bar{q}_i/q_i^{\text{max}}$ zunächst:

$$\sum_i \bar{\lambda}_i \left(\frac{\bar{\lambda}_i T}{\tau_i^0 + \bar{\lambda}_i T} \right)^{-1} < 1. \quad (5.10)$$

Nach einem Umformungsschritt

$$\sum_i \bar{\lambda}_i + \frac{1}{T} \sum_i \tau_i^0 < 1 \quad (5.11)$$

erhält man für die zulässigen Werte von T die Bedingung

$$\boxed{T > \frac{\sum_i \tau_i^0}{1 - \sum_i \bar{\lambda}_i} = T^{\text{cyc,min}}.} \quad (5.12)$$

Die hinreichende Stabilitätsbedingung (5.8) kann also auch als eine Bedingung an die Wunsch-Bedienperiode T formuliert werden. Die Bedienperiode T muss demnach mindestens so groß sein wie die Mindestumlaufzeit $T^{\text{cyc,min}}$ einer umlaufbasierten Steuerung nach Gleichung (2.8).

Das Schaltprogramm eines Knotenpunktes, nach dem eine Zufahrt i erst dann bedient wird, wenn dort mindestens $n_i^{\text{crit}} = \bar{\lambda}_i T$ Fahrzeuge mit maximaler Rate abgefertigt werden können und ansonsten alle Zufahrten sperrt, ist genau dann stabil, wenn sich der Knotenpunkt mit einer umlaufbasierten Steuerung mit der Umlaufzeit $T^{\text{cyc}} = T$ betreiben lässt. Liegt die Zeitspanne T im Extremfall sehr nahe an der Mindestumlaufzeit $T^{\text{cyc,min}}$, verschwinden die Zustände, in denen alle Zufahrten gesperrt sind, und das Schaltprogramm verhält sich ebenso wie eine umlaufbasierte Steuerung.

5.3. Spezifikation des Stabilisierungsverfahrens

Das soweit entwickelte Schaltprogramm lässt sich folgendermaßen formalisieren.

5.3.1. Priorisierungsregeln

Die Zufahrten werden in der Reihenfolge bedient, in der ihre Werte für $\hat{n}_i(t)$ den jeweiligen Schwellenwert n_i^{crit} überschreiten. Das Einhalten der Reihenfolge wird mit einer FIFO-Liste⁴⁰ Ω sichergestellt. Eine Zufahrt i wird erst dann bedient, wenn die Bedienvorgänge der vor ihr in der Liste vermerkten Zufahrten beendet sind.

Die Schaltregeln lauten wie folgt:

1. **Aufnahme:** Sobald an einer Zufahrt i mindestens n_i^{crit} Fahrzeuge mit maximaler Rate abgefertigt werden können, wird ihr Index i als letztes Element aufgenommen. Diese Regel gilt nur für Indizes i , die noch nicht in Ω enthalten sind.

$$\boxed{i \rightarrow \Omega, \quad \text{falls} \quad \hat{n}_i(t) \geq n_i^{\text{crit}} \quad \text{für alle } i \notin \Omega} \quad (5.13)$$

2. **Auswahl:** Ist die Liste Ω nicht leer, so ist aus ihr das erste Element ($\text{head } \Omega$) auszuwählen. Die entsprechende Zufahrt i ist für die Dauer der zugesicherten Freigabezeit g_i^s zu bedienen (siehe Abschnitt 5.3.2).

$$\boxed{\sigma(t) = \text{head } \Omega, \quad \text{falls} \quad \Omega \neq \emptyset} \quad (5.14)$$

Falls Ω leer ist, wird keine Zufahrt bedient.

3. **Entfernen:** Sobald die zugesicherte Freigabezeit g_i^s überschritten oder die Warteschlange vollständig geleert wurde, ist der aktuelle Bedienvorgang zu abzubrechen und der Index i der aktuell bedienten Zufahrt aus Ω zu entfernen.

$$\boxed{i \leftarrow \Omega, \quad \text{falls} \quad g_i(t) \geq g_i^s \quad \text{oder} \quad n_i(t) = 0} \quad (5.15)$$

Falls Zufahrt i weiterhin die Bedingung zur Aufnahme erfüllt, kann ihr Index unmittelbar wieder an die letzte Stelle von Ω gesetzt werden.

Nun ist noch zu klären, welche Freigabezeit g_i^s jeder Zufahrt zugesichert werden kann.

⁴⁰FIFO steht für First In First Out. Die Elemente werden in der Reihenfolge aus der Liste entnommen, in der sie aufgenommen wurden.

5.3.2. Zuweisung der Freigabezeiten

Die Zuteilung der zugesicherten Freigabezeit g_i^s soll sich aus der Lösung eines linearen Optimierungsproblems ergeben. Die Linearisierung dieses Zuteilungsproblems lässt sich darauf zurückführen, dass die Fahrzeuge während der Freigabezeit ausschließlich mit der konstanten Rate q_i^{\max} abgefertigt werden [62, 65]. Es gelten die folgenden drei linearen Nebenbedingungen und die folgende lineare Zielfunktion:

Lineare Nebenbedingungen

- (i) Die zugesicherte Freigabezeit g_i^s jeder Zufahrt i muss groß genug sein, um die während der Wunsch-Bedienperiode T ankommenden Fahrzeuge abfertigen zu können:

$$g_i^s \geq \bar{\lambda}_i T \quad \text{für alle } i. \quad (5.16)$$

- (ii) Die zugesicherte Freigabezeit g_i^s soll nicht größer sein, als zur Leerung der Warteschlange benötigt wird:

$$g_i^s \leq g_i^{s,\max} \quad \text{für alle } i. \quad (5.17)$$

Es wird zunächst angenommen, dass die zur Leerung benötigte Freigabezeit mit $g_i^{s,\max}$ gegeben ist.⁴¹ In Abschnitt 5.3.3 wird diese Nebenbedingung durch einen heuristischen Ansatz ersetzt.

- (iii) Die Summe der zugesicherten Freigabezeiten g_i^s aller Zufahrten i darf zusammen mit den Zwischenzeiten τ_i^0 nicht größer als die Wunsch-Bedienperiode T sein:

$$\sum_i g_i^s \leq T - \sum_i \tau_i^0. \quad (5.18)$$

Abbildung 5.1 veranschaulicht den zulässigen Bereich, in dem die Nebenbedingungen (i) bis (iii) erfüllt sind. Der zulässige Bereich existiert immer, wenn die Wunsch-Bedienperiode T das Stabilitätskriterium (5.12) erfüllt.

Lineare Zielfunktion

Das Ziel ist es, während der zugesicherten Freigabezeiten g_i^s eine möglichst große Anzahl von Fahrzeugen abzufertigen. Daraus ergibt sich folgende zu maximierende Zielfunktion:

$$F(g_{i_1}^s, g_{i_2}^s, \dots) = \sum_i g_i^s q_i^{\max} \quad \rightarrow \quad \max. \quad (5.19)$$

⁴¹Der Prognosewert $\hat{g}_i(t)$ kann an dieser Stelle nicht verwendet werden, weil sein Wert erst zu Beginn des Bedienvorgangs bekannt ist.

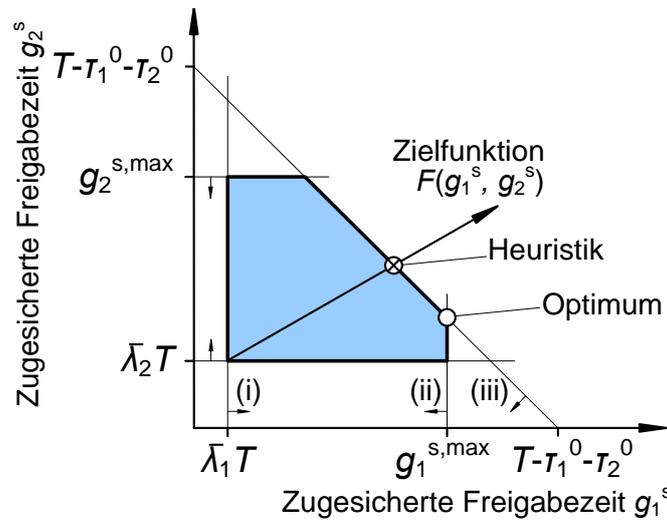


Abbildung 5.1.: Lineares Optimierungsproblem für die Zuweisung der Freigabezeiten g_i^s an einem Knotenpunkt mit zwei Zufahrten. Im zulässigen Bereich (dick umrandet) sind die Nebenbedingungen (i) bis (iii) erfüllt. Die Lösungen aus diesem Bereich garantieren, dass jede Knotenzufahrt einmal innerhalb der Wunsch-Bedienperiode T mit der geforderten Effizienz bedient wird. Die Zielfunktion F maximiert den Ausfluss und führt die Trajektorie aus Abb. 5.4 auf dem schnellsten Weg zum Grenzzyklus. Der helle Kreis markiert die optimale Lösung, der Kreis mit Kreuz die Lösung des heuristischen Ansatzes.

Optimale Lösung

Das lineare Optimierungsproblem (5.16) bis (5.19) kann mit Standardmethoden des Operations Research, beispielsweise dem Simplex-Algorithmus [52, 200], numerisch effizient und exakt gelöst werden. Das Optimum liegt immer auf dem Rand des konvexen Bereichs, der die zulässigen Lösungen markiert (siehe Abb. 5.1).

Die Nebenbedingungen stellen unter anderem das Einhalten der gewünschten Bedienperiode T sicher. Insbesondere wenn die Warteschlangen zu lang sind um innerhalb einer Periode geleert zu werden, kann es erforderlich sein, die Bedienvorgänge nach der zugesicherten Freigabezeit g_i^s abubrechen. Bei Maximierung der Zielfunktion (5.19) werden die Freigabezeiten g_i^s so zugewiesen, dass die Anzahl der innerhalb einer Periode abgefertigten Fahrzeuge maximal ist. Sind die maximalen Ausflussraten q_i^{\max} der Zufahrten nicht identisch, zeigt die optimale Lösung folgendes Verhalten: Die Zufahrt mit der größten maximalen Ausflussrate bekommt die meiste Freigabezeit zugewiesen, die anderen nur die minimale. Sobald auf der ausgewählten Zufahrt die Warteschlange abgebaut ist, erhält die mit der zweitgrößten maximalen Ausflussrate die meiste Freigabezeit.

Dieses Verhalten wurden von Gazis in den Referenzen [63, 64] ausführlich untersucht. Er interpretiert es im Kontext der optimalen Regelungstheorie nichtlinearer Systeme als Bang-Bang-Verhalten nach Pontryagins Maximumprinzip [145, 170].

5.3.3. Heuristischer Ansatz

Ausgangslösung

Je näher die Ausgangslösung, mit der das Lösungsverfahren startet, bereits am gesuchten Optimum liegt, desto weniger Iterationsschritte werden zum Auffinden des Optimums benötigt. Eine Ausgangslösung g_i^s , die bereits Nebenbedingung (i) und (iii) erfüllt und deshalb in der Nähe des gesuchten Optimums liegen muss, kann mit folgender Formel berechnet werden:

$$g_i^{s,0} = \bar{\lambda}_i T + \frac{q_i^{\max}}{\sum_{i'} q_{i'}^{\max}} T^{\text{idle}} \quad \text{mit} \quad T^{\text{idle}} = T - \sum_{i'} (\bar{\lambda}_{i'} T + \tau_{i'}^0). \quad (5.20)$$

Sie kennzeichnet den Punkt auf der Bereichsgrenze aus Nebenbedingung (iii), den eine Gerade vom Schnittpunkt aller Bereichsgrenzen aus Nebenbedingung (i) in Richtung des größten Zielfunktionswerts F durchstößt. Die Lösung entspricht folgender Zuteilungsregel: Jede Zufahrt i erhält zusätzlich zur minimal benötigten Freigabezeit $\bar{\lambda}_i T$ einen zur Ausflusskapazität q_i^{\max} proportionalen Anteil an der nicht minimal benötigten Bedienzeit T^{idle} . In Abb. 5.1 ist diese Lösung durch einen Kreis mit Kreuz markiert.

Vereinfachung

Das Lösen des Optimierungsproblems kann vollständig umgangen werden, wenn man für die zugesicherten Freigabezeiten g_i^s die Ausgangslösung aus Gleichung (5.20) verwendet. Die Lösung mit $g_i^{s,0}$ erfüllt sowohl die hinreichende Stabilitätsbedingung (5.8) als auch die aus dem gewünschten Schaltverhalten, jede Zufahrt einmal innerhalb T zu bedienen, abgeleitete Bedingung (5.12). Man verzichtet dabei auf das Schaltverhalten nach Pontryagins Maximumprinzip, wonach eine Zufahrt die maximale und die anderen die minimale Freigabezeit zugewiesen bekommen. Die Vereinfachung ist notwendig, wenn man auf eine Prognose für $g_i^{s,\max}$, der benötigten Freigabezeit eines zukünftig beginnenden Bedienvorgangs, verzichten möchte. Das in dieser Arbeit vorgeschlagene Verfahren verwendet deshalb die zugesicherten Freigabezeiten g_i^s aus Gleichung (5.20):

$$g_i^s := g_i^{s,0}. \quad (5.21)$$

Die damit vereinfachte Schaltregel garantiert sowohl die gewünschte Bedienperiode als auch netzwerkweite Stabilität.

5.4. Maximale Bedienperiode

Die Wunschperiode T stellt sich nicht durch das Messen der Zeit sondern durch das Abzählen von Fahrzeugen ein. Wenn innerhalb des Zeitraums T immer gleich viele Fahrzeuge ankommen, so wird die Wunschperiode exakt eingehalten. Bei stochastisch verteilten Ankunftszeitpunkten schwankt die tatsächliche Bedienperiode um den Mittelwert T . Das wird weiter unten in Szenario 5.iv bzw. in Anhang A.3 genauer ausgeführt.

Jede Zufahrt „im Mittel“ regelmäßig zu bedienen stellt zwar die geforderte Stabilität sicher, garantiert aber noch nicht die Sicherheit des Verkehrsablaufs [4, 5, 26]. Zu lange Sperrzeiten frustrieren die Verkehrsteilnehmer und verleiten sie zum Fahren bei Rot [152, 147, 45, 100]. Deshalb muss das zu entwickelnde Verfahren unter allen Umständen – das heißt bei Schwankungen im Verkehrsaufkommen, bei Messfehlern in der Verkehrsflusserfassung und bei Schätzfehlern in der Verkehrsprognose – maximale Sperrzeiten einhalten.

Diese strikte Zeit-Forderung muss jedoch nicht im Widerspruch stehen zum verkehrsadaptiven Reglerprinzip, die Umschaltzeitpunkte aus Fahrzeuganzahlen abzuleiten. Dies gelingt, indem man einen Teil der tatsächlich gemessenen Fahrzeuge durch virtuelle ersetzt und sicherstellt, dass auch die virtuellen Fahrzeuge allein hinreichend häufig den Bedienvorgang auslösen.

5.4.1. Virtueller Fahrzeugstrom

Adaptivitätsparameter α

Der virtuelle Fahrzeugstrom sei homogen und habe die gleiche Verkehrsstärke wie der tatsächliche. Das heißt, seine Ankunftsrate ist konstant und entspricht der bekannten mittleren Ankunftsrate \bar{q}_i^{exp} des tatsächlichen Fahrzeugstroms. Auf jeder Zufahrt i ist der Teil α der tatsächlich abzufertigenden Fahrzeuganzahl $\hat{n}_i(t)$ durch die virtuelle Fahrzeuganzahl $\hat{n}_i^{\text{virt}}(t)$ zu ersetzen. Das Kriterium (5.13) zur Aufnahme in die FIFO-Liste Ω lautet damit:

$$(1 - \alpha) \hat{n}_i(t) + \alpha \hat{n}_i^{\text{virt}}(t) \geq n_i^{\text{crit}} \quad \text{mit } \alpha \in [0, 1]. \quad (5.22)$$

Falls der tatsächliche Verkehrsstrom die konstante Ankunftsrate \bar{q}_i^{exp} hat, wird der Schwellenwert n_i^{crit} nach Gleichung (5.9) einmal innerhalb T erreicht. Das gilt auch, wenn ein beliebiger Teil α des tatsächlichen Verkehrsstroms durch den virtuellen ersetzt wird. Die unten angegebene Gleichung (5.23) erhält man analog aus Gleichung (5.9). Nun sei gefordert: Falls der tatsächliche Verkehrsstrom die Ankunftsrate null hat, soll der gleiche Schwellenwert n_i^{crit} allein durch den Teil α der virtuellen Fahrzeuge einmal innerhalb einer maximalen Bedienperiode T^{max}

erreicht werden. Das sich daraus ergebende Gleichungssystem

$$(1 - \alpha) \overbrace{\bar{q}_i^{\text{exp}} T}^{\text{real}} + \alpha \overbrace{\bar{q}_i^{\text{exp}} T}^{\text{virtuell}} = n_i^{\text{crit}} \quad (5.23)$$

$$0 + \alpha \bar{q}_i^{\text{exp}} T^{\text{max}} = n_i^{\text{crit}} \quad (5.24)$$

ist erfüllt für

$$\boxed{\alpha = T / T^{\text{max}}.} \quad (5.25)$$

Mit $T^{\text{max}} \rightarrow \infty$ beziehungsweise $\alpha = 0$ erhält man den ursprünglichen Fall, in dem die virtuellen Fahrzeuge keine Rolle spielen. Mit $T^{\text{max}} = T$ werden dagegen $\alpha = 100\%$ der tatsächlichen Fahrzeuge durch virtuelle ersetzt. Unabhängig vom Verkehrsaufkommen wird dann die gewünschte Periode T exakt eingehalten. Für Werte dazwischen behält das Schaltprogramm einen Teil seiner Adaptivität.

Im Folgenden soll der Zeitverlauf der virtuellen Fahrzeuganzahlen $\hat{n}_i^{\text{virt}}(t)$ modelliert werden.

Zuverlässige Information

Das Einhalten der maximalen Periode T^{max} ist sicherheitsrelevant und soll deshalb nicht von der Prognose der erwarteten Ankünfte abhängig sein. Entsprechend Annahme A2 wird davon ausgegangen, dass lediglich

- (i) die mittlere Ankunftsrate \bar{q}_i^{exp} und
- (ii) das Vorhandensein einer Warteschlange $n_i(t) > 0$ oder $n_i(t) = 0$

zuverlässig bekannt sind.⁴² Diese Informationen können mit entsprechenden Verkehrsdetektoren unmittelbar an der Haltelinie erhoben werden.⁴³

Die virtuellen Fahrzeuge werden mit der Rate \bar{q}_i^{exp} erzeugt. Wie viele von ihnen während der Freigabezeit entfernt werden können, ist ohne Kenntnis der tatsächlich angekommenen Fahrzeuge nicht bekannt. Gefordert sei aber: Die „virtuelle Warteschlange“ soll genau dann leer sein, wenn auch die tatsächliche Warteschlange leer ist. Deshalb soll der Wert für $\hat{n}_i^{\text{virt}}(t)$ ständig anwachsen und immer dann zurückgesetzt werden, wenn der Fall $n_i(t) = 0$ eintritt.

Nach Gleichung (3.30) entwickelt sich die abzufertigende Fahrzeuganzahl mit der konstanten Rate:

$$\frac{d\hat{n}_i^{\text{virt}}}{dt} = \frac{q_i^{\text{max}} \bar{q}_i^{\text{exp}}}{q_i^{\text{max}} - \bar{q}_i^{\text{exp}}}, \quad \text{falls } n_i(t) > 0. \quad (5.26)$$

⁴²Vgl. Abschnitt 1.2.3

⁴³Vgl. Abb. 1.2 (auf S. 7)

Sobald die tatsächliche Warteschlange $n_i(t)$ geleert wurde, wird $\hat{n}_i^{\text{virt}}(t)$ auf den Wert gesetzt, der sich für eine Zufahrt ohne Warteschlange und konstanter Ankunftsrate unmittelbar nach Abbruch des Bedienvorgangs einstellt.⁴⁴ Es gilt:

$$\hat{n}_i^{\text{virt}}(t) := \tau_i^0 \frac{q_i^{\max} \bar{q}_i^{\text{exp}}}{q_i^{\max} - \bar{q}_i^{\text{exp}}}, \quad \text{falls } n_i(t) = 0. \quad (5.27)$$

Dieser Ansatz stellt nicht nur sicher, dass jede Zufahrt mindestens einmal innerhalb T^{\max} bedient wird, sondern er garantiert auch die vollständige Leerung der Warteschlangen. Ist dies nicht in einem einzigen Bedienvorgang möglich, so wird gegebenenfalls eine Serie von Bedienvorgängen gestartet. Denn ist der Schwellenwert n_i^{crit} nach Ablauf der zugesicherten Freigabezeit q_i^s immer noch überschritten, so wird i erneut in Ω eingereiht und die Zufahrt erneut bedient.

5.4.2. Zeitabhängiger Schwellenwert

Damit die zusätzliche Forderung der Maximalperiode T^{\max} mit dem Schaltprogramm aus Abschnitt 5.3.1 konsistent ist, soll auf den virtuellen Fahrzeugstrom und den Parameter α verzichtet werden. Stattdessen wird ein zeitabhängiger Schwellenwert $n_i^{\text{crit}}(t)$ verwendet. Dieser wird im Folgenden hergeleitet.

Die Gleichung (5.22) lässt sich mit $\alpha < 1$ bzw. $T^{\max} > T$ nach $\hat{n}_i(t)$ umstellen:

$$\hat{n}_i(t) \geq \frac{n_i^{\text{crit}} - \alpha \hat{n}_i^{\text{virt}}(t)}{(1 - \alpha)}. \quad (5.28)$$

Die rechte Seite erhält somit die Bedeutung eines zeitabhängigen Schwellenwerts $n_i^{\text{crit}}(t)$. Damit kann die Schaltregel (5.7) bzw. (5.13) ersetzt werden durch

$$\boxed{\hat{n}_i(t) \geq n_i^{\text{crit}}(t)}. \quad (5.29)$$

Setzt man für die Werte von n_i^{crit} , α und $\hat{n}_i^{\text{virt}}(t)$ in der rechten Seite von Gleichung (5.28) die entsprechenden Gleichungen (5.9) und (5.25) bis (5.27) ein, so erhält

⁴⁴Zum Zeitpunkt t werde die Bedienung von Zufahrt i abgebrochen. Aufgrund der angenommenen konstanten Ankunftsrate q_i^{exp} lässt sich die Zeitreihe der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ aus (3.1) durch folgende Geradengleichung ausdrücken:

$$N_i^{\text{exp}}(t + \Delta t) = N_i^{\text{exp}}(t) + \Delta t q_i^{\text{exp}}.$$

Die Warteschlange sei vollständig geleert. Aus $n_i(t) = 0$ folgt mit (3.10) die Identität $N_i^{\text{exp}}(t) = N_i^{\text{out}}(t)$. Wird beides in die Bestimmungsgleichung (3.20) eingesetzt, so erhält man mit $\hat{n}_i(t) = q_i^{\max} \hat{g}_i(t)$ nach (3.19) die Gleichung:

$$\hat{n}_i(t) = (\tau_i^0 + \hat{n}_i(t)/q_i^{\max}) q_i^{\text{exp}}.$$

Diese lässt sich nach der gesuchten Größe $\hat{n}_i(t)$ auflösen. Man erhält Gleichung (5.27)

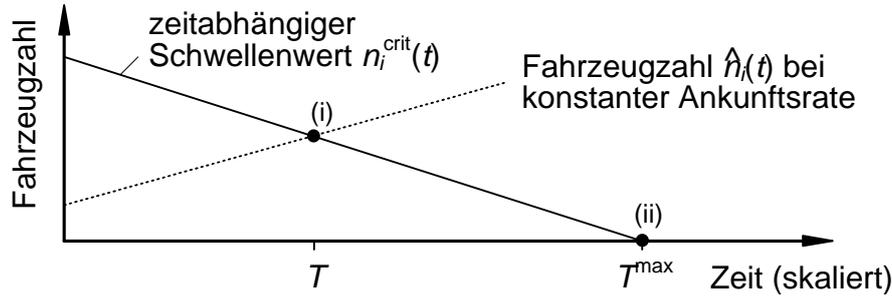


Abbildung 5.2.: Der zeitabhängige Schwellenwert $n_i^{\text{crit}}(t)$ wird durch die Geradengleichung (5.30) beschrieben. Die abfallende Gerade verbindet zwei Punkte und erfüllt damit zwei Forderungen: (i) Bei konstanter Zufussrate $\bar{q}_i^{\text{exp}}(t) = \bar{q}_i^{\text{exp}}$ wird der Schwellenwert genau einmal innerhalb der Wunschperiode T erreicht. (ii) Selbst wenn keine Fahrzeuge detektiert werden ($\hat{n}_i = 0$), wird die Zufahrt einmal innerhalb der maximalen Bedienperiode T^{max} bedient. Weil sich T und T^{max} auf die Bedienperiode einschließlich Sperr-, Zwischen- und Freigabezeit beziehen, der Schwellenwert aber bereits am Ende der Sperrzeit erreicht werden soll, wurde die Zeitachse mit $\bar{q}_i^{\text{exp}} / (q_i^{\text{max}} - \bar{q}_i^{\text{exp}})t - \tau_i^0$ skaliert.

man für den zeitabhängigen Schwellenwert $n_i^{\text{crit}}(t)$ den Ausdruck

$$n_i^{\text{crit}}(t) = \bar{q}_i^{\text{exp}} T \frac{T^{\text{max}} - a(t) / (1 - \bar{q}_i^{\text{exp}} / q_i^{\text{max}})}{T^{\text{max}} - T}. \quad (5.30)$$

Für die Hilfsvariable $a(t)$ gilt dabei:

$$\begin{cases} a(t) = \tau_i^0, & \text{falls } n_i(t) = 0 \\ \dot{a}(t) = 1, & \text{sonst.} \end{cases} \quad (5.31)$$

Der Term $a(t) - \tau_i^0$ kann als die Zeitspanne interpretiert werden, in der die Warteschlange nicht mehr leer gewesen ist. Es ist wichtig zu bemerken, dass $n_i^{\text{crit}}(t)$ bereits bei Sperrzeiten $a(t) - \tau_i^0 < T^{\text{max}}$ auf null abgefallen ist. Der Wert von T^{max} legt nämlich nicht die maximale Sperrzeit sondern die maximale Bedienperiode einschließlich Sperr-, Zwischen- und Freigabezeit fest.

Das neue Kriterium (5.29) ist damit ausschließlich von der bekannten mittleren Ankunftsrate \bar{q}_i^{exp} , dem Vorhandensein der Warteschlange $n_i(t)$ und den beiden Parametern T und T^{max} mit $T^{\text{max}} > T$ abhängig. Die virtuellen Fahrzeuge werden nicht mehr benötigt. Abbildung 5.2 zeigt die Bedeutung der Parameter T und T^{max} im Verlauf von $n_i^{\text{crit}}(t)$.

Es sei nochmals darauf verwiesen, dass die Zufahrten eines Knotenpunktes entsprechend den Priorisierungsregeln aus Abschnitt 5.3.1 in der Reihenfolge bedient werden, in der sie die Schwellenwerte $n_i^{\text{crit}}(t)$ erreichen. Die Zeitspanne, die zwischen Erreichen des Schwellenwerts und Beginn des Bedienvorgangs liegen kann, werden zunächst vernachlässigt. In den Szenarien 5.iv und 5.v wird jedoch gezeigt, dass das Stabilisierungsverfahren eine Tendenz mit sich bringt, diese Zeitspanne zu minimieren. Mit dem kombinierten Verfahren aus Kapitel 6 wird diese Zeitspanne nochmals verringert.

5.5. Schaltverhalten und Zusammenfassung

5.5.1. Schaltverhalten

Um das Schaltverhalten des Stabilisierungsverfahrens deutlich zu machen, werden im Folgenden fünf unterschiedliche Szenarien betrachtet.

Fehlende Fahrzeuge (Szenario 5.i)

Im ersten Testfall ist folgende Frage zu beantworten: Welche Zeitspanne Δt muss eine Kolonne aus

$$\hat{n}_i(t) = n_i^{\text{crit}}(t) - \Delta n \quad (5.32)$$

Fahrzeugen bis zum Erreichen des Schwellenwerts mindestens warten? Für Δt und Δn gilt aufgrund des linearen Verlaufs von $n_i^{\text{crit}}(t)$ die Beziehung:

$$\frac{\Delta n}{\Delta t} = -\frac{dn_i^{\text{crit}}}{dt}. \quad (5.33)$$

Durch Ableiten von Gleichung (5.30) und Umstellen erhält man:

$$\Delta t = \left(\frac{T^{\text{max}}}{T} - 1 \right) \left(\frac{1}{q_i^{\text{exp}}} - \frac{1}{q_i^{\text{max}}} \right) \Delta n. \quad (5.34)$$

Aus dem ersten Klammersausdruck folgt: Je kleiner T^{max} ist, desto früher wird der Kolonne freigegeben. Der zweite berücksichtigt die mittlere Verkehrsstärke. Auf einer Zufahrt mit geringer Ankunftsrate, wo innerhalb einer Periode nur wenige Fahrzeuge abgefertigt werden, führt der selbe Fehlbetrag Δn zu längeren Sperrzeiten. Der lineare Verlauf von $n_i^{\text{crit}}(t)$ lässt auch den umgekehrten Schluss zu: Δn zusätzliche Fahrzeuge führen zu einer Δt Zeiteinheiten früheren Bedienung.

Die Umschaltzeitpunkte, die zur Bedienperiode T gehören, werden somit stabilisiert. Um das genauer zu erläutern, sei nochmals auf Abb. 5.2 (auf S. 103) verwiesen : Je näher T^{max} an T rückt, desto steiler fällt die Gerade $n_i^{\text{crit}}(t)$ ab. Das hat zur Folge, dass sich die Zeitpunkte, zu denen ein beliebiger Verlauf von

$\hat{n}_i(t)$ die Gerade durchstößt, näher um den zu T gehörigen Umschaltzeitpunkt konzentrieren. Dies lässt sich für Poisson-verteilte Fahrzeugankünfte analytisch berechnen (siehe auch Abb. 5.3).

Poisson-verteilte Fahrzeugankünfte (Szenario 5.ii)

Bei zufälligen Fahrzeugankünften ist die tatsächliche Bedienperiode z_i , d.h. die Zeitspanne zwischen zwei Bedienvorgängen für die selbe Zufahrt i , als eine Zufallsgröße Z_i zu betrachten. Modelliert man die Fahrzeugankünfte mit einem stationären Poisson-Prozess, so gehorcht Z_i einer Erlang-Verteilung:

$$Z_i \sim \text{Erl}(\lambda = \bar{q}_i^{\text{exp}}, n = \lceil n_i^{\text{crit}} \rceil). \quad (5.35)$$

Die Herleitungen und Berechnungen sind in Anhang A.4 aufgeführt. Die Annahme Poisson-verteilter Fahrzeugankünfte ist nur erfüllt, wenn die Ankunftszeitpunkte der einzelnen Fahrzeuge unabhängig voneinander sind. Das ist in guter Näherung an mehrspurigen Zufahrten zu isolierten Knotenpunkten der Fall.

Bei einem konstantem Schwellenwert n_i^{crit} , d.h. für $T^{\text{max}} \rightarrow \infty$, entspricht der Erwartungswert von $E(Z_i)$ exakt der Wunschperiode T . Der Variationskoeffizient⁴⁵ $\text{VarK}(Z_i)$ ist umso kleiner, je größer die mittlere Verkehrsnachfrage \bar{q}_i^{exp} ist:

$$E(Z_i) = T \quad \text{und} \quad \text{VarK}(Z_i) = \frac{1}{\sqrt{\bar{q}_i^{\text{exp}} T}}. \quad (5.36)$$

Bei einem zeitabhängigen Schwellenwert $n_i^{\text{crit}}(t)$ mit endlichem T^{max} gehorcht die Zufallsgröße Z_i einer verallgemeinerten Erlang-Verteilung [143, 6] mit der Dichtefunktion (A.56). Abbildung 5.3 zeigt die Verteilung für verschiedene Werte von T^{max} .

Stabiler Grenzyklus (Szenario 5.iii)

Abbildung 5.4 zeigt das sich einstellende Schaltverhalten an einem Knotenpunkt mit konstanten Ankunftsrate. In diesem Fall wird die Wunschperiode T immer exakt eingehalten. Die maximale Bedienperiode T^{max} spielt keine Rolle, weil der Schwellenwert stets im Punkt (i) aus Abb. 5.2 überschritten wird. Das Schaltverhalten ist deterministisch. Die Trajektorie der abzufertigenden Fahrzeuganzahlen $\hat{n}_i(t)$ (und ebenso die der Warteschlangenlängen $n_i(t)$) läuft von jedem beliebigen Startwert auf einen stabilen Grenzyklus zu. Innerhalb jeder Periode T rückt $\hat{n}_i(t)$ dem stabilen Bereich um den gleichen Betrag näher. Der Grenzyklus selbst kann zwar verschiedene Formen annehmen, aber er liegt immer innerhalb eines Rahmens, der durch $[0, T\bar{q}_i^{\text{exp}}]$ für jede Zufahrt i aufgespannt wird. Von Beginn

⁴⁵Der Variationskoeffizient ist der Quotient aus Standardabweichung und Erwartungswert. Er bemisst damit die relativen Abweichungen vom Erwartungswert.

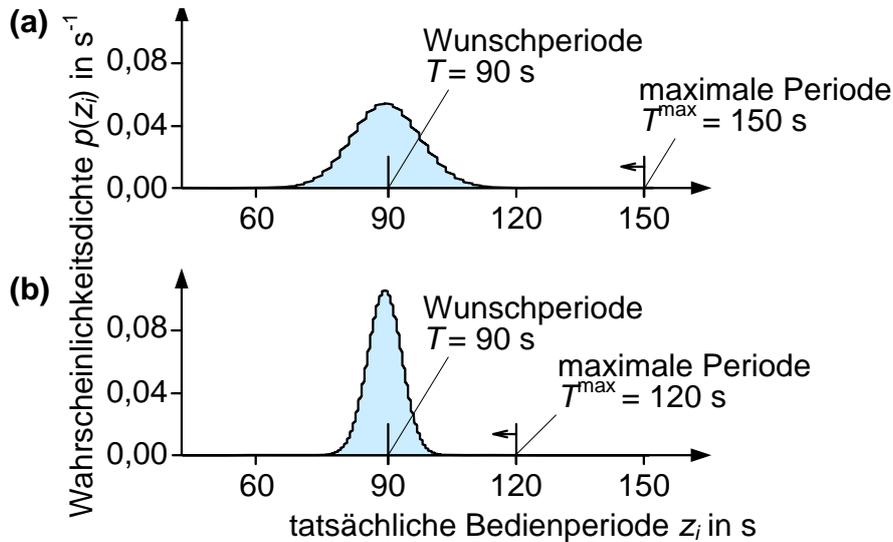


Abbildung 5.3.: Bei Poisson-verteilten Fahrzeugankünften lässt sich die Wahrscheinlichkeitsverteilung der tatsächlichen Bedienperiode z_i analytisch berechnen (Anhang A.4). Die Verteilung hat ihr Maximum bei der gegebenen Wunsch-Bedienperiode T und ist null für Werte größer als die gegebene Maximalperiode T^{\max} . Je näher T^{\max} an T liegt, desto mehr konzentriert sich auch die tatsächliche Bedienperiode z_i um T .

an, auch schon bevor der Grenzyklus erreicht ist, wird jede Zufahrt genau einmal innerhalb T bedient.

Periodische Ankunft einer Kolonne (Szenario 5.iv)

Der stabile Grenzyklus bei konstanten Ankunftsraten (siehe Abb. 5.4) hat die Eigenschaft, dass niemals zwei Zufahrten gleichzeitig in Ω vermerkt sind. Der Bedienvorgang für eine Zufahrt startet zur gleichen Zeit wie $\hat{n}_i^{\text{crit}}(t)$ den Schwellenwert $n_i^{\text{crit}}(t)$ erreicht. Das muss im Allgemeinen nicht so sein. Dennoch kann zumindest für periodische Ankunftsprofile qualitativ gezeigt werden, dass das Stabilisierungsverfahren zur Minimierung der Verweilzeiten in Ω tendiert.

Erreicht eine Zufahrt i den Schwellenwert $n_i^{\text{crit}}(t)$ zu einem Zeitpunkt, zu dem gerade eine andere Zufahrt bedient wird, verzögert sich die Bedienung für Zufahrt i . Falls durch die verzögerte Bedienung mehr als $n_i^{\text{crit}}(t)$ Fahrzeuge abgefertigt werden, wird der Schwellenwert in den nachfolgenden Perioden entsprechend später erreicht. Eine Verzögerung der Bedienung kann zu einer Verschiebung aller nachfolgenden Perioden führen. Dieser Mechanismus kann als Phasensynchronisation [144] interpretiert werden. Abbildung 5.5 zeigt diesen Mechanismus am Beispiel periodischer Kolonnenankünfte.

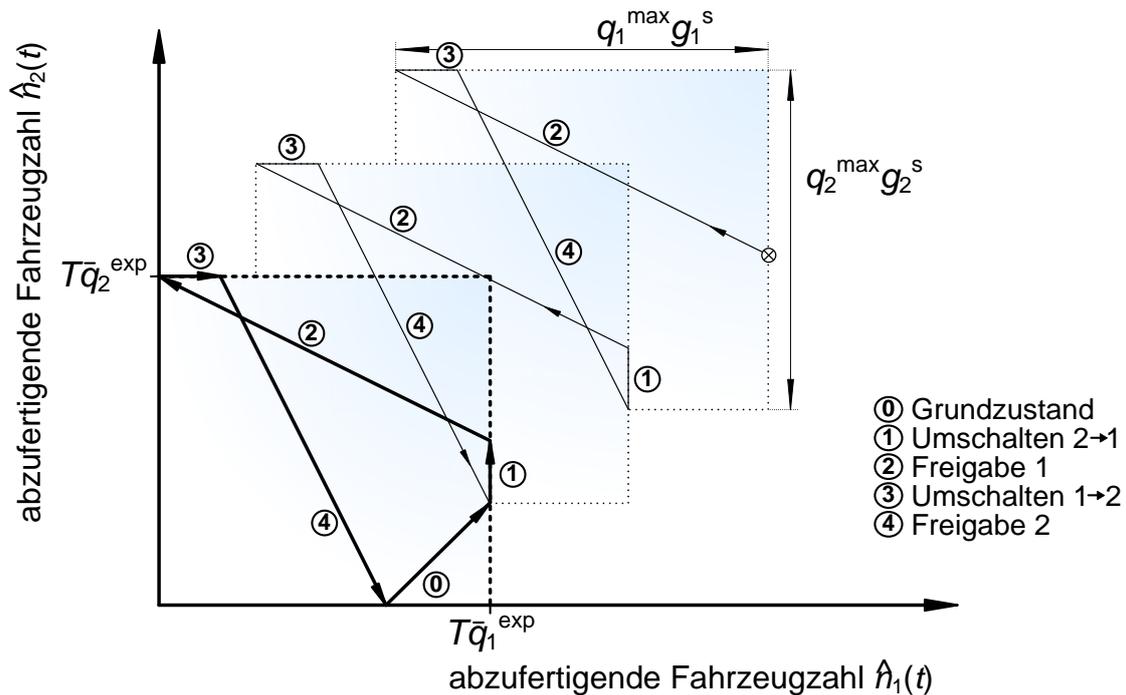


Abbildung 5.4.: Stabiler Grenzyklus (dicke Kurve) des Stabilisierungsverfahren für einen Knotenpunkt mit zwei Zufahrten und konstanten Ankunftsrate \bar{q}_1^{exp} bzw. \bar{q}_2^{exp} . Die Trajektorie der abzufertigenden Fahrzeuganzahlen (\hat{n}_1, \hat{n}_2) konvergiert von einem beliebigen Anfangszustand aus gegen einen stabilen Grenzyklus. Sowohl im Grenzyklus als auch während der Annäherung wird jede Zufahrt genau einmal innerhalb der Wunschperiode T bedient.

Beliebige periodische Ankunftsprofile (Szenario 5.v)

Das Ankunftsprofil auf den Zufahrten eines Knotenpunktes sei periodisch in T . Die einzelnen Fahrzeuge innerhalb jeder Periode seien in gleicher Weise aufsteigend nummeriert. Die Verzögerung eines Bedienvorganges kann bewirken, dass in der selben Periode mehr als n_i^{crit} Fahrzeuge abgefertigt werden. Unter keinen Umständen jedoch werden weniger als n_i^{crit} Fahrzeuge abgefertigt.

Endet also das Freigabezeitfenster in einer Periode nach der Ankunft des n -ten Fahrzeugs, so kann sie im nächsten Umlauf nach dem $(n + 1)$ -ten Fahrzeug enden – nicht jedoch nach dem $(n - 1)$ -ten. Damit das Freigabezeitfenster die Ankunft dieses $(n + 1)$ -ten Fahrzeuges mit erfassen kann, darf die Zeitlücke zwischen dem n -ten und $(n + 1)$ -ten Fahrzeug nicht größer sein als die Verzögerung des Bedienvorganges. Ist die Zeitlücke anderenfalls größer als die Verzögerung, so „haftet“ das Ende des Freigabezeitfensters am n -ten Fahrzeug. Das Ende des Freigabe-

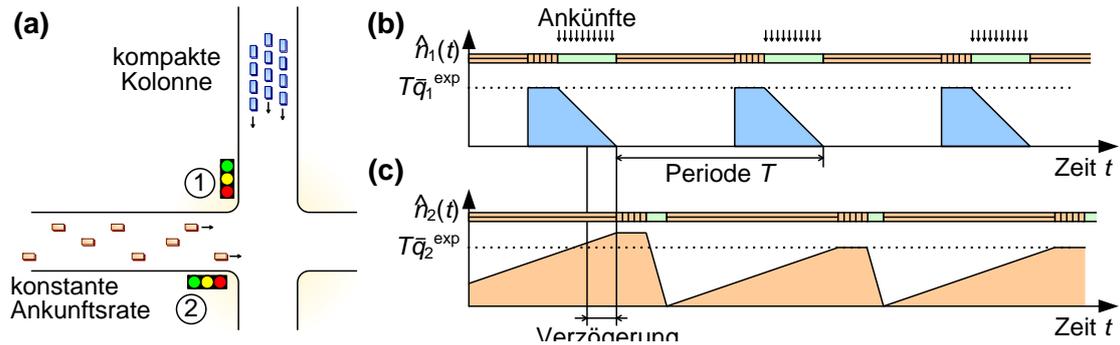


Abbildung 5.5.: (a) Knotenpunkt mit zwei Zufahrten. Auf Zufahrt 1 kommt periodisch in T eine kompakte Kolonne an, während die Ankunftsrate auf Zufahrt 2 konstant ist. (b) Die abzufertigende Fahrzeuganzahl $\hat{n}_1(t)$ springt rechtzeitig vor Ankunft der Kolonne auf $T\bar{q}_1^{\text{exp}}$ und erreicht damit den Schwellenwert $n_1^{\text{crit}}(t)$. Die Kolonne erhält bei ihrer Ankunft freie Fahrt. (c) Auf Zufahrt 2 wächst $\hat{n}_2(t)$ linear an. Wird, wenn der Schwellenwert erreicht ist, gerade Zufahrt 1 bedient, zögert sich die Bedienung von Zufahrt 2 heraus. Dadurch werden mehr Fahrzeuge abgefertigt und die nachfolgenden Perioden verschoben. Diese Phasenverschiebung muss sich nicht einstellen. Es handelt sich vielmehr um eine Tendenz des Stabilisierungsverfahrens.

zeitfensters „haftet“ dabei also vorzugsweise an dem Fahrzeug, auf das eine große Zeitlücke folgt.

Diese Beobachtung deckt sich mit den Erkenntnissen (v), (vi) und (vii) aus dem Abschnitt 2.3.4 zur optimalen Platzierung des Freigabezeitfensters: Das Ende der Freigabezeit sollte mit der Ankunft des letzten Fahrzeugs derjenigen Kolonne zusammenfallen, die von der größten Zeitlücke bis zur Ankunft der nächsten Kolonne gefolgt wird.

5.5.2. Zusammenfassung

Im Gegensatz zum Optimierungsverfahren aus Kapitel 4, das die Warteschlangen so schnell wie möglich zu leeren versucht und die Häufigkeit der Umschaltvorgänge nicht direkt berücksichtigt, geht das Stabilisierungsverfahren in erster Linie sparsam mit der Ressource Abfertigungskapazität um. Aber dieser sparsame Umgang garantiert effiziente Bedienvorgänge und damit letztlich stabiles Verhalten.

Das Stabilisierungsverfahren bedient die Zufahrten entsprechend den Priorisierungsregeln (5.13) bis (5.15) in der Reihenfolge, in der sie den kritischen Schwellenwert $n_i^{\text{crit}}(t)$ nach Gleichung (5.30) erreichen. Der Bedienvorgang wird abgebrochen, wenn die Warteschlange vollständig geleert wurde, spätestens jedoch nach Ablauf der zugesicherten Freigabezeit g_i^s nach Gleichung (5.20) bzw. (5.21).

Somit erhält jeder Verkehrsstrom die minimal benötigte Abfertigungskapazität. Außerdem werden Schwankungen in der Verkehrsnachfrage dadurch ausgeglichen, dass die Freigabezeit bei Bedarf bis auf die zugesicherte Freigabezeit g_i^s verlängert wird.

Schwellenwert und Komplexität

Die mit dem Schwellenwert $n_i^{\text{crit}}(t)$ zu vergleichende Zustandsgröße ist die abzufertigende Fahrzeuganzahl $\hat{n}_i(t)$ aus Abschnitt 3.3.1. Sie umfasst alle Fahrzeuge, die im nachfolgenden Bedienvorgang mit maximaler Rate abgefertigt werden können (siehe Abb. 3.6). Weil $\hat{n}_i(t)$ bei herannahenden Fahrzeugkolonnen schnell anwächst bzw. bei kompakten Kolonnen abrupt auf einen größeren Wert springt, werden tendenziell Fahrzeugkolonnen den Bedienvorgang auslösen. Die auslösende Kolonne wird dann mit „Grüner Welle“ bedient.

Der benötigte Rechenaufwand wächst auch mit dem Stabilisierungsverfahren lediglich linear mit der Anzahl der Knotenzufahrten.

Stabilität

Wird jeder Knotenpunkt im Straßennetzwerk mit dem Stabilisierungsverfahren betrieben und ist an jedem Knotenpunkt die hinreichende Stabilitätsbedingung (5.12) erfüllt, dann bleiben die Warteschlangen aller Knotenzufahrten beschränkt. Denn weil jeder Verkehrsstrom die minimal benötigte Abfertigungskapazität zugesichert bekommt, wird kein Verkehrsstrom im Netzwerk systematisch aufgehalten. Das dezentrale Steuerungsverfahren garantiert damit netzwerkweite Stabilität.

Die hinreichende Stabilitätsbedingung (5.12) des Stabilisierungsverfahrens ist genau dann erfüllt, wenn auch die hinreichende Stabilitätsbedingung (2.10) einer Festzeitsteuerung mit Umlaufzeit $T^{\text{cyc}} = T$ erfüllt ist. Also wann immer sich ein Straßennetzwerk bei gegebener Verkehrsnachfrage mit einer Festzeitsteuerung stabilisieren lässt, dann ist es auch mit dem Stabilisierungsverfahren stabil.

Zwei Parameter

Mit den beiden Parametern T und T^{max} lassen sich Mittel- und Maximalwert der tatsächlichen Bedienperiode festlegen. Für die Wunsch-Periode T gilt die Bedingung (5.12) und für die Maximal-Periode T^{max} die Bedingung $T^{\text{max}} > T$.

Durch den Ansatz, einen Teil der Fahrzeuge durch virtuelle zu ersetzen, wird die maximale Bedienperiode auch dann eingehalten, wenn keine Fahrzeuge detektiert werden. In diesem Fall wird die entsprechende Knotenzufahrt so lange bzw. so oft bedient, bis die Warteschlange vollständig geleert wurde. Vorausgesetzt wird lediglich, dass die mittlere Ankunftsrate und das Vorhandensein einer Warteschlange bekannt sind.

Kapitel 6.

Kombiniertes Verfahren

Kurzfassung

In den beiden vorherigen Kapiteln 4 und 5 wurden zwei grundsätzlich verschiedene Verfahren zur dezentralen Online-Steuerung entwickelt. Zwar ist das Streben des Optimierungsverfahrens nach minimalen Wartezeiten wünschenswert, die Stabilität des Stabilisierungsverfahrens dagegen unverzichtbar. Nun sollen beide Verfahren so miteinander kombiniert werden, dass sich die Vorzüge beider nutzen lassen.

Der Ansatzpunkt dabei ist, das Optimierungsverfahren in das Stabilisierungsverfahren einzubetten. Das Stabilisierungsverfahren soll nur dann in die lokale Optimierung eingreifen, wenn diese nicht von selbst zum gewünschten Schaltverhalten führt. Im Ergebnis garantiert das „kombinierte Verfahren“ netzwerkweite Stabilität entsprechend des Stabilisierungsverfahrens, und es nutzt verbleibende Abfertigungskapazitäten für lokal optimale Umschaltvorgänge entsprechend des Optimierungsverfahrens.

6.1. Schaltprogramm

6.1.1. Grundidee

Das Stabilisierungsverfahren sei dem Optimierungsverfahren übergeordnet. Dahinter steht folgende Idee:

Das Stabilisierungsverfahren garantiert der aktuellen Verkehrsnachfrage exakt die minimal benötigte Abfertigungskapazität. Es stellt sicher, dass jede Zufahrt genau so viel Freigabezeit zugewiesen bekommt, wie zur Abfertigung der in einer Wunschperiode T neu ankommenden Fahrzeuge minimal benötigt wird. Die Differenz zwischen maximal verfügbarer und minimal benötigter Abfertigungskapazität, die das Stabilisierungsverfahren an sich ungenutzt im Grundzustand verstreichen lassen würde, soll dem Optimierungsverfahren zur Verfügung gestellt werden.

Das Optimierungsverfahren strebt danach, die Gesamtwartezeiten lokal zu minimieren. Falls die sich dynamisch einstellende Schaltfolge die notwendigen Eigenschaften hat, d.h. die gesamte Verkehrsnachfrage innerhalb der Wunschperiode T abfertigt, dann braucht das Stabilisierungsverfahren niemals einzugreifen. Bei Abweichungen vom gewünschten Schaltverhalten greift das Stabilisierungsverfahren sofort und effizient ein.

Letztlich ergibt sich Folgendes: Die nicht unbedingt benötigte Abfertigungskapazität wird für zusätzliche Umschaltvorgänge und das Abwarten von Zeitlücken bei Freigabezeitverlängerungen genutzt.

6.1.2. Programmablauf

Die Grundidee lässt sich umsetzen, wenn man die beiden Verfahren in folgender Weise kombiniert. An jedem Knotenpunkt sind in jedem Zeitschritt t vier Schritte (siehe Abb. 6.1) durchzuführen:

1. **Aktualisierung der Zustandsgrößen:** Für jede Zufahrt i des betrachteten Knotenpunktes werden ihr aktueller Prioritätsindex $\pi_i(t)$ nach Gleichung (4.35) und ihr aktueller Prognosewert $\hat{n}_i(t)$ für die abzufertigende Fahrzeuganzahl nach Gleichung (3.20) berechnet.
2. **Aktualisierung der Liste Ω :** Die Liste Ω des Stabilisierungsverfahrens wird entsprechend den beiden Regeln (5.13) und (5.15) aktualisiert: Eine Zufahrt i wird in Ω aufgenommen, wenn dort mindestens $n_i^{\text{crit}}(t)$ Fahrzeuge nach (5.30) mit maximaler Rate abgefertigt werden können. Sie wird aus Ω entfernt, wenn die Warteschlange geleert wurde oder die zugesicherte Freigabezeit g_i^s nach (5.20) bzw. (5.21) abgelaufen ist.

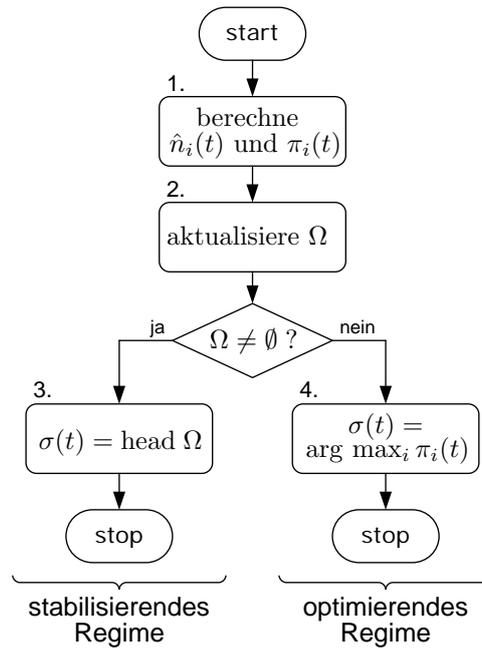


Abbildung 6.1.: Der Programmablaufgraph wird vom Schaltprogramm in jedem Zeitschritt t durchlaufen, um den Index $\sigma(t)$ der zu bedienenden Zufahrt zu bestimmen. Rechtecke stehen für Operationen, die Raute für eine Verzweigung. Der linke Pfad (stabilisierendes Regime) wird nicht durchlaufen, wenn das optimierende Regime stabil ist und die Zufahrten mit der gewünschten Häufigkeit bedient.

3. **Stabilisierendes Regime:** Falls die Liste Ω nicht leer ist, d.h. mindestens eine Zufahrt mit größerer als der minimal geforderten Effizienz bedient werden kann, so ist die zu bedienende Zufahrt entsprechend der Regel (5.14) auszuwählen:

$$\sigma(t) = \text{head } \Omega, \quad \text{falls } \Omega \neq \emptyset. \quad (6.1)$$

Das kombinierte Verfahren befindet sich im stabilisierenden Regime.

4. **Optimierendes Regime:** Ist die Liste Ω dagegen leer, wird die Zufahrt mit größtem Prioritätsindex $\pi_i(t)$ zur Bedienung ausgewählt. Es gilt die Regel (4.34):

$$\sigma(t) = \arg \max_i \pi_i(t). \quad (6.2)$$

Das kombinierte Verfahren befindet sich im optimierenden Regime.

Zusammenfassend lässt sich die Schaltregel des kombinierten Verfahrens durch folgende Gleichung ausdrücken:

$$\sigma(t) = \begin{cases} \text{head } \Omega, & \text{falls } \Omega \neq \emptyset \quad (\text{stabilisierendes Regime}), \\ \arg \max_i \pi_i(t), & \text{sonst} \quad (\text{optimierendes Regime}). \end{cases} \quad (6.3)$$

6.2. Schaltverhalten

6.2.1. Eingriff des stabilisierenden Regimes

Das optimierende Regime wird unverzüglich unterbrochen, wenn es eine Zufahrt nicht bedient, obwohl dort mehr Fahrzeuge abzufertigen sind als innerhalb der Wunschperiode T im Mittel ankommen. Dieser Zustand dient als Indikator für unerwünschtes oder instabiles Verhalten. Das stabilisierende Regime greift ein noch bevor sich die Instabilität in zu langen Warteschlangen oder zu langen Bedienperioden ausdrücken kann. Der Eingriff erfolgt jedoch auch, wenn das optimierende Regime zwar einen stabilen Grenzyklus verfolgt, dieser aber eine größere Periode als T hat. Aufgrund der wichtigeren Forderung, die Dauer der Bedienperiode und damit die Sperrzeiten zu beschränken, ist dieser Grenzyklus selbst dann unerwünscht, wenn er mit geringeren Gesamtwartezeiten verbunden ist. Der Eingriff des stabilisierenden Regimes dauert nur so lange, wie zur Sicherstellung des gewünschten Verhaltens notwendig ist.

Angenommen ein Knotenpunkt wird mit dem kombinierten Verfahren betrieben und die Ankünfte auf seinen Zufahrten sind vom Schaltprogramm unabhängig. Dann sind die entstehenden Gesamtwartezeiten niemals größer, als wenn der Knotenpunkt nur mit dem Stabilisierungsverfahren betrieben werden würde.

Die Eingriffe des stabilisierenden Regimes garantieren das gewünschte Schaltverhalten auch dann, wenn das optimierende Regime durch ein anderes Schaltprogramm ersetzt werden würde. Das neue Schaltprogramm würde von der hier vorgestellten Reglerstruktur ebenfalls stabilisiert werden. Erzeugt es von sich aus das gewünschte Schaltverhalten, greift das stabilisierende Regime niemals ein.

6.2.2. Grenzyklus bei konstanter Ankunftsrate

Szenario 6.i

Ausgangssituation ist, ebenso wie in den Szenarien 4.vi und 5.iii, ein Knotenpunkt mit zwei Zufahrten und konstanten Ankunftsrate. Auf den Zufahrten warten anfangs mehr Fahrzeuge, als innerhalb einer Bedienperiode abgefertigt werden können.

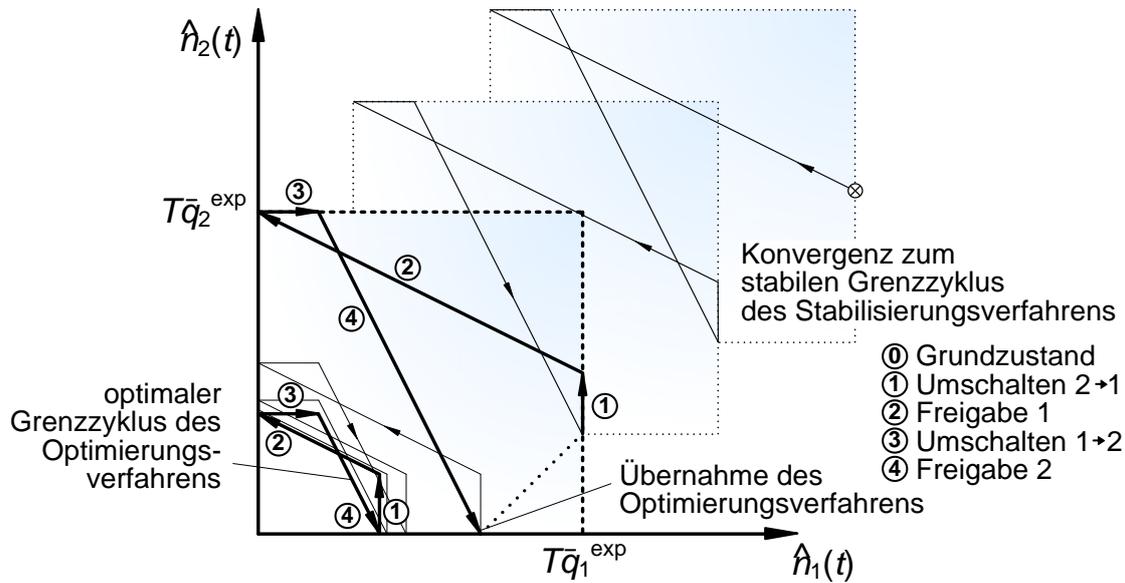


Abbildung 6.2.: Trajektorienverlauf der Zustandsgrößen (\hat{n}_1, \hat{n}_2) an einem Knotenpunkt mit zwei Zufahrten und konstanten Ankunftsrate. Von einem beliebigen Anfangszustand aus (Kreis mit Kreuz) läuft die Trajektorie auf den stabilen Grenzzyklus des Stabilisierungsverfahrens zu (siehe auch Abb. 5.4). Dabei wird die Wunschperiode T exakt eingehalten. Sobald der Grenzzyklus erreicht ist und das Stabilisierungsverfahren in den Grundzustand (dick gestrichelt) gehen würde, übernimmt das optimierende Regime. Dieses lässt die Trajektorie asymptotisch gegen den optimalen Grenzzyklus des Optimierungsverfahrens konvergieren (siehe Abb. 4.4 (auf S. 86)). Die Bedienperiode ist dann deutlich kleiner als die Wunschperiode T .

Trajektorienverlauf

Die Abbildung 6.2 zeigt, wie die Trajektorie der abzufertigenden Fahrzeuganzahlen (\hat{n}_1, \hat{n}_2) zuerst auf den stabilen Grenzzyklus des Stabilisierungsverfahrens und unmittelbar anschließend auf den optimalen Grenzzyklus des Optimierungsverfahrens geführt wird.

Während das stabilisierende Regime die Warteschlangen abbaut, wird die gesamte Periode T ausgenutzt. Zustände, in denen alle Zufahrten gesperrt sind, existieren nicht. Solange der Grenzzyklus des stabilisierenden Regimes noch nicht erreicht ist, nimmt die Anzahl der wartenden Fahrzeuge mit jedem Umlauf um den gleichen Betrag ab. Das stabilisierende Regime gibt dann an das optimierende Regime ab, wenn nach Beendigung eines Bedienvorgangs auf keiner Zufahrt mehr als $n_i^{\text{crit}}(t)$ Fahrzeuge mit maximaler Rate bedient werden können. Das Sta-

bilisierungsverfahren würde in diesem Moment in den Grundzustand übergehen und alle Zufahrten sperren.

Das optimierende Regime führt die Trajektorie asymptotisch an den Grenzzyklus des Optimierungsverfahrens heran. Dieser ist mit minimalen Wartezeiten verbunden, wie im Anhang A.3 genauer untersucht wird.

6.2.3. Kumar-Seidman-Netzwerk

Szenario 6.ii

In Anhang A.1 wird am Beispiel des Kumar-Seidman-Netzwerks⁴⁶ die Instabilität eines dezentralen Steuerungsverfahrens gezeigt. Nun wird auf das gleiche Beispiel das kombinierte Verfahren angewendet. Um das Szenario realistischer zu gestalten, werden diesmal nichtverschwindende Zwischenzeiten $\tau^0 = 5$ Sekunden gewählt. Die anderen Parameter sind gleich.

Trajektorienverlauf

Die zu Beginn anwachsenden Warteschlangen $n_i(t)$ werden durch das stabilisierende Regime begrenzt. Es stellt sich, wie Abb. 6.3 zeigt, ein periodisches Schaltmuster mit der gewünschten Umlaufzeit $T = 125$ s ein. Die Bedienung der zweispurigen Zufahrten 1 und 3 wird jeweils genau dann gestartet, wenn dort $\hat{n}_1^{\text{crit}} = 2q_A T = 75$ abzufertigende Fahrzeuge prognostiziert werden. Das ist der Fall, wenn sich während der Sperrzeit eine Warteschlange mit $n_1(t) = 50$ Fahrzeugen aufgebaut hat.

Selbstständige Versatzzeitenanpassung

Während des Übergangs zum periodischen Schaltmuster wird Zufahrt 3 mehrmals mit einer geringeren als der maximalen Rate bedient. Diese Ineffizienz, die unter der Clear-Largest-Buffer-Policy für die Instabilität verantwortlich war, trägt nun zum Ausgleich der Rückstaulängen und zur Anpassung der Versatzzeiten zwischen den beiden Knotenpunkten bei: Es wird nur derjenige Knotenpunkt mit geringerer Effizienz bedient, der weniger Fahrzeuge auf seinen Zufahrten hat und folglich dem anderen voraus eilt. Der andere Knotenpunkt bedient effizienter und kann dadurch aufholen. Die Versatzzeiten sind ausgeglichen, sobald die Freigabezeit von Zufahrt 1 (bzw. 3) innerhalb des Freigabezeitfensters von Zufahrt 2 (bzw. 4) liegt. Indem jede Abweichung ausgeglichen wird, wird das periodische Schaltmuster stabilisiert.

6.2.4. Zusammenfassung

Das kombinierte Verfahren eignet sich zur dezentralen Online-Steuerung von Lichtsignalanlagen in Straßennetzwerken. Es garantiert den Verkehrsströmen die mini-

⁴⁶Vgl. auch Abschnitt 2.2.1 und Abb. 2.6 (auf S. 22)

Kapitel 7.

Simulationsergebnisse

Kurzfassung

Nun soll das Verfahren zur dezentralen Online-Steuerung, das kombinierte Verfahren aus Kapitel 6, anhand von Simulationsstudien genauer untersucht werden.

Zuerst wird gezeigt, wie sich das kombinierte Verfahren unter verschiedenen, regelmäßigen wie auch irregulären, Verkehrslasten verhält. Danach wird seine Fähigkeit zur Koordinierung der Knotenpunkte entlang einer Arterie und abschließend in einem komplexen Netzwerk untersucht. Im letzteren verläuft der Zufluss in das Netzwerk entlang einer empirischen Tagesganglinie.

In den Testszenarien wird das kombinierte Verfahren mit einer koordinierten Festzeitsteuerung verglichen, welche an jedem Knotenpunkt für den größten Verkehrsstrom eine Grüne Welle einrichtet. In diesem Vergleich wird sich zeigen, ob die lose Kopplung über die Verkehrsströme der Festzeitkoordinierung überlegen ist.

Neben den Warteschlangenlängen und Wartezeiten wird auch der Kraftstoffverbrauch betrachtet. Für diesen spielt die Anzahl der Haltevorgänge eine entscheidende Rolle, welche mit dem kombinierten Verfahren höher ausfällt als mit der koordinierten Festzeitsteuerung. In diesem Vergleich muss das kombinierte Verfahren den Mehrverbrauch bei den Beschleunigungsmanövern durch Einsparungen des Leerlaufverbrauchs bei den Wartezeiten ausgleichen können.

7.1. Simulationsdurchführung

Das in den Kapiteln 4 bis 6 vorgestellte Verfahren wurde vollständig in Mathworks Matlab v.7.3.0 implementiert. Eine Computersimulation erlaubt nun den Vergleich mit einer koordinierten Festzeitsteuerung. Die Simulationsergebnisse sind vielversprechend, obgleich anzumerken ist, dass es sich hierbei um idealisierte Testszenarien handelt. Beispielsweise wird angenommen, dass der Verkehrsfluss exakt dem zugrunde liegendem Modell folgt und dass keine Fußgängerströme und ÖPNV-Fahrzeuge zu berücksichtigen sind. Aber weil diese Annahmen für beide Steuerungsverfahren in gleicher Weise gelten, ist die Vergleichsstudie dennoch zulässig. Zunächst wird aufgelistet, welche Parameter und welches Verkehrsflussmodell für die Simulationen verwendet werden.

7.1.1. Verwendete Parameter

Systemparameter: Die Sättigungsverkehrsstärke jeder Fahrspur beträgt 1800 Fz/h, die zulässige Höchstgeschwindigkeit aller Streckenabschnitte 50 km/h. Für die Zwischenzeiten τ^0 werden einheitlich 5 Sekunden angenommen.

Modellparameter: Die Umlaufzeit T^{cyc} und die Wunschperiode T haben den gleichen Wert von 120 Sekunden, die maximale Bedienperiode T^{max} beträgt 180 Sekunden.

Simulationsparameter: Jeder Simulationslauf beginnt mit einem leeren Straßennetz. Es werden (außer im Szenario 7.v) stets 90 Minuten simuliert. Das Simulationsergebnis entspricht dem arithmetischen Mittel der untersuchten Messgrößen über die letzten 60 Minuten. Die Simulationsschrittweite ist konstant und beträgt 0,5 Sekunden.

7.1.2. Verkehrsflussmodell

Punkt-Warteschlangenmodell

Quellen und Senken: Die Quellen der Verkehrsströme werden in den betrachteten Szenarien explizit benannt, und, wo es relevant ist, auch die Abbiegeanteile an den Knotenpunkten. Die Fahrzeuge verlassen das Straßennetzwerk ausschließlich an seinen Rändern.

Kapazitäten: Die Rückstaukapazität der Knotenzufahrten ist unbegrenzt. Der Ausfluss aus den Knotenzufahrten ist stets mit maximaler Rate möglich.

Prognose: Die Fahrzeuge werden mit identischen Geschwindigkeiten und vernachlässigbaren Anfahr- und Bremsverzögerungen angenommen. Damit kann die Prognose der erwarteten Ankünfte $N_i^{\text{exp}}(t)$ direkt auf eine Zeitverschiebung gemäß Gleichung (3.3) zurückgeführt werden. Als stromaufwärtiger Messquerschnitt wird der Zufluss in den Streckenabschnitt am vorgelagerten Knotenpunkt gewählt.

Für Prognosen, die über die Reisezeit der Streckenabschnitte hinaus gehen, wird $N_i^{\text{exp}}(t)$ beim letzten bekannten Wert beginnend mit der mittleren Ankunftsrate \bar{q}_i^{exp} extrapoliert. Die mittleren Ankunftsraten \bar{q}_i^{exp} auf den Knotenzufahrten selbst werden noch vor Beginn des Simulationslaufs aus den gegebenen Quellflüssen und den gegebenen Abbiegeanteilen berechnet.

Koordinierte Festzeitsteuerung

Freigabezeiten: Das umlaufbasierte Verfahren weist den Knotenzufahrten i die Freigabezeiten g_i^0 proportional zu ihrer mittleren Auslastung $\bar{\lambda}_i$ zu:

$$g_i^0 = \frac{\lambda_i}{\sum_{i'} \lambda_{i'}} \left(T^{\text{cyc}} - \sum_{i'} \tau_{i'} \right). \quad (7.1)$$

Damit ist die Forderung (2.5), d.h. $g_i^0 > \lambda_i T^{\text{cyc}}$, immer erfüllt, wenn auch die hinreichende Stabilitätsbedingung (2.10) erfüllt ist.

Phaseneinteilung: Ebenso wie in den vorherigen Kapiteln wird auch hier angenommen, dass sämtliche Verkehrsströme eines Knotenpunktes unverträglich zueinander sind. Die Knotenzufahrten werden einzeln der Reihe nach im Uhrzeigersinn bedient. Eine Variation der Phasenabfolge geschieht nicht.

Versatzzeiten: Es wird eine vereinfachte Version des Dominanzverfahrens⁴⁷ verwendet: Jeder Knotenpunkt ermittelt die größte Fahrzeugkolonne, die während einer Umlaufzeit ankommt. Dazu untersucht jeder Knoten die Bedienvorgänge an all seinen Nachbarknoten. Dies geschieht mit Hilfe der bekannten mittleren Flussraten, der gegebenen Abbiegeanteile und den mit (7.1) berechneten Freigabezeiten. Die Versatzzeit wird so berechnet, dass die größte Kolonne mit einer Grünen Welle bedient wird. Zu beachten ist hierbei, dass die Abhängigkeiten keine geschlossene Schleife bilden dürfen. Ist dies der Fall, richtet der Knotenpunkt, bei dem die größte Kolonne kleiner als bei den anderen ist, eine Grüne Welle für die nächstkleinere Kolonne ein.

7.2. Isolierter Knotenpunkt

Die Untersuchungen sollen mit einem isolierten vierarmigen Knotenpunkt beginnen, an dem die Verkehrsstärke von zwei Zufahrten variiert wird. Der Knotenpunkt wird mit vier verschiedenen Schaltprogrammen betrieben, dem Optimierungsverfahren, dem Stabilisierungsverfahren, dem kombinierten Verfahren und einer Festzeitsteuerung. Zunächst sind die Zuflussraten konstant. Einen besonderen Einblick in das Schaltverhalten des kombinierten Verfahrens gewährt Abb. 7.3, in der die Schaltmuster über eine Bedienperiode aufgetragen sind. Anschließend fließen dem

⁴⁷Vgl. Abschnitt 2.3.5

Knotenpunkt Fahrzeugkolonnen zufälliger Größe zu. Es zeigt sich in Abb. 7.5, dass sowohl der Mittelwert als auch die Varianz der Warteschlangenlänge mit dem kombinierten Verfahren deutlich geringer ausfällt als mit der Festzeitsteuerung.

7.2.1. Konstante Zuflussraten

Szenario 7.i

Es wird der Knotenpunkt aus Abb. 7.1 betrachtet. Die Zuflussraten sind konstant und auf den jeweils gegenüberliegenden Knotenzufahrten identisch. Sie beträgt für die beiden einspurigen Zufahrten in Nord-Süd-Richtung $q^{\text{nord}} = q^{\text{sued}} = 180$ Fz/h. Variiert werden die Zuflüsse in Ost-West-Richtung im Intervall $q^{\text{ost}} = q^{\text{west}} = [0, \dots 1440]$ Fz/h.

Mit diesen Werten ergeben sich für die Knotenzufahrten Auslastungen von $\lambda^{\text{nord}} = \lambda^{\text{sued}} = 0.1$ und $\lambda^{\text{ost}} = \lambda^{\text{west}} = [0, \dots 0, 4]$. Folglich variiert die Knotenpunktauslastung $\Lambda = \sum \lambda$ zwischen 0,2 und 1,0. Damit ist die notwendige Stabilitätsbedingung (2.3), d.h. $\Lambda < 1$, bis auf den obersten Extremwert im gesamten Parameterraum erfüllt. Nicht überall erfüllt ist jedoch die hinreichende Stabilitätsbedingung (2.10) bzw. (5.12). Sie fordert, dass die Umlaufzeit T^{cyc} und die Wunschperiode T größer als ein von der Auslastung Λ abhängiger Mindestwert ist, d.h.:

$$T^{\text{cyc}} = T > \frac{\sum_i \tau_i}{1 - \Lambda} \quad \text{bzw.} \quad 120 \text{ s} > \frac{20 \text{ s}}{1 - \Lambda}. \quad (7.2)$$

Diese Bedingung ist nur für $\Lambda < 5/6 = 0,833$ erfüllt. Das heißt, auf den zweispurigen Zufahrten muss die Auslastung $\lambda^{\text{ost}} = \lambda^{\text{west}} < 0,317$ bzw. der Zufluss $q^{\text{ost}} = q^{\text{west}} < 1140$ sein. Anderenfalls kann mindestens eine Zufahrt nicht mehr die nötige Freigabezeit innerhalb der geforderten Bedienperiode erhalten, was in der Folge zu einer Divergenz der Warteschlangenlänge führen muss. In Abb. 7.2 ist die Grenze des stabilen Bereichs mit einer gepunkteten Geraden markiert.

Beobachtung

Im Folgenden wird die Gesamtwarteschlangenlänge \bar{n} untersucht, die sich aus den zeitlichen Mittelwerten der Warteschlangenlänge aller vier Knotenzufahrten zusammensetzen soll:

$$\bar{n} = \bar{n}^{\text{nord}} + \bar{n}^{\text{ost}} + \bar{n}^{\text{sued}} + \bar{n}^{\text{west}}. \quad (7.3)$$

Sie ist ein direktes Maß für die mittlere Zuwachsrate der Gesamtwartezeiten $W(t)$. Abbildung 7.2 trägt \bar{n} für die verwendeten Schaltprogramme über verschiedene Knotenpunktauslastungen Λ auf.

Am auffälligsten ist der Kurvenverlauf des Optimierungsverfahrens. Obwohl die Warteschlangenlängen bei kleinen bis mittleren Auslastungen $\Lambda < 0,6$ am gerings-

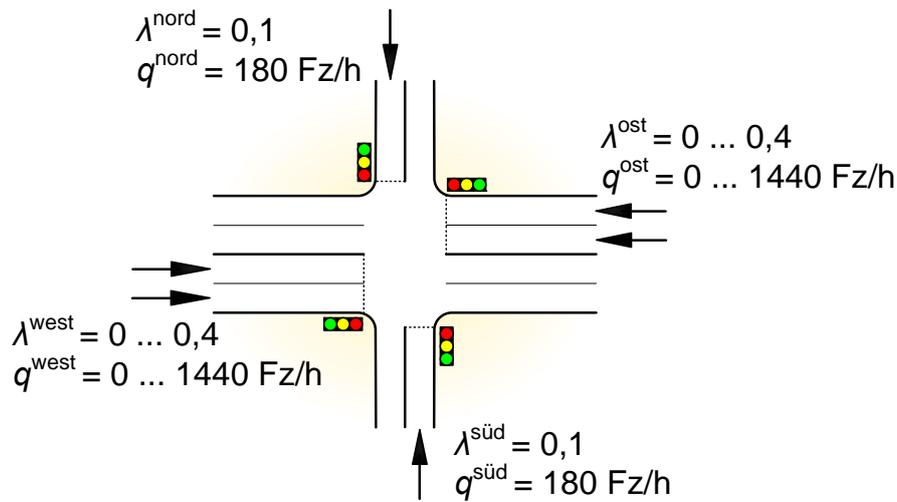


Abbildung 7.1.: Isolierter vierarmiger Knotenpunkt. Die mittleren Zuflussraten $q^{\text{nord}} = q^{\text{süd}}$ der einspurigen Zufahrten betragen 180 Fz/h. In den verschiedenen Simulationsläufen werden die Zuflüsse $q^{\text{ost}} = q^{\text{west}}$ der zweispurigen Zufahrten zwischen 0 und 1440 Fz/h variiert.

ten sind, divergieren diese bei größeren Auslastungen. Diese Instabilität tritt weit unterhalb der Stabilitätsgrenze auf.

Bemerkenswert ist auch der Verlauf des kombinierten Verfahrens. Ihm gelingt es, solange das Optimierungsverfahren stabil ist, dessen Verlauf zu folgen. Anschließend nähert es sich in Stufen von unten her der Kurve des Stabilisierungsverfahrens. Da die Stufensprünge immer größer werden, hat die Kurve einen ausgeprägt progressiven Verlauf. Das heißt, dass die hohen Werte für \bar{n} erst bei großen Auslastungen angenommen werden und vorher weit darunter liegen.

Mit dem Stabilisierungsverfahren ist die Gesamtwarteschlangenlänge stets größer als mit der Festzeitsteuerung. Mit beiden Verfahren nimmt die Gesamtwarteschlangenlänge \bar{n} im gesamten stabilen Bereich stetig und monoton mit der Knotenpunktauslastung Λ zu. Bei einer Auslastung unmittelbar unterhalb der Stabilitätsgrenze treffen sich die drei Kurven der bis dahin stabilen Verfahren in einem Punkt. Bei Auslastungen $\Lambda > 0,833$ sind alle betrachteten Schaltprogramme instabil.

Interpretation

Die **Festzeitsteuerung** bietet den Verkehrsströmen mit der Zuweisungsregel (7.1) größere Freigabezeitfenster als zur Abfertigung der Verkehrsnachfrage minimal nötig wäre. Die Freigabezeit wird über die Leerung der Warteschlange hinaus verlängert, wodurch ein Teil der Fahrzeuge überhaupt nicht verzögert wird. Die Kurve

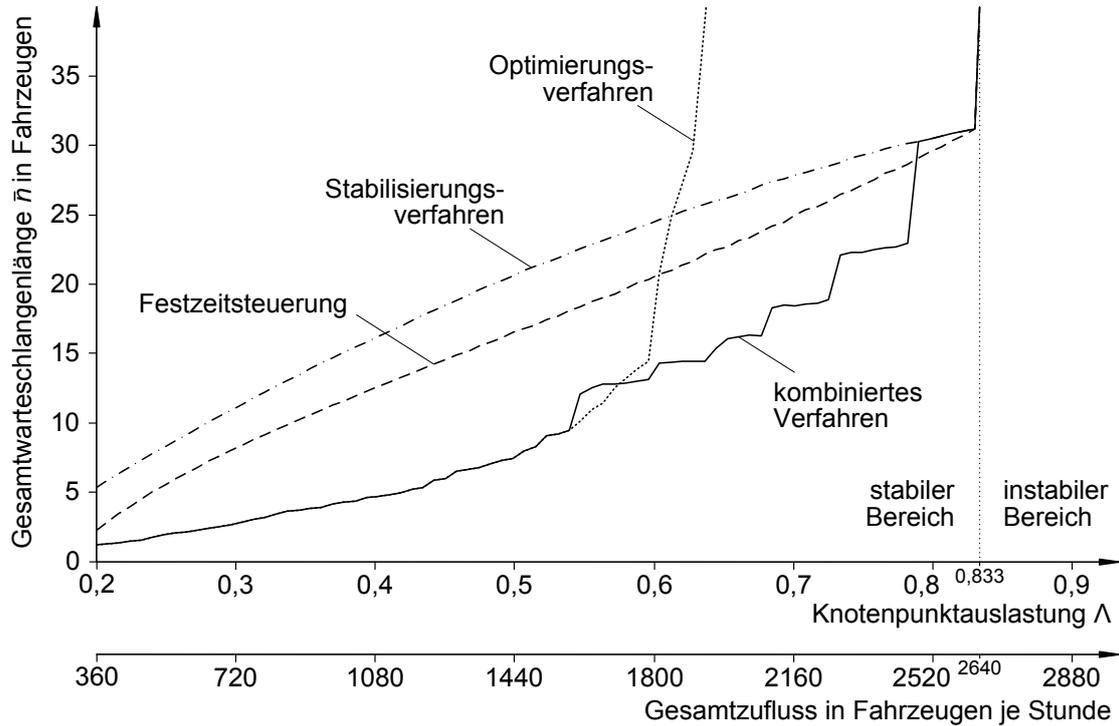


Abbildung 7.2.: Die Gesamtwarteschlangenlänge \bar{n} mit jedem der vier verwendeten Schaltprogramme ist über der Knotenpunktauslastung Λ aufgetragen. Mit der Festzeitsteuerung, dem Stabilisierungsverfahren und dem kombinierten Verfahren sind die Warteschlangen im gesamten stabilen Bereich $\Lambda < 0,833$ beschränkt. Mit dem Optimierungsverfahren divergieren die Warteschlangen dagegen bereits bei Auslastungen $\Lambda > 0,6$.

der Gesamtwarteschlangenlänge $\bar{n}(\Lambda)$ lässt sich analytisch beschreiben.⁴⁸ Bei einer Auslastung unmittelbar unterhalb der Stabilitätsgrenze kann jede Zufahrt nur noch exakt die minimal benötigte Freigabezeit erhalten. Eine andere Aufteilung der Freigabezeiten würde zu instabilem Verhalten führen. Weil sich dort die Kurve \bar{n} mit der des Stabilisierungsverfahrens und der des kombinierten Verfahrens

⁴⁸Die Freigabezeit g_i^0 einer Zufahrt i ist eine Funktion (7.1) ihrer Auslastung λ_i . Mit Hilfe von Gleichung (2.11) und der Beziehung

$$\bar{n}_i(\lambda_i) = W_i^{\text{cyc}}(g_i^0(\lambda_i))/T^{\text{cyc}} \quad (7.4)$$

lässt sich die Gesamtwarteschlangenlänge $\bar{n}(\Lambda)$ in Abhängigkeit der Knotenpunktauslastung Λ wie folgt ausdrücken:

$$\bar{n}(\Lambda) = \sum_i \frac{q_i^{\max} \lambda_i (T^{\text{cyc}} - g_i^0(\lambda_i))^2}{2T^{\text{cyc}}(1 - \lambda_i)}. \quad (7.5)$$

trifft, kann geschlussfolgert werden, dass auch die beiden anderen Verfahren die Freigabezeiten in der gleichen und einzig stabilen Weise aufteilen.

Das **Optimierungsverfahren** wendet die dynamische Priorisierungsregel (4.35) an, die aus einem heuristischen Optimierungsverfahren abgeleitet wurde. Jeder Schalteingriff soll die Wartezeiten der momentan im Wirkungsbereich⁴⁹ befindlichen Fahrzeuge minimieren. Falls neue Fahrzeuge hinzukommen, was in diesem Szenario permanent der Fall ist, wird reoptimiert. Dieser Ansatz funktioniert sehr gut bei geringen bis mittleren Verkehrsstärken. In Anhang A.3 wird gezeigt, wie das Optimierungsverfahren bei konstanten Zuflussraten zum optimalen Grenzyklus strebt.⁵⁰ Dennoch versagt das Verfahren bei höheren Verkehrsstärken. Es bedient dann abwechselnd die beiden Hauptströme und niemals mehr die Nebenstraßen. Sobald eine der beiden Hauptzufahrten geleert wurde, haben sich auf der anderen so viele Fahrzeuge aufgestaut, dass diese vor den Nebenstraßen geleert wird. Das wiederholt sich und die Nebenstraßen kommen niemals zum Zug. Die Divergenz der Gesamtwarteschlangenlänge zeigt, dass die lokale Optimierung bei größeren Auslastungen instabil ist.

Das **Stabilisierungsverfahren** reagiert ebenso wie das Optimierungsverfahren lediglich auf die lokalen Fahrzeugankünfte. Trotzdem sind die Schaltprinzipien beider Verfahren grundsätzlich verschieden. Das Stabilisierungsverfahren bedient jede Zufahrt im stationären Fall genau einmal innerhalb der Bedienperiode T und nur genau so lange bis ihre Warteschlange vollständig geleert ist. Während der verbleibenden Zeit sind alle Zufahrten gesperrt.⁵¹ Dass das Stabilisierungsverfahren nur die minimal nötige Bedienfrequenz und nur die minimal nötige Freigabezeit aufbringt, erklärt die höhere Gesamtwarteschlangenlänge⁵² \bar{n} im Unterschied zu den anderen Verfahren. Es ist der sparsame Umgang mit der Ressource Abfertigungskapazität, der dem Stabilisierungsverfahren stabiles Verhalten bis an die kritische Auslastungsgrenze heran garantiert.

Das **kombinierte Verfahren** hat beinahe im gesamten Spektrum von Λ die kleinste Gesamtwarteschlangenlänge \bar{n} . Bei geringen Auslastungen, wo das Opti-

⁴⁹Vgl. Abb. 3.6

⁵⁰Vgl. Abb. 4.4

⁵¹Vgl. Abb. 5.4 und die Schaltregeln (5.13) bis (5.15)

⁵²Die Freigabezeit g_i^0 einer Zufahrt i entspricht mit dem Stabilisierungsverfahren im stationären Fall exakt dem Teil λ_i der Bedienperiode T . Es gilt:

$$g_i^0(\lambda_i) = \lambda_i T. \quad (7.6)$$

Das lässt sich in Gleichung (2.11) einsetzen und mit der Beziehung (7.4) und der Ersetzung $T^{\text{cyc}} = T$ in folgenden analytischen Ausdruck für die Gesamtwarteschlangenlänge $\bar{n}(\Lambda)$ in Abhängigkeit der Knotenpunktauslastung Λ umformen:

$$\bar{n}(\Lambda) = \frac{T}{2} \sum_i q_i^{\text{exp}} (1 - \lambda_i). \quad (7.7)$$

mierungsverfahren stabil ist, sind die Kurven \bar{n} beider Verfahren identisch. Das kombinierte Verfahren befindet sich in diesem Bereich ausschließlich im optimierenden Regime. Um $\Lambda = 0,55$ sind die \hat{n} -Werte des kombinierten Verfahrens vorübergehend größer als die des Optimierungsverfahrens. Das deutet auf den Eingriff des stabilisierenden Regimes hin. Obwohl das optimierende Regime bei dieser Auslastung gerade noch stabil ist, bedient es die Knotenzufahrten nicht mehr innerhalb der gewünschten Bedienperiode T . Die Eingriffe des stabilisierenden Regimes stellen auch bei größeren Auslastungen das Einhalten der Bedienperiode sicher. Diese Eingriffe sind auch für den stufenförmigen Verlauf von \bar{n} verantwortlich: Der aktuelle Bedienvorgang des optimierenden Regimes kann durch diese Eingriffe jederzeit unterbrochen werden, beispielsweise auch noch während des Umschaltens oder zu Beginn der Freigabezeit. In solchen Fällen wäre es günstiger gewesen, den vorherigen Bedienvorgang fortzuführen. Das erklärt auch, warum die Kurve \bar{n} bei kleineren als der kritischen Auslastung auf die des Stabilisierungsverfahrens stößt: Das optimierende Regime startet einen neuen Bedienvorgang, kann aber den Umschaltvorgang nicht abschließen. Die resultierenden Freigabezeiten sind dann ebenso kurz wie mit dem Stabilisierungsverfahren. Einen genaueren Einblick in das Schaltverhalten des kombinierten Verfahrens erhält man im Folgenden durch die Untersuchung der stationären Schaltmuster.

Schaltmuster des kombinierten Verfahrens

Abbildung 7.3 zeigt die periodischen Schaltmuster des kombinierten Verfahrens, die sich für verschiedene Werte von Λ im stationären Fall einstellen. Jedes Schaltmuster beginnt mit einem Bedienvorgang für die südliche Zufahrt und endet mit dem Beginn des darauf folgenden. Bei Auslastungen $\Lambda \leq 0,53$ ist die tatsächliche Bedienperiode kleiner als $T = 120$ s, was bestätigt, dass sich das kombinierte Verfahren ausschließlich im optimierenden Regime befindet. Bei größeren Auslastungen wird die Wunschperiode T exakt eingehalten. In den Schaltmustern wird die Anpassungsfähigkeit des kombinierten Verfahrens an die variable Knotenpunkttauslastung besonders deutlich. Während die Zufahrten bei sehr kleinen Auslastungen $\Lambda = 0,29$ noch nacheinander abgefertigt werden, stellt sich bereits ab $\Lambda = 0,33$ ein qualitativ anderes Schaltmuster ein: Zwischen den Bedienvorgängen für die einspurigen Zufahrten (2. und 4. Streifen) werden die beiden zweispurigen Zufahrten (1. und 3. Streifen) jeweils einmal, ab $\Lambda = 0,45$ sogar in abwechselnder Folge ein- und zweimal bedient. Weiterhin geht bei größeren Auslastungen die Häufigkeit der Umschaltvorgänge zurück. Bei der größten dargestellten Auslastung $\Lambda = 0,81$ wird jede Zufahrt nur noch genau einmal innerhalb der Wunschperiode T bedient.

Besonders bemerkenswert ist, dass sich die Schaltmuster bei hohen Auslastungen über mehrere Bedienperioden hinweg erstrecken. Betrachtet wird dazu die Schaltfolge der zweispurigen Zufahrten bei $\Lambda = 0,57$. Hier wird die westliche

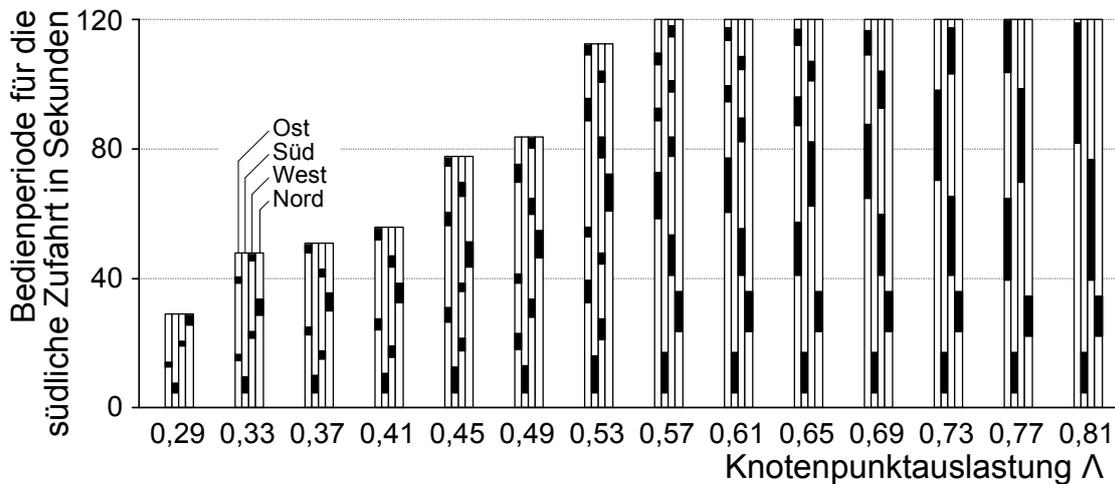


Abbildung 7.3.: Periodische Schaltmuster des kombinierten Verfahrens im stationären Fall. Die vier vertikalen Streifen innerhalb eines Balkens stehen für die Zufahrten in der Reihenfolge Ost-Süd-West-Nord. Die schwarzen Abschnitte zeigen die Lage der Freigabezeitfenster an. Jeder Balken beginnt mit einem Bedienvorgang für die südliche Zufahrt (2. Streifen). Bei kleinen Auslastungen ist die Bedienperiode für alle Zufahrten kleiner als T , bei größeren nur noch für die zweispurigen (1. und 3. Streifen). Darin, dass mit zunehmender Auslastung Λ die Häufigkeit der Umschaltvorgänge abnimmt, zeigt sich die Anpassungsfähigkeit des kombinierten Verfahrens. Weiterhin ist zu erkennen, dass die Schaltmuster selbst eine größere Periode als T haben (Erläuterung im Text).

Zufahrt (3. Streifen) sowohl als erste als auch als letzte bedient. Sobald das optimierende Regime in der sich anschließenden Periode erneut am Zuge ist, wird es die östliche Zufahrt (1. Streifen) als erste auswählen, weil sich dort eine längere Warteschlange gebildet haben wird. Diese Periodendopplung lässt sich auch in sämtlichen anderen Schaltfolgen mit hoher Auslastung beobachten. Nur bei $\Lambda = 0,81$ reicht die verbleibende Freigabezeit nicht mehr für einen neuen Bedienvorgang aus.

7.2.2. Kraftstoffverbrauch

In Abschnitt 3.2.3 wurde festgestellt, dass der Kraftstoffverbrauch für ein Beschleunigungsmanöver 70 Sekunden Warten im Leerlauf entspricht. Zwar reduziert das kombinierte Verfahren mit der höheren Bedienfrequenz die Wartezeiten der Fahrzeuge, aber es erhöht gleichzeitig die Anzahl der Haltevorgänge. Im Unterschied dazu sind mit dem umlaufbasierten Verfahren zwar die Wartezeiten größer, aber es erlaubt einem Teil der Fahrzeuge die Durchfahrt ohne anzuhalten.

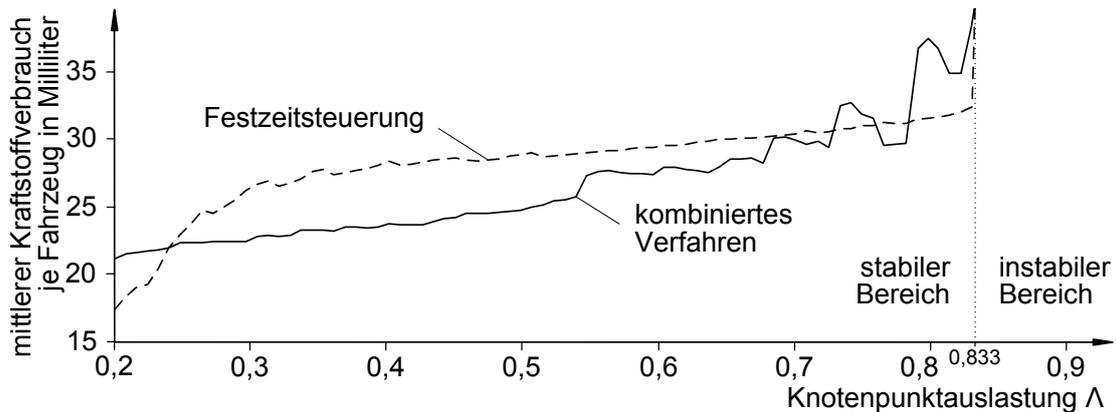


Abbildung 7.4.: Mittlerer Kraftstoffverbrauch je Fahrzeug. Für Auslastungen Λ zwischen 0,23 und 0,73 kann das kombinierte Verfahren mit den deutlich geringeren Wartezeiten die häufigeren Beschleunigungsmanöver erfolgreich ausgleichen. Bei höheren Auslastungen wird durch das Unterbrechen von Bedienvorgängen und das Nachrücken der Wartekolonnen mehr Kraftstoff verbraucht als mit der Festzeitsteuerung.

Beobachtung

Abbildung 7.4 trägt für beide Schaltverfahren den mittleren Kraftstoffverbrauch je Fahrzeug auf. Mit der Festzeitsteuerung ist der Kraftstoffverbrauch bei sehr kleinen und bei großen Auslastungen geringer. Die Kurve des kombinierten Verfahrens liegt stets oberhalb der 20ml-Marke, dem Kraftstoffverbrauch für ein Beschleunigungsmanöver, und zeigt bei höheren Auslastungen $\Lambda > 0,55$ eine zunehmende Variabilität.

Interpretation

Bereits bei Auslastungen ab $\Lambda > 0,23$ gelingt es dem kombinierten Verfahren, durch geringere Wartezeiten den Gesamtkraftstoffverbrauch zu senken. Denn mit der Festzeitsteuerung erhält nur ein geringer Teil der Fahrzeuge freie Durchfahrt. Der andere Teil muss dagegen vergleichsweise lange warten.

Einen sprunghaften Anstieg hat die Kurve des kombinierten Verfahrens um $\Lambda = 0,55$, wofür die Eingriffe des stabilisierenden Regimes verantwortlich sind. Bei jedem Eingriff wird ein bereits begonnener Bedienvorgang unterbrochen. Die Fahrzeuge, die noch nicht herausfahren konnten, müssen wieder anhalten und das nächste Mal aufs Neue beschleunigen.⁵³ Weil die Warteschlangen mit zunehmender Knotenpunktauslastung Λ immer länger werden, wird auch durch das Nachrücken

⁵³Entsprechend dem verwendeten Modell (3.18) ist der Kraftstoffverbrauch für ein in der Warteschlange nachrückendes Fahrzeug ebenso groß, wie für ein Fahrzeug, das aus dem Stand auf seine Wunschgeschwindigkeit beschleunigt.

im Mittel zunehmend mehr Kraftstoff verbraucht. Im ungünstigsten Fall wird ein Bedienvorgang kurz nach dem Beginn der Freigabezeit abgebrochen. Dann müssen für nur wenige abgefertigte Fahrzeuge alle anderen nachrücken. Diese Umstände erklären den hohen und unregelmäßigen Kraftstoffverbrauch bei großen Auslastungen. Darum ist es im Hinblick auf den Kraftstoffverbrauch unbedingt erforderlich, die Eingriffe des stabilisierenden Regimes zu moderieren. Im Ausblick 8.1.4 wird deshalb vorgeschlagen, kurz vor den Eingriffen, die sich ohne Weiteres prognostizieren lassen, keine neuen Bedienvorgänge zu beginnen.

7.2.3. Stochastische Zuflüsse

Die Fahrzeuge sollen nun nicht mehr gleichmäßig in der Zeit verteilt an den Knotenzufahrten ankommen, sondern in Kolonnen. Hierbei werden sowohl die Größe der Kolonnen als auch ihr zeitlicher Abstand als Zufallsgrößen modelliert.

Szenario 7.ii

Für die mittleren Zuflussraten \bar{q} gelten die Flusswerte q aus Szenario 7.i. Die Zuflussrate $q(t)$ sei aber nicht mehr konstant, sondern soll zwischen null und Maximalwert wechseln. Die Kolonnengröße Δn und ihre Bruttozeitlücken⁵⁴ Δt , werden hierbei als exponentialverteilte Zufallsgrößen angenommen. Für die Erwartungswerte der Verteilungen gelte:

$$\langle \Delta t_i \rangle = \frac{\langle \Delta n \rangle}{\bar{q}_i} \quad \text{und} \quad \langle \Delta n \rangle = 5 \text{ Fahrzeuge.} \quad (7.8)$$

Um den theoretischen Fall auszuschließen, dass sich zwei oder mehr Kolonnen überlagern, werden die erzeugten Fahrzeuge zunächst gepuffert. In die Streckenabschnitte fahren sie mit höchstens der maximalen Rate q_i^{\max} ein.⁵⁵

Beobachtung

Abbildung 7.5 zeigt die Verteilungen der Gesamtwarteschlangenlängen \bar{n} als Box-Whisker-Plot. Die „Box“ erstreckt sich vom 25% zum 75% Quantil, die „Whisker“ vom Minimal- zum Maximalwert. Der Strich innerhalb der „Box“ markiert das 50% Quantil, den Median. Jeder Verteilung liegen 25 unabhängige⁵⁶ Simulationsläufe zugrunde.

⁵⁴zeitlicher Abstand zwischen den jeweils ersten Fahrzeugen zweier aufeinander folgender Kolonnen

⁵⁵Auf den einspurigen Zufahrten gilt $q_i^{\max} = 1800 \text{ Fz/h}$, auf den zweispurigen $q_i^{\max} = 3600 \text{ Fz/h}$. Vgl. Systemparameter in Abschnitt 7.1.1

⁵⁶Damit keines der beiden Verfahren zufällig bevorzugt wird, sind die 25 unabhängigen Zuflusszeitreihen für beide Verfahren gleich.

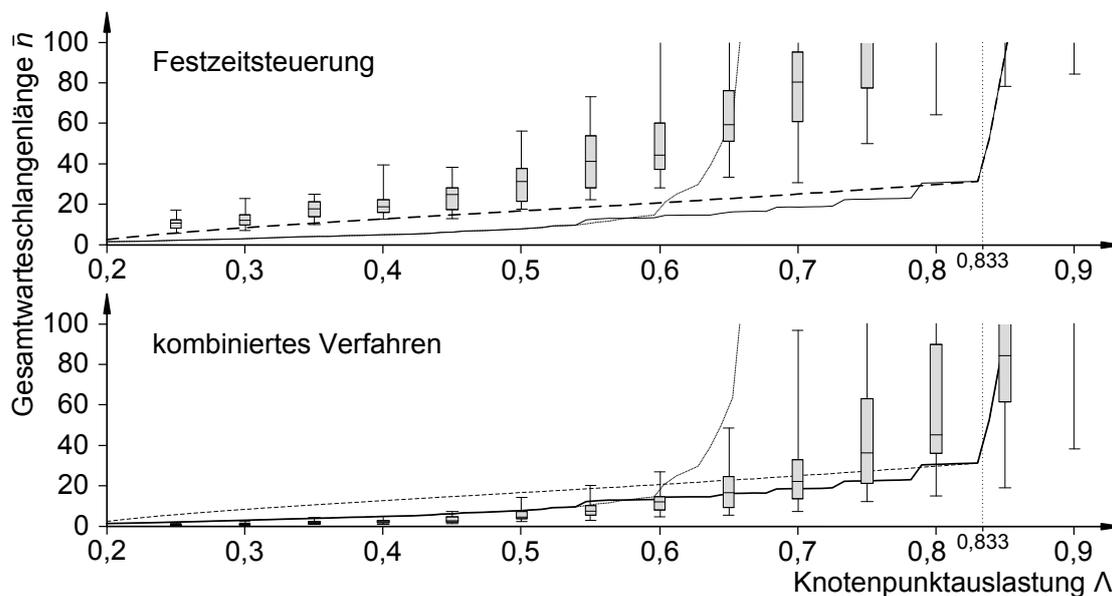


Abbildung 7.5.: Box-Whisker-Plot für die Verteilung der mittleren Warteschlangenlänge \bar{n} bei stochastischen Zuflüssen. Nicht nur der Mittelwert, sondern auch die Streuung der Warteschlangenlängen sind mit dem kombinierten Verfahren deutlich geringer als mit der Festzeitsteuerung. Die zufälligen Schwankungen im Verkehrsfluss führen mit der Festzeitsteuerung unweigerlich zu Schwankungen in den Warteschlangenlängen und den Wartezeiten. Das kombinierte Programm stellt sich auf die ankommenden Fahrzeuge ein und gleicht die Schwankungen flexibel aus.

Betrachtet wird zuerst die Festzeitsteuerung. Mit stochastischen Zuflüssen ist \bar{n} in fast jedem einzelnen Simulationslauf größer als mit konstanten Zuflussraten. Bereits bei mittleren Auslastungen kommt es vereinzelt vor, dass die Gesamtwarteschlangenlänge bei mehr als 100 Fahrzeugen liegt. Die Festzeitsteuerung scheint bereits bei unkritischen Auslastungen instabil zu sein.

Im Gegensatz dazu liegt der Median des kombinierten Verfahrens bei Auslastungen $\Lambda < 0,65$ unterhalb der Kurve für konstante Zuflussraten, anschließend vergleichsweise wenig darüber. Am auffälligsten ist jedoch die Streuung: Diese ist mit dem kombinierten Verfahren bedeutend geringer als mit der Festzeitsteuerung.

Interpretation

Das umlaufbasierte Verfahren kann sich nicht flexibel genug auf die Ankunftszeitpunkte der Fahrzeuge einstellen. Die Schwankungen im Zufluss führen zwangsläufig zu Schwankungen in den Warteschlangenlängen. Aufgrund der festen Zuteilung der Freigabezeiten können schneller angewachsene Warteschlangen erst nach meh-

renen Umläufen abgebaut werden, während gleichzeitig andere Zufahrten länger als nötig bedient werden. Deshalb dauert es sehr lange bis diese Warteschlangen geleert sind.

Das kombinierte Verfahren kompensiert die irregulären Zuflussschwankungen, indem es die Bedienvorgänge flexibel darauf anpasst. Falls es passiert, dass für die Leerung einer schnell angewachsenen Warteschlange mehrere Bedienperioden erforderlich sind, dann ist solange das stabilisierende Regime aktiv.⁵⁷ Und in dieser Zeit werden auf keiner Zufahrt die Freigabezeiten verlängert. Das führt zum schnellen⁵⁸ Abbau langer Warteschlangen. Da die Fahrzeuge nicht mit konstanter Rate ankommen, ist auch im stationären Fall kein periodisches Schaltmuster zu erwarten. Aufgrund der flexiblen Anpassung des kombinierten Verfahrens an die zufälligen Fahrzeugankünfte entsteht ein unregelmäßiges Schaltmuster, das in Konsequenz die Warteschlangenlängen und damit auch die Wartezeiten auf einem niedrigen Niveau ausgleicht.

7.3. Arterie

Im vorherigen Szenario 7.ii wurde die besondere Fähigkeit des kombinierten Verfahrens zur Anpassung an variable Zuflüsse deutlich. Nun soll sich zeigen, ob es auch zur Koordinierung im Netzwerk fähig ist. Dazu wird zunächst eine Arterie mit irregulär angeordneten Knotenpunkten betrachtet. Im ersten Schritt werden die beiden Hauptrichtungsströme in ihrer Verkehrsstärke variiert, im zweiten Schritt die Anzahl der Knotenpunkte. Während für die Festzeitsteuerung eine Grüne Welle eingerichtet wird, sind die Knotenpunkte mit dem kombinierten Programm allein über die Verkehrsströme gekoppelt. Die Untersuchung der Freigabezeitbänder in Abschnitt 7.3.2 gibt weiteren Aufschluss über das Schaltverhalten des kombinierten Verfahrens.

7.3.1. Variation der Richtungsströme

Szenario 7.iii

Es wird die in Abb. 7.6 dargestellte Arterie mit zehn irregulär angeordneten Knotenpunkten betrachtet. Die Richtungsfahrbahnen entlang der Arterie sind zwei-, die der Nebenstraßen einspurig.

Die an den Rändern zufließenden Verkehrsströme haben konstante Flussraten. Die Zuflussrate jeder Nebenstraße beträgt $q^{\text{nord}} = q^{\text{sued}} = 180$ Fz/h. Insgesamt

⁵⁷Das stabilisierende Regime ist aktiv sobald – und solange – auf einer Zufahrt mehr Fahrzeuge abgefertigt werden können als dort innerhalb der Bedienperiode T im Mittel ankommen. Vgl. Abschnitte 5.2.1 und 5.3.1.

⁵⁸Ohne den heuristischen Ansatz in Abschnitt 5.3.3 könnte man sagen: „schnellst möglichen“.

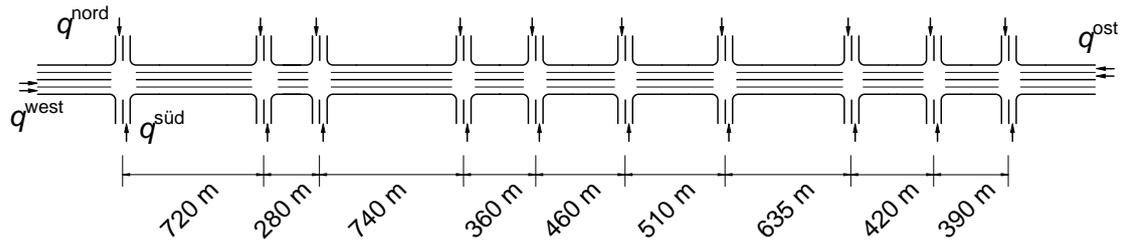


Abbildung 7.6.: Arterie mit zehn irregulär angeordneten Knotenpunkten. Die Richtungsfahrbahnen der Arterie sind zwei- und die der Nebenstraßen einspurig. An jedem Knotenpunkt biegen die Fahrzeuge zu jeweils 5% nach links und nach rechts ab. Die Zuflussraten an den Rändern sind konstant $q^{\text{nord}} = q^{\text{sued}} = 180$ Fz/h und $q^{\text{ost}} + q^{\text{west}} = 3600$ Fz/h. Ein Parameter β erlaubt es, die Verkehrsstärke der beiden Richtungsströme entsprechend den Gleichungen (7.9) und (7.10) zu variieren.

fließen im Westen und im Osten $q^{\text{west}} + q^{\text{ost}} = 1800$ Fz/h zu. Wie sich diese auf die beiden Richtungsströme aufteilen, soll mit einem Parameter β variiert werden:

$$q^{\text{west}} = (1 - \beta) \cdot 1800 \text{ Fz/h} \quad (7.9)$$

$$\text{und } q^{\text{ost}} = \beta \cdot 1800 \text{ Fz/h} \quad \text{mit } 0 \leq \beta \leq 1. \quad (7.10)$$

Damit fließen in das Netzwerk insgesamt 5400 Fz/h hinein.

An jedem Knotenpunkt biegen jeweils 5% der ankommenden Fahrzeuge nach links und nach rechts ab, 90% fahren gerade aus. Deshalb sind die Knotenpunkttauslastungen Λ nicht überall gleich. Am meisten ausgelastet ist entweder der westlichste ($\beta < 0,5$) oder der östlichste ($\beta > 0,5$) Knotenpunkt. Seine Auslastung beträgt höchstens $\Lambda = 0,7$, was bei $\beta = 0$ bzw. $\beta = 1$ der Fall ist. Damit ist die hinreichende Stabilitätsbedingung (7.2), die sich aufgrund des gleichen Knotenpunktdesigns direkt vom Szenario 7.i auf das hier betrachtete übertragen lässt, immer erfüllt.

Beobachtung

Abbildung 7.7 zeigt die Gesamtwarteschlangenlänge \bar{n} und die mittlere Wartezeit je Fahrzeug⁵⁹ \bar{w} über den Werten von β . Mit der Festzeitsteuerung haben die Wartezeiten ein Maximum um $\beta = 0,5$. Jedes Fahrzeug wird dort im Mittel

⁵⁹Die Gesamtwarteschlangenlänge \bar{n} und die mittlere Wartezeit \bar{w} je Fahrzeug sind im betrachteten Szenario direkt proportional zueinander. Das liegt daran, dass der Gesamtzufluss bezüglich β konstant ist. Es gilt:

$$\bar{w} = \frac{\bar{n}}{\sum_i q_i} = \frac{\bar{n}}{5400 \text{ Fz/h}} = \frac{\bar{n}}{1,5 \text{ Fz/s}}. \quad (7.11)$$

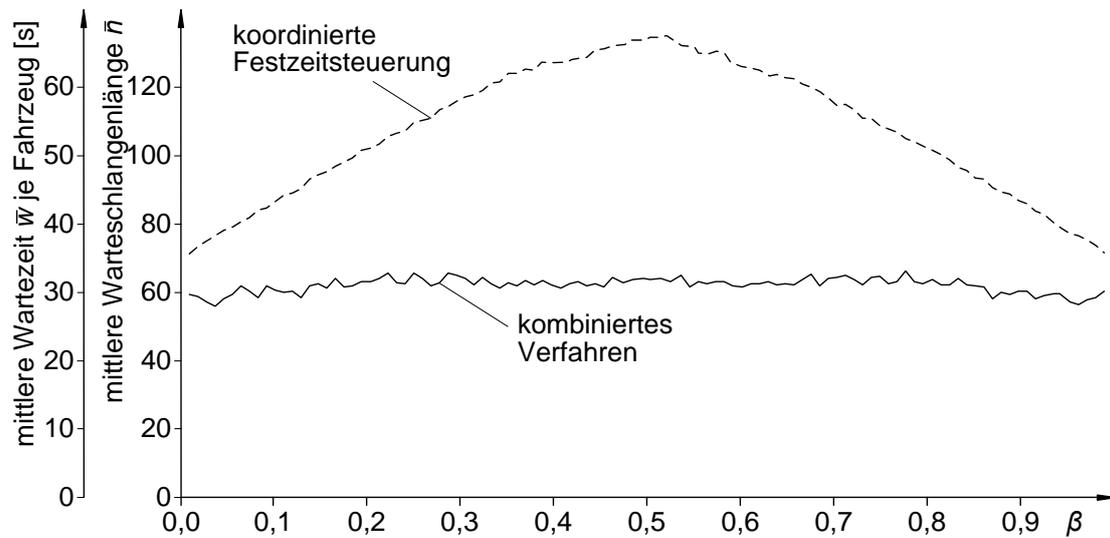


Abbildung 7.7.: Mittlere Warteschlangenlänge \bar{n} und mittlere Wartezeit \bar{w} je Fahrzeug bei Variation des Richtungsanteils β . Die Festzeitsteuerung richtet für den größten Richtungsstrom eine Grüne Welle ein und ordnet dem alle anderen Verkehrsströme systematisch unter. Je ausgeglichener die beiden Richtungsströme sind (mittlere β -Werte), desto weniger Fahrzeuge können von der Grünen Welle profitieren. Das kombinierte Verfahren versucht, die Gesamtwartezeiten aller Fahrzeuge zu minimieren. Die entstehenden Wartezeiten sind ebenso wie die Anzahl der insgesamt abzufertigenden Fahrzeuge nahezu unabhängig von β .

um $\bar{w} = 66$ Sekunden verzögert. Verglichen mit dem kombinierten Verfahren ist das mehr als das Doppelte. Mit dem kombinierten Verfahren liegt die mittlere Wartezeit nahezu unabhängig vom β bei etwa 30 Sekunden je Fahrzeug. Selbst an den Rändern bei $\beta = 0$ und $\beta = 1$ entstehen mit dem kombinierten Verfahren geringere Wartezeiten.

Interpretation

Das umlaufbasierte Verfahren richtet entsprechend dem Dominanzverfahren eine Grüne Welle ein. An jedem Knotenpunkt erhält der größte Verkehrsstrom freie Fahrt. Welche Wartezeiten dabei für die untergeordneten Verkehrsströme entstehen, hängt allein von der gegebenen Netzwerkgeometrie ab. Sie werden vom verwendeten Dominanzverfahren⁶⁰ nicht berücksichtigt. Bei sehr kleinen und sehr großen Werten für β ist der dominierende Verkehrsstrom groß und ebenso die Anzahl der Fahrzeuge, die von der Grünen Welle profitieren. Aber je mehr sich β

⁶⁰Vgl. Abschnitt 7.1.2

mittleren Werten nähert, umso mehr nimmt ihre Zahl ab und die Fahrzeuganzahl in den untergeordneten Strömen zu.⁶¹

Anstatt die Knotenpunkte mit einer Grünen Welle zu koordinieren und in Folge bestimmte Verkehrsströme systematisch unterzuordnen, fertigt das kombinierte Verfahren die lokal wartenden und herannahenden Fahrzeuge flexibel ab. Für das zugrunde liegende Optimierungsverfahren⁶² zählt die Wartezeit jedes Fahrzeugs gleich viel, ob es mit dem Hauptstrom kommt oder aus einer Nebenstraße. Das erklärt im Allgemeinen, warum die mittleren Wartezeiten – ebenso wie die im Netzwerk insgesamt abzufertigende Fahrzeuganzahl – nahezu unabhängig vom Richtungsanteil β sind.

Bei $\beta = 0$ und bei $\beta = 1$ fließen bei der Festzeitsteuerung alle Fahrzeuge des Hauptstroms mit der Grünen Welle. Dennoch sind die mittleren Wartezeiten mit dem kombinierten Verfahren geringer. Dies ist der Tatsache zu verdanken, dass es die Nebenstraßen nur so lange wie nötig bedient, und die verbleibende Zeit für den Hauptstrom nutzt. Ein genauerer Einblick in das Schaltverhalten des kombinierten Verfahrens ergibt sich über die Freigabezeitbänder.

7.3.2. Freigabezeitbänder

Beobachtung

Abbildung 7.8 zeigt die Freigabezeitbänder der beiden Hauptströme im Weg-Zeit-Diagramm für $\beta = 0, 1$ (oben) und $\beta = 0, 3$ (unten). Links ist für die koordinierte Festzeitsteuerung deutlich die von Ost nach West laufende Grüne Welle (helle Streifen) zu erkennen. Aufgrund der Abbiegeanteile werden die Bänder entlang der Arterie immer schmaler.⁶³ Der gegenläufige Verkehrsstrom (dunkle Streifen) wird an jedem Knotenpunkt gestoppt. Es sei nochmals erwähnt, dass alle Verkehrsströme eines Knotenpunktes als unverträglich angenommen wurden und deshalb entgegengesetzte Richtungsströme nicht gleichzeitig bedient werden. An der Tatsache, dass die Wartezeit des untergeordneten Stroms von der gegebenen Netzwerkgeometrie abhängt, ändert sich mit dieser Annahme nichts.

⁶¹Das abgeflachte Maximum der Wartezeitenkurve der Festzeitsteuerung aus Abb. 7.7 hat folgende Ursache: Aufgrund der Abbiegeanteile fließen bei mittleren β -Werten mehr Fahrzeuge von der Arterie in die Nebenstraßen ab als umgekehrt. Damit reduziert sich die Verkehrsstärke der Hauptströme entlang der Arterie. Die Laufrichtung der Grünen Welle schlägt folglich nicht abrupt um. Bei zunehmendem β beginnt die Umkehr der Laufrichtung am östlichen Knotenpunkt.

⁶²Vgl. Kapitel 4

⁶³Bei den Freigabezeitbändern der Festzeitsteuerung ist zu erkennen, dass die Endzeitpunkte der Freigabezeitfenster', exakt mit der Grünen Welle verlaufend, zeitversetzt angeordnet sind und nicht ihre Anfangszeitpunkte. Das ist für umlaufbasierte Verfahren optimal, wie im Punkt (v) aus Abschnitt 2.3.4 festgestellt wurde.

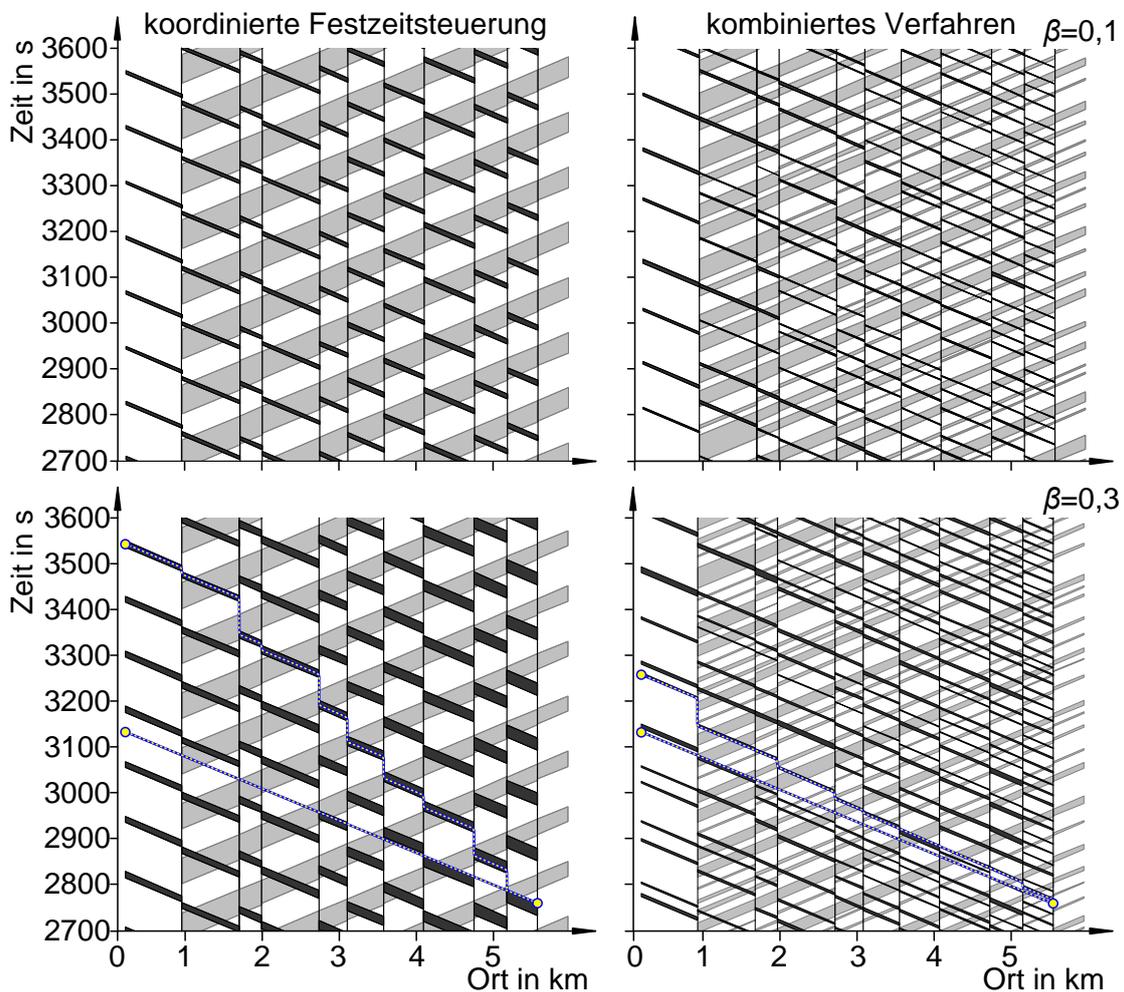


Abbildung 7.8.: Freigabezeitbänder im Weg-Zeit-Diagramm. Mit $\beta = 0,1$ (oben) und $\beta = 0,3$ (unten) ist die Hauptflussrichtung von West nach Ost. Die Festzeitsteuerung (links) bedient den Hauptstrom (helle Streifen) mit einer Grünen Welle. Der entgegengesetzte Strom (dunkle Streifen) wird an jedem Knotenpunkt verzögert. Die Freigabezeitbänder des kombinierten Verfahrens (rechts) sind irregulär. Es ist zu beobachten, wie Kolonnen wachsen und zerfallen (siehe Text). Die gepunktete Linie in den beiden unteren Bildern verfolgt ein und dasselbe Testfahrzeug. Auf seiner Fahrt entlang der Arterie wird es mit der Festzeitsteuerung 420 Sekunden verzögert, mit dem kombinierten Verfahren lediglich 120 Sekunden.

Rechts zeigt Abb. 7.8 die irregulären Freigabezeitbänder des kombinierten Verfahrens. An der höheren Anzahl der Freigabezeitbänder erkennt man die höhere Bedienfrequenz. Die Wunschperiode $T = 120$ s wird für die beiden Hauptrichtungsströme bis auf sehr wenige Ausnahmen stets deutlich unterschritten. Man kann die Entstehung und das Zerfallen von Kolonnen beobachten. Dazu wird zuerst der kleinere Richtungsstrom von Ost nach West (dunkel) für $\beta = 0,1$ betrachtet. Am ersten Knotenpunkt (rechts), wo er mit konstanter Rate zufließt, wird er häufig bedient. Es formen sich kleine Kolonnen, denen es am nächsten oder übernächsten Knoten kaum gelingt eine sofortige Bedienung auszulösen. Dort schließen sie sich zu größeren Kolonnen zusammen. Das wird zusätzlich dadurch unterstützt, dass der entgegengesetzte Richtungsstrom im Westen stärker ist. Gegen Ende (links) gibt es fast nur noch große Kolonnen. Der entgegengesetzte Richtungsstrom von West nach Ost (hell) ist bedeutend größer und wird deshalb aufgrund der prozentualen Abbiegeanteile entlang der Arterie stärker geschwächt. Die Kolonnen werden kleiner und zerfallen. An mehreren Stellen ist zu sehen, wie kleinere Kolonnenfragmente nach oben abzweigen und mit geringerer mittlerer Geschwindigkeit weiterlaufen. Unten im selben Bild sieht man, wie sich ein solches Fragment mit der nachfolgenden Kolonne verbindet.

Interpretation

Das kombinierte Verfahren versucht im optimierenden Regime, die Zeit zwischen Ankunft und Ausfahrt der Fahrzeuge zu minimieren. Die Fahrzeuge werden nicht erst innerhalb eines vorher festgelegten Zeitfensters abgefertigt, sondern sobald sich dafür eine günstige Gelegenheit bietet. Günstig ist ein Bedienvorgang dann, wenn durch diesen lokal Wartezeiten eingespart werden (optimierendes Regime) können oder wenn dieser mit mindestens der stabilitätskritischen Bedieneffizienz (stabilisierendes Regime) durchgeführt werden kann.

Das Prinzip der Grünen Welle ist, eine Fahrzeugkolonne so lange aufzuhalten, bis sie bei der Ankunft am Nachfolgerknoten unverzüglich abgefertigt werden kann. Im Gegensatz dazu leitet das kombinierte Verfahren die Fahrzeuge so bald wie möglich zum Nachbarknoten. Wann sie von dort abgefertigt werden, soll vom dortigen Verkehrszustand abhängen. Weil auch der Nachbarknoten die Zeitspanne zwischen Ankunft und Ausfahrt der Fahrzeuge minimiert, entstehen im Endeffekt deutlich geringere Wartezeiten. Das bestätigen die Abbildungen 7.7 und 7.8.

Das Koordinierungsprinzip des kombiniertes Verfahrens ist deshalb grundsätzlich verschieden zum Prinzip der Grünen Welle. Obwohl die Schaltprogramme zweier benachbarter Knotenpunkte vollständig unabhängig voneinander sind, koppelt sie der gemeinsame Verkehrsstrom. Das Schaltprinzip, Wartezeiten lokal zu minimieren und effiziente Bedienvorgänge zu bevorzugen, führt zu einer flexiblen Abfertigung der herannahenden oder bereits wartenden Kolonnen. Die Fahrzeuge lösen die Bedienvorgänge selbst aus.

7.3.3. Variation der Knotenanzahl

Im Folgenden wird die Anzahl der Knotenpunkte entlang der Arterie variiert. Bei großer Knotenanzahl verschwinden die Einflüsse der konstanten Zuflussraten an den beiden äußeren Rändern. Es soll sich herausstellen, wie gut die beiden Verfahren mit den Kolonnen innerhalb der Arterie umgehen.

Szenario 7.iv

Begonnen wird mit dem isolierten Knotenpunkt aus Abb. 7.1 (auf S. 121), dem nach und nach weitere Knotenpunkte mit zufälligem Abstand hinzugefügt werden. Die Knotenpunktabstände sind dabei gleichverteilt zwischen 200 und 800 Metern. Die bereits bestehenden Knotenpunkte behalten ihre Position bei.

Die Zuflüsse der Nebenstraßen bleiben unverändert $q^{\text{nord}} = q^{\text{sued}} = 180$ Fz/h, die beiden Hauptströme haben einen Zufluss von $q^{\text{west}} = 2400$ Fz/h und $q^{\text{ost}} = 1200$ Fz/h. Das entspricht einem Richtungsanteil von $\beta = 1/3$.

Damit es nicht passiert, dass die Hauptströme ihre Verkehrsstärke entlang der Arterie verlieren und dadurch das Ergebnis beeinflussen, seien die Abbiegeanteile null. An jedem Knotenpunkt fahren alle Fahrzeuge geradeaus. Folglich hat jeder Knotenpunkt dieselbe Auslastung $\Lambda = 0,7$. Die hinreichende Stabilitätsbedingung (7.2) ist erfüllt.

Beobachtung

Abbildung 7.9 trägt die mittleren Wartezeiten je Fahrzeug und Knotenpunkt über die Anzahl der Knotenpunkte auf. Mit beiden Verfahren nimmt dieser Wert mit zunehmender Knotenanzahl ab und konvergiert gegen einen festen Wert. Dieser beträgt für die Festzeitsteuerung 11,8 und für das kombinierte Verfahren 7,3 Sekunden.⁶⁴

Interpretation

Der abfallende Verlauf beider Kurven verdeutlicht, dass beide Verfahren die Kolonnen innerhalb der Arterie mit geringeren Wartezeiten abfertigen können als die mit konstanter Rate zufließenden Fahrzeuge an den Rändern. Beiden Verfahren gelingt eine systematische Anpassung an die Verkehrsflussprofile, der koordinierten Festzeitsteuerung mit der Grünen Welle und dem kombinierten Verfahren mit der flexiblen Abfertigung.

⁶⁴Zum Vergleich: Angenommen, die Nebenstraßen werden genau einmal innerhalb der Bedienperiode T mit genau der minimal erforderlichen Freigabezeit $g_i = \lambda_i T$ bedient. Dann beträgt nach Gleichung (7.7) auf jeder Nebenstraße die mittlere Warteschlangenlänge $\bar{n}_i = 2,7$ Fz. Dort wartet nach Gleichung (7.11) jedes Fahrzeug im Mittel $\bar{w}_i = 54$ Sekunden. Im theoretischen Fall, dass die Fahrzeuge der beiden Hauptströme nirgends verzögert werden, beträgt die mittlere Wartezeit je Fahrzeug und Knotenpunkt $\bar{w} = 4,9$ Sekunden.

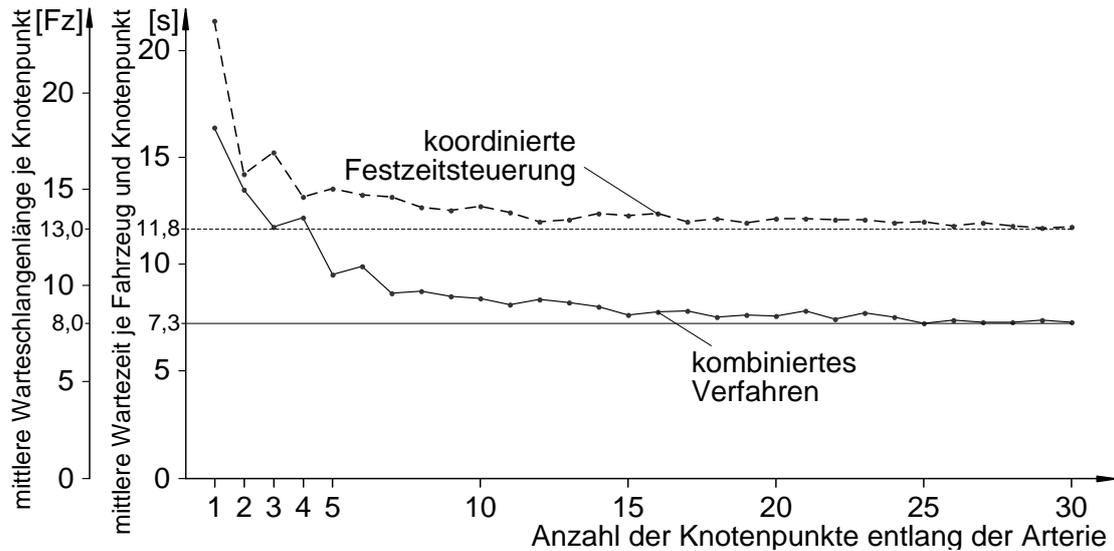


Abbildung 7.9.: Mit zunehmender Anzahl der Knotenpunkte entlang der Arterie nimmt die mittlere Wartezeit je Fahrzeug und Knotenpunkt ab. Sie konvergiert gegen einen Wert, der mit dem kombinierten Verfahren mehr als ein Drittel geringer ist als mit der Festzeitsteuerung.

Dafür, dass die flexible Abfertigung mit dem kombinierten Verfahren zu deutlich geringeren Wartezeiten führt, sind drei Gründe verantwortlich: Erstens bilden sich kompakte Kolonnen heraus. Das geschieht immer dann, wenn die Fahrzeuge der Hauptströme aufgehalten werden, beispielsweise schon am ersten Knoten. Da kein Fahrzeug der Hauptströme in die Nebenstraßen abbiegt, behalten die Kolonnen ihre maximale innere Flussrate entlang der Arterie bei. Zweitens werden die Zeitlücken zwischen den Kolonnen zur Abfertigung der Nebenstraßen genutzt. Solange die Kolonnen noch zu klein sind, um einen Bedienungsvorgang der Nebenstraße zu unterbrechen, werden sie verzögert und schließen sich mit den nachfolgenden zu größeren Kolonnen zusammen. Irgendwann haben die Kolonnen eine Größe erreicht, mit der sie von keiner Nebenstraße mehr aufgehalten werden.

Drittens werden die Kolonnen an den Knotenpunkten in der Reihenfolge ihrer Ankunft bedient. Dazu trägt sowohl das optimierende⁶⁵ als auch das stabilisierende⁶⁶ Regime bei. Dieses Verhalten ist – lokal betrachtet – wartezeitenoptimal. Das wurde in Abschnitt 4.2 (insbesondere auf Seite 76) herausgestellt: Es ist immer

⁶⁵Vgl. Szenario 4.ii

⁶⁶Entsprechend den Priorisierungsregeln (5.13) bis (5.15) werden die Zufahrten in der Reihenfolge bedient, in der ihre abzufertigende Fahrzeuganzahl $\hat{n}_i(t)$ den kritischen Schwellenwert $n_i^{\text{crit}}(t)$ überschreitet.

optimal die Abfertigung mit maximaler Rate fortzusetzen, wenn die alternativ zur Auswahl stehende Zufahrt keine größere Spurenzahl hat. Von diesem Prinzip wird nur abgewichen, wenn ein Bedienvorgang des optimierenden Regimes durch das stabilisierende Regime unterbrochen wird. Das führt im Grenzfall dazu, dass alle Kolonnen auf eine Größe über dem kritischen Wert Tq anwachsen und letztlich nur noch mit dem stabilisierenden Regime bedient werden. Das Prinzip lautet dann: Die zuerst ankommende Kolonne erhält freie Fahrt.

Wie dieses Szenario zeigt, führt das Prinzip, die Kolonnen in der Reihenfolge ihrer Ankunft und die Nebenstraße während der Zeitlücken zwischen den Kolonnen zu bedienen, zu deutlich geringeren Wartezeiten als das Prinzip der Grünen Welle. Die dafür notwendige Flexibilität kommt durch die lose Kopplung der Knotenpunkte über die Verkehrsströme zustande.

7.4. Komplexes Netzwerk

7.4.1. Tagesganglinie

Szenario 7.v

Die folgende Untersuchung bezieht sich auf den Auszug aus einem irregulären Straßennetz, wie ihn Abb. 7.10 zeigt. Am zentralen Knotenpunkt kreuzen sich zwei Hauptstraßen, deren Richtungsfahrbahnen zweispurig sind. Die Hauptstraßen sind von einem Nebenstraßennetz mit einspurigen Richtungsfahrbahnen umgeben. In die Nebenstraßen fließt an den Rändern der Strom $q(t)$ zu, in die Hauptstraßen der Strom $3q(t)$. Die Fahrzeuge biegen an jedem Knotenpunkt zu je 10% nach links und nach rechts ab, 80% fahren geradeaus.

Mit jedem der beiden Schaltprogramme wird in einem einzigen Simulationslauf das Netzwerk über einen gesamten Tag simuliert. Dabei soll der Zufluss $q(t)$ dem Verlauf der empirischen Tagesganglinie⁶⁷ aus Abb. 7.11 folgen. Links in der Abbildung ist der skalierte Zufluss $q(t)$ aufgetragen, wie er im Szenario verwendet wird. Der Skalierung wurde Folgendes zugrunde gelegt: Für einen gegebenen mittleren Zufluss \bar{q} hat der zentrale Knotenpunkt, an dem sich die beiden Hauptstraßen kreuzen, unter allen Knotenpunkten die größte Auslastung. Seine maximale Auslastung beträgt $\Lambda^{\max} = 0,833$. Dieser Maximalwert soll während der ersten Nachfragespitze am Morgen um 06:45 Uhr erreicht werden. Es soll also gelten:

$$\Lambda(\bar{q}(t = 06:45)) \stackrel{!}{=} \Lambda^{\max} = 0,833. \quad (7.12)$$

⁶⁷Die Kurve wurde am 13. Juni 2006 auf den beiden Fahrspuren der Nürnberger Straße vom Nürnberger Platz zum Fritz-Förster-Platz in Dresden aufgenommen. Die originalen Messwerte sind in der Abbildung rechts aufgetragen. Leider fehlt die letzte Stunde des Tages. (Quelle: Landeshauptstadt Dresden, Straßen- und Tiefbauamt)

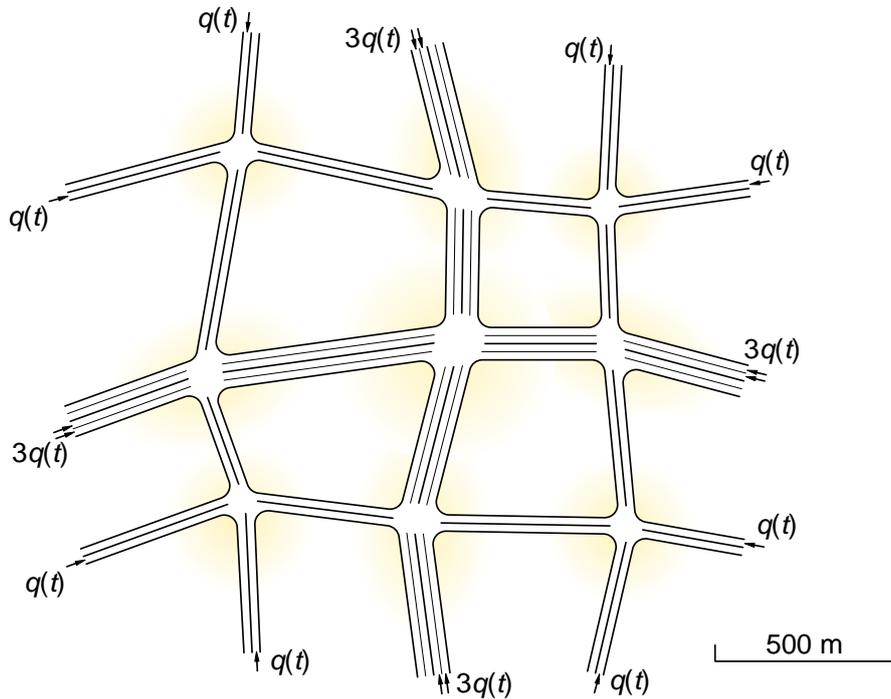


Abbildung 7.10.: Straßennetzwerk mit neun irregulär angeordneten Knotenpunkten. Die sich in der Mitte kreuzenden Hauptstraßen sind von einem Nebenstraßennetz umgeben. Die Zuflüsse $q(t)$ an den Rändern sollen einer empirischen Tagesganglinie (siehe Abb. 7.11) folgen.

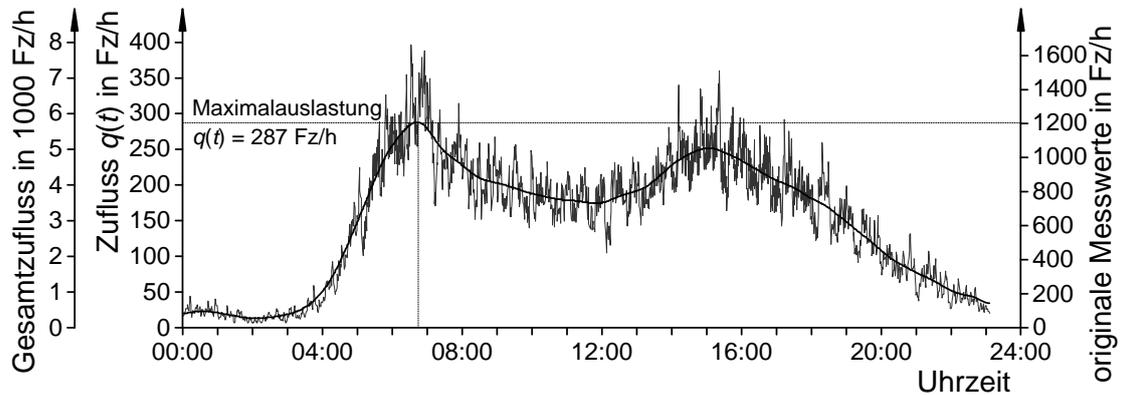


Abbildung 7.11.: Tagesverlauf der Zuflussrate $q(t)$ entlang einer empirischen Zeitreihe. Die originalen Messwerte (rechts), wie sie in Dresden auf der Nürnberger Straße erhoben wurden, werden so skaliert (links), dass der kritische Knotenpunkt zur Nachfragespitze am Morgen maximal ausgelastet ist. (Quelle: Landeshauptstadt Dresden, Straßen- und Tiefbauamt)

Numerisch lässt sich ermitteln, dass dies für $\bar{q}(t = 06:45) = 287 \text{ Fz/h}$ der Fall ist.

Schaltprogramme

Das kombinierte Verfahren gewinnt seine Schätzung der mittleren Ankunftsrate \bar{q}^{exp} aus der als bekannt angenommenen gemittelten Tagesganglinie (dicke Linie in Abb. 7.11).

Für die umlaufbasierte Steuerung soll die Umlaufzeit, wie üblich,⁶⁸ konstant bleiben. Unter den genannten Randbedingungen brauchen damit auch die Freigabezeiten g_i^0 im Laufe des Tages nicht angepasst werden. Sie bemessen sich nach Gleichung (7.1) an dem Verhältnis der Auslastungen λ_i zueinander. Weil sich aber die Flussverhältnisse an den Knotenzufahrten nicht ändern, behält die umlaufbasierte Steuerung die gleiche Freigabezeitaufteilung den ganzen Tag über bei. Es handelt sich also auch hier um eine koordinierte Festzeitsteuerung, wobei für die Koordinierung das im Abschnitt 7.1.2 erläuterte Dominanzverfahren verwendet wird.

7.4.2. Mittlere Warteschlangenlänge

Beobachtung

Abbildung 7.12 stellt die Gesamtwarteschlangenlänge $n(t) = \sum_i n_i(t)$ aller Knotenzufahrten des Netzwerks im Tagesverlauf dar.

Mit dem **kombinierten Verfahren** sind die Warteschlangen während der sehr frühen Morgenstunden nahezu konstant und liegen nur wenig über null. Auch wenn gegen fünf Uhr der morgendliche Berufsverkehr einsetzt, sind die Warteschlangen noch vergleichsweise klein. Sie steigen jedoch um halb sieben, wenn das Straßennetz gegen seine Auslastungsgrenze geht, beinahe sprunghaft an. Etwa eine halbe Stunde später, nachdem die Verkehrslast bereits wieder rückläufig ist, erreichen die Warteschlangen eine maximale Gesamtlänge von über 500 Fahrzeugen.⁶⁹ Unmittelbar danach fällt die Kurve ebenso schnell wieder ab wie sie zuvor gestiegen ist. Zwischen den späten Morgen- bis zu den frühen Nachmittagsstunden, während das Straßennetz noch immer zu etwa 70% gesättigt⁷⁰ ist, halten sich die Warteschlangenlängen auf einem sehr niedrigen Niveau von etwa 50 Fahrzeugen. Sie betragen damit ein Zehntel des Tagesmaximums. Um 15:00 Uhr, während der nachmittäglichen Nachfragespitze, steigen die Warteschlangen nochmals an. Im Mittel sind die Warteschlangenlängen dann aber nur wenig größer als mit koordinierten Festzeitsteuerung in der Mittagszeit. Sobald am Abend die Verkehrsnachfrage wieder zurück geht, konvergieren auch die Warteschlangenlängen

⁶⁸Vgl. Abschnitt 2.3.4

⁶⁹Bei insgesamt 36 Knotenzufahrten entspricht das 14 Fahrzeugen je Zufahrt.

⁷⁰Der Sättigungsgrad ist das Verhältnis aus tatsächlicher und maximaler Auslastung.

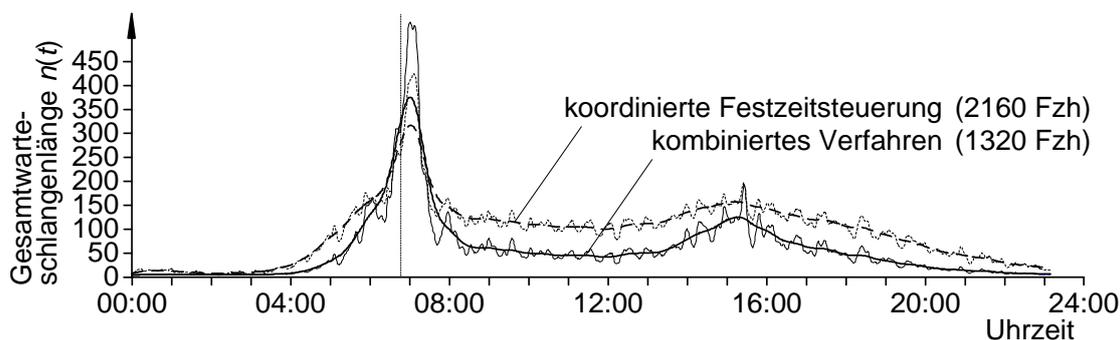


Abbildung 7.12.: Tagesverlauf der Gesamtwarteschlangenlänge $n(t)$ aller Zufahrten des Straßennetzes. Bis auf die Zeit der morgendlichen Nachfragespitze liegt die Kurve des kombinierten Verfahrens gänzlich unter der einer koordinierten Festzeitsteuerung. Über den gesamten Tag können im betrachteten Szenario mit dem kombinierten Verfahren 39% der Gesamtwartezeiten eingespart werden.

wieder gegen Werte knapp über null. Während eines gesamten Tages sind mit dem kombinierten Verfahren Gesamtwartezeiten im Umfang von 1320 Fahrzeugstunden angefallen.

Mit der **koordinierten Festzeitsteuerung** liegen die Warteschlangenlängen auch in den sehr frühen Morgenstunden deutlich über null. Dann wachsen sie zwar zwischen 04:00 und 05:30 schneller an, als mit dem kombinierten Verfahren, haben aber im Sieben-Uhr-Maximum einen um 18% geringeren Wert. Von dieser Zeitspanne abgesehen sind die Warteschlangen mit der Festzeitsteuerung den gesamten Tag lang größer als mit dem kombinierten Verfahren. Auch wenn die Warteschlangen gegen 15:00 Uhr beinahe gleich groß sind, so betragen sie doch während der gesamten Mittags- und Abendstunden mehr als das Doppelte. Insgesamt fallen mit der koordinierten Festzeitsteuerung Gesamtwartezeiten im Umfang von 2160 Fahrzeugstunden an, das sind 64% mehr als mit dem kombinierten Verfahren.

Interpretation

Die Ergebnisse sind konsistent mit den Beobachtungen am isolierten Knotenpunkt mit stochastischen Zuflüssen in Szenario 7.ii. Bei geringer Verkehrslast, d.h. in den frühen Morgenstunden und am Abend, führt die flexible Abfertigung des kombinierten Verfahrens zu sehr kleinen Warteschlangenlängen. Bei mittlerer bis hoher Verkehrslast, d.h. in den Mittags- und späten Nachmittagsstunden, kann das kombinierte Verfahren zwar nicht mehr alle Schwankungen des Zuflusses ausgleichen. Aber es gelingt ihm aufgrund der Anpassungsfähigkeit, sowohl des optimierenden als auch des stabilisierenden Regimes, die Warteschlangen auf einem niedrigen Niveau zu halten. Im Gegensatz fertigt die koordinierte Festzeitsteuerung die an-

kommenden Fahrzeuge nur innerhalb der vorher festgelegten Freigabezeitfenster ab.

Die größeren Warteschlangenlängen des kombinierten Verfahrens bei maximaler Verkehrslast haben folgende Ursache: Bereits bevor die Verkehrsnachfrage ihr Maximum erreicht, muss das stabilisierende Regime zum Einhalten der Wunschperiode T eingreifen.⁷¹ Dabei können noch nicht zu Ende geführte Bedienvorgänge des Optimierungsverfahrens unterbrochen werden, wie im Szenario 7.i erläutert wurde. Damit werden Abfertigungskapazitäten verschenkt, die später, wenn die Verkehrsnachfrage weiter ansteigt, nicht mehr zurückgewonnen werden können. Unter maximaler Verkehrslast werden mit dem kombinierten Verfahren ebenso viele Fahrzeuge abgefertigt wie mit der Festzeitsteuerung – nur können dann die zuvor entstandenen Warteschlangen nicht mehr abgebaut werden. Allerdings gelingt der Warteschlangenabbau anschließend, sobald die Verkehrslast wieder nachlässt, schneller und nachhaltiger als mit der koordinierten Festzeitsteuerung.

7.4.3. Kraftstoffverbrauch

Beobachtung

In Abbildung 7.13 wird für beide Schaltprogramme gezeigt, wie sich der momentane Kraftstoffverbrauch aller Fahrzeuge, gemessen in Liter je Sekunde, im Laufe des Tages entwickelt.

Mit dem kombinierten Verfahren liegt der Verbrauch früh um sechs und nachmittags halb vier jeweils kurzzeitig und früh gegen sieben Uhr sogar fast eine Stunde lang über dem Verbrauch mit der Festzeitsteuerung. Den Rest des Tages liegt die Kurve des kombinierten Verfahrens teils deutlich, teils geringfügig, unter der Kurve der Festzeitsteuerung. Insgesamt werden während des gesamten Tages mit dem kombinierten Verfahren 6127 Liter Kraftstoff verbraucht, mit der Festzeitsteuerung dagegen 7586 Liter. Das entspricht einem Mehrverbrauch von 23,8%.

Interpretation

Kürzere Wartezeiten reduzierten den Kraftstoffverbrauch. Allein aufgrund dieser Tatsache wird mit dem kombinierten Verfahren zu beinahe jeder Tageszeit weniger Kraftstoff verbraucht. Wegen der häufigeren Halte sind die Unterschiede zwischen diesen Kurven jedoch nicht mehr so ausgeprägt, wie zwischen den Kurven der Warteschlangenlängen. Über den gesamten Tag können mit dem kombinierten Verfahren dennoch 19,2% des Kraftstoffverbrauchs eingespart werden.

Die Ursache für den vorübergehend höheren Verbrauch liegt, wie bereits im Abschnitt 7.2.2 erläutert, in den Eingriffen des stabilisierenden Regimes. Diese

⁷¹Vgl. Abschnitt 6.2.1

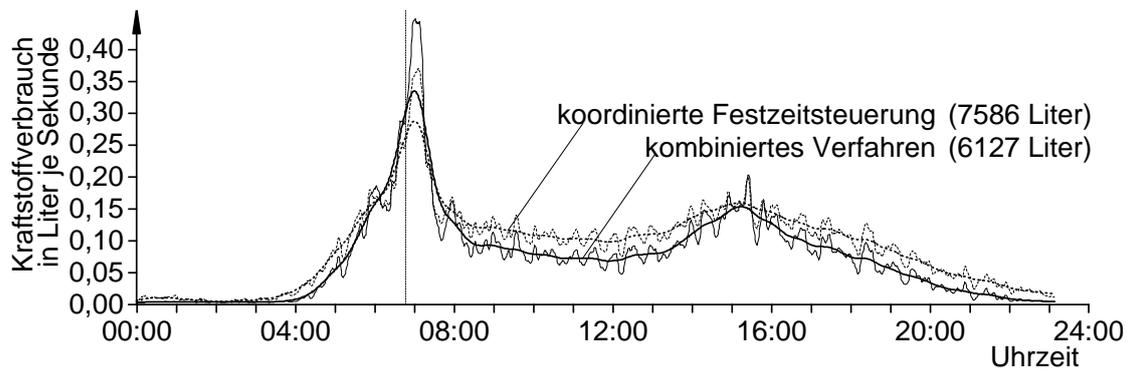


Abbildung 7.13.: Tagesverlauf des momentanen Kraftstoffverbrauchs in Liter je Sekunde. Das kombinierte Verfahren kann mit den kürzeren Wartezeiten gleichzeitig auch den Kraftstoffverbrauch reduzieren. Zwar ist der Verbrauch mit der koordinierten Festzeitsteuerung kurzzeitig geringer, aber dennoch können mit dem kombinierten Verfahren 19,2% des Gesamtverbrauchs eingespart werden.

erfolgen vorwiegend bei hohen Verkehrslasten und stellen das Einhalten der mittleren Bedienperiode T sicher. Zum einen können dadurch längere Warteschlangen und damit auch längere Wartezeiten entstehen, und zum anderen müssen nachrückende Wartekolonnen ein zusätzliches Mal anfahren. Es ist zu vermuten, dass der Gesamtkraftstoffverbrauch mit einer Moderierung dieser Eingriffe noch weiter zu senken ist. Wie dies geschehen kann, wird im Ausblick 8.1.4 beschrieben.

Kapitel 8.

Ausblick und Zusammenfassung

8.1. Ausblick

Die Grundfunktionalität des vorgestellten Verfahrens beruht auf einer vergleichsweisen einfachen Reglerstruktur⁷². Deshalb bieten sich an vielen Stellen Ansatzpunkte für mögliche Erweiterungen. Einige davon sollen im Folgenden aufgezeigt werden.

8.1.1. Heterogene Verkehrsströme

Individuelle Wartezeitengewichtung

In der Konzeption des Verfahrens wurde davon ausgegangen, dass jedes Fahrzeug die gleiche Zeit zum Überfahren der Haltelinie benötigt, und dass jedes wartende Fahrzeug den gleichen Beitrag zur Kostenfunktion „Gesamtwartezeit“ liefert.

Um einzelne Fahrzeuge individuell zu gewichten, ist die Zeitreihe der erwarteten Ankünfte⁷³ zunächst auf Pkw-Äquivalente zu beziehen. Beispielsweise entspricht dann ein Lkw, der die doppelte Zeit zum Überfahren der Haltelinie benötigt, zwei Pkw-Äquivalenten und ein Kraftrad einem halben [166, 9, 4, 149]. Damit ist die Annahme A1, dass die Ausflussrate aufgestauter Pkw-Äquivalente konstant ist, weiterhin erfüllt. Des Weiteren ist eine zweite Zeitreihe einzuführen, in der die individuellen „Wartezeitengewichte“ aufsummiert werden. Anhand dieser lässt sich für eine Wartekolonne unmittelbar angeben, mit welcher Rate die nun allgemeine Kostenfunktion in der Zeit anwächst. Auch mit individuellen Wartezeitengewichten sind die Voraussetzungen⁷⁴ für die Existenz einer optimalen Prioritätsregel erfüllt. Deshalb ist zu erwarten, dass die Formel (4.35) zur Berechnung des dynamischen Prioritätsindex ihre prinzipielle Struktur beibehält.⁷⁵ Die Verallgemeinerung des Verfahrens um individuelle Wartezeitengewichte wurde von Carsten Knoll im Rahmen einer Studienarbeit durchgeführt [96].

Fahrplanabhängige Priorisierung des ÖPNV

Bezüglich der Lichtsignalsteuerung unterscheidet sich der öffentliche Personennahverkehr (ÖPNV) in zweierlei Hinsicht von den anderen Verkehrsströmen: Zum Ersten kommen die Busse und Straßenbahnen einzeln und vergleichsweise selten an, und zum Zweiten sind diese an einen Fahrplan gebunden.

⁷²Vgl. Abb. 6.1 (auf S. 112)

⁷³Vgl. Abschnitt 3.1.1

⁷⁴Vgl. Abschnitt 4.1.1

⁷⁵Die benötigte Freigabezeit \hat{g}_i berechnet sich weiterhin mit Gleichung (3.20). Allerdings ergibt sich die Zuwachsrates \hat{n}_i der zu minimierenden Kostenfunktion \hat{W}_i in Gleichung (3.48) nun aus der Zeitreihe der Wartezeitengewichte. Damit sind \hat{g}_i und \hat{n}_i bekannt und die Schritte zur Herleitung der Priorisierungsregel direkt übertragbar.

Der ÖPNV wird in Deutschland mittels sogenannter „*Beschleunigungsprogramme*“ [181] bevorzugt bedient. Obwohl die Ziele unstrittig sind, erzeugen die – teils bedingungslosen – Vorrangschaltungen auf der anderen Seite Frust. Deshalb fordert der ADAC Dresden: „*Das müsste viel ausgewogener gehandhabt werden*“ [28]. Einen Ansatz zur fahrplanabhängigen Priorisierung des ÖPNV ohne systematische Benachteiligung der anderen Verkehrsströme bietet das in dieser Dissertationschrift vorgestellte Verfahren. Hierzu können die Priorisierungsregeln des Stabilisierungsverfahrens aus Kapitel 5 verwendet werden. Die Aufnahme der Fahrzeuge des ÖPNV in die Prioritätsliste Ω ist dann beispielsweise an die Bedingung zu knüpfen, dass nur so der Fahrplan eingehalten werden kann. Das garantiert weiterhin jedem Verkehrsstrom die minimal benötigte Freigabezeit innerhalb der gewünschten Bedienperiode.

Bedingungslose Vorrangschaltungen für Einsatzfahrzeuge

Die Einsatzfahrzeuge von Rettungsdiensten, Feuerwehr, Katastrophenschutz und Polizei sollen unter allen Umständen freie Fahrt erhalten. Vorausgesetzt, diese Fahrzeuge lassen sich als solche erkennen, kann hierfür auf das Prognoseverfahren aus Kapitel 3 zurückgegriffen werden. Mit diesem lässt sich die benötigte Zeit bis zur Leerung einer Warteschlange und damit bis zum Herstellen freien Verkehrs an der Haltelinie abschätzen. Der Wirkungsbereich⁷⁶ ist in Abhängigkeit der Geschwindigkeit der Einsatzfahrzeuge so anzupassen, dass diese bei ihrer Ankunft an der Haltelinie freie Fahrt haben.

Fußgängerströme

Fußgänger lassen sich nur schwer detektieren. Auch wenn der Anforderungstaster gedrückt wurde ist der Steuerung nicht bekannt, wie viele Fußgänger angekommen sind. Zwar gibt es technische Lösungen zur Detektion von Fußgängerströmen auf Video- und Laserbasis [190, 178, 78], aber diese werden in der Praxis bislang noch nicht eingesetzt. Deshalb bietet es sich an, die Fußgänger als einen virtuellen Verkehrsstrom zu modellieren. In gleicher Weise, wie dieser Ansatz im Abschnitt 5.4.1 für das Stabilisierungsverfahren verwendet wurde, kann er auch auf das Optimierungsverfahren übertragen werden. Damit fügt sich die Bedienung der Fußgänger flexibel in die dynamische Priorisierung der anderen Verkehrsströme ein.

8.1.2. Stauraumbegrenzung

Das unkontrollierte Anwachsen von Staus kann in Straßennetzwerken zu drastischen Kapazitätseinbrüchen führen [43], die sich unter Umständen kaskadenartig

⁷⁶Vgl. Abb. 3.6 (auf S. 57)

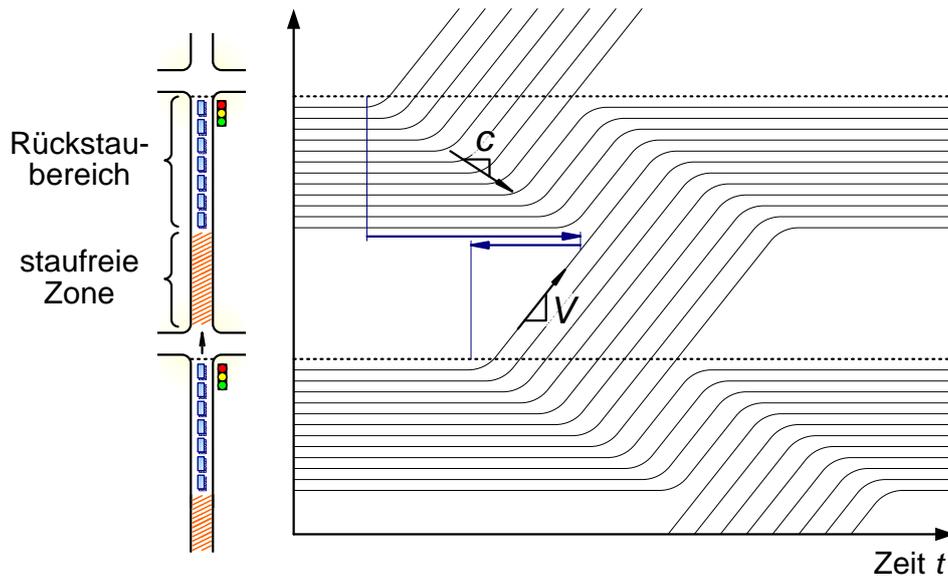


Abbildung 8.1.: Fahrzeugtrajektorien auf zwei übersättigten Richtungsfahrbahnen mit begrenztem Stauraum. Olsson [136] schlägt vor, nicht mehr Fahrzeuge in einen Streckenabschnitt einfahren zu lassen, als dort auch wieder herausfahren können. Damit lässt sich vermeiden, dass vorgelagerte Knotenpunkte infolge eines Rückstaus blockiert werden. Die Pfeile deuten an, wie sich der Zufluss in den Streckenabschnitt aus dem bekannten Ausfluss mittels Zeitverschiebung berechnen lässt [77, 74].

über große Teile des Netzwerks ausbreiten [199]. Zur Vermeidung solcher Effekte schlägt Olsson [136] folgenden dezentralen Ansatz vor:

*„The fundamental principle for traffic control is:
don't let more traffic in than what can pass out.“*

Um diesen Ansatz in das hier vorgestellte Verfahren zu integrieren, benötigt man zunächst ein Prognosemodell für die zeitliche Entwicklung der Rückstaulänge. Hierfür eignet sich das „*section-based queueing-theoretical traffic model*“ [77], welches von Helbing, dem Autor dieser Dissertationschrift und Lebacque in Referenz [74] zur Modellierung von Straßennetzwerken verallgemeinert wurde. Das Modell beruht auf der Beobachtung, dass im gebundenen Verkehr jedes Fahrzeug zeitversetzt die Geschwindigkeitsänderungen seines Vorderfahrzeugs adaptiert [13] und sich infolgedessen Stauwellen mit einer typischen Geschwindigkeit $c \approx 20$ km/h entgegen der Fahrtrichtung ausbreiten [108, 153, 193, 75, 76]. Diese Erscheinung kann man übrigens auch beim Anfahren einer Wartekolonne nach dem Umschalten auf Grün beobachten,⁷⁷ wie Abb. 8.1 zeigt. Die Wellengeschwindigkeit c im gebun-

⁷⁷Empirische Untersuchungen von Long [113] ergaben einen mittleren Zeitversatz von 1,15 Sekunden je Fahrzeug.

denen Verkehr und die mittlere Reisegeschwindigkeit V im freien Verkehr sind die beiden charakteristischen Geschwindigkeiten, mit denen sich das Verkehrsgeschehen im Inneren eines homogenen Streckenabschnitts auf die Zeitreihen seiner Zu- und Ausflüsse zurückführen lässt.

Ausgangspunkt ist eine Richtungsfahrbahn, wie sie in Abb. 8.1 links dargestellt ist. Sie habe einen Rückstaubereich und eine staufreie Zone. Es sei gefordert, dass sich der Rückstau nicht in die staufreie Zone hinein ausbreitet. Dann dürfen am Fahrbahnquerschnitt zwischen den beiden Bereichen nicht mehr Fahrzeuge aus der staufreien Zone ankommen, als in den Rückstaubereich nachrücken können. Weiterhin sei angenommen, dass die Steuerung des unmittelbar davor befindlichen Knotenpunktes Zugriff auf die Zeitreihe der abgefertigten Fahrzeuge hat. Dann lässt sich mit Hilfe der charakteristischen Geschwindigkeiten c und V durch Zeitverschiebung abschätzen, wie viele Fahrzeuge bis zu einem bestimmten Zeitpunkt höchstens in diese Richtungsfahrbahn einfahren dürfen. Die Pfeile in Abb. 8.1 stellen dies schematisch dar. Letztlich kann man mit den bekannten Abbiegeanteilen darauf schließen, wie viele Fahrzeuge auf jeder Zufahrt des vorgelagerten Knotenpunktes bis zu einem bestimmten Zeitpunkt höchstens abgefertigt werden dürfen.⁷⁸

Folglich wird eine Zufahrt, auf der zwar viele Fahrzeuge warten, diese jedoch aufgrund des stromabwärtigen Rückstaus ohnehin nicht abfließen können, auch nicht zur Bedienung ausgewählt. Die dadurch freiwerdenden Abfertigungskapazitäten stellt das kombinierte Verfahren stattdessen den Verkehrsströmen auf den Umgehungsrouen zur Verfügung. Olsson [136] vermutet, dass dieser Ansatz das Routenwahlverhalten der Verkehrsteilnehmer beeinflusst und auf lange Sicht zu einer gleichmäßigen Auslastung der Streckenkapazitäten führt.

8.1.3. Mehrphasenbetrieb

In Anlehnung an Websters Ausspruch, „*each phase can be represented by one approach only*“⁷⁹, könnte man eine beliebige zulässige Phaseneinteilung⁸⁰ wählen und folgenden Ansatz vorschlagen: Das vorgestellte Verfahren wählt nur die kritischen Verkehrsströme⁸¹ zur Bedienung aus, während die jeweils dazu verträglichen Verkehrsströme automatisch mit bedient werden. Das allein würde ausreichen, um jeder Knotenzufahrt die benötigte Abfertigungskapazität zuzusichern, netzwerk-

⁷⁸Diese Zeitreihe der maximal abzufertigenden Fahrzeuganzahlen lässt sich bei der Prognose der Freigabezeiten \hat{g}_i durch einen Minimumoperator in Gleichung (3.20) berücksichtigen. Die so berechneten Prognosewerte für \hat{g}_i und \hat{n}_i erfüllen dann die Forderung, auf den nachfolgenden Richtungsfahrbahnen maximale Rückstaulängen einzuhalten.

⁷⁹Vgl. Referenz [192]

⁸⁰Es muss die hinreichende Stabilitätsbedingung (2.10) erfüllt sein.

⁸¹Vgl. Definition auf Seite 28

weite Stabilität zu garantieren und gleichzeitig die gewünschte Bedienperiode und die geforderte Maximalperiode einzuhalten.

Um jedoch alle Verkehrsströme in die Auswahl einzubeziehen, sind die vier Lösungsschritte⁸² dieser Dissertationsschrift erneut anzuwenden und dabei um den kombinatorischen Aspekt der gleichzeitigen Bedienung mehrerer Zufahrten zu verallgemeinern. Weil es hierfür nicht nötig ist, von einer gegebenen Phaseneinteilung auszugehen, sondern alle zulässigen Schaltkombinationen einbezogen werden können, ist sogar zu erwarten, dass sich das vorgestellte Steuerungsverfahren noch besser an das tatsächliche Verkehrsgeschehen anpasst. Denn um beispielsweise einer herannahenden Kolonne freie Fahrt zu geben, brauchen nicht die Bedienvorgänge der zu ihr verträglichen Verkehrsströme unterbrochen zu werden. Auch wenn ein Verkehrsstrom besonders lange Freigabezeiten benötigt, lassen sich währenddessen die zu ihm verträglichen Ströme flexibel abfertigen.

8.1.4. Weitere Vorschläge

Eingriffe des stabilisierenden Regimes

Die Grundidee⁸³ des vorgestellten Steuerungsverfahrens ist es, bei jedem Anzeichen unerwünschten Schaltverhaltens sofort ins stabilisierende Regime zu wechseln. Der abrupte Wechsel hat sich jedoch im Hinblick auf den Kraftstoffverbrauch als ungünstig erwiesen.⁸⁴ Die Eingriffe des stabilisierenden Regimes lassen sich aber ohne Weiteres prognostizieren.⁸⁵ Mit folgendem Ansatz kann der Übergang „sanfter“ gestaltet werden: Sobald prognostiziert wird, dass eine Zufahrt i in beispielsweise spätestens zehn Sekunden vom stabilisierenden Regime zur bevorzugten Bedienung ausgewählt wird, darf das optimierende Regime zu keiner anderen Zufahrt als i umschalten. Ob bis dahin der aktuelle Bedienvorgang fortgesetzt oder bereits früher die betrachtete Zufahrt i bedient wird, hängt aufgrund der dynamischen Priorisierungsregel (4.34) davon ab, wo die meisten Wartezeiten eingespart werden können. Es ist zu erwarten, dass diese Modifikation des kombinierten Verfahrens zu einer weiteren Senkung des Kraftstoffverbrauchs führt.

Variable Kopplungsstärke

Mit dem Ansatz, einen Teil der tatsächlich detektierten Fahrzeuge durch virtuelle zu ersetzen, konnte in Abschnitt 5.4.1 mit einem Parameter α die Variabilität des Stabilisierungsverfahrens eingestellt werden. Der virtuelle Fahrzeugstrom wurde

⁸²Für einen Überblick siehe Abschnitt 1.2.2.

⁸³Vgl. Abschnitt 6.1.1

⁸⁴Vgl. Abschnitte 7.2.2 und 7.4.3

⁸⁵Das stabilisierende Regime wählt Zufahrt i in spätestens Δt Zeiteinheiten zur Aufnahme in die FIFO-Liste Ω aus, wenn der Prognosewert $\hat{n}_i(t, \tau_i + \Delta t)$ aus den Gleichungen (3.19) und (3.20) größer ist als der kritische Schwellenwert $n_i^{\text{crit}}(t + \Delta t)$ aus Gleichung (5.30).

dabei mit konstanter Ankunftsrate modelliert. Es ist aber auch denkbar, die Ankunftsrate der virtuellen Fahrzeuge einer beliebigen – beispielsweise auch periodischen – Funktion folgen zu lassen. Das würde die Wahrscheinlichkeit bestimmter Zeitpunkte zum Umschalten gezielt erhöhen. Ebenso ist es denkbar, einen Teil der virtuellen Fahrzeuge wieder aufzulösen, um die Wahrscheinlichkeit anderer Umschaltzeitpunkte gezielt zu senken.

Damit ist es möglich, die herkömmlichen Verfahren zur Netzkoordinierung in das hier vorgestellte Verfahren zu integrieren. Aus diesen Verfahren können die virtuellen Verkehrsströme so abgeleitet werden, dass sich das Schaltverhalten mit einem einzigen Parameter α stufenlos zwischen flexibel und koordinierend variieren lässt. Der Parameter α erhält mit diesem Ansatz die Rolle einer variablen „Kopplungsstärke“ zwischen den Knotenpunkten.

8.2. Zusammenfassung

Die vorliegende Dissertationsschrift widmete sich der Entwicklung eines neuartigen Verfahrens zur verkehrsabhängigen Steuerung von Lichtsignalanlagen in Straßennetzwerken.

Ausgangspunkt war eine Analyse der herkömmlichen Steuerungsverfahren. Es wurde deutlich, dass das größte Verbesserungspotential in dem Ansatz liegt, die Schaltprogramme flexibler an das tatsächliche Verkehrsgeschehen anzupassen. Dieser Ansatz kann jedoch nur mit einem dezentralen Verfahren umgesetzt werden, das beschränkte Prognosehorizonte und netzwerkweite Rückkopplungseffekte in Kauf nimmt. Von diesen Umständen ausgehend wurden schrittweise folgende – zunächst unabhängigen – Verfahren entwickelt:

- ein lokales Prognoseverfahren,⁸⁶ das aus einer Zeitreihe der erwarteten Fahrzeugankünfte den Verkehrsablauf während eines nachfolgenden Bedienvorgangs prognostiziert,
- ein lokales Optimierungsverfahren,⁸⁷ das jeder Phase einen Prioritätsindex zuweist und die Phase mit größter Priorität zur Bedienung auswählt und
- ein lokales Stabilisierungsverfahren,⁸⁸ das einen Bedienvorgang nur dann durchführt, wenn dies mit einer minimal erforderlichen Effizienz möglich ist.

Erst durch Kombination⁸⁹ dieser drei Verfahren stellt sich das gewünschte Ergebnis ein: Jeder Verkehrsstrom bekommt die minimal benötigte Abfertigungskapazität

⁸⁶Vgl. Zusammenfassung in Abschnitt 3.4

⁸⁷Vgl. Zusammenfassung in Abschnitt 4.3.1

⁸⁸Vgl. Zusammenfassung in Abschnitt 5.5.2

⁸⁹Vgl. Zusammenfassung in Abschnitt 6.2.4

garantiert, während die verbleibende Kapazität zur lokalen Wartezeitenminimierung genutzt wird. Mit anderen Worten heißt das: In die lokale Wartezeitenminimierung wird nur dann korrigierend eingegriffen, wenn diese nicht von selbst zum gewünschten Schaltverhalten führt.

Aus regelungstechnischer Sicht besonders herauszustellen ist das Ergebnis, dass trotz der dezentralen Architektur, trotz dem die Steuerung verkehrabhängig ist, und trotz dem die Prognosehorizonte beschränkt sind, instabiles Verhalten ausgeschlossen werden kann. Dass dies nicht selbstverständlich ist, wurde anhand der Schaltregel „Längste Schlange zuerst“ erläutert.⁹⁰ Die hinreichende Stabilitätsbedingung ist die gleiche wie für ein umlaufbasiertes Verfahren.⁹¹ Die Garantie netzwerkweiter Stabilität ist überhaupt erst die Voraussetzung dafür, Lichtsignalanlagen mit dezentralen Verfahren betreiben zu können.

Das vorgestellte Verfahren setzt keine Kommunikation mit benachbarten Knotenpunkten oder mit einer Verkehrszentrale voraus. Vorausgesetzt wird dagegen, dass sich das Verkehrsgeschehen mit Detektoren erfassen lässt, wie in der Skizze aus Abb. 1.2 (auf S. 7) dargestellt ist. Hierfür ist es ausreichend, einen Detektor unmittelbar vor den Haltelinien und einen anderen hinreichend weit stromaufwärts davon zu platzieren. Der Abstand zwischen den beiden Detektoren sollte etwa so groß sein wie die Reisezeit bei freiem Verkehr während der Dauer eines Bedienvorgangs.⁹² Die Möglichkeit, dass die Prognose der Fahrzeugankünfte unter Umständen fehlerbehaftet sein kann, wurde beim Reglerentwurf explizit berücksichtigt. Aus den Detektordaten leitet die lokale Recheneinheit die Schaltsignale für die Signalgeber ab. Den prinzipiellen Aufbau der lokalen Recheneinheit zeigt Abb. 8.2. Zuerst sind aus den Messdaten der Detektoren die Zeitreihen der erwarteten Ankünfte⁹³ zu konstruieren. Auf Grundlage dieser Zeitreihen können die relevanten Zustandsgrößen, wie zum Beispiel die zur Warteschlangenleerung benötigte Freigabezeit⁹⁴ \hat{g}_i oder die Zuwachsrate⁹⁵ \hat{n}_i der zu minimierenden Gesamtwarzeiten berechnet werden. Auf diese lassen sich die Schaltregeln des kombinierten Verfahrens⁹⁶ anwenden. Das Ergebnis dieser Schaltregeln ist der Index σ der zu bedienenden Zufahrt. Wie mehrere Zufahrten gleichzeitig bedient werden können, erläutert Abschnitt 8.1.3. Mit jeder neuen Berechnung der Zustandsgrößen und jeder neuen Anwendung der Schaltregeln wird die Auswahl der zu bedienenden Verkehrsströme reoptimiert. Deshalb sind diese Berechnungen so häufig wie möglich, beispielsweise sekundlich, durchzuführen.

⁹⁰Vgl. Anhang A.1

⁹¹Vgl. Abschnitt 5.2.2

⁹²Zahlenbeispiel: Bei einer Reisegeschwindigkeit von 55 km/h und einer Bediendauer von 20 Sekunden sollte der Abstand etwa 300 Meter betragen.

⁹³Vgl. Abschnitt 3.1.1 und Abb. 3.1 (auf S. 41)

⁹⁴Vgl. Abschnitt 3.3.1, Gleichung (3.20) und Abb. 3.4 (auf S. 52)

⁹⁵Vgl. Abschnitt 3.3.3 und Gleichung (3.48)

⁹⁶Vgl. Abschnitt 6.1.2, Gleichung (6.3) und Abb. 6.1 (auf S. 112)

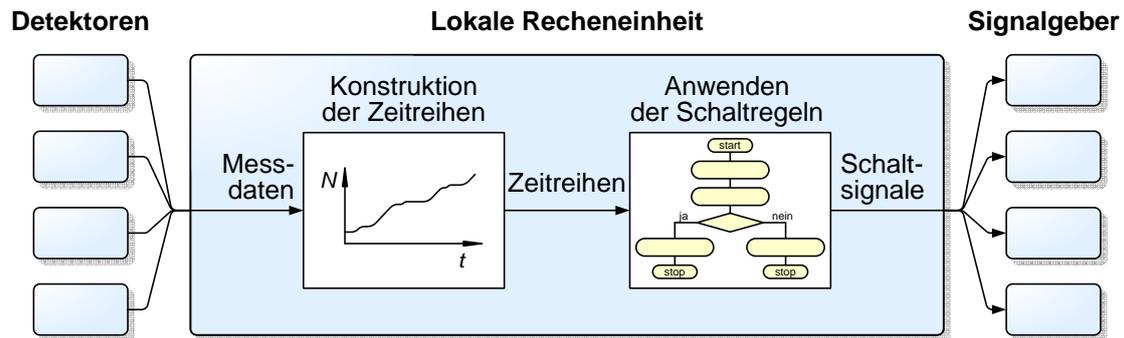


Abbildung 8.2.: Die lokale Recheneinheit (vgl. Abb. 1.2 (auf S. 7)) besteht aus zwei Funktionseinheiten. Die erste konstruiert aus den Messdaten der Detektoren die Zeitreihen der erwarteten Ankünfte (vgl. Abb. 3.1 (auf S. 41)), die zweite leitet aus diesen mit den Schaltregeln des kombinierten Verfahrens (vgl. Abb. 6.1 (auf S. 112)) die Schaltsignale für die Signalgeber ab.

Die Umschaltzeitpunkte ergeben sich damit unmittelbar aus den Ankunftszeitpunkten der Fahrzeuge. Zwar sind die resultierenden Schaltmuster scheinbar irregulär,⁹⁷ aber gerade dies gleicht die irregulären Fahrzeugankünfte aus, was in der Konsequenz zu einer beachtlich geringen Streuung⁹⁸ der Wartezeiten führt. Die Zeitspanne zwischen zwei Bedienvorgängen für die selbe Zufahrt kann als Zufallsgröße interpretiert werden. Ihr Mittel- und ihr Maximalwert lassen sich durch jeweils einen Parameter vorgeben. Diese beiden Parameter, die Wunschperiode⁹⁹ T und die Maximalperiode¹⁰⁰ T^{\max} , sind die beiden einzigen Parameter des vorgestellten Verfahrens. Auch darin zeigt sich seine besonders hohe Anpassungsfähigkeit.

Da sich das vorgestellte Verfahren so flexibel an das tatsächliche Verkehrsgeschehen anpasst, können an einzelnen Knotenpunkten wie auch in komplexen Straßennetzwerken Wartezeiten und Kraftstoffverbrauch deutlich gesenkt werden. Und das obwohl – oder gerade weil – die Knotenpunkte über nichts anderes als die Verkehrsströme selbst gekoppelt sind.

⁹⁷Vgl. Abb. 7.3 (auf S. 125) und Abb. 7.8 (auf S. 133)

⁹⁸Vgl. Abb. 7.5 (auf S. 128)

⁹⁹Vgl. Abschnitt 5.2 und Gleichung (5.9)

¹⁰⁰Vgl. Abschnitt 5.4 und Gleichung (5.24)

Anhang A.

Herleitungen

A.1. Instabilität der Clear-Largest-Buffer-Policy

A.1.1. Kumar-Seidman-Netzwerk

Clearing-Policies können in nicht-maschenfreien Netzwerken zu instabilem Verhalten führen.¹⁰¹ Das folgende Beispiel stammt aus der Arbeit von Kumar und Seidman [98] zur Steuerung von Produktionsnetzwerken. Es wird eine spezielle Clearing-Policy betrachtet, nämlich die Clear-Largest-Buffer-Policy [142, 29, 150, 104, 86]. Sie entspricht der lokalen Schaltregel „Wähle die Zufahrt mit der längsten Warteschlange aus und bediene sie solange dies mit maximaler Rate möglich ist.“

Szenario A.i

Das Beispiel von Kumar und Seidman aus Abb. A.1(a) wird auf ein einfaches Straßennetzwerk mit zwei benachbarten Knotenpunkten und zwei symmetrisch verschränkten Verkehrsströmen A und B übertragen. Dieses zeigt Abb. A.1(b). Um die Betrachtung einfach zu halten und andere Ursachen für Instabilitäten auszuschließen, werden sowohl die Zwischenzeiten vernachlässigt ($\tau_i^0 = 0$), als auch Reisezeiten im freien Verkehr. Für die maximalen Ausflussraten soll $q_1^{\max} = q_3^{\max} = 2q^{\max}$ und $q_2^{\max} = q_4^{\max} = q^{\max}$ gelten. Die Zuflussraten q_A und q_B werden als konstant angenommen. Die Kapazität der Knotenzufahrten reiche aus, beide Verkehrsströme zu bedienen. Das heißt, die Stabilitätsvoraussetzung (2.3) sei sowohl am linken Knotenpunkt erfüllt,

$$\lambda_1 + \lambda_4 = \frac{q_A}{2q^{\max}} + \frac{q_B}{q^{\max}} < 1, \quad (\text{A.1})$$

als auch am rechten Knotenpunkt,

$$\lambda_2 + \lambda_3 = \frac{q_A}{q^{\max}} + \frac{q_B}{2q^{\max}} < 1. \quad (\text{A.2})$$

Anfangszustand

Das soeben beschriebene System kann – abhängig vom Anfangszustand – sowohl stabiles als auch instabiles Verhalten zeigen [98]. Wir betrachten den instabilen Fall. Er tritt beispielsweise auf, wenn für die Warteschlangenlängen $\hat{n}_i(t) = (n_1(t), \dots, n_4(t))^T$ zum Zeitpunkt t_0 folgendes gilt:

$$\vec{n}(t_0) = (n_1(t_0), 0, 0, 0)^T. \quad (\text{A.3})$$

Dann wird zunächst an beiden Knotenpunkten Verkehrsstrom A bedient, das heißt Zufahrt 1 und Zufahrt 2 sind im Zustand Freigabe.

¹⁰¹Vgl. Abschnitt 2.2.1

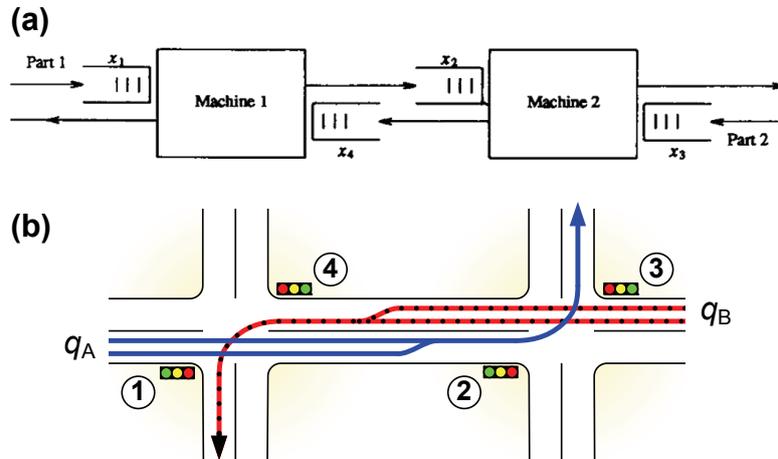


Abbildung A.1.: (a) Nicht-maschenfreies Produktionsnetzwerk mit zwei Bedienstationen wie es von Kumar und Seidman in Referenz [98] untersucht wurde. (b) Entsprechendes Straßennetzwerk mit zwei verschränkten Verkehrsströmen A und B.

A.1.2. Zeitliche Entwicklung der Warteschlangen

Die folgenden Rechenschritte können in Abb. A.2 nachvollzogen werden.

Aus der Warteschlange von Zufahrt 1 fahren Fahrzeuge mit der maximalen Rate $2q^{\max}$ heraus, während neue Fahrzeuge mit der Rate q_A ankommen. Die heraus gefahrenen Fahrzeuge kommen mit der gleichen Rate $2q^{\max}$ auf Zufahrt 2 an, wo sie aber lediglich mit q^{\max} abbiegen können. Im Zeitraum $[t_0, t_0 + \Delta t]$ wird die Warteschlange 1 geleert, während die Warteschlangen 2 und 3 anwachsen:

$$\vec{n}(t_0 + \Delta t) = \begin{pmatrix} n_1(t_0) \\ 0 \\ 0 \\ 0 \end{pmatrix} + \Delta t \begin{pmatrix} q_A \\ 2q^{\max} \\ q_B \\ 0 \end{pmatrix} - \Delta t \begin{pmatrix} 2q^{\max} \\ q^{\max} \\ 0 \\ 0 \end{pmatrix}. \quad (\text{A.4})$$

Diese Entwicklungsgleichung gilt genau so lange, bis Zufahrt 1 vollständig geleert ist. Dies ist zum Zeitpunkt $t_1 = t_0 + n_i(t_0)/(2q^{\max} - q_A)$ der Fall. An der linken Kreuzung sind beide Warteschlangen $n_1(t_1) = n_4(t_1) = 0$ vollständig geleert, aber auf Zufahrt 1 können die ankommenden Fahrzeuge weiter bedient werden. Auf Zufahrt 4 sind noch keine Fahrzeuge angekommen, weil an der rechten Kreuzung noch immer Zufahrt 2 bedient wird. Die linke Kreuzung ist gezwungen, Zufahrt 1 mit einer geringeren als der maximalen Rate zu bedienen – und damit, indem sie fortwährend Fahrzeuge auf Zufahrt 2 weiterleitet, das Umschalten der rechten Kreuzung weiter hinauszuzögern. Die Warteschlangen entwickeln sich im Zeitraum

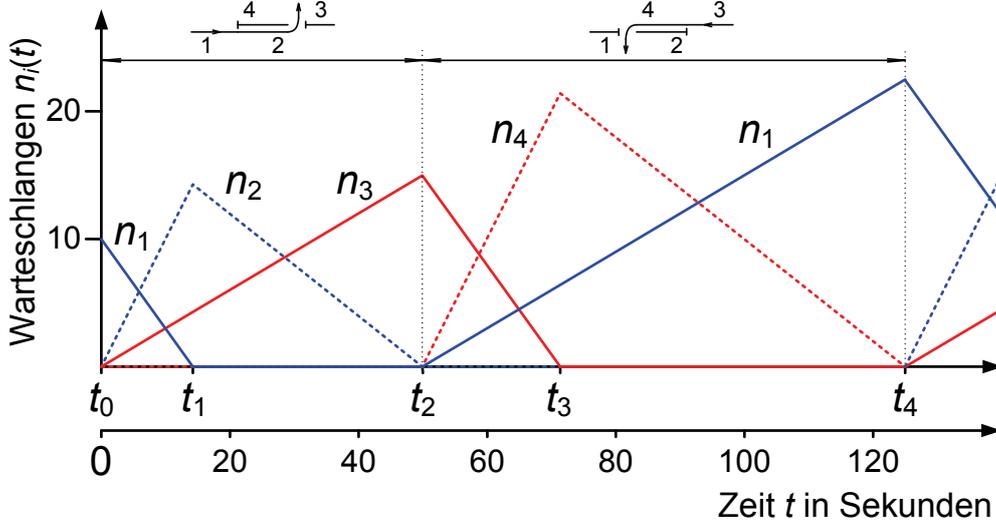


Abbildung A.2.: Zeitliche Entwicklung der Warteschlangenlängen für das Beispiel aus Abb. A.1. Der Anfangszustand $\vec{n}(t_0)$ stellt sich zur Zeit t_4 wieder ein – jedoch mit $G = 2,25$ mal so langen Warteschlangen. Das Szenario nimmt konstante Zuflussraten $q_A = q_B = 0,6$ Fz/s an, maximale Ausflussrate von $q_1^{\max} = q_3^{\max} = 2q^{\max}$ und $q_2^{\max} = q_4^{\max} = q^{\max}$ mit $q^{\max} = 2$ Fz/s, sowie den Anfangszustand $\vec{n}(t_0) = (10, 0, 0, 0)^T$ Fz.

$[t_1, t_1 + \Delta t]$ gemäß:

$$\vec{n}(t_1 + \Delta t) = n_1(t_0) \begin{pmatrix} 0 \\ q^{\max}/(2q^{\max} - q_A) \\ q_B/(2q^{\max} - q_A) \\ 0 \end{pmatrix} + \Delta t \begin{pmatrix} q_A \\ q_A \\ q_B \\ 0 \end{pmatrix} - \Delta t \begin{pmatrix} q_A \\ q^{\max} \\ 0 \\ 0 \end{pmatrix}. \quad (\text{A.5})$$

Zum Zeitpunkt $t_2 = t_0 + n_1(t_0)/(q^{\max} - q_A)$ ist Zufahrt 2 vollständig geleert. Nun wird entsprechend der Clear-Largest-Buffer-Policy an der rechten Kreuzung zur Zufahrt 3 umgeschaltet. Unter Vernachlässigung der Zeitverzögerungen bildet sich zum gleichen Zeitpunkt auf Zufahrt 4 eine Warteschlange, so dass auch die linke Kreuzung zur Bedienung von Zufahrt 4 umschaltet. Nun wird Verkehrsfluss B bedient. Zum Umschaltzeitpunkt t_2 haben die Warteschlangen die Länge:

$$\vec{n}(t_2) = (0, \quad 0, \quad n_1(t_0) q_B/(q^{\max} - q_A), \quad 0)^T. \quad (\text{A.6})$$

Da dies symmetrisch zur Anfangsbedingung (A.3) ist, können die Entwicklungsgleichungen (A.4) und (A.5) mit paarweise vertauschten Indizes ebenso für die Berechnung der Systemzustände zu den Zeitpunkten t_3 und t_4 herangezogen werden. Zeitpunkt t_4 spiegelt den Anfangszustand wider, jedoch mit einem Vielfachen

G der anfänglichen Warteschlangenlänge:

$$\vec{n}(t_4) = G \vec{n}(t_0) \quad \text{mit} \quad G = \frac{q_A}{q^{\max} - q_A} \frac{q_B}{q^{\max} - q_B}. \quad (\text{A.7})$$

Mit $G < 1$ nimmt die Warteschlange von Periode zu Periode exponentiell ab (stabil), anderenfalls wächst sie exponentiell an, was dem instabilen Fall entspricht.

A.1.3. Diskussion

Die Stabilitätsbedingung $G < 1$ lässt sich durch Anwenden der binomischen Formel auf Gleichung (A.7) in folgenden Ausdruck umformen:

$$q_A + q_B < q^{\max}. \quad (\text{A.8})$$

Obwohl die Voraussetzungen (A.1) und (A.2) erfüllt sind, beispielsweise mit $q_A = q_B = 0.6 q^{\max}$, kann die Stabilitätsbedingung (A.8) verletzt sein. Die Ursache liegt darin, dass in den Zeiträumen $[t_1, t_2]$ und $[t_3, t_4]$ Abfertigungskapazitäten systematisch verschenkt werden. Der linke bzw. rechte Knotenpunkt bedient mit einer geringeren Rate und zögert damit sogar das Umschalten in den effizienteren Bedienvorgang heraus. Diesen Zustand bezeichnen Kumar und Seidman mit „*being starved of input*“. Das heißt im übertragenen Sinne, der eine Knotenpunkt lässt den anderen „verhungern“.

Wie dieses Beispiel zeigen soll, sind die Schaltprogramme beider Knotenpunkte über die Verkehrsströme gekoppelt und damit nicht mehr unabhängig voneinander. Obwohl kein einziges Fahrzeug im Kreis fährt, entstehen aufgrund der vernetzten Verkehrsströme Rückkopplungsschleifen. Beispielsweise konnte Reinschke in Referenz [151] zeigen, dass in linearen Netzwerken allein diese Rückkopplungsschleifen, bzw. auch „Zyklusfamilien“, für die resultierende Dynamik relevant sind. Im betrachteten Szenario ist die Rückkopplung positiv, das heißt die Wirkung verstärkt ihre Ursache. Dies kann eine Ursache für instabiles Verhalten lokal optimierender Steuerungsverfahren sein. Eine andere Ursache wird in Abschnitt 4.3.3 diskutiert. Dort ist an einem isolierten Knotenpunkt der beschränkte Optimierungshorizont für die Entstehung instabiler Verhaltens verantwortlich.

A.2. Optimale Platzierung des Freigabezeitfensters bei periodischen Ankünften

Es wird ein umlaufbasiertes Schaltprogramm betrachtet. Für einen einzelnen Knotenpunkt, an dem die Zeitreihe $q_i^{\text{exp}}(t)$ der Fahrzeugankünfte an jeder Zufahrt i bekannt und periodisch in der gegebenen Umlaufzeit T^{cyc} ist, sollen die Frei-

gabezeitfenster so platziert werden, dass die Gesamtwartezeiten aller Fahrzeuge minimal sind.¹⁰²

A.2.1. Gesamtwartezeiten an einer Zufahrt

Das Optimierungskriterium wird aus der Betrachtung des Ankunfts- und Ausfahrzeitpunktes eines einzelnen Fahrzeugs α an einer Knotenzufahrt i abgeleitet. Der Ankunftszeitpunkt t_α^{exp} sei im Folgenden der Zeitpunkt, zu dem es die Haltelinie bei freiem Verkehr erreichen würde, d.h. zu dem es frühestens bedient werden kann.¹⁰³ Je später das Fahrzeug nach dem Umschalten auf Sperrung ankommt, desto später wird es im nachfolgenden Freigabezeitfenster bedient. Es wird angenommen, dass aus der Lage des Ankunftszeitpunktes t_α^{exp} innerhalb des Umlaufs mit Hilfe einer linearen Abbildung auf die Lage des Ausfahrzeitpunktes t_α^{out} innerhalb des Freigabezeitfensters geschlossen werden kann. Das heißt, falls das Fahrzeug zum Zeitpunkt $t_\alpha^{\text{exp}} = t_i^{\text{red}} + h_\alpha T^{\text{cyc}}$ mit $h_\alpha \in [0, 1]$ ankommt, so wird es in der nächsten Freigabezeit zum Zeitpunkt $t_\alpha^{\text{out}} = T^{\text{cyc}} + t_i^{\text{red}} + (1 - h_\alpha)g_i^0$ herausfahren. Die Hilfsvariable

$$h_\alpha = (t_\alpha^{\text{exp}} - t_i^{\text{red}})/T^{\text{cyc}} \quad (\text{A.9})$$

gibt dabei an, welcher Anteil der Umlaufzeit T^{cyc} zwischen dem letzten Umschalten auf Sperrung t_i^{red} und dem Ankunftszeitpunkt t_α^{exp} verstrichen ist. Die Wartezeit w_α dieses einzelnen Fahrzeuges α ist die Zeitspanne zwischen seiner Ankunft t_α^{exp} und seiner Ausfahrt t_α^{out} :

$$w_\alpha = t_\alpha^{\text{out}} - t_\alpha^{\text{exp}} = (T^{\text{cyc}} - g_i^0)(1 - h_\alpha) \quad (\text{A.10})$$

Unter dieser Annahme sollen nun die Gesamtwartezeiten W_i^{cyc} aller Fahrzeuge an Zufahrt i während einer Umlaufzeit T^{cyc} berechnet werden. Dazu wird, beim Umschaltzeitpunkt t_i^{red} beginnend, das Produkt aus den individuellen Wartezeiten w_α und der Ankunftsrate $q_i^{\text{exp}}(t^{\text{exp}})$ über einen gesamten Umlauf integriert. Für einen gegebenen Umschaltzeitpunkt t_i^{red} lassen sich die Gesamtwartezeiten W_i^{cyc} mit Hilfe der Vorschrift

$$W_i^{\text{cyc}}(t_i^{\text{red}}) = \int_{t_i^{\text{red}}}^{t_i^{\text{red}} + T^{\text{cyc}}} w_\alpha \cdot q_i^{\text{exp}}(t^{\text{exp}}) dt^{\text{exp}} \quad (\text{A.11})$$

$$= \int_{t_i^{\text{red}}}^{t_i^{\text{red}} + T^{\text{cyc}}} [T^{\text{cyc}} - g_i^0] [1 - (t^{\text{exp}} - t_i^{\text{red}})/T^{\text{cyc}}] \cdot q_i^{\text{exp}}(t^{\text{exp}}) dt^{\text{exp}} \quad (\text{A.12})$$

¹⁰²Vgl. Abb. 2.10 (auf S. 32)

¹⁰³Vgl. Abschnitt 3.1.1

berechnen. Mit dem Ansatz, den Ausfahrzeitpunkt eines Fahrzeugs mit einer linearen Funktion des Ankunftszeitpunkts zu modellieren, bleibt das tatsächliche Ausflussprofil, wie es genauer in Abschnitt 3.1.2 untersucht wird, unberücksichtigt. Unter den hier getroffenen Annahmen ergibt sich das Profil der Ausflüsse innerhalb des Freigabezeitfensters durch Stauchung des Profils der Ankünfte innerhalb der Umlaufzeit. Der Fehler bei der Abschätzung der Wartezeiten ist genauso groß wie die Differenzen zwischen den tatsächlichen und den angenommenen Ausfahrzeitpunkten.

A.2.2. Konstruktion der Wartezeitenkurve

Als freie Variable zur Minimierung der geschätzten Gesamtwartezeiten $W_i^{\text{cyc}}(t_i^{\text{red}})$ soll der Zeitpunkt t_i^{red} , zu dem das Freigabezeitfenster für die Zufahrt i endet, variiert werden. Der optimale Umschaltzeitpunkt liegt an der Stelle, an der die Kurve der geschätzten Gesamtwartezeiten $W_i^{\text{cyc}}(t_i^{\text{red}})$ ihr Minimum hat. Die Kurve lässt sich beispielsweise dadurch konstruieren, dass man die Berechnungsvorschrift (A.12) auf jeden möglichen Umschaltzeitpunkt $t_i^{\text{red}} \in [0, T^{\text{cyc}})$ anwendet. Dies wäre jedoch sehr aufwändig. Einen günstigeren Zugang verspricht der Ansatz, die Gleichung (A.12) nach dem Umschaltzeitpunkt t_i^{red} abzuleiten. Hierzu kann man die Beziehung

$$\frac{d}{da} \int_a^{a+b} y(a, x) dx = \int_a^{a+b} \frac{\partial}{\partial a} y(a, x) dx + y(a, a+b) - y(a, a) \quad (\text{A.13})$$

mit den Substitutionen $a = t_i^{\text{red}}$, $b = T^{\text{cyc}}$ und $x = t^{\text{exp}}$ ausnutzen. Als Ergebnis erhält man einen analytischen Ausdruck für den Betrag dW_i^{cyc} , um den die Gesamtwartezeiten zunehmen, wenn der Umschaltzeitpunkt t_i^{red} um eine infinitesimale Einheit dt_i^{red} verschoben wird:

$$\frac{dW_i^{\text{cyc}}}{dt_i^{\text{red}}} = [T^{\text{cyc}} - g_i^0] [\bar{q}_i^{\text{exp}} - q_i^{\text{exp}}(t_i^{\text{red}})]. \quad (\text{A.14})$$

Hierbei steht \bar{q}_i^{exp} für die mittlere Ankunftsrate auf Zufahrt i , d.h. es gilt:

$$\bar{q}_i^{\text{exp}} = \frac{1}{T^{\text{cyc}}} \int_{T^{\text{cyc}}} q_i^{\text{exp}}(t) dt. \quad (\text{A.15})$$

Mit dieser Umformung ist zweierlei gewonnen: Zum einen lässt sich beim Entwurf einer Schaltfolge unmittelbar abschätzen, wie sich kleine Verschiebungen der Freigabezeitfenster auf die Gesamtwartezeiten auswirken. Zum anderen lässt sich der qualitative Kurvenverlauf von $W_i^{\text{cyc}}(t_i^{\text{red}})$ auf die Integration des Ausdrucks $[\bar{q}_i^{\text{exp}} - q_i^{\text{exp}}(t_i^{\text{red}})]$ zurückführen. Weder die Vorfaktoren noch die Integrationskonstante haben einen Einfluss auf die Lage des Minimums. Für quantitative Untersu-

chungen lässt sich die fehlende Integrationskonstante durch einmaliges Anwenden der Berechnungsvorschrift (A.12) bestimmen.

Die Gesamtwartezeiten W^{cyc} sämtlicher Fahrzeuge am Knotenpunkt innerhalb eines Umlaufs erhält man durch Addition der Gesamtwartezeiten W_i^{cyc} jeder einzelnen Zufahrt i . Indem die Phasenabfolge als gegeben angenommen wird, ist die freie Variable der gemeinsame Nullpunkt t^0 . Nun ist die Frage, wie die gesamte Zeitscheibe aus Abb. 2.10 (auf S. 32) zu drehen ist. Durch Substitution von t_i^{red} durch $t_0 + t_p^{\text{red}}$ lässt sich die Kurve $W^{\text{cyc}}(t^0)$ für die Gesamtwartezeiten in Abhängigkeit des Nullpunkts t^0 wie folgt konstruieren:

$$\frac{dW^{\text{cyc}}}{dt^0} = \sum_p \sum_{i \in p} \frac{dW_i^{\text{cyc}}}{dt_i^{\text{red}}} \quad (\text{A.16})$$

$$= \sum_p \sum_{i \in p} [T^{\text{cyc}} - g_i^0] [\bar{q}_i^{\text{exp}} - q_i^{\text{exp}}(t^0 + t_p^{\text{red}})] . \quad (\text{A.17})$$

Die Integration von (A.17) kann kontinuierlich und gleichzeitig mit dem Durchlaufen der Zeitscheibe durchgeführt werden. Die wartezeitenoptimalen Umschaltzeitpunkte lassen sich dann in den unmittelbar nachfolgenden Umläufen anwenden. Die optimale Lage des Freigabezeitfensters für ein gegebenes Ankunftsprofil zeigt Abb. 2.11 (auf S. 33).

A.3. Konvergenz zum optimalen Grenzyklus

Es soll das Szenario 4.vi aus Abschnitt 4.3.2 genauer untersucht werden. Abb. A.3(a) zeigt den betrachteten Knotenpunkt mit konstanten Ankunftsraten $q_1^{\text{exp}}(t) = q_1$ und $q_2^{\text{exp}}(t) = q_2$. Der optimale Grenzyklus für diese Situation wurde in der Literatur ausführlich untersucht, beispielsweise in den Referenzen [34, 56, 83, 173, 174, 171, 172]. Der optimale Grenzyklus zeichnet sich durch eine minimale Bedienperiode aus, falls die Auslastungen der beiden Zufahrten nicht in einem extremen Verhältnis zueinander stehen. Das heißt, nur wenn die Auslastung einer Zufahrt sehr groß und die der anderen verschwindend klein ist, kann es optimal sein, die Freigabezeit der mehr ausgelasteten Zufahrt zu verlängern. Die Referenzen [34, 56] gehen auf diesen Sonderfall detailliert ein. Die folgende Betrachtung geht von annähernd symmetrischen Auslastungen aus. Es soll gezeigt werden, ob und wie das flexible Programm von beliebigen Anfangszuständen aus gegen den optimalen Grenzyklus läuft.

A.3.1. Zeitliche Entwicklung der Systemgrößen

Die folgenden Rechenschritte können in Abb. A.3(b) und Abb. 4.4 (auf S. 86) nachvollzogen werden.

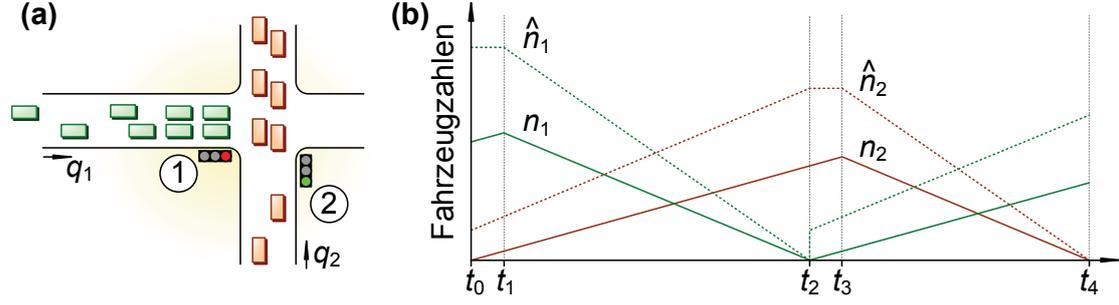


Abbildung A.3.: (a) Knotenpunkt mit zwei Zufahrten und konstanten Ankunftsrate q_1 und q_2 aus Szenario 4.vi. (b) Die Zustandsgrößen $\hat{n}_i(t)$ und $n_i(t)$ konvergieren mit dem flexiblen Programm gegen eine periodische stationäre Lösung, die im betrachteten Fall optimal ist. Den zugehörigen Verlauf der Prioritätsindizes $\pi_i(t)$ zeigt Abb. 4.4 (auf S. 86).

Anfangszustand

Zum Zeitpunkt t_0 werde Zufahrt 2 bedient ($\sigma(t_0) = 2$). Während ihre Warteschlange soeben vollständig geleert wurde ($n_2(t_0) = 0$), warten auf der ersten Zufahrt n_1^0 Fahrzeuge. Die auf Zufahrt 1 abzufertigende Fahrzeuganzahl $\hat{n}_1(t_0)$ berechnet sich aus Gleichung (3.20). Sie vereinfacht sich mit der Substitution

$$N_1^{\text{exp}}(t_0 + \tau_i^0 + \hat{g}_1) = N_1^{\text{out}}(t_0) + (\tau_i^0 + \hat{g}_1)q_1 \quad (\text{A.18})$$

zu

$$\hat{n}_1(t_0) = n_1^0 + (\tau_i^0 + \hat{n}_1(t_0)/q_1^{\text{max}})q_1. \quad (\text{A.19})$$

Diese kann dann explizit nach $\hat{n}_i(t_0)$ aufgelöst werden. Zusammenfassend gilt zum Zeitpunkt t_0 :

$$\begin{pmatrix} n_1(t_0) \\ n_2(t_0) \end{pmatrix} = \begin{pmatrix} n_1^0 \\ 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} \hat{n}_1(t_0) \\ \hat{n}_2(t_0) \end{pmatrix} = \begin{pmatrix} \hat{n}_1^0 \\ 0 \end{pmatrix} \quad \text{mit} \quad \hat{n}_1^0 = \frac{n_1^0 + \tau_1^0 q_1}{1 - q_1/q_1^{\text{max}}}. \quad (\text{A.20})$$

Für den Prioritätsindex $\pi_1(t_0)$ erhält man mit der zweiten Zeile aus Gleichung (4.35) und mit $\tau_2^{\text{pen}} = \tau_2^0/2$ den Wert

$$\pi_1(t_0) = \frac{\hat{n}_1^0}{\tau_2^0/2 + \tau_1^0 + \hat{n}_1^0/q_1^{\text{max}}}. \quad (\text{A.21})$$

Zur Berechnung des Prioritätsindex $\pi_2(t)$ der zweiten Zufahrt ist die erste Zeile aus Gleichung (4.35) anzuwenden. Entsprechend den Überlegungen in der Fußnote auf Seite 88, d.h. weil aufgrund der konstanten Ankunftsrate $\hat{g}_2(t, \tau) = \tau/(1 - q_2/q_2^{\text{max}})$

für $\tau > 0$ gilt, vereinfacht sich der Ausdruck zu:

$$\pi_2(t_0) = q_2. \quad (\text{A.22})$$

Umschalten zur Zufahrt 1

Es wird $\pi_1(t_0) > \pi_2(t_0)$ angenommen. Diese Annahme ist bei nicht extrem verschiedenen Zuflussraten erfüllt. Die Schaltregel (4.34) löst damit, sofort nach dem die Warteschlange auf Zufahrt 2 geleert wurde, das Umschalten zur Zufahrt 1 aus. Während des Umschaltvorgangs im Zeitraum $t = [t_0, t_0 + \tau_1^0]$ sind die Ausflüsse beider Zufahrten gesperrt, die Warteschlangen entwickeln sich entsprechend Gleichung (3.11) mit:

$$\begin{pmatrix} n_1(t_0 + \Delta t) \\ n_2(t_0 + \Delta t) \end{pmatrix} = \begin{pmatrix} n_1^0 \\ 0 \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \Delta t \quad \text{mit} \quad \Delta t = [0, \tau_1^0]. \quad (\text{A.23})$$

Die abzufertigende Fahrzeuganzahl $\hat{n}_1(t)$ bleibt im Zustand ‚Umschalten‘ nach Gleichung (3.31) konstant. Dagegen springt $\hat{n}_2(t_0)$ zum Umschaltzeitpunkt t_0 von null auf den Wert

$$\tau_2^0 \frac{q_2^{\max} q_2}{q_2^{\max} - q_2}, \quad (\text{A.24})$$

der sich aus Gleichung (3.20) mit der Substitution (A.18) berechnen lässt. Anschließend wächst \hat{n}_2 nach Gleichung (3.30) mit der konstanten Rate

$$\frac{d\hat{n}_2}{dt} = \frac{q_2^{\max} q_2}{q_2^{\max} - q_2}. \quad (\text{A.25})$$

Zusammenfassend gelten für die abzufertigenden Fahrzeuganzahlen die Entwicklungsgleichungen:

$$\hat{n}_1(t_0 + \Delta t) = \hat{n}_1^0 \quad (\text{A.26})$$

$$\text{und} \quad \hat{n}_2(t_0 + \Delta t) = (\tau_2^0 + \Delta t) \frac{q_2 q_2^{\max}}{q_2^{\max} - q_2}. \quad (\text{A.27})$$

Der Prioritätsindex $\pi_1(t)$ gehorcht der ersten Zeile aus Gleichung (4.35) und entwickelt sich deshalb mit

$$\pi_1(t_0 + \Delta t) = \frac{\hat{n}_1^0}{\tau_1^0 - \Delta t + \hat{n}_1^0 / q_1^{\max}}. \quad (\text{A.28})$$

Die Differenz im Nenner macht deutlich, dass $\pi_1(t)$ mit zunehmender Umschaltzeit Δt gegen den Maximalwert q_1^{\max} strebt, der bei $\Delta t = \tau_1^0$ erreicht wird. Der Prioritätsindex $\pi_2(t)$ berechnet sich mit der dritten Zeile aus Gleichung (4.35).

Wegen $\tau_1^{\text{pen}} = \Delta t/2$ erhält man:

$$\begin{aligned}\pi_2(t_0 + \Delta t) &= \frac{\hat{n}_2(t_0 + \Delta t)}{\Delta t/2 + \tau_2^0 + \hat{n}_2(t_0 + \Delta t)/q_2^{\text{max}}} \\ &= q_2^{\text{max}} \frac{(\tau_2^0 + \Delta t)q_2}{(\Delta t/2 + \tau_2^0)(q_2^{\text{max}} - q_2) + (\tau_2^0 + \Delta t)q_2}.\end{aligned}\quad (\text{A.29})$$

Für den Verlauf von π_2 kann man aus Gleichung (A.29) Folgendes ablesen: Unmittelbar zu Beginn des Umschaltvorgangs behält π_2 seinen vorherigen Wert $\pi_2(t_0)$ bei:

$$\pi_2(t_0 + 0) = \pi_2(t_0). \quad (\text{A.30})$$

Das lässt sich durch Einsetzen der Gleichungen (A.22) und (A.29) zeigen. Anschließend, für $\Delta t > 0$, wächst $\pi_2(t)$ mit $d\pi_2/dt > 0$ kontinuierlich an.

Zum Zeitpunkt $t_1 = t_0 + \tau_1^0$ ist der Umschaltvorgang abgeschlossen und die Freigabe von Zufahrt 1 kann beginnen.

Freigabe von Zufahrt 1

Aus der Warteschlange von Zufahrt 1 fahren die Fahrzeuge mit maximaler Rate q_1^{max} heraus, während neue Fahrzeuge mit der Rate q_1 hinzukommen. Die Warteschlange der zweiten Zufahrt wächst mit der Rate q_2 . Die Entwicklungsgleichungen für beide Warteschlangen lauten:

$$n_1(t_0 + \tau_1^0 + \Delta t) = q_1 \tau_1^0 + (q_1^{\text{max}} - q_1) \Delta t \quad (\text{A.31})$$

$$\text{und} \quad n_2(t_0 + \tau_1^0 + \Delta t) = q_2 \tau_1^0 + q_2 \Delta t. \quad (\text{A.32})$$

Die Warteschlange n_1 der ersten Zufahrt ist zu einem Zeitpunkt t_2 vollständig geleert, für den gilt:

$$t_2 = t_0 + \tau_1^0 + \Delta t \quad \text{mit} \quad \Delta t = \frac{n_1^0 + q_1 \tau_1^0}{q_1^{\text{max}} - q_1}. \quad (\text{A.33})$$

Die Anzahl der abzufertigenden Fahrzeuge $\hat{n}_1(t)$ nimmt während der Freigabe im Zeitraum $[t_1, t_2]$ nach Gleichung (3.32) mit der Rate q_1^{max} ab. Auf der zweiten Zufahrt wächst $\hat{n}_2(t)$ weiter mit der Rate $q_2 q_2^{\text{max}}/(q_2^{\text{max}} - q_2)$ an. Man erhält die Entwicklungsgleichungen:

$$\hat{n}_1(t_0 + \tau_1^0 + \Delta t) = \hat{n}_1^0 - q_1^{\text{max}} \Delta t \quad (\text{A.34})$$

$$\text{und} \quad \hat{n}_2(t_0 + \tau_1^0 + \Delta t) = (\tau_2^0 + \tau_1^0 + \Delta t) \frac{q_2 q_2^{\text{max}}}{q_2^{\text{max}} - q_2}. \quad (\text{A.35})$$

Zur Probe kann der Zeitpunkt t_2 , für den $\hat{n}_1(t_2) = 0$ gelten soll, in gleicher Weise auch aus Gleichung (A.34) berechnet werden:

$$t_2 = t_0 + \tau_1^0 + \Delta t \quad \text{mit} \quad \Delta t = \frac{\hat{n}_i^0}{q_1^{\max}} = \frac{n_1^0 + q_1 \tau_1^0}{q_1^{\max} - q_1}. \quad (\text{A.36})$$

Die Übereinstimmung mit Gleichung (A.33) bestätigt, dass n_1 und \hat{n}_1 gleichzeitig null werden.

Der Prioritätsindex $\pi_1(t)$ ist während der Warteschlangenleerung maximal, danach fällt er auf den Wert q_1 ab:

$$\pi_1(t_0 + \tau_1^0 + \Delta t) = \begin{cases} q_1^{\max}, & \text{falls } t_0 + \tau_1^0 + \Delta t < t_2 \\ q_1, & \text{sonst.} \end{cases} \quad (\text{A.37})$$

Der Prioritätsindex $\pi_2(t)$ wächst weiterhin mit der zweiten Zeile aus Gleichung (4.35). Mit $\tau_1^{\text{pen}} = \tau_1^0/2$ gehorcht sein Verlauf folgendem Ausdruck:

$$\begin{aligned} \pi_2(t_0 + \tau_1^0 + \Delta t) &= \frac{\hat{n}_2(t_0 + \tau_1^0 + \Delta t)}{\tau_1^0/2 + \tau_2^0 + \hat{n}_2(t_0 + \tau_1^0 + \Delta t)/q_2^{\max}} \\ &= q_2^{\max} \frac{(\tau_2^0 + \tau_1^0 + \Delta t)q_2}{(\tau_1^0/2 + \tau_2^0)(q_2^{\max} - q_2) + (\tau_2^0 + \tau_1^0 + \Delta t)q_2}. \end{aligned} \quad (\text{A.38})$$

Es sei davon ausgegangen, dass aufgrund des Abfalls von $\pi_1(t)$ zum Zeitpunkt t_2 die Bedingung $\pi_1(t_2) < \pi_2(t_2)$ erfüllt ist. Damit wird nun zur Bedienung von Zufahrt 2 umgeschaltet.

Bedienung von Zufahrt 2

Die Bedienung von Zufahrt 2 kann man mit vertauschten Indizes in gleicher Weise berechnen. Der Umschaltvorgang ist zum Zeitpunkt

$$t_3 = t_2 + \tau_2^0 = t_0 + \tau_1^0 + \frac{\hat{n}_i^0}{q_1^{\max}} + \tau_2^0 \quad (\text{A.39})$$

abgeschlossen. Zufahrt 2 ist zum Zeitpunkt

$$t_4 = t_3 + \frac{\hat{n}_2(t_2)}{q_2^{\max}} = t_0 + \tau_1^0 + \frac{\hat{n}_i^0}{q_1^{\max}} + \tau_2^0 + \frac{\hat{n}_2(t_2)}{q_2^{\max}} \quad (\text{A.40})$$

vollständig geleert und die Schaltfolge wiederholt sich.

Zum Zeitpunkt t_4 erhält man für $\hat{n}_1(t_4)$ folgenden Wert:

$$\hat{n}_1(t_4) = \frac{q_1 q_1^{\max}}{q_1^{\max} - q_1} \left(\tau_1^0 + \tau_2^0 + \frac{\hat{n}_2(t_2)}{q_2^{\max}} \right) \quad (\text{A.41})$$

$$= \frac{q_1 q_1^{\max}}{q_1^{\max} - q_1} \left(\tau_1^0 + \tau_2^0 + \frac{\frac{q_2 q_2^{\max}}{q_2^{\max} - q_2} \left(\tau_1^0 + \tau_2^0 + \frac{\hat{n}_1^0}{q_1^{\max}} \right)}{q_2^{\max}} \right) \quad (\text{A.42})$$

$$= \underbrace{q_1^{\max}(\tau_1^0 + \tau_2^0)c_1(1 + c_2)}_{=:a} + \underbrace{c_1 c_2 \hat{n}_1^0}_{=:b} \quad \text{mit} \quad c_i = \frac{q_i}{q_i^{\max} - q_i}. \quad (\text{A.43})$$

Das heißt, der Wert für $\hat{n}_1(t_4)$ ergibt sich mit einer linearen Iterationsvorschrift aus dem Wert im selben Schaltzustand der zurückliegenden Periode. Im Folgenden wird das Konvergenzverhalten dieser Iterationsvorschrift genauer untersucht.

A.3.2. Konvergenz der Iterationsvorschrift

Der Wert für \hat{n}_1^{k+1} in Periode $k+1$ ergibt sich gemäß Gleichung (A.43) mit folgender linearen Iterationsvorschrift aus dem Wert \hat{n}_1^k der vorherigen Periode k :

$$\hat{n}_1^{k+1} = a + b \hat{n}_1^k. \quad (\text{A.44})$$

Die Werte der Konstanten a und b sind in Gleichung (A.43) definiert.

Der Zustand, in dem

$$\hat{n}_1^{k+1} = \hat{n}_1^k \quad (\text{A.45})$$

gilt, stellt einen Gleichgewichtszustand dar. Die zugehörige stationäre Lösung ist gekennzeichnet durch:

$$\hat{n}_1^{\text{stat}} = \frac{a}{1 - b}. \quad (\text{A.46})$$

Für diese erhält man durch Einsetzen:

$$\hat{n}_1^{\text{stat}} = q_1 T^{\text{cyc}, \min} \quad \text{mit} \quad T^{\text{cyc}, \min} = \frac{\tau_1^0 + \tau_2^0}{1 - q_1/q_1^{\max} - q_2/q_2^{\max}}. \quad (\text{A.47})$$

Im stationären Fall werden also in einem Bedienvorgang so viele Fahrzeuge mit maximaler Rate abgefertigt, wie während der minimal möglichen „Umlaufzeit“ aus Gleichung (2.8) ankommen. Folglich hat der Grenzyklus die minimal mögliche Periode.

Wie eingangs gesagt, stellt dieser Grenzyklus mit minimaler Periode die optimale Lösung dar. Nun ist noch zu prüfen, ob dieser Gleichgewichtszustand auch

stabil ist. Er ist genau dann stabil, wenn gilt:

$$b = c_1 c_2 < 1. \quad (\text{A.48})$$

Durch Einsetzen stellt man fest, dass diese Bedingung genau dann erfüllt ist, wenn auch die notwendige Stabilitätsbedingung (2.3) erfüllt ist. Das heißt, man kann die Bedingung (A.48) durch folgende Umformungsschritte in notwendige Stabilitätsbedingung (2.3) überführen:

$$\frac{q_1}{q_1^{\max} - q_1} \frac{q_2}{q_2^{\max} - q_2} < 1 \quad (\text{A.49})$$

$$q_1 q_2 < q_1^{\max} q_2^{\max} - q_1^{\max} q_2 - q_1 q_2^{\max} + q_1 q_2 \quad (\text{A.50})$$

$$\frac{q_1}{q_1^{\max}} + \frac{q_2}{q_2^{\max}} < 1. \quad (\text{A.51})$$

Da die Iterationsvorschrift (A.44) linear ist, handelt es sich um globale Stabilität. Damit ist gezeigt, dass das flexible Programm im betrachteten Szenario 4.vi aus Abb. A.3(a) von jedem beliebigen Anfangszustand aus gegen die optimale Lösung strebt.

A.4. Theoretische Verteilung der Bedienperiode

Im Allgemeinen wird die Wunschperiode T aufgrund stochastischer Einflüsse nicht exakt eingehalten. Die tatsächliche Umlaufzeit variiert mit der Anzahl der Fahrzeuge, die innerhalb der Zeitspanne T ankommen. Im Folgenden werden die stochastischen Eigenschaften der tatsächlichen Umlaufzeit untersucht.

A.4.1. Definition

Tatsächliche Bedienperiode

Die tatsächliche Bedienperiode z_i sei die Zeitspanne gemessen vom Ende eines Bedienvorgangs für Zufahrt i bis zum Ende des darauf folgenden Bedienvorgangs für die selbe Zufahrt. Es sei angenommen, dass der Bedienvorgang sofort beginnt nachdem $\hat{n}_i(t)$ den Schwellenwert $n_i^{\text{crit}}(t)$ aus Gleichung (5.30) erreicht. Die Warteschlange wird dann vollständig geleert, d.h. es werden genau $n_i^{\text{crit}}(t)$ Fahrzeuge abgefertigt. Damit entspricht die tatsächliche Bedienperiode z_i genau der Zeit, die während der Ankunft von $n_i^{\text{crit}}(t)$ Fahrzeugen vergeht. Somit erfüllt z_i folgende Gleichung:

$$n_i^{\text{crit}}(t) = \int_{t-z_i}^t q_i^{\text{exp}}(t') dt'. \quad (\text{A.52})$$

Die zu z_i gehörige Zufallsgröße sei Z_i . Die Zeitabhängigkeit von z_i muss nicht berücksichtigt werden, weil der Zufallsprozess, der die Fahrzeugankünfte beschreibt, als stationär angenommen wird.

Zufallsprozess der Fahrzeugankünfte

Die Ankunftszeitpunkte der Fahrzeuge seien unabhängig voneinander. Diese Annahme ist auf mehrspurigen Knotenzufahrten zu isolierten Knotenpunkten erfüllt. Damit sind die Zeitlücken zwischen zwei aufeinander folgenden Fahrzeugen exponentialverteilt. Die zufällige Anzahl der innerhalb eines gegebenen Zeitintervalls ankommenden Fahrzeuge ist dann Poisson-verteilt. Dagegen gehorcht die zufällige Zeitspanne bis zum Eintreffen von n Fahrzeugen der Erlang-Verteilung [20, 121].

A.4.2. Theoretische Verteilung mit $T^{\max} \rightarrow \infty$

Zunächst wird davon ausgegangen, dass der Schwellenwert n_i^{crit} mit $T^{\max} \rightarrow \infty$ konstant ist.

Während einer Bedienperiode z_i werden genau n_i^{crit} Fahrzeuge abgefertigt. Damit gehorcht die Zufallsgröße Z_i der Erlang-Verteilung:

$$Z_i \sim \text{Erl}(\lambda, n), \quad \text{mit} \quad \lambda = \bar{q}_i^{\text{exp}} \quad \text{und} \quad n = \lceil n_i^{\text{crit}} \rceil. \quad (\text{A.53})$$

Der Parameter λ ist die Ereignisrate und entspricht im vorliegenden Fall der mittleren Ankunftsrate \bar{q}_i^{exp} . Der Parameter n ist die Anzahl der beobachteten Ereignisse, hier die Ankunft von n_i^{crit} Fahrzeugen. Da n eine positive natürliche Zahl sein muss, wird n_i^{crit} mit $\lceil \cdot \rceil$ aufgerundet. Die zugehörige Wahrscheinlichkeitsdichte $p(z_i)$ wird durch die Funktion

$$p(z_i) = \begin{cases} \frac{\lambda^n}{(n-1)!} z_i^{(n-1)} e^{-\lambda z_i}, & \text{falls } z_i \geq 0 \\ 0, & \text{sonst} \end{cases} \quad (\text{A.54})$$

beschrieben.

Erwartungswert und Varianz

Für Erwartungswert und Varianz gelten $E(Z_i) = n/\lambda$ bzw. $\text{Var}(Z_i) = n/\lambda^2$. Ersetzt man n_i^{crit} entsprechend Gleichung (5.9) durch $\bar{q}_i^{\text{exp}} T$, leiten sich daraus zwei wesentliche Erkenntnisse ab:

$$E(Z_i) = \frac{n}{\lambda} = \frac{n_i^{\text{crit}}}{\bar{q}_i^{\text{exp}}} = T \quad \text{und} \quad \text{VarK}(Z_i) = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{\bar{q}_i^{\text{exp}} T}}. \quad (\text{A.55})$$

Zum Ersten deckt sich die im Mittel zu erwartende Bedienperiode $E(Z_i)$ genau mit der gegebenen Wunschperiode T . Zum Zweiten sind die relativen Abweichungen vom Erwartungswert T , die sich durch den Variationskoeffizienten $\text{VarK}(Z_i) = \sqrt{\text{Var}(Z_i)}/E(Z_i)$ ausdrücken, umso kleiner, je größer die mittlere Verkehrsnachfrage \bar{q}_i^{exp} ist.

A.4.3. Theoretische Verteilung mit endlichem T^{max}

Theoretische Verteilung

Nun wird der Einfluss des Parameters α aus Gleichung (5.22) auf die Verteilung der tatsächlichen Umlaufzeit Z_i untersucht. Dem homogenen Poisson-Prozess mit der Intensität $\lambda = (1 - \alpha)\bar{q}_i^{\text{exp}}$, der die Fahrzeugankünfte beschreibt, ist nun eine deterministische lineare Drift $\alpha\bar{q}_i^{\text{exp}}$ überlagert. Nach z_i Zeiteinheiten hat die lineare Drift bereits $\alpha\bar{q}_i^{\text{exp}}z_i$ virtuelle Fahrzeuge zum Erreichen des Schwellenwerts n_i^{crit} beigetragen. Der kritische Schwellenwert für den Poisson-Prozess, d.h. die Anzahl der verbleibenden $n(z_i) = \lceil n_i^{\text{crit}} - \alpha\bar{q}_i^{\text{exp}}z_i \rceil$ Fahrzeuge, ist somit eine linear abfallende Funktion in z_i . Mit diesen Parametern λ und $n(z_i)$ gehorcht die Zufallsgröße Z_i einer verallgemeinerten Erlang-Verteilung [143, 6] mit der Wahrscheinlichkeitsdichte

$$p(z_i) = \begin{cases} \frac{\lambda^{n(z_i)} z_i^{[n(z_i)-1]}}{(1-\alpha)^{[n(z_i)-1]}!} e^{-\lambda z_i}, & \text{falls } 0 \leq z_i \leq T/\alpha \\ 0, & \text{sonst .} \end{cases} \quad (\text{A.56})$$

Mit $\alpha = 0$ geht Gleichung (A.56) direkt in Gleichung (A.54) über. Die obere Schranke T/α des Definitionsbereichs rührt daher, dass der Schwellenwert n_i^{crit} in jedem Fall, auch ohne Beiträge des stochastischen Anteils, nach einer maximalen Zeit allein von der deterministischen Drift erreicht wird. Deshalb wird diese obere Schranke $T^{\text{max}} = T/\alpha$ auch Maximalperiode genannt.

Symbolverzeichnis

Symbol	Beschreibung	Seite
α	Relativer Anteil des virtuellen Fahrzeugstroms. Vgl. Gl. (5.22).	100
β	Richtungsanteil im Simulationsszenario 7.iii. Vgl. Gln. (7.9) und (7.10).	130
$C_i^{\text{fuel}}(t)$	Kumulierter Gesamtkraftstoffverbrauch auf Zufahrt i bis zur Zeit t . Vgl. Gl. (3.18).	50
$C_i^{\text{stop}}(t)$	Kumulierte Anzahl der Haltevorgänge aller Fahrzeuge auf Zufahrt i bis zur Zeit t . Vgl. Gl. (3.15).	49
η_i	Effizienz eines Bedienvorgangs. Vgl. Gl. (5.2).	91
$\hat{\eta}_i(t)$	Prognostizierte Bedieneffizienz. Vgl. Gl. (5.5).	93
$\hat{g}_i(t, \tau)$	Benötigte Freigabezeit zur Leerung der Warteschlange auf Zufahrt i . Der Bedienvorgang beginnt zur Zeit t , die Freigabe zur Zeit $t + \tau$. Vgl. Abb. 3.4 und Gln. (3.19) und (3.20).	51
g_i^0	Freigabezeit einer Knotenzufahrt i im umlaufbasierten Schaltprogramm. Vgl. Gl.(2.5).	27
g_i^s	Zugesicherte Freigabezeit. Vgl. Abb. 5.1 und Gl. (5.15).	96
$\gamma_i(t)$	Freigabezustand. Einheitenlos. Wertebereich: $\gamma_i(t) \in \{0, 1\}$. Vgl. Abb. 2.7(c).	26
$\bar{\gamma}_i$	Mittlerer Freigabezeitanteil für Knotenzufahrt i .	27
i	Index einer Knotenzufahrt.	40
$\lambda_i(t)$	Auslastung von Knotenzufahrt i . Verhältnis aus tatsächlicher zu maximaler Zuflussrate. Vgl. Gl. (2.4).	27
$\bar{\lambda}_i$	Mittlere Auslastung von Knotenzufahrt i . Vgl. Gl. (2.4).	27
Λ	Knotenpunktauslastung. Vgl. Gl. (2.9).	28
Λ^{\max}	Maximalauslastung eines Knotenpunktes. Vgl. Gl. (2.10).	28
$n_i(t)$	Warteschlangenlänge von Knotenzufahrt i zur Zeit t . Einheit: Fahrzeuge. Vgl. Abb. 3.3 und Gl. (3.10).	46

Symbol	Beschreibung	Seite
$\hat{n}_i(t, \tau)$	Abzufertigende Fahrzeuganzahl zur Leerung der Warteschlange auf Zufahrt i . Der Bedienvorgang beginnt zur Zeit t , die Freigabe zur Zeit $t + \tau$. Vgl. Abb. 3.4 und Gln. (3.19) und (3.20).	51
$n_i^{\text{crit}}(t)$	Schwellenwert. Vgl. Gln. (5.9) und (5.30).	103
$N_i^{\text{exp}}(t)$	Zeitreihe der erwarteten Ankünfte. Vgl. Abb. 3.1 und Gl. (3.1).	40
$N_i^{\text{out}}(t)$	Anzahl der abgefertigten Fahrzeuge. Vgl. Abb. 3.3 und Gl. (3.5).	44
Ω	Liste der Indizes jener Zufahrten, die vorrangig zu bedienen sind. Vgl. Gln. (5.13) bis (5.15).	96
p	Index einer Phase.	25
$\pi_i(t)$	Dynamischer Prioritätsindex. Einheit: Fz/s. Wertebereich $\pi_i(t) \in [0, q_i^{\text{max}}]$. Vgl. Abb. 4.4 und Gl. (4.35).	83
$q_i^{\text{exp}}(t)$	Ankunftsrate an der Haltelinie von Knotenzufahrt i zur Zeit t unter Annahme freien Verkehrs. Einheit: Fz/s. Wertebereich: $q_i(t) \in [0, q_i^{\text{max}}]$. Vgl. Abb. 3.1 und Gl. (3.1).	40
q_i^{max}	Sättigungsverkehrsstärke. Maximale Ausflussrate einer Knotenzufahrt i . Vgl. Gln. (3.2) und (3.6).	40
$q_i^{\text{out}}(t)$	Ausflussrate aus Knotenzufahrt i zur Zeit t . Vgl. Gl. (3.5) und Abb. 3.2.	44
$\sigma(t)$	Index der zur Zeit t bedienten Knotenzufahrt.	83
T	Wunsch-Bedienperiode. Vgl. Gl. (5.9).	94
T^{cyc}	Umlaufzeit. Vgl. Gl. (2.7).	28
$T^{\text{cyc, min}}$	Mindestumlaufzeit. Vgl. Gl. (2.8).	28
$T^{\text{cyc, opt}}$	Wartezeitenoptimale Umlaufzeit bei stochastischen Ankünften. Vgl. Abb. 2.9 und Gl. (2.12).	31
T^{max}	Maximalperiode. Vgl. Abb. 5.2 und Gl. (5.24).	101
$\tau_i(t)$	Verbleibende Umschaltzeit. Vgl. Abb. 3.5.	55
τ_i^0	Zwischenzeit.	25
V_i^0	Zulässige Höchstgeschwindigkeit auf Knotenzufahrt i .	42
$W_i(t)$	Kumulierte Gesamtwarezeiten aller Fahrzeuge auf Knotenzufahrt i bis zur Zeit t . Einheit: Fahrzeugsekunden. Vgl. Abb. 3.3 und Gl. (3.12).	46

Symbol	Beschreibung	Seite
$\hat{W}_i(t)$	Prognostizierte Gesamtwartezeiten bis zum Ende des nachfolgenden Bedienvorgangs. Vgl. Abb. 3.7 und Gl. (3.34).	58
W^{cyc}	Gesamtwartezeiten aller Fahrzeuge während eines Umlaufs auf allen Zufahrten eines Knotenpunktes.	157
W_i^{cyc}	Gesamtwartezeiten aller Fahrzeuge während eines Umlaufs auf einer Knotenzufahrt i . Vgl. Gl. (2.10).	30
z_i	Tatsächliche Bedienperiode. Vgl. Abb. 5.3 und Gl. (A.56).	105

Literaturverzeichnis

- [1] New Traffic Signal Installed. In: *The Motorist* 8 (1914), S. 28–29
- [2] To Rule All Traffic from Times Square. In: *New York Times* (1922), 5. Januar, S. 1
- [3] *Merkblatt über Detektoren für den Straßenverkehr*. Köln : Forschungsgesellschaft für Straßen- und Verkehrswesen e. V., 1991
- [4] *Richtlinien für Lichtsignalanlagen – RiLSA*. Köln : Forschungsgesellschaft für das Straßen- und Verkehrswesen e. V., 1992
- [5] *Handbuch für die Bemessung von Straßenverkehrsanlagen*. Köln : Forschungsgesellschaft für Straßen- und Verkehrswesen e.V., 2005
- [6] ABDEL-HAMEED, M. : Optimal control of a dam using $P_{\lambda, \tau}^M$ policies and penalty cost when the input process is a compound Poisson process with positive drift. In: *Journal of Applied Probability* 37 (2000), Nr. 2, S. 408–416
- [7] ADAMS, W. F.: Road traffic considered as a random series. In: *Journal of the Institute of Civil Engineers* 4 (1936)
- [8] ADOLPH, J. : Verfahren zur Regelung des Wagenverkehrs in Straßen mittels elektrischer Signallampen. In: *Reichspatentamt* DE 439255 A (1925), S. 1–6
- [9] AHN, M. : Veränderung der Leistungsfähigkeit städtischer Hauptverkehrsstraßen über die Tageszeit / Ruhr-Universität Bochum. 1987. – Forschungsbericht
- [10] ALLSOP, R. E.: Delay at a Fixed Time Traffic Signal-I: Theoretical Analysis. In: *Transportation Science* 6 (1972), Nr. 3, S. 260–305
- [11] ALMASRI, E. ; FRIEDRICH, B. : Online offset optimisation in urban networks based on cell transmission model. In: *Proceedings of the 5th European Congress on Intelligent Transport Systems and Services*. Hannover, Juni 2005
- [12] BÄCK, T. : *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, 2004

- [13] BANKS, J. H.: Investigation of some characteristics of congested flow. In: *Transportation research record* 1678 (1999), S. 128–134
- [14] BELL, M. C. ; BRETHERTON, R. D.: Ageing of fixed-time traffic signal plans. In: *2nd International Conference on Road Traffic Control*. London, 1986
- [15] Benson Ford Research Center (Dearborn, Michigan, USA): *Museum archives, Henry Ford Museum*
- [16] BERNHARD, J. ; RIEDEL, T. : Erkennung von Stau mit kurzen Schleifendetektoren. In: *Straßenverkehrstechnik* 7 (1999), Nr. 99, S. 329–332
- [17] BIELEFELDT, C. ; BUSCH, F. : MOTION - a new on-line traffic signal network control system. In: *Seventh International Conference on Road Traffic Monitoring and Control*. London, 1994, S. 55–59
- [18] BIELEFELDT, C. ; DIAKAKI, C. ; PAPAGEORGIOU, M. : TUC and the SMART NETS project, 2001, S. 55–60
- [19] BLANK, M. ; BUNIMOVICH, L. : Switched flow systems: pseudo billiard dynamics. In: *Dynamical Systems: An International Journal* 19 (2004), Nr. 4, S. 359–370
- [20] BOHLEY, P. : *Statistik*. Oldenbourg, 2000
- [21] BOILLOT, F. ; BLOSSEVILLE, J. ; LESORT, J. ; MOTYKA, V. ; PAPAGEORGIOU, M. ; SELLAM, S. : Optimal signal control of urban traffic networks. In: *Sixth International Conference on Road Traffic Monitoring and Control* Bd. 355, 1992, S. 75–79
- [22] BOILLOT, F. ; MIDENET, S. ; PIERRELEE, J. C.: Real-life CRONOS evaluation. In: *Tenth International Conference on Road Transport Information and Control* Bd. 472, 2000, S. 182–186
- [23] BOSE, S. K.: *An Introduction to Queueing Systems*. Springer, 2001
- [24] BRETHERTON, D. ; BODGER, M. ; BABER, N. : SCOOT - the future. In: *12th IEE International Conference on Road Transport Information and Control* Bd. 501, 2004, S. 301–306
- [25] BRILON, W. ; WU, N. : Capacity at unsignalized Two-Stage Priority Intersections. In: *Transportation Research A* 33 (1999), Nr. 3, S. 275–289
- [26] BRILON, W. ; GROSSMANN, M. ; BLANKE, H. : *Verfahren für die Berechnung der Leistungsfähigkeit und Qualität des Verkehrsablaufes auf Straßen*. Bonn - Bad Godesberg : Bundesministerium für Verkehr, Abteilung Straßenbau, 1994 (Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik 669)

-
- [27] BROWN, D. E.: *Inventing Modern America: from the Microwave to the Mouse*. The MIT Press, 2002
- [28] BUHL, P.-A. : Bus und Bahn haben Vorfahrt. In: *Sächsische Zeitung* (2007), 23. Januar
- [29] BURGESS, K. ; PASSINO, K. M.: Stable scheduling policies for flexible manufacturing systems. In: *IEEE Transactions on Automatic Control* 42 (1997), Nr. 3, S. 420–425
- [30] BUSCH, F. ; KRUSE, G. : MOTION - Ein neues Verfahren für die städtische Lichtsignalsteuerung und seine Erprobung in Rahmen des EG-Programms ATT. In: *Heureka!* 93 (1993), S. 79–93
- [31] BUSCH, F. ; KRUSE, G. : MOTION for SITRAFFIC - a modern approach to urban traffic control. In: *IEEE Intelligent Transportation Systems Conference Proceedings*. Oakland, 2001, S. 61–64
- [32] CHANG, K. C. ; SANDHU, D. : Mean waiting time approximations in cyclic-service systems with exhaustive limited service policy. In: *Performance Evaluation* 15 (1992), Nr. 1, S. 21–40
- [33] CHANG, K. C. ; SANDHU, D. : Delay Analyses of Token-Passing Protocols with Limited Token Holding Times. In: *IEEE transactions on communications* 42 (1994), S. 2833–2842
- [34] CHASE, C. ; RAMADGE, P. J.: On real-time scheduling policies for flexible manufacturing systems. In: *IEEE Transactions on Automatic Control* 37 (1992), Nr. 4, S. 491–496
- [35] CHEN, H. ; COHEN, S. L. ; GARTNER, N. H. ; LIU, C. C.: Simulation study of OPAC: A demand-responsive strategy for traffic signal control. In: GARTNER, N. H. (Hrsg.) ; WILSON, N. H. M. (Hrsg.): *Transportation and Traffic Theory: Proceedings of the Tenth International Symposium on Transportation and Traffic Theory*. Elsevier Applied Science, 1987, S. 233–249
- [36] CHEN, H. ; YANG, P. ; YAO, D. D.: Control and scheduling in a two-station queueing network: Optimal policies and heuristics. In: *Queueing Systems* 18 (1994), Nr. 3-4, S. 301–332
- [37] CHIOU, S.-W. : TRANSYT derivatives for area traffic control optimisation with network equilibrium flows. In: *Transportation Research Part B* 37 (2003), Nr. 3, S. 263–290

- [38] CLINTON, B. : Ansprache zum National Defense Transportation Day, The White House, Office of the Press Secretary, 10. Mai 1997
- [39] COTTLE, R. W. ; PANG, J.-S. ; STONE, R. E.: *The Linear Complementarity Problem*. Academic Press, 1992
- [40] COX, D. ; SMITH, W. : *Queues*. Methuen, 1961
- [41] CREMER, M. ; LANDENFELD, M. : A Mesoscopic Model for Saturated Urban Road Networks. In: SCHRECKENBERG, M. (Hrsg.) ; WOLF, D. E. (Hrsg.): *Traffic and Granular Flow '97*. Springer, 1997
- [42] DAGANZO, C. F.: Traffic Delay at Unsignalized Intersections: Clarification of Some Issues. In: *Transportation Science* 11 (1977), Nr. 2, S. 180–189
- [43] DAGANZO, C. F.: Queue spillovers in transportation networks with a route choice. In: *Transportation Science* 32 (1998), Nr. 1, S. 3–11
- [44] DAI, J. G. ; JENNINGS, O. B.: Stabilizing Queueing Networks with Setups. In: *Mathematics of Operations Research* 29 (2004), Nr. 4, S. 891–922
- [45] DATTA, T. K. ; FEBER, D. ; SCHATTLER, K. ; DATTA, S. : Effective Safety Improvements Through Low-Cost Treatments. In: *Transportation Research Record* 1734 (2000), S. 1–6
- [46] DEUFLHARD, P. ; HOHMANN, A. : *Numerische Mathematik I. Eine algorithmisch orientierte Einführung*. Berlin : de Gruyter, 2002
- [47] DIAKAKI, C. ; DINOPOULOU, V. ; ABOUDOLAS, K. ; PAPAGEORGIOU, M. ; BEN-SHABAT, E. ; SEIDER, E. ; LEIBOV, A. : Extensions and new applications of the traffic signal control strategy TUC. In: *Transportation Research Board* 1856 (2003), S. 202–211
- [48] DIAKAKI, C. ; PAPAGEORGIOU, M. ; ABOUDOLAS, K. : A multivariable regulator approach to traffic-responsive network-wide signal control. In: *Control Engineering Practice* 10 (2002), Nr. 2, S. 183–195
- [49] DINOPOULOU, V. ; DIAKAKI, C. ; PAPAGEORGIOU, M. : Simulation investigations of the traffic-responsive urban control strategy TUC. In: *Intelligent Transportation Systems* (2000), S. 458–463
- [50] DINOPOULOU, V. ; DIAKAKI, C. ; PAPAGEORGIOU, M. : Applications of the urban traffic control strategy TUC. In: *European Journal of Operational Research* 175 (2006), S. 1652–1665

-
- [51] DÖGE, K.-P. : *Ein Beitrag zur videobasierten Verkehrszustandsidentifikation*. Institut für Verkehrsinformationssysteme, Professur für Verkehrsleitsysteme und -Prozessautomatisierung, TU Dresden, Diss., 2004
- [52] DOMSCHKE, W. ; DREXL, A. : *Einführung in Operations Research*. Springer, 2004
- [53] DONATI, F. ; MAURO, V. ; RONCOLINI, G. ; VALLAURI, M. : A Hierarchical Decentralised Traffic Light Control System. The First Realisation: 'Progetto Torino'. In: *9 th IFAC Triennial World Congress Bd. II 11.G/A-1*. Budapest, 1984
- [54] DUENYAS, I. ; VAN OYEN, M. P.: Heuristic Scheduling of Parallel Heterogeneous Queues with Set-Ups. In: *Management Science* 42 (1996), Nr. 6, S. 814–829
- [55] DUPUIS, P. : Explicit Solution to a Robust Queueing Control Problem. In: *SIAM Journal on Control and Optimization* 42 (2003), Nr. 5, S. 1854–1875
- [56] VAN EEKELEN, J. A. W. M. ; LEFEBER, E. ; ROODA, J. E.: State feedback control of switching server flowline with setups. In: *American Control Conference, 2007*
- [57] FRIEDRICH, B. : *Ein verkehrsadaptives Verfahren zur Steuerung von Lichtsignalanlagen*, Technische Universität München, Diss., 1997
- [58] FRIEDRICH, B. : Steuerung von Lichtsignalanlagen, BALANCE - ein neuer Ansatz. In: *Straßenverkehrstechnik* 7 (2000), S. 1–44
- [59] GARTNER, H. N.: OPAC: A demand-responsive strategy for traffic signal control. In: *Transportation Research Record* 906 (1983), S. 75–84
- [60] GARTNER, N. H. ; LITTLE, J. D. C. ; GABBAY, H. : Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming Part I: The Network Coordination Problem. In: *Transportation Science* 9 (1975), November, Nr. 4, S. 321–343
- [61] GARTNER, N. H. ; LITTLE, J. D. C. ; GABBAY, H. : Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming Part II: The Network Synchronization Problem. In: *Transportation Science* 9 (1975), Nr. 4, S. 344–363
- [62] GAZIS, D. C.: Optimum Control of a System of Oversaturated Intersections. In: *Journal of the Operations Research Society of America* 12 (1964), Nr. 6, S. 815–831

- [63] GAZIS, D. C.: Network Modelling and Control: Store-and-Forward Approach. In: PAPAGEORGIOU, M. (Hrsg.): *Concise Encyclopedia of Traffic & Transportation Systems*. Pergamon, 1991 (Advances in Systems Control and Information Engineering), S. 278–284
- [64] GAZIS, D. C.: The Origins of Traffic Theory. In: *Operations Research* 50 (2002), Nr. 1, S. 69–77
- [65] GAZIS, D. C. ; POTTS, R. B.: The oversaturated intersection. In: ALMOND, J. (Hrsg.): *Proceedings of the 2nd International Symposium on the theory of traffic flow*. London, 1965, S. 221–237
- [66] GEROLIMINIS, N. ; SKABARDONIS, A. : Prediction of arrival profiles and queue lengths along signalized arterials by using a markov decision process. In: *Transportation Research Record* 1934 (2005), S. 116–124
- [67] GLEBER, A. : Ampeln lernen die Grüne Welle. In: *Der Spezialist - das Magazin für Technik und Management* (2007), Februar, S. 46–49
- [68] GÖPFERT, Y. : Intelligente Ampeln - Projekt “Grüne Welle”. In: *Süddeutsche Zeitung* (2007), 30. Januar
- [69] GRACE, M. J. ; POTTS, R. B.: A Theory of the Diffusion of Traffic Platoons. In: *Operations Research* 12 (1964), Nr. 2, S. 255–275
- [70] GUBERINIC, S. ; SENBORN, G. : Optimal Sequences of Fixed Time Signal Plans. In: *Transportation Science* 12 (1978), Nr. 1, S. 48–58
- [71] GUBERINIĆ, S. ; LAZIĆ, B. ; ŠENBORN, G. : Optimal traffic control on a signalized intersection: structural constraints. In: MLADENOVIĆ, N. (Hrsg.): *30th Yugoslavian symposium on operations research*. Yugoslavia, September 2003, S. 599–602
- [72] HABBARD, B. B.: Hybrid systems: The control theory of tomorrow? In: *SIAM News* (1995), S. 12–13
- [73] HEIDEMANN, D. : Queue length and waiting-time distributions at priority intersections. In: *Transportation Research B* 25 (1991), Nr. 4, S. 163–174
- [74] HELBING, D. ; LÄMMER, S. : Supply and production networks: From the bullwhip effect to business cycles. In: ARMBRUSTER, D. (Hrsg.) ; MIKHAILOV, A. (Hrsg.) ; KANEKO, K. (Hrsg.): *Networks of Interacting Machines: Production Organization in Complex Industrial Systems and Biological Cells*. Singapore : World Scientific, 2005, S. 33–66

-
- [75] HELBING, D. : *Verkehrsdynamik. Neue physikalische Modellierungskonzepte*. Berlin : Springer, 1997
- [76] HELBING, D. : Traffic and related self-driven many-particle systems. In: *Review of Modern Physics* 73 (2001), S. 1067–1141
- [77] HELBING, D. : A section-based queueing-theoretical traffic model for congestion and travel time analysis in networks. In: *Journal of Physics A* 36 (2003), S. L593–L598
- [78] HELBING, D. ; JOHANSSON, A. ; AL-ABIDEEN, H. Z. ; AL-BOSTA, S. : *Evaluations of video data of crowds around the jamarat bridge and their safety implications*. 2007. – eingereicht
- [79] HELBING, D. ; LÄMMER, S. ; LEBACQUE, J.-P. : Self-Organized Control of Irregular or Perturbed Network Traffic. In: DEISSENBERG, C. (Hrsg.) ; HARTL, R. F. (Hrsg.): *Optimal Control and Dynamic Games*. Dordrecht : Springer, 2005, S. 239–274
- [80] HENRY, J. J. ; FARGES, J. L. ; TUFAL, J. : The PRODYN Real Time Traffic Algorithm. In: *4th IFAC/IFIP/IFORS International Conference on Control in Transportation Systems*. Baden Baden, 1983, S. 307–311
- [81] HENRY, J.-J. : PRODYN Tests and Future Experiments on ZELT. In: *IEEE Conference on Vehicle Navigation and Information Systems*. Toronto, 1989
- [82] HOFFMANN, S. ; ERNST, H. : Und wenn die Ampel länger grün ist? Verkehrsplanung mit Hilfe eines mikroskopischen Abgas-Simulationsmodells. In: *Unimagazin Hannover* 1 (2005), Nr. 2, S. 36–39
- [83] HOFRI, M. ; ROSS, K. : On the optimal control of two queues with server setup times and its analysis. In: *SIAM journal on computing* 16 (1987), Nr. 2, S. 399–420
- [84] HOGE, J. B.: Municipal traffic control system. In: *United States Patent and Trademark Office* 01251666 (1918), S. 1–12
- [85] HOPCROFT, J. E. ; MOTWANI, R. ; ULLMANN, J. D.: *Einführung in die Automatentheorie. Formale Sprachen und Komplexitätstheorie*. Pearson Studium, 2002
- [86] HUMES, C. : A regulator stabilization technique: Kumar-Seidman revisited. In: *IEEE Transactions on Automatic Control* 39 (1994), Nr. 1, S. 191–196

- [87] HUNT, P. B. ; ROBERTSON, D. L. ; BRETHERTON, R. : The SCOOT on-line traffic signal optimisation technique. In: *Traffic Engineering & Control* 23 (1982), S. 190–192
- [88] HUNT, P. B. ; ROBERTSON, D. ; BRETHERTON, R. D. ; WINTON, R. I.: SCOOT: a traffic responsive method of coordinating signals / Transport and Road Research Laboratory. Crowthorne, 1981 (TRRL Report No. LR 1041). – Forschungsbericht
- [89] HURDLE, V. F.: Signalized intersection delay models - A primer for the uninitiated. In: *Transportation Research Record* 971 (1984), S. 96–105
- [90] IMPROTA, G. : Signal Control at Individual Junctions: Phase-Based Approach. In: PAPAGEORGIOU, M. (Hrsg.): *Concise Encyclopedia of Traffic & Transportation Systems*. Pergamon, 1991 (Advances in Systems Control and Information Engineering), S. 473–478
- [91] INOSE, H. ; FUJISAKI, H. ; HAMADA, T. : Road traffic control system. In: *United States Patent and Trademark Office* US 3660812 (1970), S. 1–25
- [92] JAHN, P. . *www.info-lsa.de*
- [93] KALMAN, R. : A new approach to linear filtering and prediction problems. In: *Journal of Basic Engineering* D 82 (1960), S. 35–45
- [94] KANG, Y.-S. : *Delay, stop and queue estimation for uniform and random traffic arrivals at fixed-time signalized intersections*, Virginia State University, Diss., 2000
- [95] KHALIL, H. K.: *Nonlinear Systems*. Prentice Hall, 2001
- [96] KNOLL, C. : *Entwurf und Umsetzung eines Verkehrs-Zustands-Beobachters für dezentral gesteuerte Knotenzufahrten*. Fakultät Verkehrswissenschaften „Friedrich List“, Technische Universität Dresden, Studienarbeit, 2007
- [97] KRUMKE, S. O. ; NOLTEMEIER, H. : *Graphentheoretische Konzepte und Algorithmen*. B.G. Teubner Verlag, 2005
- [98] KUMAR, P. R. ; SEIDMAN, T. I.: Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. In: *IEEE Transactions on Automatic Control* 35 (1990), Nr. 3, S. 289–298
- [99] KUMAR, S. ; KUMAR, P. R.: Performance bounds for queueing networks and scheduling policies. In: *IEEE Transactions on Automatic Control* 39 (1994), Nr. 8, S. 1600–1611

-
- [100] KUMAR, S. ; MITRA, S. : Self-organizing traffic at a malfunctioning intersection. In: *Journal of Artificial Societies and Social Simulation* 9 (2006), Nr. 4
- [101] LÄMMER, S. ; GEHLSSEN, B. ; HELBING, D. : Scaling laws in the spatial structure of urban road networks. In: *Physica A* 363 (2006), Nr. 1, S. 89–95
- [102] LÄMMER, S. ; KORI, H. ; PETERS, K. ; HELBING, D. : Decentralised control of material or traffic flows in networks using phase-synchronisation. In: *Physica A* 363 (2006), Nr. 1, S. 39–47
- [103] LAN, C.-J. : Adaptive turning flow estimation based on incomplete detector-information for advanced traffic management. In: *Intelligent Transportation Systems IEEE*, 2001, S. 830–835
- [104] LAN, W.-M. ; OLSEN, T. L. *Multi-Product Systems with Both Setup Times and Costs: Fluid Bounds and Schedules*. Operations Research. 2003
- [105] LEFEBER, E. ; ROODA, J. E.: Controller design for switched linear systems with setups. In: *Physica A* 363 (2006), Nr. 1, S. 48–61
- [106] LEHNHOFF, N. : *Überprüfung und Verbesserung der Qualität von automatisch erhobenen Daten an Lichtsignalanlagen*, Universität Hannover, Diss., 2005
- [107] LIBERZON, D. ; MORSE, A. S.: Basic Problems in Stability and Design of Switched Systems. In: *IEEE Control Systems Magazine* 19 (1999), S. 59–70
- [108] LIDTHILL, M. ; WHITHAM, G. : On Kinematic Waves: II. A Theory of Traffic Flow on Long Crowded Roads. In: *Proceedings of the Royal Society A* 229 (1955), Nr. 1178, S. 317–345
- [109] LINDSEY, R. ; VERHOEF, E. : Congestion Modelling. In: HENSHER, D. A. (Hrsg.) ; BUTTON, K. J. (Hrsg.): *Handbook of Transport Modelling*. Pergamon, 2002, S. 353–374
- [110] LITTLE, J. D. C.: A Proof of the Queueing Formula $L = \lambda W$. In: *Operations Research* 9 (1961), S. 383–387
- [111] LITTLE, J. D. C. ; KELSON, M. D. ; GARTNER, N. H.: MAXBAND: A program for setting signals on arteries and triangular networks. In: *Transportation Research Record* 795 (1981), S. 40–46
- [112] LITTLE, J. D. C.: The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. In: *Operations Research* 14 (1966), Nr. 4, S. 568–594

- [113] LONG, G. : Start-up delays of queued vehicles. In: *Transportation Research Record* 1934 (2005), S. 125–131
- [114] MANDELBAUM, A. ; STOLYAR, A. L.: Scheduling Flexible Servers with Convex Delay Costs: Heavy-Traffic Optimality of the Generalized $c\mu$ -Rule. In: *Operations Research* 52 (2004), Nr. 6, S. 836–855
- [115] MARSH, B. W.: Traffic Control. In: *Annals of the American Academy of Political and Social Science* 133 (1927), S. 90–113
- [116] MATHIAS, P. : *Statische und dynamische Verkehrsumlegung mit Rekurrenten Neuronalen Netzen*. Herzogenrath : Shaker Verlag, 1999
- [117] MAURO, V. : Road Network Control. In: PAPAGEORGIOU, M. (Hrsg.): *Concise Encyclopedia of Traffic & Transportation Systems*. Pergamon, 1991 (Advances in Systems Control and Information Engineering), S. 361–366
- [118] MAURO, V. ; DI TARANTO, C. : UTOPIA. In: PERRIN, J.-P. (Hrsg.): *Control, Computers, Communications in Transportation*. Pergamon, 1990, S. 245–252
- [119] McDONALD, M. ; HOUNSELL, N. B.: Road Traffic Control: TRANSYT and SCOOT. In: PAPAGEORGIOU, M. (Hrsg.): *Concise Encyclopedia of Traffic & Transportation Systems*. Pergamon, 1991 (Advances in Systems Control and Information Engineering), S. 400–408
- [120] McNAUGHTON, R. : Scheduling with Deadlines and Loss Functions. In: *Management Science* 6 (1959), Nr. 1, S. 1–12
- [121] MEDHI, J. : *Stochastic Models in Queueing Theory*. Academic Press, 2002
- [122] MENEGUZZER, C. : Review of models combining traffic assignment and signal control. In: *Journal of Transportation Engineering* 123 (1997), Nr. 2, S. 148–155
- [123] MICHALOPOULOS, P. G. ; PISHARODY, V. : Platoon Dynamics on Signal Controlled Arterials. In: *Transportation Science* 14 (1980), Nr. 4, S. 365–396
- [124] MILLER, A. J.: Settings for Fixed-Cycle Traffic Signals. In: *Operational Research Quarterly* 14 (1963), Nr. 4, S. 373–386
- [125] MIRCHANDANI, P. ; HEAD, L. : A real-time traffic signal control system: architecture, algorithms, and analysis. In: *Transportation Research Part C* 9 (2001), Nr. 6, S. 415–432

-
- [126] MIRCHANDANI, P. ; WANG, F.-Y. : RHODES to Intelligent Transportation Systems. In: *IEEE Intelligent Systems* 10 (2005), Nr. 1, S. 10–15
- [127] MOREIRA, J. ; BISPO, C. : Stability or Stabilizability? Seidmans FCFS example revisited. In: *10th Mediterranean Conference on Control and Automation*, 2002
- [128] MORGAN, G. A.: Traffic Signal. In: *United States Patent and Trademark Office* 01475024 (1923), S. 1–6
- [129] MORGAN, J. T. ; LITTLE, J. D. C.: Synchronizing traffic signals for maximal bandwidth. In: *Journal of the Operations Research Society of America* 12 (1964), Nr. 6, S. 896–912
- [130] MUELLER, E. A.: Aspects of the History of Traffic Signals. In: *IEEE Transactions on Vehicular Technology* 19 (1970), Nr. 1, S. 6–17
- [131] NEWELL, G. F.: A theory of platoon formation in tunnel traffic. In: *Operations Research* 7 (1959), Nr. 5, S. 589–598
- [132] NEWELL, G. F.: *Applications of queueing theory*. Chapman and Hall, 1971
- [133] NEWELL, G. F.: The effect of queues on the traffic assignment to freeways. In: SASAKI, T. (Hrsg.) ; YAMAOKA, T. (Hrsg.): *Proceedings 7th International Symposium on Transportation and Traffic Theory*. Kyoto, Japan : Institute of Systems Science Research, 1977, S. 311–340
- [134] OBERBÜRGERMEISTER DER STADT MÜNSTER: Öffentliche Beschlussvorlage an den Haupt- und Finanzausschuss. 549 (2003), S. 1–5
- [135] OLSEN, T. L.: A Practical Scheduling Method for Multiclass Production Systems with Setups. In: *Management Science* 45 (1999), Nr. 1, S. 116–130
- [136] OLSSON, K. : Method and means for network control of traffic. In: *United States Patent and Trademark Office* US 6496773 B1 (2002), S. 1–25
- [137] VAN OYEN, M. P. ; PANDELIS, D. G. ; TENEKETZIS, D. : Optimality of index policies for stochastic scheduling with switching penalties. In: *Journal of Applied Probability* 29 (1992), Nr. 4, S. 957–966
- [138] PACEY, G. M.: The Progress of a Bunch of Vehicles released from a Traffic Signal / Road Research Laboratory. London, 1956 (RN/2665/GMP). – Forschungsbericht
- [139] PANWALKAR, S. S. ; ISKANDER, W. : A Survey of Scheduling Rules. In: *Operations Research* 25 (1977), Nr. 1, S. 45–61

- [140] PAPAGEORGIOU, M. (Hrsg.): *Concise Encyclopedia of Traffic & Transportation Systems*. Pergamon, 1991 (Advances in Systems Control and Information Engineering)
- [141] PAPAGEORGIOU, M. ; DIAKAKI, C. ; DINOPOULOU, V. ; WANG, A. K. Y.: Review of Road Traffic Control Strategies. In: *Proceedings of the IEEE* 91 (2003), Nr. 12, S. 2043–2067
- [142] PERKINS, J. ; KUMAR, P. R.: Stable, distributed, real-time scheduling of flexible manufacturing/assembly/diassembly systems. In: *IEEE Transactions on Automatic Control* 34 (1989), Nr. 2, S. 139–148
- [143] PERRY, D. ; STADJE, W. ; ZACKS, S. : Contributions to the theory of first-exit times of some compound processes in queueing theory. In: *Queueing Systems* 33 (1999), Nr. 4, S. 369–379
- [144] PIKOVSKY, A. ; KURTHS, J. ; ROSENBLUM, M. : *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2003
- [145] PONTRYAGIN, L. S. ; BOLTYANSKII, V. G. ; GAMKRELIDZE, R. V. ; MISHCHENKO, E. F.: *The Mathematical Theory of Optimal Processes*. Interscience Publishers, 1962
- [146] PORCHE, I. ; SAMPATH, M. ; SENGUPTA, R. ; CHEN, Y.-L. ; LAFORTUNE, S. : A decentralized scheme for real-time optimization of traffic signals. In: *Proceedings of the 1996 IEEE International Conference on Control Applications*, 1996, S. 582–589
- [147] PORTER, B. E. ; ENGLAND, K. J.: Predicting Red-Light Running Behavior: A Traffic Safety Study in Three Urban Settings. In: *Journal of Safety Research* 31 (2000), Nr. 1, S. 1–8
- [148] RAM, R. ; VISWANADHAM, N. : GSPN models for versatile multi-machine workcenters with finite buffers. In: *International Conference on Systems, Man and Cybernetics* 2 (1993), October, S. 186–191
- [149] RAMSAY, E. D. ; BUNKER, J. M. ; TROUTBECK, R. J.: Signalised Intersection Capacity Reduction of Trucks. In: MAO, B. (Hrsg.) ; TIAN, Z. (Hrsg.) ; SUN, Q. (Hrsg.): *Proceedings Fourth International Conference on Traffic and Transportation Studies*, Dalian, 2004, S. 793–802
- [150] REIMAN, M. I. ; WEIN, L. M.: Dynamic scheduling of a two-class queue with setups. In: *Operations Research* 35 (1998), Nr. 4, S. 532–547

-
- [151] REINSCHKE, K. : Graph-theoretic characterization of fixed modes in centralized and decentralized control. In: *International Journal of Control* 39 (1984), Nr. 4, S. 715–730
- [152] RETTING, R. A. ; GREENE, M. A.: Influence of Traffic Signal Timing on Red-Light Running and Potential Vehicle Conflicts at Urban Intersections. In: *Transportation Research Record* 1595 (1997), S. 1–7
- [153] RICHARDS, P. I.: Shock Waves on the Highway. In: *Operations Research* 4 (1956), Nr. 1, S. 42–51
- [154] ROBERTSON, D. I.: TRANSYT: A Traffic Network Study Tool / Transport and Road Research Laboratory. Crowthorne, Berkshire, 1969 (LR 253). – Forschungsbericht
- [155] ROBERTSON, D. I.: TRANSYT method for area traffic control. In: *Traffic Engineering & Control* 10 (1969), S. 276–281
- [156] ROBERTSON, D. I. ; BRETHERTON, R. D.: Optimizing networks of traffic signals in real time - the SCOOTmethod. In: *IEEE Transactions on Vehicular Technology* 40 (1991), Nr. 1, S. 11–15
- [157] ROTHKOPF, M. H. ; SMITH, S. A.: There are no undiscovered priority index sequencing rules for minimizing total delay costs. In: *Operations Research* 32 (1984), Nr. 2, S. 451–456
- [158] ROUPHAIL, N. M. ; TARKO, A. ; LI, C. : Traffic flow at signalized intersections. In: GARTNER, N. (Hrsg.) ; MAHMASSANI, H. (Hrsg.) ; MESSER, C. J. (Hrsg.) ; LIEU, H. (Hrsg.) ; CUNARD, R. (Hrsg.) ; RATHI, A. K. (Hrsg.): *Traffic Flow Theory: A State-of-the-Art Report*. Transportation Research Board, 1997, S. 8.1–8.32
- [159] RUBENSSON, M. ; LENNARTSON, B. : Stability of limit cycles in hybrid systems using discrete-time Lyapunov techniques. In: *IEEE Conference on Decision and Control, 2000. Proceedings of the 39th, 2003*
- [160] SAVKIN, A. V.: Controllability of complex switched server queueing networks modelled as hybrid dynamical systems. In: *Proceedings of the 37th IEEE Conference on Decision and Control* Bd. 4, 1998, S. 4289–4293
- [161] SAVKIN, A. V.: Optimal distributed real-time scheduling of flexible manufacturing networks modeled as hybrid dynamical systems. In: *Proceedings of the 42nd IEEE Conference on Decision and Control* Bd. 5, 2003, S. 5468–5471
- [162] SAVKIN, A. V. ; EVANS, R. J.: *Hybrid Dynamical Systems*. Bosten : Birkhäuser, 2002

- [163] SCHILLER, C. ; VRTIC, M. ; LOHSE, D. ; TEICHERT, H. ; FRÖHLICH, P. ; SCHÜSSLER, N. ; AXHAUSEN, K. W.: Ein zweiseitig gekoppeltes Modell zur simultanen Berechnung der Verkehrserzeugung, Verkehrsverteilung und Verkehrsaufteilung: Theoretischer Hintergrund und praktische Anwendung für ein nationales Modell der Schweiz. In: *Verkehrsforschung Online* 3 (2006)
- [164] SCHNABEL, W. ; LÄTZSCH, L. ; BRÜCKNER, N. : Kraftstoffverbrauch im Straßenverkehr. In: *Straßenverkehrstechnik* 12 (2001), S. 584–595
- [165] SCHNABEL, W. : Probleme, Effizienz und Erfahrungen beim Betrieb von Grünen Wellen. In: *Straßenverkehrstechnik* 4 (2006), S. 193–199
- [166] SCHNABEL, W. ; LOHSE, D. : *Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung, Band 1 Straßenverkehrstechnik*. Verlag Bauwesen, 1997
- [167] SCHNABEL, W. ; LOHSE, D. : *Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung, Band 2 Verkehrsplanung*. Verlag Bauwesen, 1997
- [168] SCHÖNEMANN, R. : *Abschätzung der Leistungsfähigkeit herkömmlicher Lichtsignalsteuerungen*, Technische Universität Dresden, Seminararbeit, 2006
- [169] SCHRANK, D. ; LOMAX, T. : *The 2005 urban mobility report*. Texas Transportation Institute, 2005
- [170] SCHULZ, M. : *Control Theory in Physics and Other Fields of Science: Concepts, Tools, and Applications*. Springer, 2006
- [171] DE SCHUTTER, B. : Optimal control of a class of linear hybrid systems with saturation. In: *SIAM Journal on Control and Optimization* 39 (2000), Nr. 3, S. 835–851
- [172] DE SCHUTTER, B. : Optimizing acyclic traffic signal switching sequences through an extended linear complementarity problem formulation. In: *European Journal of Operational Research* 139 (2002), Nr. 2, S. 400–415
- [173] DE SCHUTTER, B. ; DE MOOR, B. : Optimal traffic light control for a single intersection. In: *European Journal of Control* 4 (1998), Nr. 3, S. 260–276
- [174] DE SCHUTTER, B. ; DE MOOR, B. : The extended linear complementarity problem and the modeling and analysis of hybrid systems. In: ANTSAKLIS, P. (Hrsg.) ; KOHN, W. (Hrsg.) ; LEMMON, M. (Hrsg.) ; NERODE, A. (Hrsg.) ; SASTRY, S. (Hrsg.): *Hybrid Systems V* Bd. 1567. Berlin Heidelberg : Springer-Verlag, 1999, S. 70–85

-
- [175] SCHWARZ, H. R. ; KÖCKLER, N. : *Numerische Mathematik*. Teubner, 2004
- [176] SEN, S. ; HEAD, K. L.: Controlled optimization of phases at an intersection. In: *Transportation Science* 31 (1997), S. 5–17
- [177] SHIMASAKI, H. : Method of signal cycle calculation which is based on the saturation flow rate. In: *International Conference on Intelligent Transportation Systems*, 1999, S. 891–896
- [178] SICK AG: *Lasermesssystem LMS221*
- [179] SIRRINE, E. E.: Street traffic system. In: *United States Patent and Trademark Office* 00976939 (1910), S. 1–6
- [180] SMITH, W. E.: Various optimizers for single-stage production. In: *Naval Research Logistics Quarterly* 3 (1956), S. 59–66
- [181] STEIERWALD, G. ; KÜNNE, H.-D. ; VOGT, W. : *Stadtverkehrsplanung. Grundlagen, Methoden, Ziele*. Berlin : Springer, 2005
- [182] STROBEL, H. : *Computer controlled urban transportation - a survey of concepts, methods and international experiences*. John Wiley & Sons, 1982
- [183] SZENTPÉTERY, V. : Die perfekte grüne Welle. In: *Technology Review* 10 (2006), S. 79237
- [184] TANNER, J. C.: A theoretical analysis of delays at an uncontrolled intersection. In: *Biometrika* 49 (1962), Nr. 1,2, S. 163–170
- [185] TOOMEY, C. G. ; FRIEDRICH, B. ; CLARK, M. : BALANCE - a European field trial. In: *9th International Conference on Road Transport Information and Control*, 1998, S. 95–99
- [186] Transportation Association of Canada: *Traffic Signal Guidelines for Bicycles*. 2004
- [187] TROUTBECK, R. J. ; BRILON, W. : Unsignalized intersection theory. In: GARTNER, N. (Hrsg.) ; MAHMASSANI, H. (Hrsg.) ; MESSER, C. J. (Hrsg.) ; LIEU, H. (Hrsg.) ; CUNARD, R. (Hrsg.) ; RATHI, A. K. (Hrsg.): *Traffic Flow Theory: A State-of-the-Art Report*. Transportation Research Board, 1997, S. 8.1–8.47
- [188] VOIGT, W. ; KÜHNELT, W.-D. ; PACKROFF, K. : *Organisation des städtischen Straßenverkehrs*. Berlin : Transpress VEB Verlag für Verkehrswesen, 1982

- [189] VOLK, W. : *Dresden. Historische Straßen und Plätze heute*. Berlin : VEB Verlag für Bauwesen, 1974
- [190] WALKOW, K. : *Empirische Querschnittsdaten von Fußgängerströmen: Erfassung und Analyse, Verwendung in der Verkehrsmodellierung und ökonomische Bedeutung*, Technische Universität Dresden, Diplomarbeit, März 2006
- [191] WALRAND, J. : *An Introduction to Queuing Networks*. Prentice Hall, 1988
- [192] WEBSTER, F. V.: Traffic Signal Settings. In: *Road Research Technical Paper* 39 (1958), S. 1–44
- [193] WHITHAM, B. G.: *Linear and Nonlinear Waves*. Wiley-Interscience, 2006
- [194] WILLUMSEN, L. G.: Travel Networks. In: HENSHER, D. A. (Hrsg.) ; BUTTON, K. J. (Hrsg.): *Handbook of Transport Modelling*. Pergamon, 2002, S. 165–180
- [195] VAN WOENSEL, T. ; VANDAELE, N. : Modelling traffic flows with queuing models: A review. In: *Asia-Pacific Journal of Operational Research* eingereicht (2006)
- [196] WRIGHT, C. : Some characteristics of traffic leaving a signalized intersection. In: *Transportation Science*, 4 (1970), Nr. 4, S. 331–346
- [197] WU, N. : Verteilung der Rückstaulängen an Knotenpunkten ohne Lichtsignalanlagen. In: *Straßenverkehrstechnik* 8 (1997)
- [198] ZELL, A. : *Simulation Neuronaler Netze*. Oldenbourg, 1994
- [199] ZHENG, J. F. ; GAO, Z. Y. ; ZHAO, X. M.: Modeling cascading failures in congested traffic and transportation networks. In: *Europhysics Letters* in print (2007)
- [200] ZIMMERMANN, W. ; STACHE, U. : *Operations Research. Quantitative Methoden zur Entscheidungsvorbereitung*. Oldenbourg, 2001

Stichwortverzeichnis

- Abbiegeanteile, 43
- Abfertigungskapazität, 108
- Abflussgesetz, 6, 45
- abgefertigte Fahrzeuganzahl, 44
- Adaptivitätsparameter, 100, 149
- Anfahrverluste, 26
- Ankunftsrate, erwartete, 40
- Arterie, 129
- Ausflussrate, 44
 - maximale, 45
 - Zustandsabhängigkeit, 44
- Auslastung, 24, 28, 120

- Bündelungseffekt, 29, 85
- BALANCE, 19
- Bang-Bang-Steuerung, 99
- Bedieneffizienz, 91
 - prognostizierte, 93
- Bedienperiode, 166
 - gewünschte, 94–97, 124
 - maximale, 100–104
- Bedienvorgang, 50
 - Abbruch, 62, 80
 - Prognose, 50–63
- Binomialverteilung, 30
- Bisektion, 53
- Box-Whisker-Plot, 127
- Branch-and-Bound-Verfahren, 16

- charakteristische Geschwindigkeit, 147
- Clearing Policies, 21
- COP-Algorithmus, 19
- CRONOS, 17

- Dezentralisierung, 3
- Dominanzverfahren, 36, 119
- Durchlassfähigkeit, 8, 29, 121

- Effizienzkriterium, 92
- Einsatzfahrzeuge, 145
- Erlang-Verteilung, 30, 105, 166
 - verallgemeinerte, 105, 167
- erwartete Ankünfte
 - Definition, 40
 - Prognose, 41
- exhaustive, 21, 70, 73
- Exponentialverteilung, 30, 127, 166

- Fahrzeugklassen, 45, 144
- Fahrzeugkolonne, 4, 34, 50, 55, 56, 65, 109
 - Entstehung und Zerfall, 134
- Festzeitsteuerung, 24, 119
- FIFO, 96
- Floating-Car-Daten, 41
- Freigabe, 26, 44, 45
- Freigabeband, 15, 132
- Freigabezeit, 26
 - erforderliche, 51
 - Freigabezeitanteil, 27, 92
 - Mindestfreigabezeit, 27
- Freigabezeitfenster, 33, 115, 125, 157
- Freigabezeitverlängerung, 8, 51, 56, 77, 109, 111, 159
- Fußgängerströme, 145

- genetische Algorithmen, 36
- geometrische Verteilung, 42
- Gesamtwartezeiten, 120

- Grüne Welle, 15, 37, 85, 132
- Grenzyklus, 24
 - des kombinierten Verfahrens, 113
 - des Optimierungsverfahrens, 86, 159
 - des Stabilisierungsverfahrens, 98, 105
- Haltelinie, 7, 26, 101, 144
- Haltevorgänge, 49
- Henry Ford Museum, 12
- hybrides dynamisches System, 8, 46, 55, 64, 65
- Index-Policies, 23
- Instabilität, 4
 - der Clear-Largest-Buffer-Policy, 4, 153
 - des Optimierungsverfahrens, 87
- Isolierter Knotenpunkt, 119
- Kalman-Filter, 43
- kombiniertes Verfahren, 110–116, 121
 - Programmablaufgraph, 111
 - Schaltverhalten, 113
- Kommunikation, 5, 150
- komplexes Netzwerk, 137
- Koordinierung, 32, 35, 37
- Koordinierungsprinzip
 - der Grünen Welle, 134
 - des kombiniertes Verfahrens, 132
- Kraftstoffverbrauch, 49, 125, 141
- kritische Fahrzeuganzahl, 94
- kritischer Verkehrsstrom, 24, 28, 147
- Kumar-Seidman-Netzwerk, 4, 87, 115, 153
- Längste-Schlange-Zuerst-Regel, 4, 21, 153
- lineares Komplementaritätsproblem, 21
- lineares Optimierungsproblem, 97, 98
- lokale Recheneinheit, 7, 151
- Lyapunov-Funktion, 23
- Markov-Prozess, 42
- Masche, 3, 21
- Matlab, 118
- MAXBAND, 15
- μ c-Regel, 69
- Mindestfreigabezeit, 28, 92
- MOTION, 19
- Nachfragespitze, 137
- neuronale Netze, 43
- Newton-Raphsonsche Methode, 53
- Offline-Optimierung, 16
- OPAC, 17
- Operations Research, 98
- ÖPNV-Priorisierung, 17, 20, 144
- Optimalität
 - des kombinierten Verfahrens, 113
 - des Optimierungsverfahrens, 83
 - Priorisierungsregeln, 68
 - umlaufbasierter Verfahren, 38
- optimierendes Regime, 112
- Optimierungsverfahren, 67–89, 120
 - Komplexität, 84
 - Optimalität, 83
 - Schaltverhalten, 84
 - Spezifikation, 83
- Phasenbedienung, 6, 25, 147
- Phaseneinteilung, 119
- Point-Queue-Models, 45
- Poisson-Verteilung, 30, 31, 105, 166
- Pontryagins Maximumprinzip, 99
- präemptiv, 73
- PREDICT, 19
- Priorisierung
 - dynamische, 73–82
 - statische, 69–72
- Prioritätsbasierte Verfahren, 23, 68–72
- Prioritätsindex, 23, 68, 83

- PRODYN, 17
 prognosebasierten Verfahren, 17
 Prognosehorizont, 4, 17, 40, 43, 89
 Prognoseverfahren, 50–63

 Rückkopplung, 3, 21, 35, 156
 Rückstaulänge, 146
 Reaktionszeit, 26
 Reisezeitenverteilung, 42
 RHODES, 18
 Richtungsströme, 129
 Rolling Horizon Dynamic Programming, 17
 Routenwahlverhalten, 43, 147

 Sättigungsverkehrsstärke, 26, 40, 118
 Schaltmuster, 124
 Schwellenwert
 kritischer, 94, 100
 zeitabhängiger, 102
 SCOOT, 16
 Signalfügel, 12
 Signalgeber, 6, 151
 Simplex-Algorithmus, 98
 Simulation
 Arterie, 129
 isolierter Knotenpunkt, 119
 komplexes Netzwerk, 137
 Simulationsergebnisse, 117–142
 Sitraffic Concert, 19
 stabilisierendes Regime, 112
 Eingriff, 111, 124, 141, 148
 Stabilisierungsprinzip, 91–93
 Stabilisierungsverfahren, 90–109, 121
 Komplexität, 109
 Schaltregel, 92, 96
 Schaltverhalten, 104–108
 Spezifikation, 96–99
 Stabilität, 21, 23, 109
 Stabilitätsbedingung
 hinreichende, 92, 94, 95
 notwendige, 24

 Stage-Skipping, 17
 Stand der Forschung, 11–38
 staufreie Zone, 147
 Staurationbegrenzung, 145
 Stauwellen, 146
 Steuerbarkeit, 24
 stochastische Einflüsse, 29, 127
 stochastische Warteschlangenmodel-
 le, 29
 Strafterm, 80
 Streuung, 128
 Synchronisation, 106

 Tagesganglinie, 137
 TRANSYT, 16, 42
 Travolution, 3, 18, 42
 TUC, 20

 Überfahrzeiten, 26
 Umlaufbasierte Steuerungen, 24–36
 Umlaufzeit, 28
 Mindestumlauftzeit, 28, 30
 wartezeitoptimale, 31
 Umschaltzeit, *siehe* Zwischenzeit
 Umweltbelastung, 2
 UTOPIA, 18

 Varianz, 128
 Variationskoeffizient, 105
 verkehrsabhängige Verfahren, 3, 18,
 149
 verkehrsadaptive Verfahren, 2, 16
 Verkehrsdetektor, 7
 Verkehrsflusserfassung, 40–45
 Verkehrsturm, 14
 Verkehrszentrale, 2, 5, 150
 Versatzzeitenanpassung, 32, 108, 115,
 119
 Verträglichkeitsgraph, 25
 virtueller Verkehrsstrom, 100–102,
 145, 148
 Vorrangschaltung, 145

- Warteschlangenlänge, 45, 72, 120
- Warteschlangenmodell, 45–50
- Warteschlangennetzwerk, 20, 23
- Wartezeiten, 46
 - Gesamtwartezeiten, 46
 - Gewichtung, 144
 - individuelle, 47
 - prognostizierte, 58
 - Entwicklungsgleichung, 62, 63
- Wartezeitenkurve, 31, 33
- Wirkungsbereich, 57
- Wunschtrajektorie, 22

- Zeitlücken, 35, 81, 108
 - Brutto-, 30, 127
 - Netto-, 81
- Zeitscheibe, 32
- Zustandsbeobachtung, 65
- Zwischenzeit, 7, 25, 56, 78
 - verbleibende, 51, 56, 73