Het verkeerskundig
laboratorium
voor studenten

ITS EDU LAB

Route Guidance and Signal Control Based on the Back-Pressure Algorithm

experiments on integrated control

Joost van Kampen

19 February 2015

Rijkswaterstaat
Ministerie van Infrastructuur en Milieu

TU Delft

Colophon

Author	J. J. L. van Kampen
Graduation Committee	Prof. dr. ir. S. P. Hoogendoorn Chairman TU Delft, CiTG, Transport & Planning Dr. ir. H. Taale Daily supervisor Rijkswaterstaat, dienst WVL TU Delft, CiTG, Transport & Planning Prof. H. L. Vu External member Swinburne University of Technology Ir. P. B. L. Wiggenraad Graduation coordinator TU Delft, CiTG, Transport & Planning

Published by	ITS Edulab, Delft
Date	February 19th, 2015
Status	Final report
Version number	1.0
Information	Henk Taale
Telephone	+31 88 798 2498

ITS Edulab is a cooperation between Rijkswaterstaat Centre for Transport and Navigation and Delft University of Technology

TU Delft

Rijkswaterstaat
Ministerie van Infrastructuur en Milieu

Preface

"No star is ever lost we once have seen,
We always may be what we might have been."

—Adelaide Anne Procter

This thesis is the graduation work that concludes my Master of Science programme in Civil Engineering at the TU Delft, following the track Transport & Planning. I was offered the chance to work on this research at the ITS Edulab, a traffic and transport laboratory for students and partnership between TU Delft and Rijkswaterstaat.

Since I couldn't have done it without them, I would like to mention and thank the many people who supported me along the way.

First, I would like to express my gratitude to the members of my graduation committee. I would like to thank my daily supervisor and ITS Edulab coordinator Henk Taale, for giving me the opportunity to join the ITS Edulab, and especially for your guidance at Rijkswaterstaat. Your questions and feedback during our meetings often gave me the right push to continue my research. Special thanks go out to chairman professor Serge Hoogendoorn, you helped me with new ideas to get the project back on track. The meetings with you and Henk made me feel at ease and more confident. I also would like to thank graduation coordinator Paul Wiggenraad in particular for his help to (re)start my graduation process. Lastly, I would like to thank professor Hai Vu from Swinburne University of Technology, for joining the committee as an external member, and for his thorough review of my preliminary report.

Furthermore, I would like to thank the people at Rijkswaterstaat, my fellow students at the ITS Edulab, and in particular my room mates Emiel and Mark. I enjoyed your company and it was nice to share our study experiences and some random chat over a cup of coffee.

Last but not least, I would warmly thank my friends who kept supporting me all those years as well as my family. Mum, sis and brother, thanks for encouraging me to bring this long journey to a good end. I dedicate this thesis to them, and to the memory of my father.

Joost van Kampen

Delft, February 2015

Summary

Bad traffic conditionscan be improved by regulating traffic demand and network capacity. Dynamic Traffic Management (DTM) systems allocate temporal and spatial utilization of infrastructures and vehicle fleets by means of dynamic signals. By timely response to changing traffic conditions DTM goals, in terms of effective, safe and reliable use of the infrastructure, can be met. Two important trends in the field of DTM are: a shift from local control measures towards network-wide control; a shift from collective traffic information towards individual advice.

Individual advice can be transmitted via in-car technology, however, up until now in-car systems are typically used to improve the (route) choice of the individual road user, whereas DTM aims at improving the network performance as a whole. Network-wide traffic control is complicated by the dynamic nature of traffic itself and the dynamic effects of control measures, especially if the network and DTM structure is complex.

Route guidance can make a significant contribution to network-wide DTM. Using in-car navigation to transmit individual directions, it is expected to improve the network performance (better use of the available capacity, higher throughput and stability, and less spill-back), and reduce the travel time for the individual road user as well. Finding the right route guidance configuration is a complex task, that should take potential unfavourable effects and coordination into account. Moreover, integrating traffic signal control at intersections with route guidance is a logical step.

Considering the complex nature of the problem and the need for (real-time) responsive control actions, a simplified traffic control model is suggested. The concept of back-pressure control has recently been applied to the problem of controlling a network of signalized intersections. The development of a route guidance algorithm based on back-pressure control, integrated with a back-pressure model for signalized intersections, is to be considered. The main research problem in this thesis is to determine the feasibility and potential benefits of such a system.

The main objective of this thesis is to develop a framework that integrates route guidance and signal control based on the back-pressure principle, and determine its feasibility and potential benefit.

Conditions that further specify the objective include a focus on network-wide control, on optimizing network conditions (and secondarily find a trade-off with road user benefits), on real-time control, and on a generic approach.

The literature review starts with a brief overview route guidance. Applying route guidance has advantages and pitfalls. Route guidance can be modelled as a way to manipulate route choice. Often route guidance is used to minimize the travel time of the individual road user, using route guidance to improve the conditions for the network and road users as a whole isn't often done in operational route guidance.

The back-pressure control methodology originated from communication networks, where it is applied to the problem of delivering data packets via a network of nodes and servers. The back-pressure algorithm assigns the servers to be activated and the data they should transfer, based on server ca-

pacity and queue differences (queues for specific data groups). The main strengths of back-pressure control are maximization of throughput and network stability. The basic back-pressure algorithm can been extended to cope with finite queues and delay.

On the application of traffic signal control back-pressure control has been used in literature. The concept aims to activate the signalling phase that has the highest total weight, a summation of the weights of the allowed traffic streams. The weight is the product of pressure and service rate (saturation flow). The pressure of a traffic stream is the difference between the queues at the incoming link and outgoing links (weighted by proportion). There are several variants to this basic model.

In communication networks the use of routes is a result of the algorithm, whereas for traffic signal control the route choice is used as an input from a separate process. The back-pressure concept can however also be used as a method for route guidance. Instead of determining the right phase to be activated, the task is to determine the ratio to direct traffic to following links. Route pressure values can be formulated to expresses how filled up the route is. Route pressure and service rate are less straightforward to define than for the intersection control. A choice model is to be used to determine the ratio of routes, based on the product of route pressure and service rate.

The chosen control approach is a traffic control loop in which the traffic process is monitored, controllers for traffic signal control and route guidance calculate new settings which are implemented to the traffic process by means of the traffic signalling system and in-car navigation system. The traffic signal controller can be fed by route guidance information to improve the estimate turn probabilities. The model is a reactive feedback control type.

The research approach includes a list of design choices and limitations for traffic signal control and route guidance, as well as requirements related to the experiment to be conducted and the simulation tools.

A general back-pressure algorithm for traffic signal control has been designed, with two main variants: one with a fixed cycle time that assigns phase durations for one or more cycles, and one with short time slots that repeatedly activates the dominant phase. Pressure values are based on representative link densities, are normalized to the jam density and can be extended by a power function that increases the relative weight of higher pressures. The necessary turning probabilities are based on measurements or on route guidance settings, which integrates traffic signal control and route guidance.

For route guidance a general algorithm has been proposed as well. As a first attempt of service rate, the link capacity of the first outgoing link of the route is used. Furthermore a number of variants for the algorithm have been considered. The first variant is simple but myopic and observes only the first link of each route, which also limits the number of routes that need to be considered. An important limitation of this variant is that congestion on links further downstream is not taken into account. Other variants observe complete routes (from the en-route position). A path size logit choice model is used to determine the proportions for each route. A pressure value that represents the whole route needs to be determined. This can be done by taking the (weighted) average of the link pressures, or by a method that takes outliers specifically into account. Travel time as a measure of user satisfaction, allowing road users to use the shortest routes, can be incorporated into the method. The overall utility function of a route can be written as: $BP_r = \alpha_1 P_{\text{route},r} + \alpha_2 P_{\text{firstlink},r} + \alpha_3 P_{\text{user},r}$. The three terms represent the (total) route pressure, the pressure of the first link, and the travel time value.

The simulation environment is based on the macroscopic simulation model DSMART. It uses a

kinematic wave model, based on the fundamental diagram, space and time discretization. Aggregated flows carry traffic along the links and System Class Dynamics (SCD) ensure the separation of changes in traffic composition and correct handling of traffic dynamics at the nodes.

Periodically, traffic controllers are called to update the activated control strategies, based on the proposed traffic control and route guidance algorithms. Traffic signal control objects directly influence the traffic flow and route guidance is enforced by manipulating turn ratios at intersections and diverges.

The simulation experiments are divided into three parts: case 1 examines traffic signal control, case 2 focuses on route guidance, and case 3 combines traffic signal control and route guidance.

For traffic signal control, back-pressure control is a good way to generate high throughput at the intersections, while keeping the queues evenly distributed and within boundaries. Back-pressure signal control based on time slots is more effective than the cycle time based version. Some aspects of the algorithm and its practical use require special attention.

For route guidance an optimal use of back-pressure has not yet been found, as the results are on a most part lacking compared to the standard route choice model. Two main aspects are the definition of a representative route pressure, and the capacity of routes related to the assignment of (destination specific) traffic to each route. The performance is expected to be better if the pressure were (partly) based on the critical links, instead of on average density. Yet the basic algorithm that was formulated works to some extent, and so far the simulated effects can be understood and are up for improvement. It was also found that a 'pressure' function that includes not only route densities, but also travel time can give good results.

In conclusion, this thesis demonstrates the possibility to create a methodology, based on back-pressure control, that integrates traffic signal control and route guidance. Traffic signal control based on back-pressure control performs well in the simulations, especially the variant with time slots. Throughput is high and queues remain within reasonable boundaries. It is difficult to fully integrate traffic signal control and route guidance, contrary to the straightforward algorithm in the field of wireless communication networks. A modest step of integration is to use route guidance settings to determine the necessary turning probabilities at the intersection. Using back-pressure for route guidance requires (artificial) design choices. The challenge is to define a representative function of route pressure or utility, and to combine this with a service rate value, in order to obtain a high throughput and stability with minimum delays. Several ideas have been presented. The average density has turned out not a good measure for route pressure. It is possible to combine factors of pressure based on density and travel time, in order to use the shortest routes in case of low traffic, and shift to the routes with open capacity if needed.

Future research is recommended on the fine-tuning of the back-pressure traffic signal model, and on further integration and coordination of the control strategies. On the part of route guidance, especially finding representative route pressure values and making the model applicable of larger networks (distributed approach) require more research. The control approaches should further be simulated and tested with a microscopic simulation model, on more networks and scenarios. Necessary developments for a practical use include the availability of in-car systems and traffic state estimation.

Samenvatting

Gebrekkige verkeersomstandigheden kunnen worden verbeterd door het reguleren van verkeersvraag en netwerkcapaciteit. Systemen voor dynamisch verkeersmanagement (DVM) gebruiken dynamische verkeerssignalen om tijd- en ruimteafhankelijk gebruik van infrastructuur en groepen voertuigen toe te wijzen. Door tijdige respons op veranderende verkeersomstandigheden kunnen DVM-doelstellingen, in termen van effectief, veilig en betrouwbaar gebruik van de infrastructuur, worden gehaald. Twee belangrijke trends op het gebied van DVM zijn: een verschuiving van lokale regelingen naar netwerkbrede verkeersregelingen; een verschuiving van collectieve verkeersinformatie naar individueel reisadvies.

Individueel reisadvies kan worden overgebracht door middel van in-car technologie. Tot nu toe zijn in-car systemen echter vooral gericht op het verbeteren van routekeuze voor de individuele weggebruiker, terwijl DVM het verbeteren van de netwerkprestatie as geheel nastreeft. Het netwerkbreed regelen van verkeer is complex door de dynamische aard van verkeersstromen en de dynamische effecten van verkeersmaatregelen, vooral als het netwerk en het DVM-systeem complex zijn.

Route geleiding kan een significante bijdrage leveren aan netwerkbreed DVM. Het gebruik van in-car navigatie om individuele aanwijzingen over te brengen verbetert naar verwachting de netwerkprestatie (betere benutting van de beschikbare capaciteit, hogere doorvoer en stabiliteit en minder fileterugslag), en het kan tevens de reistijd voor de individuele weggebruiker verminderen. Het vinden van de juiste configuratie van routegeleiding is een complexe taak, waarbij met potentiële nadelige effecten en met coördinatie rekening gehouden moet worden. Bovendien is daarbij het integreren van verkeerslichtregelingen en routegeleiding een logische stap.

De complexe aard van het probleem en de behoefte aan in real time reagerende regelingen in acht nemend, wordt een simpel verkeersregelmodel geopperd. Het concept 'back-pressure' regeling is recentelijk toegepast op het probleem van een netwerk met geregelde kruispunten. Er wordt voorgesteld een routegeleidingalgoritme dat op 'back-pressure' gebaseerd is te ontwikkelen en te integreren met een back-pressure model voor verkeersregelingen. De belangrijkste onderzoeksopgave in deze thesis is om de haalbaarheid en potentiële voordelen van een dergelijk systeem te bepalen.

De doelstelling van deze thesis is het ontwikkelen van een raamwerk dat routegeleiding en kruispuntregelingen integreert, gebaseerd op het back-pressure principe, en het bepalen van de haalbaarheid en potentiële voordelen ervan.

Voorwaarden die de doelstelling verder specificeren omvatten een focus op netwerkbrede regelingen, op het optimaliseren van de netwerktoestand (en secundair een afweging met voordeel voor de individuele weggebruiker), op een real-time regeling en op een generieke aanpak.

De literatuurstudie begint met een kort overzicht van routegeleiding. Het toepassen van routegeleiding biedt voordelen en gevaren. Routegeleiding kan worden gemodelleerd als een manier om (vrije) routekeuze te beïnvloeden. Vaak wordt routegeleiding gebruikt om de reistijd van een individuele weggebruiker te minimaliseren, het gebruik van routegeleiding om de toestand van het netwerk

en de weggebruikers als geheel te verbeteren wordt komt niet vaak voor in oprationele toepassingen van routegeleiding.

De methode voor regelingen met back-pressure komt voort uit de wereld van communicatienetwerken, waar het wordt toegepast op het probleem van datatransport door een netwerk van knopen en servers. Het back-pressure algoritme wijst te activeren van servers toe en de data die overgebracht moet worden, gebaseerd op beschikbare servercapaciteit en wachtrijverschillen (wachtrijen voor specifieke groepen data). De kracht van het algoritme ligt in het maximaliseren van datadoorvoer en netwerkstabiliteit. Het basisalgoritme kan worden uitgebreid om met begrensde wachtrijen en omleidingen rekening te kunnen houden.

De back-pressure regeling is in de literatuur reeds toegepast op het regelen van kruispunten. Het concept streeft naar het activeren van de verkeersregelfase met het hoogste gewicht, een optelsom van de gewichten van de in die fase toegestane verkeersbewegingen. Het gewicht is een product van druk en behandelingsintensiteit. De druk van een verkeersbeweging is het verschil tussen de wachtrijen op de inkomende link en de uitgaande links (proportioneel naar richting). Er zijn verschillende varianten voor dit basismodel.

Bij communicatienetwerken is het gebruik van routes een uitkomst van het algoritme, terwijl voor geregelde kruispunten de routekeuze de invoer is vanuit een losstaand proces. Toch kan het concept van back-pressure ook worden toegepast op een methode voor routegeleiding. In plaats van het bepalen van de te activeren regelfase is nu de taak om de verdeling te vinden om verkeer over de voorliggende links te sturen. Waarden voor routedruk kunnen geformuleerd worden om uit te drukken hoe vol de route is. Routedruk en 'behandelingscapaciteit' zijn minder voor de hand liggend te definiëren als voor kruispuntregelingen. Een keuzemodel wordt gebruikt om de verdeling van routegebruik te bepalen, gebaseerd op het product van routedruk en behandelingscapaciteit.

De gekozen aanpak behelst een regellus waarin het lopende verkeersproces wordt gemeten, regelaars voor kruispuntregelingen en voor routegeleiding nieuwe instellingen bepalen, die in werking worden gebracht in het verkeersproces door het verkeersregelsysteem (VRI's) en in-car navigatiesystemen. De kruispuntenregelaar kan gevoed worden met informatie over routegeleiding, om de ingeschatte richtingkeuze op het kruispunt te verbeteren; waarmee kruispuntregeling en routegeleiding deels geïntegreerd worden. Het regelmodel is van het type reactief en feedback.

De onderzoeksaanpak bevat verder een lijst met ontwerpkeuzen en beperkingen voor het regelmodel, alsook aan de voorwaarden die aan het uit te voeren experiment en bijbehorende simulatiemiddelen worden gesteld.

Een algemeen back-pressure algoritme voor kruispuntregelingen is opgesteld, met twee hoofdvarianten: één met een vaste cyclustijd die fasetijden toewijst voor een of meer cycli, en één met korte tijdsloten die steeds opnieuw aan de dominante fase worden toegewezen. Drukniveaus worden gebaseerd op representatieve linkdichtheden, zij worden genormaliseerd naar de vastloopdichtheid (jam density) en kunnen worden uitgebreid met een machtsfunctie die ervoor zorgt dat het relatieve gewicht van hoge drukwaarden zwaarder meeweegt. De benodigde schattingen voor richtingkeuze zijn gebaseerd op metingen of op instellingen van routegeleiding.

Voor routegeleiding is ook een algemeen algoritme opgesteld. Als eerste aanzet voor de behandelingscapaciteit wordt de linkcapaciteit van de eerste uitgaande link gebruikt. Verder is een aantal algoritmevarianten beschouwd. De eerste variant is simpel maar 'bijziend' en beschouwt alleen de eerste link van elke route, waarmee het aantal keuzemogelijkheden erg klein is. Een belangrijke beperking van deze variant is dat congestie verderop op de route niet wordt meegewogen. Andere varianten

beschouwen gehele routes (vanaf de huidige positie). Een 'path size logit' model wordt gebruikt om het aandeel van elke route te bepalen, op basis van een nutsfunctie die in eerste instantie op 'druk' is gebaseerd. Het is nodig om een representatieve drukwaarde voor de gehele route te bepalen. Dit kan door een (gewogen) gemiddelde van de druk op alle links, of door een methode die (maatgevende) uitschieters meeweegt. Reistijd als een maat voor tevredenheid van weggebruikers, hen de snelste route toestaand, kan ook worden ingepast in het model. De algehele nutsfunctie van een route kan worden geschreven als: $BP_r = \alpha_1 P_{\text{route},r} + \alpha_2 P_{\text{firstlink},r} + \alpha_3 P_{\text{user},r}$. De drie termen representeren de drukwaarde voor de gehele route, de druk van de eerste link apart, en de waarde voor reistijd.

De gebruikte simulatieomgeving is gebaseerd op het macroscopische simulatiemodel DSMART. Het gebruikt een kinematisch golfmodel, gebaseerd op het fundamentele diagram en discretisering van tijd en ruimte. Geaggregeerde verkeersstromen leiden het verkeer over de link heen, 'System Class Dynamics' (SCD) verzorgen de strikte scheiding van veranderingen in verkeerscompositie en een correcte afhandeling van verkeersdynamica bij de knopen.

Periodiek worden verkeersregelsystemen, gebaseerd op de voorgestelde algoritmen voor kruispuntregeling en routegeleiding, aangeroepen om de geactiveerde strategie te herzien. Met deze strategie kunnen kruispunten direct het verkeer beïnvloeden of kan routegeleiding worden afgedwongen door het manipuleren van de richtingverdeling bij kruispunten en andere splitsingen.

De simulatieexperimenten zijn verdeeld over drie delen: case 1 onderzoekt kruispuntregelingen, case 2 richt zich op routegeleiding en case 3 combineert de twee.

Voor kruispuntregelingen is de methode met back-pressure een goede manier om een hoge doorvoer op de kruispunten te verkrijgen, terwijl de wachtrijen redelijk verdeeld en binnen grenzen blijven. De variant van de back-pressure regeling met phasetoedeling per tijdslot is iets effectiever dan de variant die gebaseerd is op vaste cycli met variabele faselengten. Sommige punten van het aloritme, waaronder de toepassing in praktijk vereist wel nadere aandacht.

Voor routegeleiding is een optimale variant met back-pressure nog niet gevonden, aangezien de resultaten grotendeels tegenvallen ten opzichte van standaard routekeuze. Twee belangrijke aspecten hieraan zijn de keuze voor representatieve routedruk en de (ongecoördineerde) manier waarop capaciteit wordt toegedeeld aan het verkeer. Naar verwachting is de prestatie al beter als de routedruk (gedeeltelijk) op de toestand van kritieke links wordt gebaseerd, in plaats van op de gemiddelde dichtheid. Toch werkt het opgestelde algoritme tot op zekere hoogte, en de gesimuleerde effecten kunnen worden begrepen en aangewend voor verbeteringen. Een andere uitkomst was dat een nutsfunctie die naast dichtheden ook gebruik maakt van reistijden goede resultaten kan bieden.

Concluderend toont deze thesis de mogelijk aan voor een methode, gebaseerd op back-pressure regeling, die kruispuntregelingen en routegeleiding combineert. Kruispuntregeling op basis van back-pressure presteert goed in simulaties, met neme de variant met tijdsloten. De doorvoer is hoog en wachtrijen blijven binnen de perken. Het is moeilijk om kruispuntregelingen en routegeleiding volledig te integreren, in tegenstelling tot het eenvoudige algoritme op het gebied van communicatienetwerken. Een bescheiden stap is om routegeleiding te gebruiken voor de benodigde geschatte richtingkeuze bij kruispunten. Het gebruik van back-pressure voor routegeleiding vereist (kunstmatige) ontwerpkeuzes. Het is de uitdaging om een representatieve functie te vinden voor routedruk (en in het algemeen het nut van de route), en om dit te combineren met een juiste waarde voor behandelingscapaciteit, teneinde een system met hoge doorvoer, hoge stabiliteit en lage vertragingen te verkrijgen. Diverse ideeën komen in de thesis aan bod. De gemiddelde dichtheid over een route is

echter geen goede maat voor routedruk gebleken. Het is mogelijk om factoren van dichtheid en reistijd te combineren, om zo bij weinig verkeer de snelste routes aan te kunnen bevelen en bij drukte meer gewicht toekennen aan de routes met restcapaciteit.

Toekomstig onderzoek is aanbevolen om de back-pressure kruispuntregeling op punten te kunnen verbeteren en om verdere integratie en coördinatie van maatregelen mogelijk te maken. Op het punt van routegeleiding vereist vooral het vinden van een goede waarde voor routedruk meer onderzoek, en ook het toepasbaar maken van de methode op grotere netwerken (gedistribueerde aanpak). De voorgestelde methoden zouden ook nader getest moeten worden met een preciezer, microscopisch simulatiemodel, en op meer netwerken en scenario's. Benodigde ontwikkelingen voor gebruik in de praktijk omvatten vooral de beschikbaarheid van geschikte in-car systemen (en communicatiemiddelen) en een betrouwbaar systeem voor snelle verkeersmetingen.

Contents

Lists of figures and tables

List of Figures

List of Tables

Chapter 1

Introduction

Only one century ago, the 'auto-mobile' was a modern marvel, but soon the car became a product of mass production and affordable for an increasing number of people. Roads were expanded and motorway networks were built to facilitate inter city transportation. As mobility, together with urban sprawl, kept increasing, the limits to growth became apparent when road networks began reaching their capacities during the second half of the twentieth century. This is manifested by the arise of congestion. Apart from the time (economic) loss, safety risks and environmental damage became problems as well. Dynamic traffic management aims at better utilizing the existing infrastructure, by offering traffic information and applying control measures.

The introduction of this thesis starts off in Section 1.1 with an outline of existing traffic problems, and a survey of dynamic traffic management and its current trends and challenges. Section 1.2 discusses problem aspects to be dealt with and formulates the problem statement. The objective and the research questions depict the purpose of this thesis and are stated in Section 1.3. The methods and steps of of the research is explained in Section 1.4. Section 1.5 outlines the organization of the remaining chapters in this thesis.

1.1 Background

Before getting to the research problem of this thesis in the next section, first a general background of road traffic management is sketched out. This section starts by giving an overview of traffic problems, and then describes possible solutions and introduces dynamic traffic management in particular.

1.1.1 Background of traffic problems: congestion

In many parts of the world the need for mobility has surpassed the development of road networks. An obvious indicator of the unbalance between traffic demand and network capacity is congestion. Road users in congested parts of the network experience a delay, in terms of additional travel time. Apart from a nuisance for travelling people, lost time is costly from an economical point of view.

Congestion often occurs at bottlenecks (where traffic faces a decrease in capacity) and points of conflicting traffic streams, for example merges, diverges, lane drops, (signalized) intersections, or even slower vehicles that obstruct the traffic flow. The onset of congestion also depends on variability in traffic demand, road and weather conditions, and driving behaviour (human factors). If the traffic density is high enough, a small disturbance can cause congestion. Congestion can be recurrent, when it occurs at regular traffic patterns, such as the daily commute, or non-recurrent, as a result of incidents, large public events, roadworks and detours.

The discharge rate of congested areas is often lower than their maximum capacity. This capacity drop deteriorates the achievable network performance. Furthermore, when the congestion back-

propagates to upstream intersections (spill back) it obstructs other parts of the network and might even cause a gridlock situation.

As a consequence of congestion, not only do travel times increase, they become unreliable, if the traffic process becomes unstable and less predictable.

In a nutshell, traffic conditions could be improved if supply and demand are matched appropriately along space and time. Simply expanding the network (the supply side) is often not desired or even possible. Therefore, regulating traffic demand and network capacity is often a more feasible option. This can be done by means of traffic control and dynamic traffic management.

1.1.2 Dynamic Traffic Management

Introduction to Dynamic Traffic Management

Expanding road networks generally has a negative impact on safety and environmental issues, and may induce new mobility needs as well. This is therefore often not the right way to reduce congestion. Moreover, building new infrastructure is often not beneficial from a financial point of view or even possible by limited available space. In many cases, using the existing infrastructure in a smarter, more efficient, way is often preferable[1].

Traffic flow can be considered the result of traffic supply (capacity offered by infrastructure) and demand (vehicles willing to use that infrastructure). If demand exceeds supply, 'excess' vehicles are being held up, and a queue (congestion) forms behind. The key to make better use of the network is to somehow regulate demand and supply.

Mobility as a whole can be regarded as a market system in which choices are made to make a trip, at what time, using which modality and, which route. The decision making of potential travellers can be affected by encouraging or discouraging policies, often by means of taxes or subsidies, or by providing information.

In the traffic market the demand side is defined by actual transport needs (vehicles on the road) and meets traffic supply (the available infrastructure) to generate the actual traffic flows, including route choice. The way traffic is distributed and progresses along the network, and the regulation of traffic at intersections and other discontinuities in the network, is a result of road geometry, traffic regulations and traffic signs, and vehicle and driver properties. Where and when needed, traffic management operators influence the traffic stream by means of traffic control signals and other dynamically set devices. These are part of Dynamic Traffic Management (DTM), which is the collection of all traffic control measures that allocate temporal and spatial utilization of infrastructures and vehicle fleets by means of dynamic signals. DTM measures can timely respond to changing traffic conditions, in order to meet the goals of traffic management, in terms of effective, safe and reliable use of the infrastructure.

The variety of DTM services is developing rapidly, especially due to innovationin the field of Intelligent Transportation Systems (ITS).

Dynamic traffic management architecture and tools

In order to effectively implement Dynamic Traffic Management, it needs to be organized in an appropriate architecture. In general the DTM architecture depends on the organization of the road authorities. Each road authority is represented by traffic operators and is responsible for the infrastructure (or certain road types) in its (jurisdictional) area. The operators have goals for their (part

[1]The Dutch Ministry of Infrastructure has a policy programme on this theme: "Beter Benutten" (Better Use)

of the) network, that are translated into policies and measures. They operate various DTM systems, to collect measurement data, to assess the current situation and identify problems, and to choose the right policies and measures to improve the situation. These systems are increasingly automated, which is necessary as the problems and possible solutions get more complex and the bar of performance standards is raised. Section 1.2 further investigates the organization of (comprehensive) DTM systems.

DTM is a matter of controlling vehicle movements, in order to affect traffic flows towards the intended speeds, intensities and routes. DTM tools on the road include traffic control signals, ramp metering systems, variable messages signs, lane signalling (closure, speed restriction, congestion warning), and dynamic parking guidance signs. Apart from these roadside systems, in-car systems offer increasing possibilities to facilitate DTM services.

A (non-complete) list of DTM services that can be implemented:

- optimizing throughput on highways to prevent congestion;
- providing route guidance information;
- safety warnings (bad weather, congestion);
- incident management;
- using buffer space in networks;
- dynamic lane management (and peak-hour lanes);
- efficient tapering;
- optimizing intersection control.

DTM consists of a wide range of measures and policies, and sometimes a specific tool can be used for multiple policies, or a specific policy can be executed by multiple tools.

DTM trends and loose ends

The twenty-first century faces highly complex traffic situations as well as an increasing set of possible solutions and performance standards. Traffic management based on predefined scenarios limits the number of possible actions, and often doesn't match the actual traffic state, or doesn't account for special cases, such as incidents. Traffic management operators need to be able to anticipate and respond to changing situations and choose a coordinated set of measures, out of all the instruments at their disposal, to improve bad traffic or even prevent it from occurring.

Innovative approaches are required, as well as cooperation between various instruments and actors. There are several developments that can be identified in the field of DTM, two important trends are:

- a shift from local control measures towards network-wide control;
- a shift from collective traffic information towards individual advice.

Local traffic control (for example signalized intersections or ramp metering) usually consists of standalone systems. Each system optimizes its actions, based on local measurements, to improve the local situation. If these local systems are integrated, larger areas can be managed as a whole. An example is a system that combines neighbouring signalized intersections into one system that coordinates the traffic lights and passing vehicle platoons. Various types of traffic control and DTM services could be conducted to work as one system. An example is the integration of ramp metering, route guidance

and intersection control of an area. Network-wide traffic management involves integrating the data collection and measures of a larger network area. A trade-off between the interests of multiple road authorities can be necessary, such as the performance of urban versus highway network, or the performance in various neighbourhoods. Integrated traffic control can significantly improve the overall network performance (van den Berg et al., 2007).

Besides collective DTM by roadside systems, there are measures that can be directed at the individual road user. Individual advice can be transmitted via in-car technology, such as navigational devices or radio (RDS). This advice can consist of directions for a route to follow, or warnings in case of congestion, weather or dangerous situations. Innovations in vehicle and communications technology makes vehicles increasingly 'smart' and connected, which creates the possibility of interactions between vehicles, road side systems, and DTM authorities. In-car technology could be beneficial for various applications within DTM, such as individual route guidance, effective car-following, lane-changing, or dynamic speed limits. These applications would contribute to an efficient traffic flow and safety, and support DTM measures. Moreover, dynamic vehicle data could be added to conventional data collection (such as loop detectors) for better estimation of the traffic state .

However, up until now in-car systems are typically used to improve the (route) choice of the individual road user (user benefit), whereas DTM aims at improving the network performance as a whole (social benefit). These different interests are potentially in conflict. New systems should provide a trade-off between user and system utility, user acceptance and social improvements. Potentially, in the future all vehicles will be automated and the infrastructure and vehicle fleet together create the 'optimal' traffic situation. More about the future of DTM can be read in (Hoogendoorn et al., 2012).

The trends of network-wide DTM and in-car systems can be illustrated by two projects that are currently being carried out in The Netherlands. The "Praktijkproef Amsterdam" (Field Operational Test Integrated Network Management Amsterdam) is a large-scale trial for coordinated DTM, where DTM measures are integrated stepwise and hierarchically organized. It involves infrastructure-based measures, and in later stages road-side and in-care systems will interact as well (Hoogendoorn, Landman, & van Kooten, 2013).

The second project is the Action Programme "Beter geïnformeerd op weg" (BGOW, Better informed on the road). Initiated by the Ministry of Infrastructure, "business sector, public authorities and knowledge institutes all collaborate closely in order to further develop the services required to provide road traffictravel information and traffic management" (Connekt, 2013). This 10-year programme aims to provide better service to travellers and reaching policy objectives for accessibility, quality of life and safety. The projected transition paths include a 'smart mix' of collective and individual services, coordinating standalone roadside systems, and expanding the coverage of DTM network areas (from local/regional to national).

The trends, described above, highly depend on technological innovations. Control algorithms should ideally be predictive, pro-active and anticipatory. Those characteristics ensure that traffic problems can be anticipated, and that the impact of possible measures can be evaluated fast enough to be timely put into action. Moreover, in order to make network-wide DTM possible, coordination and cooperation are very important.

This thesis considers only a limited number of aspects of individual and network-wide control. The next section introduces these aspects and formulates the problem statement for this thesis.

1.2 Problem description

After explaining the background and basics of dynamic traffic management in Section 1.1, this section discusses some main aspects of the problem in this thesis and proposes a starting point for the research approach. Subsection 1.2.3 condenses the analysis into the Problem statement.

1.2.1 Problem aspects

The research of this thesis takes some elements of the trend towards network-wide DTM and infrastructure–vehicle communication, and puts them into a system with an innovative DTM approach. The primary concept in this approach is *route guidance*, in itself a network-wide means of DTM. Secondly, *traffic signal control* is used from a non-local perspective. A rather simple approach using the concept of *back-pressure* control is argued to have potential benefits and will be the main topic of research in this thesis.

The aspects that are discussed (and may overlap here and there) are: complications of dynamic network-wide control, route guidance, integrated control, and coordination.

Complications of dynamic network-wide control

In Section 1.1 Dynamic Traffic Management was described as a way to regulate traffic flows, by dynamically allocating supply and demand, in order to improve traffic conditions in a network. The term *dynamic* is important, as traffic is a dynamical process, defined by certain laws of (behavioural) physics, and thus the impact of regulating traffic has a time component as well.

The state of traffic in a network is dynamic in a number of ways. Firstly, traffic demand is dynamic as it has periodic variability: daily peeks, day-of-the-week patterns, seasonal variation. Traffic demand often follows predictable patterns, but can be significantly influenced, for example by major events inside the network, or by big disturbances outside of the network. Secondly, on a lower level of scale, traffic flow usually is not an evenly distributed stream of vehicles, but rather a series of platoons,and in a congested area stop-and-go waves exist. Traffic control can be another source of varying, oscillating traffic flow, for example waves of traffic as a result of traffic lights. However, traffic control can also be prepared to deal with the variability of traffic flows. Lastly, the supply of infrastructure can be vary over time, not only by DTM measures but also by a capacity decrease due to an incident.

Applying DTM measures implies dealing with dynamic effects as well. Traffic control often solves local problems by allowing a certain level of traffic streams. The actions taken, however, also influence traffic downstream, and can have a delayed negative impact on neighbouring intersections. Another example is that route guidance relieves the traffic state on one route and causes problems of overload on the other route. In both cases the problem switches from one location to the other.

The more complex a network is, and the more complex the control structure, the more complicated will it be to achieve 'optimal' control. The stability of the network state is a point of consideration, especially if the traffic demand in the network is high. This shows the need for and the complexity of network-wide DTM. It is necessary to regard ways to anticipate on changing traffic, to stabilize oscillations, and to coordinate the control actions.

Route guidance

Route guidance is a typical example of network-wide DTM, as it influences the traffic loads on routes, and thus ideally can deviate traffic from problem areas.

Route guidance can be regarded as a way to influence the route choice of road users. There are several stages where route guidance can be applied: before making the trip from A to B, at the start of the trip, or during the trip (*en-route*). This thesis focusses on the last category.

The characteristics of route guidance highly depend on the means for transmission of route guidance information to the road user. Variable Message Signs (VMS)[2] are the current standard method, but they come with restrictions, such as the limited number of locations, and the limited amount of space for information, for the collective traffic stream. In-car navigation devices typically accommodate individual advice. However up to now they support the user to choose its individual shortest route, rather than the DTM advice for the system as a whole. Furthermore, the information is often incomplete or based on static information. As discussed in 1.1.2, developments are heading to a situation where in-car devices can be implemented for DTM measures.

The *impact* of route guidance depends on several aspects. To start with, there should be relevant route alternatives to choose from, otherwise there are no guiding options. Then, the quality of the route advice should be good, and based on solid information, and although the intention is to improve the state of the whole network, the advice for individual road users shouldn't be unfair (for example, when one is guided via a large detour). Another aspect is user compliance or obedience. This depends on a lot of things, such as the type and quality of information, the benefit to the user, the way the user is forced to follow the advice. Although defining the impact of route guidance is complicated, in this thesis assumptions are made to enable practical calculations.

Route guidance can make a significant contribution to network-wide DTM, but there are *potential negative effects* as well. As described in the previous paragraph (Complications of dynamic network-wide control) route guidance might shift traffic from a problematic route to an alternative route, where it creates a new problem. A related problem occurs when two alternative routes turn by turn get congested and relieved by route guidance, manifesting oscillating behaviour. This is an example of overreaction. This kind of negative effects should be investigated in a literature review.

A last problem aspect of route guidance is that route guidance can only be effective if it is calculated in time. This means that the route guidance settings should be computed fast enough to make real-time decisions.

Integrated control

The integration of various types of control is an increasingly important aspect of DTM. In this thesis route guidance is taken as the main control strategy, and traffic signal control as a second type of control. Integrated control implies the organization of individual control measures in a way that they can cooperate, aiming at (possibly synergetic) system improvements. Route guidance and traffic signal control influence each other. To illustrate this, the queue of waiting vehicles depends on the duration of the 'red phase' at the associated intersection. In turn, route guidance actions depend on the condition of the roads it considers, including those waiting queues. This is an example of traffic signal control influencing the route guidance, but route guidance has an effect on the intensity of traffic streams, which are in turn used to determine the traffic signal control settings.

It is possible to bidirectionally optimize route guidance and signal control, but this might be computationally difficult. Another way to approach the integrated problem, is to start from the network level, and determine the optimal route assignment, and then adapt the traffic signals to adequately accommodate the resulting traffic load.

[2] In the Netherlands called DRIPs: dynamic route information panels

Coordination

The effectiveness of individual measures is affected by other measures, especially if they are in each other's vicinity. By means of *coordination* DTM measures can be properly set to function as an integrated system. Coordination means that the settings of multiple instruments are adapted to achieve a common goal, and often involves coordinated time planning. This way, a group of DTM measures can function as one cooperating system, aiming at exploiting the strengths of the measures, or even synergetic effects. More on coordination in (van Katwijk & Taale, 2012).

There are different types of coordination. Standalone DTM systems of the same type can be coordinated, such as ramp metering systems, or traffic signals on neighbouring intersections. A next step is to coordinate all traffic control systems in an area, and then the coordination of multiple areas. Route guidance is typically a type of measure where coordination could support beneficial effects, as it partly determines the quantity of traffic flows in different parts of the network.

Coordination requires not only technical facilities, but might involve multiple authorities as well, for example when the control of urban roads is connected to the control of a provincial road or (national) highway.

The bigger and more complex a network and DTM system is, the more important it is to support coordination by a suitable structure. There are various hierarchy models to do this (Zuurbier, 2010):

- centralized control;
- decentralized control;
- distributed control, and;
- hybrid control.

Centralized control refers to a system with one controller that collects all traffic data and decides on all control actions to be taken in the network. This type of control has optimal performance potential, but faces computational complexity in large systems. Furthermore, its monolithic structure makes it hard to adapt the system to network modifications, or to cope with local failure.

In a *decentralized control* system several controllers with their own area of responsibility supervise the sensors and instruments in their area. The controllers don't interact, therefore the total performance is theoretically lower compared to the centralized case. Expanding the decentralized system implies adding or modifying one controller, without having to consider the rest of the system, making it more flexible, and robust to local failure. The *distributed control* approach also consists of individual controllers, but unlike the decentralized system it enables controllers to communicate with each other. Since the controllers share knowledge and assistance, this approach has potential gains in performance.

Hybrid control approaches combine the properties of the other approaches into a multi-level hierarchy of controllers, and aim to combine advantageous properties. There are various ways to organize a hybrid system, although often there is a central controller with the role of conductor and the other controllers perform delegated tasks.

It can be very complicated to design, build and maintain a well-coordinated control structure. In this thesis it is assumed that coordination can be achieved by adopting simplifications (of which decentralization is an example), at least as a first step.

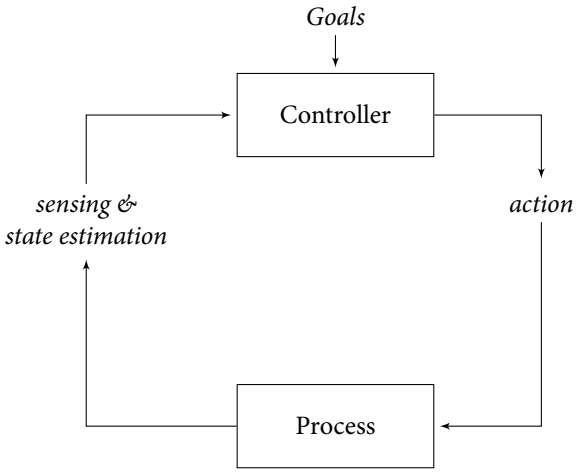


Figure 1.1 – Basic control loop

1.2.2 Research angle and starting points

The problem aspects described above cover a wide area of traffic control problems and possible solutions. This subsection narrows down the perspective in order to determine a clear path for the problem and objective of this thesis.

Starting points for system performance

Before starting to consider a methodology, there are a few starting points, things this thesis aims to achieve:

- a model that includes both route guidance and traffic signal control;
- a model that improves the system performance of a network;
- besides reducing travel time and travelled distance, special focus on:
  - high throughput (getting more travellers to finish their trip within a time period);
  - high stability in the network (to what extent can delays can be bounded, and reducing oscillations, variety and fluctuations), decrease of spill-back;
- a model that improves the performance for individual road users as well, if possible;
- a model that can be used for real-time control, which implies efficient computations.

Control approach

Like many control problems, traffic control can be condensed to a simple *closed control loop* process (Figure 1.1). The ongoing traffic state (process) is monitored by sensors. Sensor data is used to estimate the state state of the system, as an input for the controller. The controller comes up with actions and sends these to the actuators (the dynamic signal). The actuators regulate the traffic process and thus the loop closes.This control loop can be expanded depending on the way the applied control approach.

There is a wide range of possible control approaches. Keeping in mind the problem aspects and the starting points of the previous paragraph, a control approach that is fundamentally *simple* is to

be preferred. A simple control structure presumably comes with efficient computations. Especially if networks grow larger, and control systems get more complex and integrated, a simple control structure is eligible, provided that it delivers the desired performance. Looking at coordination of actions and the hierarchy models (page 7), a *distributed* or *decentralized* approach seems promising.

Control approaches exist in various shapes, and can be defined and categorized in different ways (Lin, 2011; Schreiter, 2013; Zuurbier, 2010). Here just a general overview is presented. A first distinction can be made between reactive control and predictive control.

In the case of *reactive* control, a controller determines control actions primarily based on the current state of the network (measured and translated to a model). The overall reactive control system is often a feedback process; the control output influences the traffic process and is lead back into the system as input (by contrast, open loop or feed-forward control only considers system input but isn't confronted with the output).

There are several types of reactive control. There are control algorithms that calculate the settings of the actuator based on the current traffic state. In a simple it defines a control setting as a function of an input ($output = f(input)$). Other ways of reactive control include rule based control — where 'if–then' rules are used to infer the right control settings —, and case based control — the current traffic state is compared to the case base, and control settings associated with the closest match are chosen. Reactive control methods are in general computationally fast, as they require a limited number of computations that are applied to the current traffic state.

Predictive control approaches usually apply a traffic flow model to the current traffic state, in order to estimate the traffic state in the near future. Simple predictive feedback control (including algorithms, rule based and case based control variants) makes a 'one shot' prediction, on which new control settings are based. Possible advantages over reactive control are that near future problems can be expected (onset and progress of congestion) and the control settings can be adapted to (prevent) that situation. This potential benefit however comes with a cost of a higher computational load.

Optimal control is a much more extensive method of predictive control, that is designed to minimize a performance function (for example, the total time spent by vehicles in the network), by executing an iterative process of traffic simulations. Each simulation is characterized by a trajectory of settings for its DTM measures, the effects of which are taken into account. Thus, the simulation with the optimal performance determines the best dynamic DTM settings up to the projected time horizon. The calculation of optimal control can be highly time consuming, depending on the complexity of the network and DTM system. Moreover, although this approach theoretically provides optimal control settings, there is no re-evaluation within the projected time horizon (feed-forward instead of feedback control). This makes the method mainly useful for offline simulation and planning purposes.

The concept of Model Predictive Control (MPC) is based on optimal control, but it adds a rolling horizon (see Hegyi (2004)). This means that with a small time interval the traffic state is re-estimated and an optimal control method is used to determine the optimal control settings trajectory, making it an application of feedback control, able to adapt to unexpected traffic behaviour and events. Theoretically, MPC is suited to operational traffic control, but the high computational load is a barrier for real-time applications.

Anticipatory control (Taale, 2008) is yet another approach. Here, a prediction is made for the effects of DTM, but for anticipating driver response as well. If one route becomes less attractive, road users will redistribute towards more appealing alternatives. This concept of prediction has many benefits and is probably suitable for future systems of traffic control, in which infrastructure based DTM and individual vehicles are regarded as one control system. An important drawback is the computation

time, as the prediction now involves even more factors; perhaps simplification methods will offer an increase of speed.

Back-pressure as a starting point

Besides the control concepts of the previous paragraph, there are many variants and hybrid systems. One approach, that has recently gotten attention in the field of traffic control, is *back-pressure* control (first mentioned by Tassiulas and Ephremides (1992)). This control concept has the properties of reactive and feedback control, and can be implemented as a distributed or decentralized system. It is a simple and flexible approach, making it potentially appropriate for real-time traffic control.

Back-pressure control typically consists of set of controllers, each belonging to a node in the network. In case of road traffic, each intersection could be controlled by a back-pressure controller. Each controller optimizes its service to waiting queues, according to a back-pressure algorithm. This illustrates the decentralized nature of the approach.

Typical for the back-pressure algorithm is that it is based on the difference in traffic load between roads leading into and roads leading out of the intersection. Contrary, other distributed traffic signal control algorithms, such as the P_0 policy of (Smith, 1980) or the work of (Lämmer & Helbing, 2008), only consider the expected traffic load on roads leading into the intersection.

The literature on back-pressure claims that this control method maximizes the throughput of each node and the network as a whole, while maximally stabilizing the network, keeping the queues bounded (which is explained further in the Literature review). These properties of back-pressure control appear to fit the system performance aims on page 8, therefore back-pressure control is chosen as a starting point for traffic control in this thesis.

Because the chosen approach is reactive, it won't use prediction and therefore can't look into the future to prevent problems. Nevertheless, the fact that network stability is a prominent feature of back-pressure control, one can expect the approach to be proactive, because of its distributive approach of keeping the network stable.

Although back-pressure control has been applied to traffic signal control (in theory and simulation, not in practice), it has hardly been used for route guidance of road traffic. It is worth to do research on this application. Normally route guidance is based on travel times, the back-pressure principle would use the traffic loads on routes, and would improve network throughput and stability. Furthermore, the opportunity for a distributed approach is attractive. Lastly, the original back-pressure algorithm (Tassiulas & Ephremides, 1992) for communication networks naturally combines control of switches and routing, which might be possible, to some extent, for DTM as well.

1.2.3 Problem statement

Previous subsections explored the problem aspects as well as the direction of research. The findings can be condensed into the following problem statement.

Route guidance can make a significant contribution to network-wide DTM. Using in-car navigation to transmit individual directions, it is expected to improve the network performance (better use of the available capacity, higher throughput and stability, and less spill-back), and reduce the travel time for the individual road user as well. Finding the right route guidance configuration is a complex task, that should take potential unfavourable effects and coordination into account.

Moreover, integrating *traffic signal control* at intersections with route guidance is a logical step.

Considering the complex nature of the problem and the need for (real-time) responsive control actions, a simplified traffic control model is suggested. The concept of *back-pressure* control has recently been applied to the problem of controlling a network of signalized intersections. The development of a route guidance algorithm based on back-pressure control, integrated with a back-pressure model for signalized intersections, is to be considered. The main research problem in this thesis is to determine the feasibility and potential benefits of such a system.

1.3 Objective and research questions

Section 1.2 formulated the problem statement for this thesis. This section further discusses the purpose of the research. It states the main objective, and then expresses the research questions, which compose a guiding plan for the research to achieve the objective.

1.3.1 Objective

Following the problem statement (1.2.3), the aim of the research in this thesis is formulated.

The main objective of this thesis is:

- to develop a framework that integrates route guidance and signal control based on the back-pressure principle, and determine its feasibility and potential benefit.

The following conditions are stated to further specify the objective:

- The method for route guidance should have characteristics of *network-wide* DTM (thus take into account network-wide effects), and be suited to give advice to individual road users.
- The method should consider a *trade-off* between *optimizing network conditions* and finding *shortest routes* for individual road users.
- The approach focuses on *real-time control*, implying the requirement to adequately cope with the dynamics of traffic. A computationally *efficient* model is required.
- The method should be *generic* in nature, be applicable to a wide range of network types.

Other conditions, assumptions and points of departure that make up the scope of this thesis are formulated in the chapter on the research approach (Chapter 3).

1.3.2 Research questions

The objective of the thesis is extensive and can be divided into a series of research questions. These research questions give guidance for the necessary steps to carry out, in order to accomplish the larger objective.

1. What are the characteristics of route guidance and what are its strengths and weaknesses?
2. How can traffic signal control complement route guidance into an integrated approach?
3. How can the concept of back-pressure control be used as an algorithm for traffic signal control?

4. How can the concept of back-pressure control be used as an algorithm for route guidance?
5. What is the purpose of the new control approach, and what are the design requirements?
6. What are the proposed control algorithms and their alternatives?
7. How can simulation be used to evaluate the performance of the control algorithms?
8. How should the simulation experiment be conducted, in order to get meaningful results to evaluate?
9. Based on simulation results, what are the differences in performance, between various variants in various scenarios, and compared to the prior expectations?
10. Considering the strengths and weaknesses found, what is the value of the proposed control approach,and what room is there for improvement?

These research questions are addressed by the chosen research method in Section 1.4. The thesis outline, Section 1.5, explains which questions are considered in the various chapters of this report.

1.4 Research method

The previous section defined the objective and research questions, which indicate the points of focus for this thesis. In this section, to cover these points, several steps are formulated that make up the research method. Although the steps are listed as a linear process, it turns out that some can be done in parallel, others need a kind of iteratively moving back and forth to get the desired result.

Step 1: determine the state-of-the-art
The first phase of the research consists of finding and analysing relevant literature related to traffic signal control and route guidance, but especially on the topic of back-pressure control. This leads to a survey of the state-of-the-art, and finds out how back-pressure control could be extended to the application of route guidance. Research questions 1–4 are part of this literature review.

Step 2: set up the research framework, design principles and scope
Then, the next step is to further define the framework, or playing field, for the main part of the research. Based on the objective, research questions, and the theoretic background from the literature review, a basic concept for the proposed control model is formulated, together with chosen design principles and a definition of the scope, including the assumptions and limitations for the research. A distinction is made between the control model and the experimental framework to test and evaluate this model. Research question 5 is mainly related to this step, but 7 and 8 as well.

Step 3: propose control models
In this step the control methodology is developed, starting from the basic concept in step 2. The control model is explained as a whole, but the main part consists of discussing the two separate 'modules' for traffic signal control and route guidance. The back-pressure concept will be used as a central concept, and the main control approach is fixed, however several alternative 'versions' are examined. This step covers research questions 3 and 4, but particularly research question 6.

Step 4: prepare the simulation environment
The preparation of the simulation environment means that a traffic simulation model is chosen and modified where needed, in order to execute the experiments. Not only does the model need to simulate the traffic process, also the control model should be incorporated, and the required output, results, should be produced. To prevent simulation problems as much as possible, the model is developed step-by-step, where each modification is tested. This step answers to research question 7.

Step 5: test the control model
The control model for traffic signal control and route guidance are to be tested independently, to determine if they work as expected, and to identify the differences of alternative versions and settings. Also is tested if the combined control model works. This step (and the next) are related to research question 8.

Step 6: prepare and run an experiment, show the results
The final experiment can be done if the simulation and control model are sufficiently tested. Based on the research approach of step 2 an experiment plan is set up, and then executed. The simulation results are then evaluated and presented in relevant tables and diagrams.

Step 7: discuss the results and draw conclusions
The last step of the research is to critically discuss the simulation results. Determine the main findings, and if the results meet the expectations. Describe possible reasons for deviations and shortcomings of the model. Then, draw conclusions of the research as a whole, and formulate recommendations, such as further research to improve the methodoloy. Research questions 9 and 10 are dealt with in this final step.

1.5 Thesis outline

The thesis, following the described approach, is split up into a list of chapters. Figure 1.2 illustrates the outline and shows the main relations between the individual chapters. The body of the thesis consists of eight chapters.

The first (and current) chapter is the *Introduction*, that introduces the problem, objective and general method of this thesis.

Chapter 2–7 are the middle chapters, where the main research is done. Chapter 2 *Literature review* describes the state-of-the-art and room for this research to extend the existing theories. Chapter 3 *Research Approach* lays out a basic concept for the proposed control model, and determines a playing field of design principles and the scope. Chapter 4 *Control model* is concerned with the proposed traffic control methodology and the algorithms for traffic signal control and route guidance in particular. Chapter 5 *Simulation environment* motivates and explains the used simulation model, and the modifications necessary to incorporate the new control model. Chapter 6 *Simulation experiments* presents the tests of the control model (and its main modules) and a comprehensive experiment of the complete model, including the produced results. Chapter 7 *Discussion* makes a critical analysis of the experimental results and defines strengths and weaknesses of the proposed control model.

Finally, chapter 8 *Conclusions and recommendations* draws conclusions from the researched topics, and proposes recommendations for possible future research.

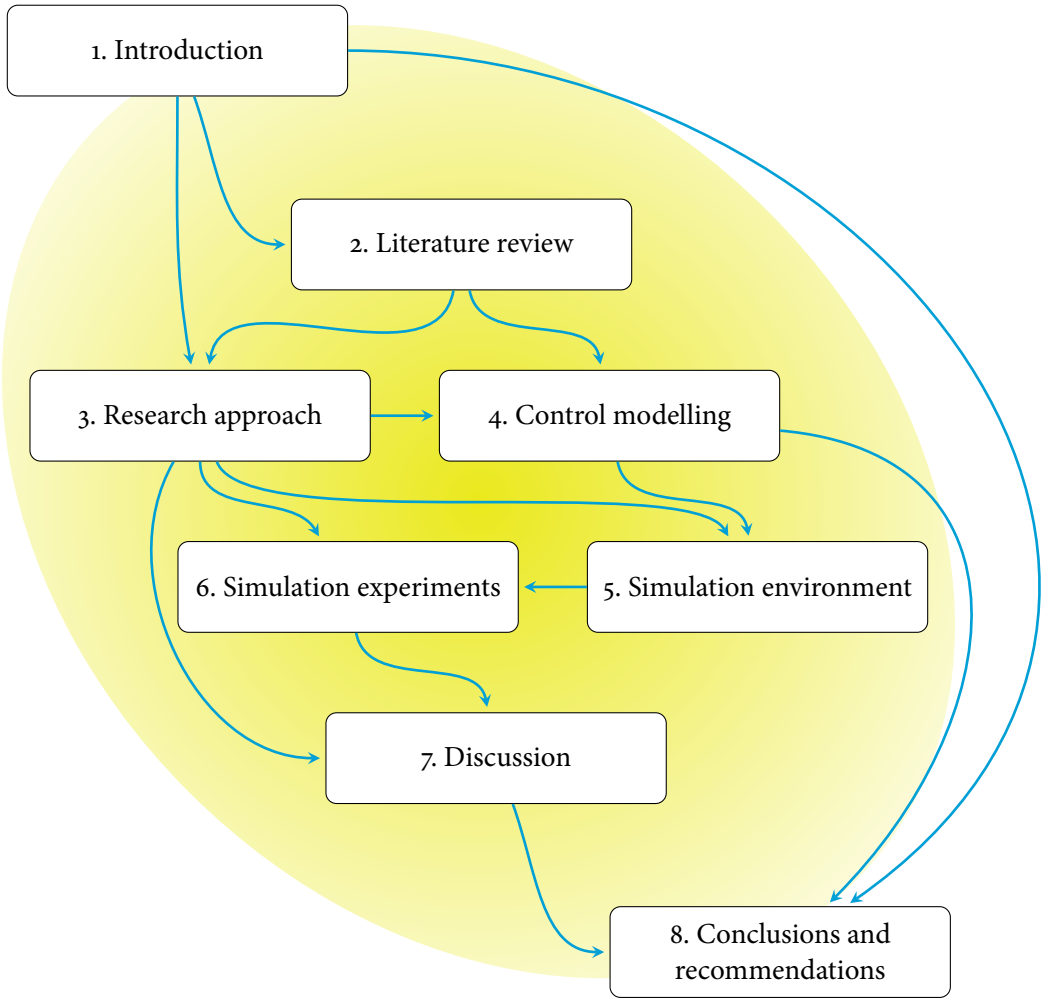


Figure 1.2 – Thesis outline

Chapter 2

Literature review

In the Introduction (Chapter 1), the problem aspects related to dynamic traffic management and route guidance in particular have been explored. The first step to reach the objective of this research is to get a deeper insight of the state-of-the-art of the theories and research on the relevant topics.

Route guidance is discussed in Section 2.1, which gives a general overview of route guidance concepts. As it is intended to be used as a starting point for the approach of a new methodology in this thesis, the concept of back-pressure, its origins, development and recent use in traffic engineering are the main topic of Section 2.2. Section 2.3 summarizes all findings.

2.1 Route guidance

The first chapter Introduction gave an overview of dynamic traffic management and indicated some of the challenges. This section of the literature review gives an outline of the way route guidance can be used as a means of traffic management. The thesis objective states that a framework to integrate route guidance and traffic signal control should be developed. Traffic signal control is not reviewed in detail, but in the next section it is described on the application of back-pressure control.

Route guidance is a widely studied topic in traffic flow networks, here only a few important aspects are discussed.

2.1.1 Route guidance in general

Route guidance can be considered as a way to influence or override the route choice behaviour. The goal of route guidance can be to minimize the total travel time for the network as a whole, a system optimum situation, or a user optimum where no road user can change its own route to a faster route. A route guidance system can be of use in everyday traffic conditions, but especially when the traffic conditions are irregular, or in case of an incident. Then, people can benefit from the information provided by the route guidance system (Papageorgiou et al., 2003).

There are three ways to receive route information. The first is pre-trip information, for example by means of radio, tv or internet. Traffic updates or route planners can provide the first routing advice (or another advice, e.g. to go by public transport). Secondly there is roadside collective route information displayed by variable message signs at strategic points in the network. The third type is what is considered in this thesis, en-route route guidance, which can be provided by in-car navigation systems (Papageorgiou et al., 2003).

Apart from the goal of a route guidance system, there are also strategy types that can be distinguished. An iterative strategy performs a repeated process of simulations, to reach the optimal setting at convergence (either system or user optimum). This approach can be put into the group model predictive control. It is hard to put into practice for real-time operations, it requires many computations.

The other type is the group of one-shot strategies. This group holds reactive control approaches or (less common) predictive approaches (where a model is used to predict a near future state to react on)(Papageorgiou et al., 2003). Examples of feedback reactive routing strategies that are based on differences in instantaneous travel time include the bang-bang, P, PI or LQI types (Wang, Papageorgiou, & Messmer, 2003, Freeways, High-Occupancy Vehicle Systems, and Traffic Signal Systems 2003). But also route choice models can be applied for the reactive type.

2.1.2 Route choice models

Route choice models can be used to estimate route choice behaviour and the probability of chosen routes, but also to assign route shares based on the value of a route (which can be based on travel time or on other route properties).

The simplest decision rule is the choice for the shortest path (Ben-Akiva & Bierlaire, 1999). In most cases a enabling spread of used routes is a better approach, which can be achieved by stochastic models based on the random utility model. The multinomial logit model can be used to model route choice when route alternatives are independent. If there is overlap, a path size logit model should better be used. This method takes into account that a route is less distinct if its links are used in other routes as well (Ben-Akiva & Bierlaire, 1999). The C-logit model (Cascetta et al., 1996) does a similar thing by using a commonality factor, but has less theoretic support (Frejinger, 2008).

In route choice models, it is often not straightforward to choose a good set of routes, that represent all possibilities, in particular when networks are bigger and more complex.

If the network gets bigger and more complex, and the number of routes increases, a recursive logit model could be a promising model (Fosgerau, Frejinger, & Karlstrom, 2013). It is a distributed approach and can also be used to take overlap into account, much like the path size logit model.

2.1.3 Aspects of route guidance

Compliance

Route guidance can make a significant contribution to network-wide DTM, but the impact of route guidance depends on the compliance of the road users. The type of route guidance messages, are they informing, advising or imposing is one aspect. The share of users that is equipped with the needed in-car technology, penetration rate, is another. A part of the drivers is that much familiar with their usual route, that they would not switch to another if that one is a bit faster.

Routing traffic from a system point of view means optimizing network flows, but individual road users shouldn't suffer too much detour and delays (fairness).

In this thesis the assumption is made that all traffic complies to the given advice.

Oscillation behaviour

Traffic control measures, and route guidance in particular, can be confronted with problems of oscillating behaviour. A simple example of this is when route guidance is based on the shortest time path. In the beginning, route A is faster than route B, so all traffic is routed through route A. But route A gets congested by the high amount of traffic flow, so now route B becomes faster, and all traffic is routed via route B. This unstable situation is an example of undesired oscillation by overreaction.

The oscillating behaviour as a result of feedback route guidance is caused by the time lag between measure and effect (Wahle et al., 2000; Landman, Hegyi, & Hoogendoorn, 2012). Furthermore there is the self load effect (Gao, Dovrolis, & Zegura, 2006), the route guidance measure doesn't take its

own added route cost into account. Oscillation also depends on the sensitivity of the control function and the speed at which it reacts on changes (Feldmann et al., 2009).

2.2 Back-pressure algorithm

According to the previous section, the development of route guidance algorithms is a field of ongoing research, and the effectiveness of a method might depend on the situation it is applied to. This section reviews and assesses literature on a recently emerged algorithm: the back-pressure algorithm. As no route guidance applications of back-pressure control exist[1], other relevant literature is critically reviewed, to serve as basis for a new algorithm.

> Loosely speaking, the idea of back-pressure control is to compute pressure at every node based on node occupancy and to open flows which have a high input pressure and a low output pressure, like opening a tap.

This statement from (Gregoire et al., 2014) describes what the concept of back-pressure is about. The back-pressure algorithm originated in the early 1990's as a means to maximize throughput in communication networks. The back-pressure policy allocates server activity to flows with the highest 'weight', based on the queue length difference between two connected points. This policy can result in maximum throughput, under stable network conditions. A key benefit of back-pressure is that it can be implemented in a distributed manner, making the architecture and control algorithm very simple and flexible. However, as described in Subsection 2.2.1, back-pressure routing also has some shortcomings, some of which have been dealt with by modifications in later research.

In recent years, back-pressure control has successfully been applied to the problem of traffic control at signalized intersections. For simplified cases of network theory and simulations, the properties of maximum throughput, network stability, and reduction of congestion back-propagation have been pointed out.

This section reviews the original back-pressure algorithm and later work in communications networks, as well as literature on applications to signalized intersections. Then, considering the found characteristics of back-pressure, starting points are formulated for translating the back-pressure concept to the route guidance problem of this thesis.

2.2.1 Back-pressure routing in communication networks

This subsection describes the concept back-pressure as it has been developed in the field of communication networks (wired and wireless multi-hop networks).

Back-pressure origins

The concept of back-pressure[2] routing was introduced by Tassiulas and Ephremides (1992) and their 'maximum throughput policy'. They consider multi-hop radio networks, where customers are routed from origin to destination. At any node in the network, represented as a directed graph, customers may enter. Each customer belongs to a class (commodity), associated to a subset of all nodes, which defines its destination.

[1] To the knowledge of the author, back-pressure hasn't been studied yet, for the type of route guidance problem in this thesis.

[2] Alternative names for back-pressure: max pressure, max weight, maximum throughput, maximum differential backlog.

The customers are sent to their destination via multiple nodes that are interlinked by service providers. As these servers (or links) are interdependent and can't all be activated simultaneously, it's only possible to activate particular combinations of servers: activation sets. Here, the control problem is to schedule the activation sets in a way that throughput is optimized. Additionally, the network state should be stabilized. The stability region is defined by all possible arrival rates that can be served in a stable manner, meaning that all queues do not tend to increase boundlessly. The existence of the stability region is proven in the paper, and is based on the Lyapunov drift theorem.

A maximum throughput policy, the back-pressure policy, π_0 decides which servers are activated at each discrete time slot.

At the beginning of each time slot, messages (classes) are coupled to outgoing links, assuring routes to the destination exist. Then the policy π_0 determines in three stages which set of servers to activate (somewhat simplified).

1. For each server i, and for each customer class j, D_{ij} is defined as the difference in queue length X between the current waiting queue at node n and the queue at node h, downstream of the server, multiplied by a service rate m_i at which customers can be served. The weight D_i for each server is defined as the maximum value of D_{ij}, over all customer classes $j \in J$.

$$D_{ij}(t) = \left(X_{n(i)j}(t) - X_{h(i)j}(t)\right) m_i\,; \quad D_i(t) = \max_{j \in J}(D_{ij}(t)),$$

   Note that nodes with a class destination have no queue for that class.[3]

2. Out of activation sets S, the one vector ($\mathbf{c}$) that maximizes served weight, is selected:

$$\hat{\mathbf{c}} = \arg\max_{\mathbf{c} \in S}(\mathbf{D}^T(t)\mathbf{c})\,,$$

   where $\mathbf{D}^T(t)$ is the transposed vector of server weights $D_i(t)$.

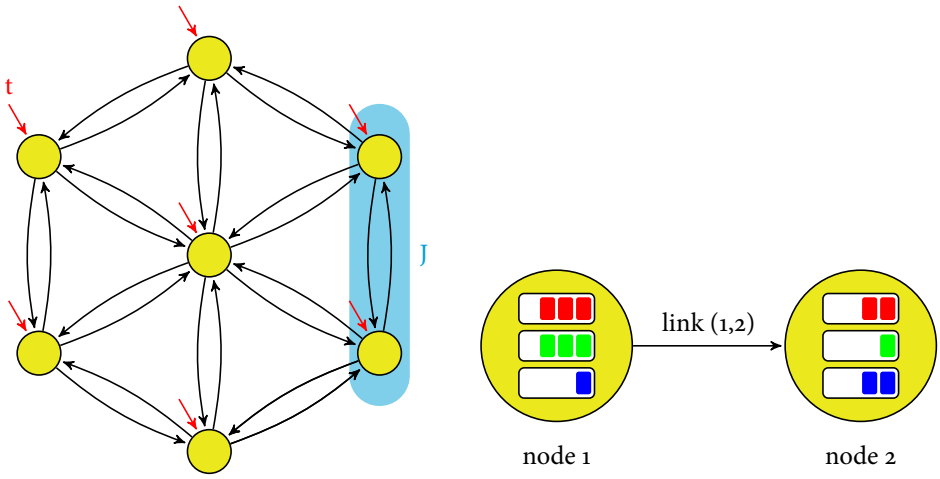
3. The customer classes $\hat{j}$ for which $D_i(t) = D_{i\hat{j}}(t)$ for each server are selected to be served.

Figure 2.1 shows an example of a multi-hop network with destination J (Fig. 2.1a), and a simple example of two nodes connected by a server (Fig. 2.1b). In this example the green (middle) customer class has the highest weight.

Using the algorithm in a centralized manner implies the need for concentration of all queue information and rather complex constraint sets and computations, especially for larger networks. Tassiulas (1995) introduces a distributed version of the algorithm (based on the same principles, using only local data), and extends it to continuous time and fluid arrival rates. In the 'adaptive back-pressure' (ABP) the server's frequency for switching queues depends on an adaptive parameter α.

The basic back-pressure algorithm can't be directly implemented for many applications, therefore several issues have been solved in later literature, building on the groundwork laid out in the paper of (Tassiulas & Ephremides, 1992). A few will be discussed next: issues on wireless networks, issues on delay capturing, the complexity of the model, and the issue of infinite buffer sizes.

[3] This is a basic means to 'pull' customers to their destination.



(a) Multi-hop network **(b)** A close-up

Figure 2.1 – Back-pressure principle in a multi-hop communication network

Wireless networks: power allocation and multi receivers

Neely, Modiano, and Rohrs (2005) use the concept of (Tassiulas & Ephremides, 1992) for the case of wireless networks with time-varying channels. Here, the rate of data transmission along links depends on power allocation by the nodes, and the channel state (influence of noise and fading). For example, if the channel state is optimal, the least amount of power is needed to transmit a certain amount of data. Nodes may simultaneously divide their available power to multiple links, and it is assumed that data can be split continuously. The joint problem of power allocation and routing can be solved with an adapted version of back-pressure routing, where each node has a destination specific queue. The algorithm (Dynamic Routing and Power Control) can maximize throughput within the capacity region (stability region). Instead of finding the best activation set (as in Tassiulas and Ephremides (1992)), the optimal power allocation, and thus the transmission rates, is determined. The authors mention that the policy uses back-pressure in an effort to equalize differential backlog, as that is a main characteristic of back-pressure routing.

Since the power allocation can be decoupled, the DPRC algorithm can be decentralized for use in ad hoc mobile networks. If channels (links) are modelled as being independent, this is straightforward, but when interference is considered, approximations can be used.

Back-pressure algorithms are not well-suited to deal with delay in a network. As stability properties imply that queues are bounded, expected delays are bounded as well (Little's theorem). Counterintuitively, in situations of sufficient load, the average delay typically is lower than for low intensities, where packets are scattered around the network. At (very) low demand, there is no build-up of pressure gradients towards the destination. Individual flows can meet unnecessarily high delays, when traversing the network via long detours or even cycles.

Instead of reducing detours by restricting the set of possible routes, Neely et al. (2005) suggest the use of *bias* parameters for their Enhanced DRPC algorithm, to account for the (shortest path) distance to destinations and to otherwise prioritize certain commodities.

The DIVBAR (Diversity Backpressure Routing) algorithm of (Neely & Urgaonkar, 2009) extends the algorithm, by enabling a sent packet to be received by multiple nodes in the proximity (in case of a transmission error). The confirming node with the largest positive differential backlog takes over

responsibility of the packet. An enhanced version, E-DIVBAR, includes a shortest path bias. In (Neely & Urgaonkar, 2008) simulation shows that this bias results in significant improvements for delay and congestion .

Delay issues

According to Gupta and Javidi (2007), the maximum throughput algorithm of DIVBAR has some shortcomings. Sending packets based on queue backlogs, could send them away from their destination. Including a measure of closeness, or cost, to the destination improves the algorithm in many of these cases (as E-DIVBAR). Another problem is that both algorithms don't look beyond one step ahead. The overall utility of a routing decision, involving congestion on the whole downstream path, reliability and the existence of routing alternatives, is not taken into account. Gupta and Javidi (2007) propose the policy that uses the algorithm "Opportunistic Routing with Congestion Diversity" (ORCD). This algorithm starts at destinations and combines the concepts of back-pressure and shortest path finding, in order to determine the attractiveness of the next route decision, by setting the expected cost for the neighbouring nodes. This method computes throughput in a centralized manner. A distributed version includes a recursive procedure that uses the node cost of the previous time slot and current local queues to determine the new node costs. The results from a simulation show that this method results in far less delays than DIVBAR and E-DIVBAR[4]. The results seem promising, the method intuitive, and computational overhead is claimed to be not much more than traditional back-pressure.

Elaborating on ORCD, Naghshvar and Javidi (2010) propose various modifications, because of its large overhead of ORCD. They compare the algorithm to shortest path algorithms and (E-)DIVBAR. As expected, ORCD results in small delays, whereas DIVBAR performs significantly worse, and although E-DIVBAR compares well to ORCD for a regular grid network, it performs poorly when the network topology becomes irregular. The modifications to ORCD include Infreq-ORCD (lowering the computational load by computing a node's 'utility' only once every n time slots), P-ORCD (using partial diversity to decrease overhead cost, by limiting the number of nodes that may confirm receiving a sent packet), and D-ORCD (the distributed version of Gupta and Javidi (2007)). All modifications lead to some decrease in performance, but still exceed (E-)DIVBAR.

Note that these algorithms apply to wireless communication networks, where power use and signal strength play role in choosing the next node. Road traffic networks have different characteristics, not all concepts can be converted to road traffic networks, but they do give insight in the way back-pressure based algorithms might be used and modified. The ORCD method is less straightforward than back-pressure, and it is to be seen if the concepts can be translated to multiple commodities, to other network types and to route guidance for road traffic.

Another approach to dealing with delay is using a shadow architecture as formulated by Bui, Srikant, and Stolyar (2009). They regard a multi-hop wireless communication network (without diversity considerations), where packets are routed in a fixed way (routes are determined upon entering the network). Back-pressure with fixed routing can lead to upstream accumulating queues, and increasingly constraint service rates. If each link would have a fixed service rate, queues at each hop would be averaged. However, all arrival rates should be know a priory, to make such an allocation possible. Counters, called shadow queues, could be used to this end. The back-pressure is applied to the shadow packets, and when flows have been determined, the 'real queues' are served accordingly.

[4] In the simulated example network, DIVBAR performed even better than E-DIVBAR.

This architecture has benefits for delay handling, but also for reducing complexity, because the real queues are cumulated for each neighbouring node, instead of for each destination.

In (Athanasopoulou et al., 2013) this shadow architecture is extended with a decoupled routing component, enabling adaptive routing. At low intensities, delay is completely reduced when using a parameter M. However, the correctness of using this constant parameter seems questionable, when compared to the bias as in Neely and Urgaonkar (2008) (as a remark, the logic seems questionable, that adding a constant to the backlog could prevent choosing long routes). In light traffic loads, M will prevent flows and keep queues growing; as a remedy extra link activation is used, based on the shadow queues. Simulations show that it works.

Another algorithm that focuses on reduction of delays (especially with light traffic loads) is proposed by Ying, Shakkottai, et al. (2011). Their back-pressure algorithm for routing and scheduling dynamically combines optimal throughput with minimization of route lengths. Compared to the EDRPC method of (Neely et al., 2005), this algorithm provably minimizes average path lengths. The objective function is similar to the one in (Bui et al., 2009), but the proposed algorithm different. Also the work of (Gupta & Javidi, 2007) is related. The algorithm reduces the average number of hops, and also significantly reduces delays, given the right parameter settings. Queue lengths per node are smaller for light traffic and for higher traffic loads approaches traditional back-pressure. It is mentioned that hop counts could be replaced by weights based on propagation time (travel time) or distance.

Reducing complexity

Also, based on the work of (Bui et al., 2009) the use of virtual (shadow) queues is proposed to replace the per-hop queues, and thus reduce complexity.

Complexity seems an important issue in this algorithms, as all possible hop counts from a node, to all destinations, have their separate queue. There should be an easier way?

Because hop counts are taken for shortest path, it is hard to get a good trade off (parameters) for also reducing end-to-end delays. Using travel time might improve that, so to prefer above hop count or distance in km?

A final way to decrease complexity in back-pressure by reducing the number of queues, is the use of clustering. Ying, Srikant, et al. (2011) divide groups of nodes into logical clusters. Each cluster contains gateway nodes, linked to nodes outside of the cluster, and interior nodes. This way, destinations can be specified by the gateway nodes, as long as the packet is still outside the cluster of its destination, and the amount of queues per node can be reduced to queues for interior nodes in the cluster and gateway nodes. The traditional back-pressure algorithm is expanded by a traffic control (routing) part and a regulator part to maintain real queues and regulated queues at the gateways.

Finite queues

Besides poor delay handling, another limitation of the original back-pressure algorithm is the use of queues with infinite buffer size (also called vertical, or point queues). Giaccone, Leonardi, and Shah (2005) address this issue and extend the work of (Tassiulas & Ephremides, 1992) by introducing finite queue sizes. They use a function that trades-off throughput and queue size, which leads to higher throughput for queues with smaller buffer sizes. The algorithm won't be explained here, but the concept is intuitive and will be used by later described work of (Gregoire et al., 2013).

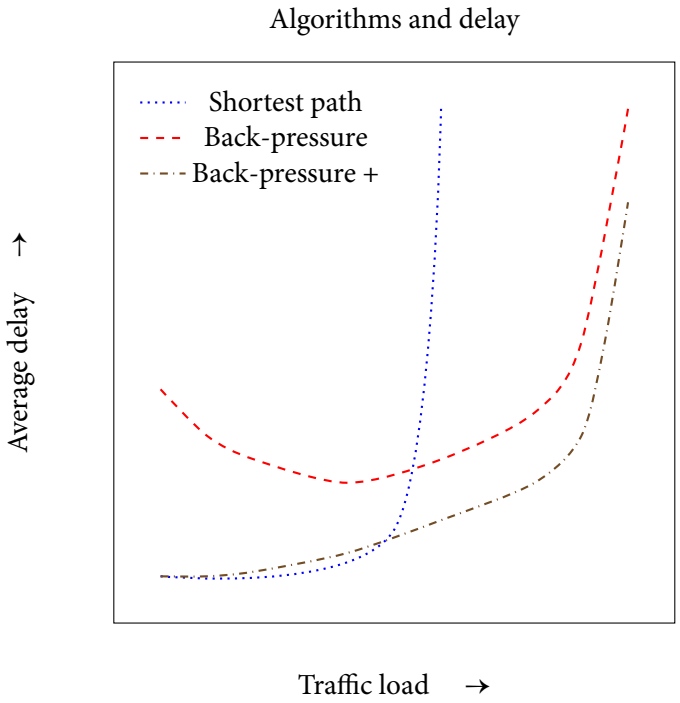


Figure 2.2 – Average delay for shortest path algorithms, back-pressure, and delay enhanced back-pressure adapted (indicative)

Summarizing back-pressure for communication networks

The back-pressure principle originates from the 'maximum throughput policy' by Tassiulas and Ephremides (1992), who used it to send data with various destination through a radio network, based on the pressure gradients of queued data. Since then it has been further developed. It turns out to be suitable for distributed control and in wireless networks. New methods have dealt with delay issues, for example by adding a distance bias to the pressure values, or with finite queues. Research is also done to reduce the complexity in case of bigger networks and with a large amount of data commodities.

2.2.2 Back-pressure for signalized intersection control

The back-pressure algorithm and its modifications have mainly been applied to (wireless) multi-hop data communication networks, as described in the previous subsection. Recently, the concept has been introduced to road traffic control, and to the problem of signalized intersections in particular.

The main design problem of a signalized intersection controller is to "determine the phase for each junction to be activated during each time slot such that the network throughput is maximized" (Wongpiromsarn et al., 2012). Other objectives could be reducing delays, minimizing fuel consumption, or improving traffic safety. To account for traffic heterogeneity priorities could be assigned to certain roads, directions, or modalities (e.g. public transport).

In most countries, conventional signal control systems work with fixed, pre-timed schedules for isolated intersections. Since they don't suffice in case of high and changing traffic loads, more advanced systems have been developed: enhancing fixed plans, adaptive control by vehicle actuation, and coordinated systems for synchronizing phases and offsets. However, the more advanced traffic

control systems usually come with increasingly complex optimizations methods (linear or dynamic programming, model predictive control, etc.). Many systems solve the optimization problem in a centralized way, making it barely scalable (Le et al., 2013).

The main benefit of back-pressure control for signalized intersections lies in the fact that it allows for achieving stabilized maximum network throughput[5], by means of decentralized calculations with low complexity involved. For each intersection, reactively chosen control policies are based on (estimated) queues on roads leading to the intersection as well as queues on outgoing roads that lead to neighbouring intersections. This last term contributes to the stability, and implicitly enables coordination[6] between junctions. Furthermore, back-pressure is 'myopic', it doesn't require a priori knowledge of turn ratios and demands or predictions of near-future demands [7]. Besides, traffic networks with back-pressure control are highly scalable, as it allows for incremental expansion and local modification.

Back-pressure control has the potential to reach a high performance compared to other advanced traffic control systems (Varaiya, 2013a). The advantageous properties are in line with the benefits of the earlier works on back-pressure. In contrast to wireless networks, where transmission interference causes computational complexity, road networks can be distributed straightforward (Le et al., 2013). However, translating the back-pressure concept to road traffic networks reveals some other complications and comes with adaptations.

Basic models of back-pressure for signal control

Despite the similarities, the network topology of a road traffic is quite different from the (wireless) communication network. In the latter, nodes are routers (or hops) containing queuing packets, and links are service providers or wireless connections to move packets (almost instantly) from one hop to the next. The (macroscopic) road traffic network model consists of (more or less) uniform road sections, which are connected by (virtual) nodes, that usually put constraints onto the traffic flow (capacity, merging and diverging). In common traffic modelling, links (road sections) are used to store the traffic while moving, instead of the nodes (in wireless networks). Therefore, caution has to be taken when modelling a road traffic network for back-pressure control. In literature on traffic control back-pressure, there is not yet a clear consensus on what topology to use.

Here, the following network model is adopted[8], see Figure 2.3. Road sections are called nodes (with dimensions!) and denoted by $N_a \in \mathcal{N}$, for the case of node a. Each intersection contains incoming and outgoing nodes, connected by links, let $l_{ab} \in \mathcal{L}$ be the link connecting node a and b. Nodes and links are unidirectional, making the network a directed graph. Q_{ab} is the queue in N_a going to N_b. The total amount of queueing vehicles in N_a is $\sum_c Q_{ac} = Q_a$. Vehicles enter the network at entrance nodes $\mathcal{O} \in \mathcal{N}$, follow there way through the network along nodes (and queues), routed by intersections, and leave the network at their destination node $D \in \mathcal{D} \in \mathcal{N}$. In the majority of literature, entrance nodes are specific nodes at the border of the network, however in (Gregoire et al., 2013, 2014) every node has an arrival rate (think of this as an abstract way to include traffic from minor roads that feed the main network from neighbouring origins); the same goes for exit nodes.

[5] At least in stationary conditions

[6] Coordination in the sense of balancing traffic flows, not as in optimizing signal off-sets

[7] Opmerking over maken. Dit geldt, als BP als simulatie wordt gebruikt gebruikt het zichzelf voor updating. En opmerking over dat resultaten wel beter kunnen zijn als je de demand en turn ratio's wel weet, maar dat geldt voor het moment zelf, niet a priori of voorspellend.

[8] The models and algorithms out of the various literature will be conformed as much as possible with this network.

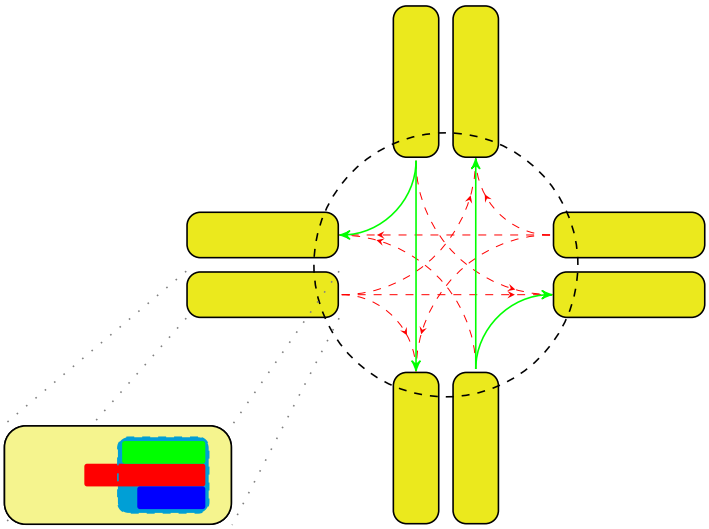


Figure 2.3 – Back-pressure structure representing an intersection

Note that also the composition of network elements is different from multi-hop networks. Compare the irregular grid of 2.1a with the grouped formation of nodes and links of the intersection. The elements of the intersection belong to one controller, that determines which links to 'activate'.

Interference in wireless multi-hop networks prohibits packets to be sent simultaneously. Likewise, in the traffic context, vehicles may not make simultaneous movements if these can cause collisions (Varaiya, 2013a). Each intersection has a collection of activation sets $\mathcal{P}$, that represent possible (service) phases. A phase $p \in \mathcal{P}$ consists of traffic movements without conflicting streams.

When a link l_{ab} gets activated, it transmits a traffic flow with a saturation rate of s_{ab}. The service rate $\mu_{ab}(p)$ is typically either zero or the saturation flow: $\mu_{ab}(p) \in \{0, s_{ab}\}\ \forall p$, assumed that all offered traffic can be accepted by the node downstream.

According to (Varaiya, 2013b; Varaiya, 2013a) the back-pressure[9] algorithm is a sequence of three main steps:

1. At the beginning of time slot t, determine the *weight* of each traffic movement l_{ab}:

$$W_{ab}(t) = Q_{ab}(t) - \sum_{c} r_{bc} Q_{bc}(t),$$

   where c represents the next road after b, and r_{bc} the portion of traffic through b that goes to c. This is a measure of queue backlog differential, the difference between the upstream and downstream queue lengths.

2. Calculate the (summed) pressure y for each admissible phase[10]:

$$y_p(t) = \sum_{a,b} \mu_{ab,p} W_{ab}(t).$$

[9] They call it "max pressure control", and in a previous work (Varaiya, 2009) "universal feedback control policy" (UCP)

[10] Varaiya (2013b), Varaiya (2013a) describe the use of a signal control matrix, which makes notation a bit more complex. The underlying principle is the same.

This is the sum of the service rate of each traffic movement in this phase, times the associated weight.

3. Activate the phase with the highest pressure, p^*, in time slot t:

$$p^*(t) = \arg\max\{\gamma_p(t) | p \in \mathcal{P}\}.$$

If two phases result in an equal pressure, the policy should make a random choice.

Ad 1. Notice the use of an *average* queue on the downstream road b. For road networks, it is hard to compare queues because dimensions and buffer space should be taken into account (more on alternative pressure functions in following paragraphs). Furthermore, a distinct difference with the 'traditional' back-pressure is that destinations are not taken into account explicitly. Lastly, notice that traffic leaving the network in node b is not included in the downstream queue, which makes sense.

Independently of the work of Varaiya, Wongpiromsarn et al. (2012) has come up with a similar algorithm. Some differences in modelling aside, their calculation for the weight is:

$$W_{ab}(t) = Q_a(t) - Q_b(t).$$

Although the first method intuitively results in better performance (because Q_{ab} is more specific than Q_a, the second term Q_b is basically the same for both), as explained in (Gregoire et al., 2013), in practise only aggregated queue lengths might be available. Besides, in the described model, distinct signalling for all directions is assumed, which is also not always the case in real situations.

The approach taken in (Le et al., 2013) is probably a more correct way for defining the weight, by adopting a turn probability to estimate the relevant downstream pressure. In terms of the previous formulations, it would look like:

$$W_{ab}(t) = Q_a(t) - \sum_b r_{ab} Q_{ab}(t),$$

All these algorithms are simple and can be implemented *locally*, for each controlled intersection. Moreover, both policies have been proven to be stable, whenever demand remains inside the capacity region.

In a simulation by Wongpiromsarn et al. (2012) the results of applying the algorithm are promising. Compared to the widely used adaptive SCATS system, both average and maximum queues were significantly smaller. Insufficient supply of split plans by SCATS seems to be a main reason for the growing queues in that simulation. One could argue that SCATS would have performed better if the set of split plans were larger, but this also shows the versatility of the back-pressure algorithm.

Limitations of the basic model

Uit Wongpiromsarn et al. (2012): The algorithms of (Varaiya, 2013b; Varaiya, 2013a; Wongpiromsarn et al., 2012) are the basic formulation of back-pressure for intersection control, but these approaches face some limitations.

Firstly, the assumption of a *single-commodity* network is unrealistic, as it has no concept of multiple destinations with O-D specific traffic demand. Instead, fixed turn probabilities are adopted. In stationary situations, these calibrated turn probabilities can mimic route choice behaviour, however O-D flows are dynamic, and so is route choice in general. The authors propose an adaptive approach,

where turn probabilities are updated by measurements, and actual service rates μ (deviating from the theoretic value) as well.

Determining the *(turn-specific) queue lengths* is another weakness. As already mentioned, turn-specific lanes are usually short, on the major part of the road the queue get mixed. Related to this, in the basic model, using point (or vertical) queues, there are *infinite buffers*. In reality, vehicle storage on lanes in a node is limited, causing queues to back-propagate onto previous nodes. This should be taken into account.

Not mentioned by (Varaiya, 2013b; Varaiya, 2013a; Wongpiromsarn et al., 2012) is the fact that delays are not taken into account when moving traffic through the network. The only objective is to maximize throughput.

Pressure functions are fully determined by absolute queue numbers. There is not yet a way to add weights, to cope with small buffer sizes or give priority to certain phases or flows.

When making signal plans, *fairness* plays a role. The current algorithm only allows 'green' to the phase with highest pressure, which might prevent some traffic movements, leading to long waiting times (and driver frustration). Each phase could have a minimum duration to solve this.

In real networks roads are of different sizes, and travel times over each node varies. This is not taken into account, time slots are chosen to be uniform. Varaiya (2013b), Varaiya (2013a) propose to split up long links into smaller ones. From another point of view, if you only use the model for control and not for simulation of traffic itself, it doesn't really matter. As long as you can incorporate valid queue estimates. Adding intermediate nodes (links) would, however, provide a delay in reaction to downstream circumstances (which could be positive or negative).

The basic models do not include features to coordinate offsets and cycle times between intersections.

Next paragraphs will discuss solutions found in literature for some of these limitations: modelling traffic demand, queues and commodities, routing, delays, finite queues, pressure functions, and cycle times.

Traffic demand, queues and commodities

An important modelling choice, when setting up back-pressure control, is how do you choose the representation of traffic demand, and of queues?

As opposed to the case of wireless networks, back-pressure approaches for traffic control don't usually treat destination commodities specifically. Instead, the traffic demand is routed with fixed or adaptive turn fractions, and for each node a distinct queue is kept for each traffic movements, representing the lane(s) going to a specific outlink.

Notice that there are situations where one lane serves two or more outgoing links, for example going right and going straight. There are also cases where a phase allows a certain amount of crossing traffic. These cases are not specifically dealt with in literature.

In (Gregoire et al., 2014) an approach is proposed that uses aggregated queues for each node, and for each traffic movement a specific demand based on detector measurements close to the intersection. This should add to more realistic assumptions on the available queue measurements, because dedicated lanes for turning vehicles typically only exist near the intersection. Further upstream, traffic (and queues) are mixed for all turning lanes, making it impossible to determine each vehicle's route. Their detector variable $d_{ab}(t) \in [0,1]$ is defined as:

$$d_{ab}(t) = \min(1, Q_{ab}/s_{ab})$$

where s_{ab} is the saturation flow, and Q_{ab} is the queue on node a to b, at proximity of the intersection. d_{ab} is a measure for the amount of traffic demand, bounded by the saturation flow. In a traffic simulation of a (simplified) grid network, this approach performs not as good as the setup with all routing rates (queues Q_{ab}) known (regarding the range of stably controlled arrival rates). However, its performance is still sufficient in many cases, and its measurements model much more realistic.

In literature queues are either aggregated (Q_a) or specified for the next node (Q_{ab}), but the nodes always represent the total road section. In (Le et al., 2013), a different approach is used. Here, roads with multiple lane groups (going to different turns at the next intersection) are modelled like separate parallel roads. This way, upon leaving an intersection vehicles are explicitly routed for a specific turn at the next intersection. This approach is the opposite of the one in (Gregoire et al., 2014).

Routing and back-pressure

Unlike scheduling in wireless networks, where packets with a certain destination are guided along the paths that provides them service (without getting in a physical queue to one of the links), car traffic gets split over a number of lanes, each connected to a specific path.

In a large part of the studies found in literature, fixed splitting rates are used to guide traffic along the network to their destination. Since this is a simplified version of reality, various attempts have been made to enhance the model of routing.

Gregoire et al. (2014), as already discussed, uses a detector to measure traffic demand and make realistic assumptions on the queue for each turn. Another way to estimate actual splitting rates is to apply a function that online updates the estimation with measurements (Le et al., 2013).

These approaches try to capture traffic flows and splitting rates that take place, reactively. Yet, they do not attempt to influence route choice behaviour, by guiding vehicles to a certain direction. However, this is a projected feature of the algorithm that is investigated in this thesis. This direction of research is supported by Gregoire et al. (2013), who recommend to consider the feedback loop between signal control and driver behaviour, and in particular rout choice. The work of (Zhang, 2012) presents a concept for 'active' routing (without predetermined splitting rates, more in the sense of traditional back-pressure), and also considers user satisfaction of the chosen routes. In practice, active routing can only be implemented up to a certain degree, because the necessary technical facilities are not yet (fully) available, and a certain amount of vehicles will choose their route by habit.

Delays

Zhang (2012) is, to the author's knowledge, the first and so far the only to take 'user satisfaction' (travel time delay) into account. In their algorithm routes are chosen for all vehicles in each node and take into account traffic load (as in the basic back-pressure algorithm) and user satisfaction (shortest paths). The logical idea behind this is that by balancing the two components, great performance on both users and the traffic system can be achieved. Drivers will only take detours when they are necessary.

User satisfaction can be expressed as the degree to which an (individually) estimated travel time is exceeded, and traffic load as the variance of all links' density. The "User Satisfaction-Traffic Load" problem is then defined as maximizing the user satisfaction for all users (as function of chosen routes) minus the traffic load, both terms scaled to a trade-off parameter. A new back-pressure algorithm, BPR-US, can approach the optimal solution of this problem.

In their model, queues are aggregated and each vehicle being first in the queue gets routed following the BPR-US algorithm. First, the car is assigned to each possible node with a probability propor-

tional to its pressure difference. A degree of urgency is calculated, based on travel time spent on the route so far and the predetermined expected travel time ($\min(1, \frac{\text{used travel time}}{\text{expected travel time}})$). The probability for each node is then determined as a trade-off between the 'pressure probability' and the 'urgency', where the urgency value becomes (1 – urgency) for all routes longer than the shortest.

This algorithm seems not the best one possible. User satisfaction starts to significantly influence route choice only after a considerable part of the route has been travelled, which is of little use if route choice is decisive at an early stage. Secondly, positively relating the urgency to the shortest route only seems insufficient if routes are comparable in travel time. A comparative choice of realistic travel times seems more plausible.

Nevertheless, the concept of trading off traffic load and user satisfaction is interesting and should be investigated further on. A simulation by (Zhang, 2012) shows that results are as expected, average and particularly individual travel times are less than the traditional back-pressure, at the cost of a little bit more congestion (but preventing detour delays). A shortest path algorithm performs worse in all cases, except where traffic flows are low.

Finite queues

In the basic algorithm of back-pressure queues have infinite buffer sizes. On road traffic networks, queues do have dimensions and a finite size, that can't be neglected. A queue will start to block throughput on the previous intersection as soon as it outgrows its limits. Gregoire et al. (2013) consider this problem of finite buffer sizes, and limited throughput of 'admissible flows'. To quantify the buffer size, they use the concept of queue values for a congestion threshold and the maximum capacity. Basing their algorithm on (Gregoire et al., 2014), they introduce a generalized pressure function, to replace the queue size. Next paragraph will discuss these pressure functions.

Pressure functions

Limited queue capacities, under back-pressure control with linear pressure functions, can lead to inefficiency, when activated services can't be transferred to the next node. In worst cases, deadlock can be the undesirable result. That's why Gregoire et al. (2013) introduce capacity-aware pressure functions, to mitigate congestion propagation.

They choose the pressure functions to be convex, normalized and fair. *Normalized* functions account for node capacity. A relative pressure function ($P_a(Q_a) = Q_a/Q_a^{\lim}$) would be a first step. As the queue grows, each additional vehicle adds more to the problem (assuming buffer overload is to be avoided). Using a marginal pressure, justifies the use of a *convex* function. Regarding *fairness* to choose any path if densities are low, then the marginal pressure should be uniform over the nodes. Finally, to recover the *stability conservation* guarantees, at low densities pressure functions should be linear.

An example function, fulfilling these requirements, would be:

$$P_a(Q_a) = \min\left(1, \frac{\frac{Q_a}{C_\infty} + \left(2 - \frac{Q_a^{\lim}}{C_\infty}\right)\left(\frac{Q_a}{Q_a^{\lim}}\right)^m}{1 + \left(\frac{Q_a}{Q_a^{\lim}}\right)^{m-1}}\right),$$

where m and C_∞ are parameters that 'shape' the function. m determines the stretch of the linear regime and C_∞ influences the slope at small densities, and the steepness near capacity. The authors consider C_∞ a constant, not related to the actual buffer, though intuition tells it should be related. Figure 2.4 shows an example of the function.

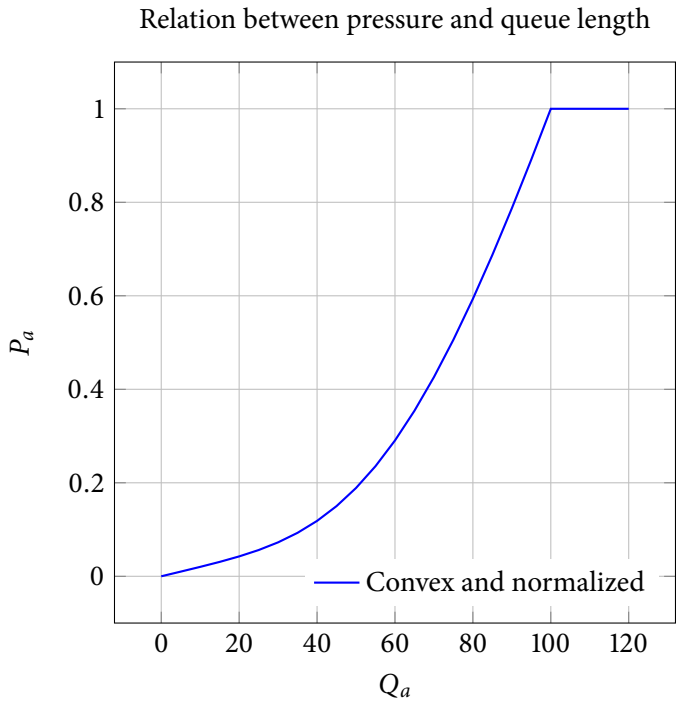


Figure 2.4 – Example of the pressure function by (Gregoire et al., 2013). Parameter $C_\infty = 500$, $Q_a^{lim} = 100$, and $m = 4$.

Using this technique, average queue lengths will be smaller for nodes with small buffer storage, and larger for nodes with more storage. The proposed back-pressure control doesn't guarantee stability under finite capacity constraints, however it will improve the behaviour of the queueing network under finite capacities. Blocking of intersections will occur significantly less. Notice that, also with (only) relative pressures the performance would have probably been increased.

Simulations are executed on a grid network with all nodes having equal capacity, except for a few (bottleneck) areas. Compared to the basic linear pressure function, the new method does a much better job at diverging traffic around the bottlenecks, whereas the linear functions cause system deadlock, back-propagating from the bottlenecks. The probability of instability and propagation of congestion has thus been strongly decreased.

Pressure functions could include more than just finite buffer sizes. For example, one could also add delay values, or priority parameters for preferable directions, modalities (e.g. public transport).

Signal cycle times

One of the drawbacks of the basic back-pressure algorithm is that, since it doesn't work with a fixed cycle time, only the phase with highest pressure gets to be served, which might lead to erratic, unpredictable ordering of phases, where some traffic movements are prevented, leading to long waiting times (and driver frustration). One way to solve this is to set a minimum duration for each phase.

Le et al. (2013) follow another approach and modify back-pressure to have a fixed (and uniform) cycle time, and to allocate a non-zero amount of time to each phase within that cycle.

The algorithm determines the weights (comparable to the basic method of (Varaiya, 2013b)). Then each phase p of the intersection is given a portion of the cycle time, ω_p:

$$\omega_p(t) = \frac{\exp(\eta\gamma_p(t))}{\sum_{\pi\in\mathcal{P}}(\exp(\eta\gamma_\pi(t)))},$$

where γ_p is the pressure of phase p. The exponential function is used to scale the possibly negative pressure values to appropriate positive values, using η as a model parameter (more parameters could probably be added for fine-tuning).

Within a simulation four policy types are considered: non-fixed cycle proportional policy[11], fixed cycle proportional policy, non-fixed cycle (basic) back-pressure , and this paper's fixed cycle back-pressure policy. As expected, both back-pressure algorithms outperform their proportional counterparts on throughput. The difference between the two back-pressure variants is not obvious. For practical applications the fixed cycle time back-pressure appears to be a better choice. Furthermore, other variants could be proposed, such as optimizing the cycle time per intersection, designing cycles with custom phase ordering, using flexible cycle times, etc., and issues such as loss time influence on the performance should be further investigated.

Summarizing back-pressure for traffic control

Back-pressure control for signalized intersections is based on the same concepts as back-pressure control in multi-hop communication networks. There are differences in the network and queueing model, but in the way the algorithm works as well. In the found literature on traffic control, focus is on finding the right phase, which signals to turn green, whereas in multi-hop networks focus is more on moving the right packet through the right servers. Both result in maximum throughput.

The back-pressure methods discussed in this subsection are compared in Table 2.1. The basic back-pressure algorithm has been expanded, and the modifications tend to improve the practical use of the algorithm.

For the application of route guidance, probably a new algorithm should be formulated, that combines existing routing methods with the relevant features of back-pressure, where active routing, delay consideration, and pressure functions for finite buffer sizes will play an important role. Next subsection further examines the possibilities for the back-pressure algorithm for route guidance.

2.2.3 Back-pressure and route guidance

So far, there is no literature on applying back-pressure to the kind of route guidance that is the topic of this thesis. However, there are elements in existing work that can be used. In this subsection characteristics of combining back-pressure and route guidance are discussed, and a set of old and new elements that will be the starting point for the new algorithm is presented.

Similarities and differences with back-pressure for signal control

For both multi-hop communication networks and road traffic networks there is a range of back-pressure models, which have their own strength and weaknesses. In the previous subsections, these have been discussed.

[11] The proportional policies base their decisions on the amount of traffic entering the intersection, without considering the queues on roads going out.

Table 2.1 – Comparison of various types of backpressure

Source	Queue in	Queue out	Pressure sfunction	Routing	Control mech.	Special features
(Varaiya, 2013b; Varaiya, 2013a)	Sp[1]	Av[4]	Qd[5]	Ft[7]	BP[10]	none
(Wongpiromsarn et al., 2012)	Ag[2]	Ag	Qd	Ft	BP	none
(Zhang, 2012)	Ag	Ag	Qd	Ar[8]	BPR-US[11]	Delays, active routing
(Gregoire et al., 2014)	Ags[3]	Ag	Qd	Ft	BP	Detector measurements
(Gregoire et al., 2013)	Ags	Ag	Adv[6]	Ft	BP	Finite queues, pressure functions
(Le et al., 2013)	Sp	Av	Qd	Eo[9]	Fct[12]	Fixed cycle time, proportional green time for each phase

[1] Sp: Splitted
[2] Ag: Aggregated
[3] Ags: Aggregated queues and split detector values
[4] Av: Average
[5] Qd: Queue difference
[6] Adv: Advanced means here: normalized convex pressure functions
[7] Ft: Fixed turn prob.
[8] Ar: Active routing, Routed with BP, probability based on pressure and urgency (delay)
[9] Eo: Estimated online
[10] Activate phase with highest pressure, each time slot
[11] Each individual vehicle routed with BP, probability based on pressure and urgency
[12] Fct: Fixed cycle time, phase splitting

In general, the back-pressure algorithms for multi-hop communication networks solve the question of which particle (of which commodity) to send to which servers, based on the queue length differences. For road traffic and signalized intersection control, the main task of the back-pressure algorithm is to decide which signalling phase to allow green. The main difference here is that the signal control application doesn't include routing, instead it assumes a certain splitting rate (or estimates it from measurements). For the application in this thesis, route guidance is a main feature, so the examples of the multi-hop communication networks might be more applicable than the road traffic examples. An exception is the work of (Zhang, 2012).

A characteristic of the signalized intersection is that each intersection has one controller for all the traffic movements at the intersection, to switch them on or off. This concept doesn't match to route guidance. In case of route guidance, the only control possible, is to influence route choice. Traffic can't be stopped by route guidance, unless by congestion. A consequence is, that a good route guidance controller considers the 'conflicting' traffic streams, that might merge a shared road ahead. Also related to this, is that coordination between points of route guidance is needed, to reach a good 'balance' between the flows.

Requirements for route guidance back-pressure

This list of requirements is a ground for collecting the elements that can be used for existing algorithms, and for pointing out the elements that need to be added.

- *Maximum throughput*: This a key feature of back-pressure control. For the algorithm in this

thesis, it is important to have a high throughput in the network, but it is not all decisive, as other factors are important as well.

- *Vehicle delay*: Vehicle delays need to be taken into account, and in particular the (O-D) travel time delay compared to the fastest possible route. Delays at individual links (bottlenecks) are expected to decrease by the spreading nature of back-pressure. This relates to the concept of *fairness*, or *exploitation*.
- *Routing*: Route guidance is about influencing route choice and thereby influencing traffic flows and the network state. To take this into account, turn probabilities should be variable. Also, a factor of *compliance* needs to be considered. O-D routing implies that the traffic model uses a *multi-commodity* representation of traffic flow.
- *Proportional route assignment*: Specifying route guidance, the proposed back-pressure algorithm should have proportional route assignments as outcome. For example, at route guidance location *A* 60 % of the incoming (macroscopic) traffic flow is directed to route 1, and 40 % to route 2.
- *Finite queues and pressure functions*: As each road section has its own properties (flow capacity, maximum vehicle buffer size), and congestion can spill back from one section to the previous, it is important to take the finiteness of queues into account and include it in the pressure functions. Furthermore, the pressure functions shouldn't be based on queues (vehicle standing at a node) but on measures that are realistic for highway sections, such as *density or the NFD*.

  Additional weight factors or priorities could be included in the algorithm, but have no priority at this point.
- *Long distance pressures*: Taking only the first next road section into account is not sufficient, when routing vehicles, especially on a large scale highway network, where routing can only be done at a limited amount of locations. Therefore, the whole route (including possible 'siblings') should be taken into account, and be represented in the pressure function.
- *Other aspects of realistic traffic*: The algorithm should comply to realistic traffic flow as much as possible. An example is that vehicles need to be served in the right order (often FIFO). The macroscopic modelling approach might make this difficult.
- *Coordination*: Ideally, route guidance controllers at various locations coordinate their actions, to result in a better balance of traffic flows. Part of the desired behaviour is expected to emerge already by applying 'isolated' route guidance. Moreover, back-pressure has a benefit of working in a decentralized manner. However, some attention should be given to coordination.

Translating the concepts

The basics of the new model will consist of a macroscopic (multicommodity) simulation model. Furthermore, the basic algorithm of back-pressure will be followed (determining pressures and split them over the network). Based on the requirements above, the following elements can be taken from literature.

- Vehicle delay and long distance pressures: Zhang (2012) shows that traffic routing is possible, but does it for individual vehicles and here another route choice model will be adopted. Taking the whole route into account, the (D-)ORCD approach of (Gupta & Javidi, 2007; Naghshvar & Javidi, 2010) seems to have potential, otherwise elements of the shadow queues architecture might be chosen (Bui et al., 2009; Athanasopoulou et al., 2013).

- Finite queues: The way (Giaccone et al., 2005) and particularly (Gregoire et al., 2013) apply the finiteness of queues, resulting in enhanced pressure functions, is a good example to follow.
- Reducing complexity. Clustering of destinations (Ying, Shakkottai, et al., 2011) can reduce complexity of the network flows, and could be used, although it is not a requirements in the first steps of modelling.
- Proportionality: Le et al. (2013) shows how cycle time can be divided among phases. The concept of scaling pressures is something that can be converted to a model of determining splitting rates (turn probabilities).

Apart from these useful concepts, the algorithm for route guidance is still an open question. The route choice or assignment model to be applied is still to be chosen, as well as the way to determine route alternatives, and their properties. It's probably a good thing to choose a basic method and expand it from there.

Summarizing

In this section the back-pressure and its various variants has been discussed. Although the back-pressure algorithm is not inherently suitable to model route guidance for highway traffic, with some modifications and extensions it could be a modelling approach with a lot of potential. It should be taken care of, though, that the main benefit (maximum throughput) is not affected too much by the proposed modifications.

2.3 Summary of the literature review

The literature review starts with a brief overview route guidance. Route choice is used to manipulate route choice to reach a system or user optimum. Applying route guidance has advantages and pitfalls. Often route guidance is used to minimize the travel time of the individual road user; route guidance to improve the conditions for the system as a whole isn't often done in operational route guidance.

The back-pressure control methodology originated from communication networks, where it is applied to the problem of delivering data packets via a network of nodes and servers. The back-pressure algorithm assigns the servers to be activated and the data they should transfer, based on server capacity and queue differences (queues for specific data groups). The main strengths of back-pressure control are maximization of throughput and network stability. The basic back-pressure algorithm can been extended to cope with finite queues and delay.

On the application of traffic signal control back-pressure control has been used in literature. The concept aims to activate the signalling phase that has the highest total weight, a summation of the weights of the allowed traffic streams. The weight is the product of pressure and service rate (saturation flow). The pressure of a traffic stream is the difference between the queues at the incoming link and outgoing links (weighted by proportion). There are several variants to this basic model.

In communication networks the use of routes is a result of the algorithm, whereas for traffic signal control the route choice is used as an input from a separate process. The back-pressure concept can however also be used as a method for route guidance. Instead of determining the right phase to be activated, the task is to determine the ratio to direct traffic to following links. Route pressure values can be formulated to expresses how filled up the route is. Route pressure and service rate are less straightforward to define than for the intersection control. A choice model is to be used to determine the ratio of routes, based on the product of route pressure and service rate.

Chapter 3

Research approach

The research approach takes a central position in this thesis. The introduction (Chapter 1) formulated the problem, the objective and research questions. Then, the literature review (Chapter 2) provided the theoretical background. It presented the origins and state-of-the-art, and defined the missing parts necessary to develop a new approach of traffic control. This chapter takes these previous two as an input to further define and demarcate the research and specify its direction.

Section 3.1 defines the general framework of the research, including its purpose, the basic concept, and a introduction to the scope and requirements. In Section 3.2 design principles and the scope for the traffic control methodology in this thesis are stated . The evaluation of the proposed traffic control methodology is to take place within a certain scope, which is described in Section 3.3, specified for the experiment set-up and the simulation model that goes along with it.

3.1 General research framework

As a first step of defining the research approach, this section provides an overview of the system to be developed and the requirements it should satisfy. It starts with the objective and purpose of the system, then presents the components of basic system design, followed by structuring the system requirements, which are specified in the next sections.

3.1.1 Starting points

Research question 5 holds the main task for this chapter, namely, to define the purpose of the new control approach, and the design requirements. The main thesis objective can be used as a starting point to elaborate on. In Section 1.3 the objective has been presented as:

- to develop a framework that integrates route guidance and signal control based on the back-pressure principle, and determine its feasibility and potential benefit.

To further specify the objective some general conditions have been formulated. The chosen approach should be generic, computationally efficient and apply (potentially) real-time control, that should take into account individual routing advice and network-wide traffic flow effects. Improving network conditions is the main angle, however, a trade-off with individual road user satisfaction should be considered.

3.1.2 System design concept

The general concept for the traffic control system is outlined in Figure 3.1.

The *traffic process* represents the traffic flows in the network and their dynamics. It is a continuously repetitive process in time, influenced by variability of traffic demand (to enter and leave the network), incidents and traffic control measures. Traffic signal control and route guidance are the two types of traffic control that this thesis focusses on, as shown in the diagram.

With a certain time interval *state estimation* is carried out to update the traffic control measures. The *traffic signal controller* determines the new settings for the traffic signals at intersections, and the *Route guidance controller* does the same for route advice. Both controllers can have a different time interval for their updates.

The *Traffic signalling system* and *navigation system* are responsible for executing the chosen control policy by implementing the given settings onto the actuators. This directly influences the traffic process, which closes the control loop.

In the ideal situation route guidance and traffic signal control work together as one traffic control system. In the literature study (Chapter 2) it was found that there is not a 'standard' way to integrate the two control parts. As a first step to full integration, traffic signal control can be made to depend on route guidance. If the routes of all vehicles (as a result of route guidance) are 'known', the traffic signal controls can be set to optimize the accommodation of these routes.

The aim for the traffic control system is, corresponding to the idea of back-pressure, to optimize throughput at intersections (traffic signal control) and the network as a whole (route guidance). Secondly, stability properties, the amount of traffic that can be handled without escalating queues, should be improved. This implies that the angle of control is system performance. On the other hand, individual road user satisfaction should be taken into account when developing the control method.

3.1.3 Categories of requirements

The project requirements can be divided into several categories (see Figure 3.2). First there is the objective and main principles, as described in Section 3.1. The design principles for the control methodology further specify the objective and concept of Section 3.1.2. These design priciples are explained in Section 3.2, and define the base for the traffic control part of the research, where a new theoretic model for the traffic control method will be developed.

The control method should be evaluated by means of experiments, in order to determine if it works as it should, and what the effects on traffic are. The simulation environment needed to do the experiments is subject to requirements to make the experiments possible, and the experiment requirements depend on the traffic control method and the research questions to answer. The scope of the experiment and simulation is described in Section 3.3.

3.2 Design priciples for the control methodology

In this section the 'design principles' for the control methodology are explained. These include the type of control, the main features, and other design choices and assumptions. Also, the limitations of this approach are briefly discussed.

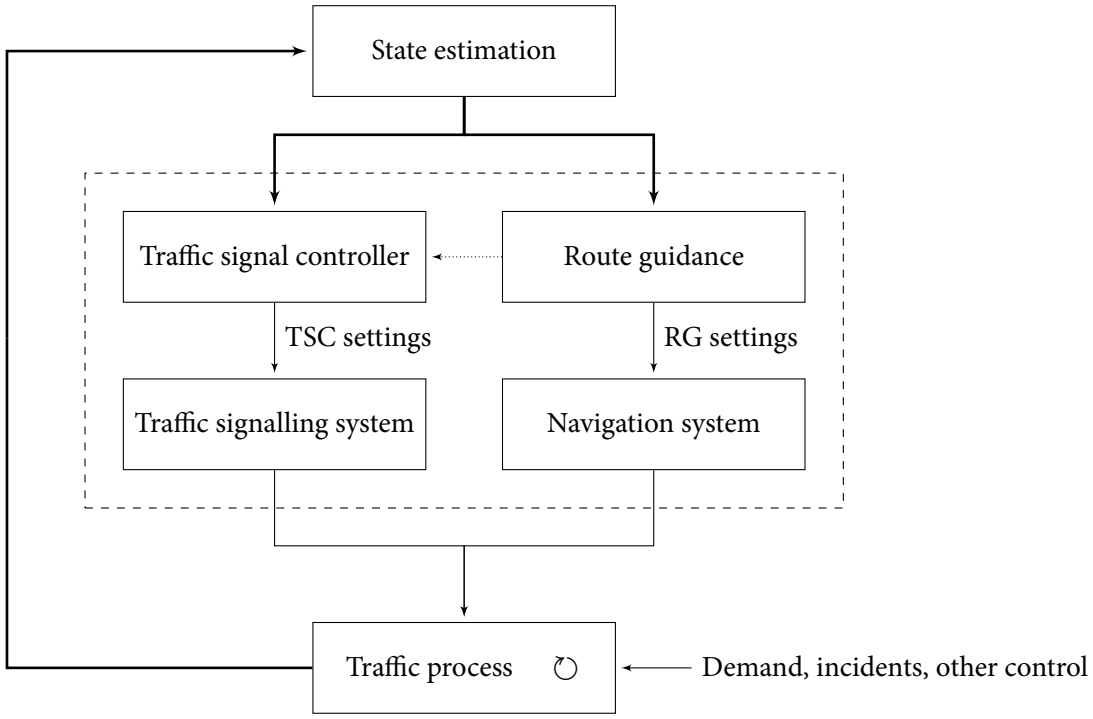


Figure 3.1 – Control approach flow chart

3.2.1 Type and features of the control approach

The traffic control method to be developed in this thesis is based on the concept of *dynamic traffic management* and *real time* control, which means that it (at least theoretically) can be used for on-line, operational traffic management. For each time period (typically 1–15 minutes) the traffic control settings are updated and implemented (see the outer loop of Figure 3.1). This is different from off-line optimization, where control settings for each time period are found by iteratively doing complete simulations and keeping the best results.

In the Introduction, the paragraph on 'Control approach' (page 8) described the range of possible types of control. The choice is made to use the concept of *reactive feedback control*, instead of others, such as model predictive control. Reactive feedback control is a simple approach, which naturally fits to the *back-pressure* type of control. Although predictive methods can anticipate on traffic dynamics and predict upcoming problems, back-pressure control has the characteristic to stabilize traffic flows in an *adaptive* way, and thus is likely to prevent many potential problems. Not only recurrent situations should be solved adequately, also in case of incidents the control approach should come up with solutions that improve the traffic state.

As mentioned earlier, the purpose of the traffic control approach is to improve the *system performance*. User satisfaction should be considered, but doesn't have the highest priority in this research, where examining the working of back-pressure algorithms is the main task.

Looking at the applicability and general use of the model, it is important to develop a *generic* model that can be used on various types of networks. Furthermore, the *scalability* of the control method is an issue, as network-wide control is part of the objective. Back-pressure has characteristics of distributed control, but the challenge is to translate this feature to route guidance.

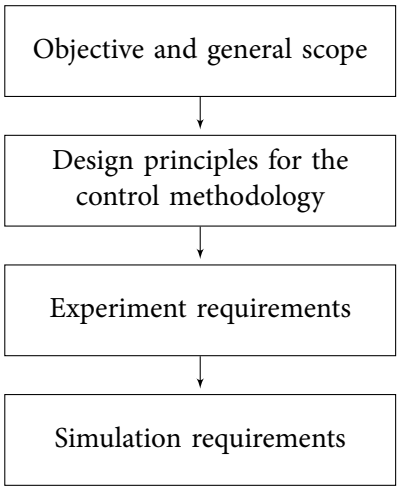


Figure 3.2 – Relations of dependence between categories of requirements

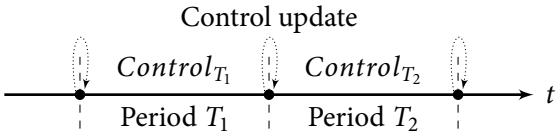


Figure 3.3 – Concept of real-time control

As a last feature, the control approach should be overall *efficient* at making calculations, in order to make it suitable for real-time control.

3.2.2 Design choices for traffic signal control

Regarding traffic signal control a few simplifications are made.

- Cycle times are fixed, as well as the order of signalling phases.
- As a different tactics, signalling phases can be assigned per time slot, without considering cycle times and phase ordering.
- There is a predetermined set of signal control phases for each intersection.
- Loss times to clear the intersection are ignored (assumed zero).
- Minimum green times, and maximum waiting times can be used.
- Coordination between intersections is not taken into account.
- Measurements of traffic flows are assumed to be available as needed.

For the evaluation of traffic flows at intersections (and their dynamics) it is necessary that each signalling phase can be simulated distinctively. This implies simulation steps of several seconds at maximum.

The control parameters for traffic signal control are timings for each signalling phase.

The way back-pressure is applied to traffic signal control is determined in Chapter 4. The approach in this thesis is to use macroscopic measurements to determine the pressures.

3.2.3 Design choices for route guidance

Route guidance in this thesis complies to the following design choices.

- Vehicles are assumed to be equipped with navigation systems (in-car systems) that can communicate with the network's traffic control system.[1]
- Route guidance is used in an en-route manner, which means that along its trip from A to B a vehicle can switch routes at every location where that's possible.
- At each link where traffic has (at least) two follow-up links to choose from route guidance can be applied.
- The action of route guidance on a link can be modelled as a split fraction (or turn fraction) for each destination.
- Coordination of route guidance is a complicated topic, and will considered but not fully addressed. The problem of overreaction is one issue to deal with. A distributed approach of coordination is preferred (suits the back-pressure concept, makes the control system scalable to large networks).
- Optimizing the network performance is a main characteristic of the back-pressure approach. User preference (shortest routes) has no priority in this research, but incorporation needs to be considered to some extent.

The effect of route guidance measures highly depends on the degree of vehicles following the given advice. The number of 'followers' is mainly determined by the penetration rate of the in-car system (how many percent of the vehicle is 'connected') and the extent to which drivers are obedient to the advice. Thorough research of driver choice behaviour is outside the scope of this research. Therefore assumptions are made. As a first step, to test the effect of the methodology, all vehicles are assumed to respond to the route guidance, 100% compliance. Later on the effect at other compliance rates will be looked at.

Another simplification is that traffic flows, including their destinations and split fractions, are known to the route guidance system, so they are highly controllable. The necessary information can be received from the simulation model. To put this into practice requires further research on estimation of origin–destination flows and route choice. In the future however, more and more cars can be expected to be equipped with advanced in-car systems, and their data could be useful to systems like the route guidance system studied here.

The control parameters for route guidance are the split fractions on each (relevant) link specified for each destination.

The main challenge is to translate the concept of back-pressure control to the application of route guidance, which is not an obvious task, and is discussed in Section 2.2.3. Just as for the case of traffic signal control, it seems logical to use macroscopic measurements to determine pressures, and these pressures should represent the routes rather than one link.

3.2.4 Limitations

Some aspects are outside the scope of this research.

[1] The necessary requirements that stem from technology, organization and social acceptance, are not further investigated, rather the aim is to clarify potential impact and benefits.

- This thesis doesn't focus on the practical implementation of the model.
- The control method doesn't consider prediction, but uses a reactive approach.
- Coordination between intersections is not taken into account.
- Full coordination of route guidance is also outside of the scope.
- Only uncomplicated intersection configurations are modelled.
- Effects on environment and safety are not considered explicitly, focus is on traffic performance.

3.3 Scope of simulation and experiment

3.3.1 Experiment requirements

In Chapter 5 the experiments will be set up and carried out. Here, a few general requirements for the experiments are stated.

- The experiments are used to evaluate the performance of the proposed traffic control algorithms. Questions to answer are: do they work as expected, what is their effect, what are difference between variants?
- First the algorithms for traffic signal control and route guidance need to be tested independently, then they can be combined into one system and evaluated.
- The results of each simulation should at least include overall network performance indicators, and more detailed indicators to show that and how the algorithm works.

3.3.2 Simulation requirements

In short, the simulation model should be well-equipped to mimic the real-world traffic process, in order to evaluate the performance of the proposed traffic control model. Its desired precision and necessary output is mainly derived from the requirements of the traffic control model and experiments (Figure 3.2).

An overview of the simulation requirements:

- Modelling of *traffic flows*:
  - realistic traffic flows along space and time, including congestion dynamics (congestion onset, spillback, densities ans speeds), for example by applying the fundamental diagram;
  - FIFO (first-in-first-out) behaviour, to prevent unrealistic overtaking and diffusion;
  - multi-commodity traffic flows, specified to origins and destinations (to enable traffic control for specific groups);
  - advanced topics such as capacity drop, multi-lane models, hysteresis, and multi-phase traffic flow regimes have no priority.
- Network related:
  - realistic behaviour at nodes (merges, diverges, intersections);
  - applicable to urban and highway networks, of big and small size;

  - route choice should not be based on fixed, general split fractions, but be specific for each destination;
  - the network is not static, but can have dynamic properties, for example to model an incident.
  - Traffic *demand* can be dynamic, and each OD-pair has its own demand.

- On traffic control:
  - traffic signal control can be applied, with discrete green phases;
  - route guidance can be applied on each relevant link and for each destination.
  - traffic flow measurements are suitable as control input;
  - the traffic control methods of this thesis should be implemented in this structure.

- Related to experiments:
  - all required results should be generated (or easily derived) from the simulation;
  - custom networks and scenarios can be built and used in the model;
  - scripts can be used to (partly) automate the experiment process;
  - a warming up period can be added to the simulation, to exclude the influence of starting up the simulation.

- Miscellaneous requirements:
  - the model is transparent enough to analyse how it works, and well documented;
  - the model is accessible to make changes;
  - a proven, reliable, and easily available model is preferred;
  - a simple and computationally fast model is preferred.

3.4 Summary of the research approach

The chosen control approach is a traffic control loop in which the traffic process is monitored, controllers for traffic signal control and route guidance calculate new settings which are implemented to the traffic process by means of the traffic signalling system and in-car navigation system. The traffic signal controller can be fed by route guidance information to improve the estimate turn probabilities. The model is a reactive feedback control type.

The research approach includes a list of design choices and limitations for traffic signal control and route guidance, as well as requirements related to the experiment to be conducted and the simulation tools.

Chapter 4

Control modelling

Now that the research approach in Chapter 3 has outlined the 'playing field' for the research, the development of a new methodology for traffic control can commence. Research questions 3 and 4, considering the use of back-pressure control to apply to traffic signal control and route guidance, have been explored in the literature review of Chapter 2, but need to be elaborated in order to deal with research question 6, the main assignment in this chapter: *'What are the proposed control algorithms and their alternatives?'*

Section 4.1 gives an overview of the methodology as a whole, by presenting the overall system, its individual parts, and the approach to develop the associated control algorithms. Next, Section 4.2 focusses on the theory and model for signalized intersections, and Section 4.3 elaborates the more extensive theory and model for route guidance. Both use back-pressure control as a theoretic foundation. Then, Section 4.4 presents a method to combine the models of traffic signal control and route guidance into one system.

The methodology developed in this chapter will be subject to a simulation experiment in the following chapters.

4.1 Overview of the methodology

Figure 4.1 highlights the main traffic control part that includes the controllers and their in- and outputs. In the next two sections of this chapter the methodology for each controller is developed separately. Then, the two are combined into one system.

Before describing the two subsystems of traffic signal control (TSC) and route guidance (RG), the relation between the two is outlined. Figure 4.2 illustrates the difference in area of influence. In the case of (uncoordinated) TSC each controller serves a local area and aims to best serve its intersection. Route guidance is a control measure that takes into account the whole network (or at least the part that is relevant to get from origin to destination). Although routes can be altered at the local level, and traffic signals can have indirect influence on other parts of the network, route guidance can be considered to have a wider view than traffic signal control. Therefore in this thesis route guidance is treated as the network level control measure, and traffic signal control as the local control measure.

4.2 Traffic signal control based on back-pressure control

Back-pressure control has been applied to traffic signal control before, as found in literature (Chapter 2). In this section a new approach is described. The main difference is the definition of pressures, which is on a macroscopic rather than vehicle level. Furthermore, some alternative methods are considered, to get a broad sense of which approach and what characteristics would work best.

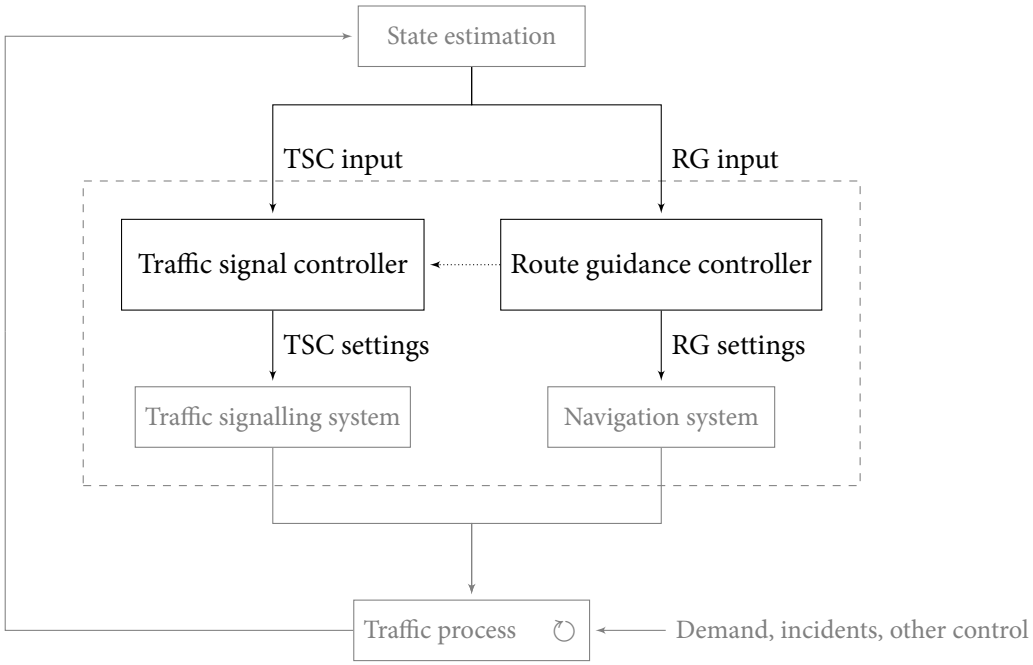
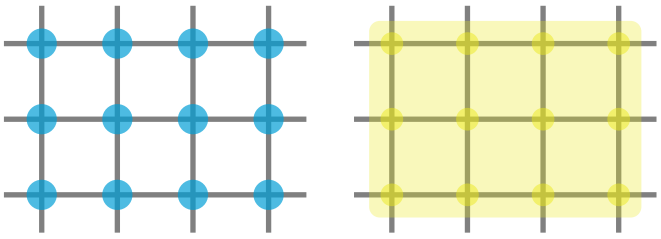


Figure 4.1 – Control approach flow chart



(a) Traffic signal control **(b)** Route guidance

Figure 4.2 – Difference in area of influence (showing a grid network with 12 intersections)

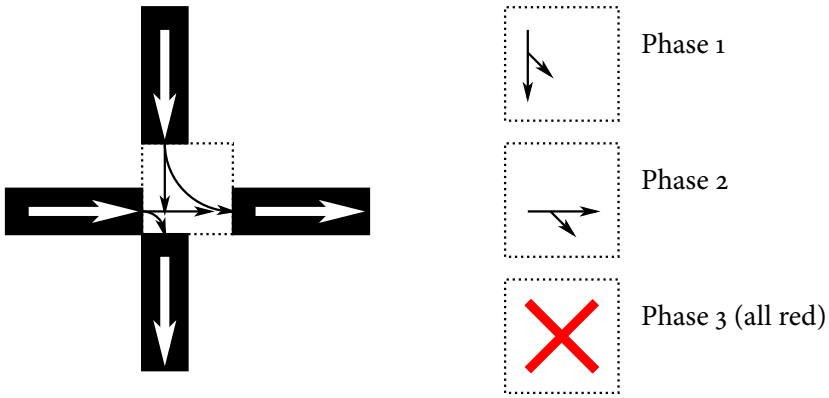


Figure 4.3 – Phases of a simple intersection

4.2.1 Basics of the method and differences with other methods

An intersection, with conflicting traffic streams, needs to be regulated in order to allow traffic flows passing it safely. For cases with light traffic road markings and right-of-way can be sufficient, but heavily used intersections require a roundabout or traffic control by traffic lights (traffic signal control).

Traffic signal control plans are typically based on two steps. First, all traffic streams are defined and conflict groups are determined, which traffic streams can't be given green at the same time. Then, phases are formulated. A phase consists of a group of traffic streams that can have green together. By ordering the phases inside a traffic control cycle, all traffic streams can be served. The time share for each phase depends on the related traffic load. This representation of signal control is very simple and limited, as there are more advanced ways to define signal control plans. In this thesis the choice is made to have simple signal control plans with strictly defined traffic phases. Another simplification is to neglect clearance times in between phases.

The set of phases is $\mathcal{P}$ for every intersection $j \in \mathcal{J}$, but since each intersection is regarded independently, index j is omitted and the phase set is formulated as:

$$\mathcal{P} = \{p_1, \ldots, p_n\} \tag{4.1}$$

Before the concept of back-pressure control is applied to traffic signal control, two other approaches are briefly discussed: fixed and vehicle actuated traffic signal control.

Fixed traffic signal control

Fixed traffic signal control is the most conventional way of controlling traffic on intersections. In this case the signal plan has a fixed cycle time with an order of phases, each having a fixed amount of green time. The (off-line) optimization of this type of signal plan is usually based on a representative traffic demand.

Vehicle actuated traffic signal control

Vehicle actuated (VA) control is a type of signal control where phase times are not fixed, but depend on the actual traffic demand measured at intersection approaches (e.g. with induction loops). Usually

each phase has a minimum green time that gets extended to further release the queues.

In the simulation experiments in Chapter 6 VA control will be compared to back-pressure control. A simple way to model VA control is the proportional policy (Le et al., 2013). Each phase has a summed weight W_p to account for the traffic demand on each approach. This traffic demand usually is the number of queueing vehicles, but in this thesis the relative vehicle density is used (similar to the way of pressure, Equation 4.8). The share of green time for the phase is assigned proportionally.

$$\xi_p(t) = \frac{W_p(t)}{\sum_{i \in \mathcal{P}} W_i(t)} \tag{4.2}$$

Back-pressure traffic signal control

The main control policy to be studied here is back-pressure (BP) control. Based on the findings in Section 2.2 of the literature study an algorithm is developed. The procedure of assigning phase times according to BP control can be summarized as follows:

1. Each traffic stream has a back-pressure value that is based on the pressure difference between the incoming link a and the outgoing one(s), b.

$$BP_{ab}(t) = P_a(t) - \sum_b r_b P_b(t) \tag{4.3}$$

   The pressure of an outgoing link has a weight according to the turning ratio r of that link. The *turning probabilities* need to be determined first.[1]

2. Determine the accumulated back-pressure values for each phase. This is done by multiplying the BP values of step 1 by the available service rate (saturation flow) for each phase p.

$$\gamma_p(t) = \sum_{a,b} \mu_{ab,p} BP_{ab}(t). \tag{4.4}$$

3. *Allocate* the right amount of time to the phase(s) to be activated. There are two main approaches here:

   a) Use a fixed cycle time and divide it among the phases according to a function of their pressures (similar to described for VA control).

$$\Xi_p(t) = \frac{f\left(\gamma_p(t)\right)}{\sum_{i \in \mathcal{P}} f\left(\gamma_i(t)\right)} \tag{4.5}$$

   b) Use time slots and assign the current slot to the phase with the highest pressure (p^*), and repeat the BP procedure every time slot. If two phases result in an equal pressure, the policy should make a random choice.

$$p^*(t) = \arg\max\{\gamma_p(t) | p \in \mathcal{P}\}. \tag{4.6}$$

[1] If incoming links are (modelled) to be connected to just one outgoing link, this step isn't needed.

In case of all negative pressures at the intersection, theoretically all signals can be put to red. This depends on choices for the control method, and will be elaborated on in the next subsections.

The steps and variations on this general method are further explained in Subsection 4.2.2–4.2.4. This description of the back-pressure algorithm ends with a short discussion on the turning probabilities.

Turning probabilities

In step 1 of the algorithm turning probabilities are used to determine a representative value for the pressure of the downstream links. This turning probability can be determined in various ways. One could use historic data, for example: at this approach of the intersection 60% goes right and 40% goes left. For dynamic control it is better to estimate the turning probabilities from real-time data. In this thesis we consider two methods.

The first method counts traffic on the intersection. After each interval of 2 minutes the turn probabilities are updated based on counts of the last 10 minutes. The number of minutes is chosen after a trade-off of fast response (interval as short as possible) and reliable counts (bigger interval).

The second method makes use of the route guidance information of approaching traffic. The main advantage is that it is the most responsive way, but it only works with complete information about the traffic flows. If you know the route choice on a link for each destination, it is easy to calculate the turning probability. However, there is one problem. The approaching traffic has a specific order, for example, there could be a group with destination B followed by a group with destination C. In general, beforehand it is not known which part can pass the intersection in the next control cycle, if the amount of green time is unknown. Therefore, an assumption is made to consider all traffic that could potentially pass the intersection in the next minute. [2]

4.2.2 Methods to determine pressure

The pressure on a link is a measure for its degree of occupation. For an outgoing link, a low pressure implies that there is much space left for vehicles to enter, with high pressure it is considered full. For the incoming link low pressure means that the amount of vehicles willing to access the intersection is low, and the other way around. By subtracting the outgoing pressure from the incoming pressure (Eq. 4.3) the value for back-pressure is calculated.

Relative and normalized pressure values

Pressures can be expressed as absolute numbers, such as the number of queueing vehicles. As stated before (see Chapter 2, page 28) an approach with relative values is probably more suitable in general. To illustrate this, a 100 m long link with 10 vehicles on it is more densely used than a 1 km long link with 10 vehicles. A higher density, higher degree of occupation, implies a higher pressure. Therefore a first choice is to use the occupation of a link relative to its length.

Relating the occupation to the length is generally not enough, as a link with many lanes can store more vehicle per km than a link with one lane. A second choice is made, to normalize the pressure per link. Thus the pressure is defined as a function of its total storage capacity.

[2] In the alternative of time slotted control, Subsection 4.2.4, the time slot value can be used instead of the 1 minute, and it would be a correct value

The basic pressure function

Following the previous motivation, to normalize the pressure function, and using density as the measure of occupation, the pressure on a link (link index omitted in following equations) and time t can be simply defined by dividing density over the jam density.

$$P^I = \frac{k}{k_{jam}}. \tag{4.7}$$

Alternative pressure functions

The first alternative to the basic pressure function is to make a generalization that makes the function convex. The idea behind is that the pressure value will rise increasingly when density is higher; the more occupied a link is, the more of a 'burden' will the extra entering traffic be, or the more 'eager' is it to release traffic.[3] This pressure value can be obtained by putting the basic pressure function to the power of m_1 (with values like 2 or 3, in P^I m_1 is 1). As P^I has a domain of $[0,1]$, the same holds for P^{II}.

$$P^{II} = \left(P^I\right)^{m_1} \tag{4.8}$$

Another alternative can be derived from the equation on page 28.

$$P^{III} = \min\left(\frac{\frac{k}{k_\infty} + \left(2 - \frac{k_{jam}}{k_\infty}\right)\left(\frac{k}{k_{jam}}\right)^{m_2}}{1 + \left(\frac{k}{k_{jam}}\right)^{m_2 - 1}}\right) \tag{4.9}$$

It claims to add more fairness at low densities, but the overall effect is not expected to be big. Therefore, in the remainder of this thesis, only the first two (simple) functions are further used.

One last idea to tweak the pressure functions could be to adjust the jam density by a scaling factor. This can be useful, for example if a link needs to be considered 'full' at 90% of its (theoretical) storage capacity.

Practical issues

The first issue is related to *timing and defining pressure*. Control actions are taken for a certain period of time, the control period, that can be a cycle time of the intersection (typically one minute) or the time slot for the next phase activation (typically around ten seconds). The pressures should be representative for this duration. As a first try, the pressure is determined based on the average density of the last control period. This attempt seems reasonable. Using instantaneous densities wouldn't account for traffic dynamics (waves of low and high density), and observing longer periods could make the control algorithm not responsive enough to the actual demand.

The second issue is related to the *length of links*. Controlling intersections is about serving the traffic that approaches the intersection, and in case of back-pressure also about traffic on roads leaving the intersection. If the links in between intersections are very long, it is perhaps not useful to look at the entire links, but only to the parts close to intersections. This is something to keep in mind, however in the experiments only relatively short links will be used.

[3] However it is questionable if these two (incoming and outgoing) should be treated equally.

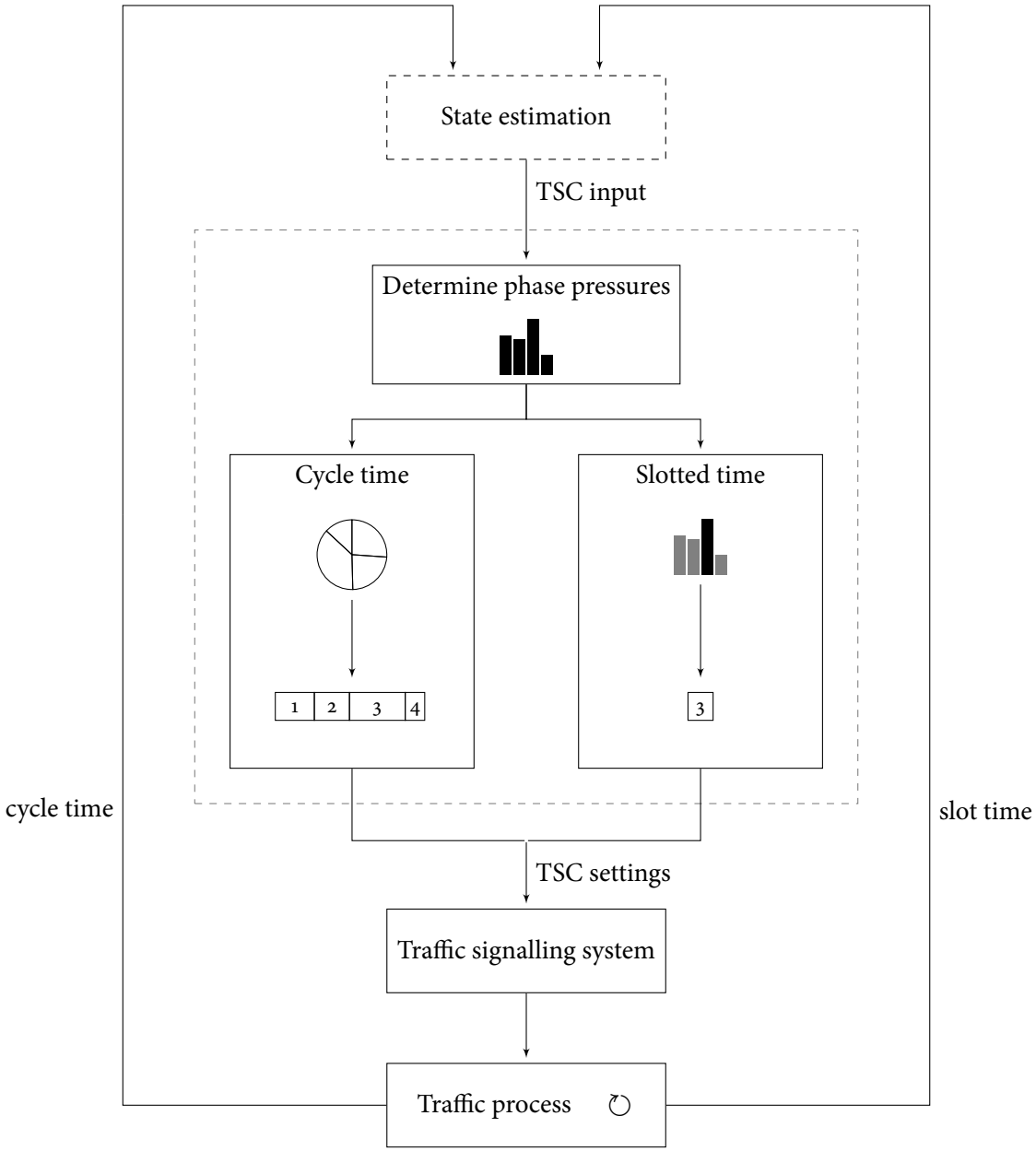


Figure 4.4 – Fixed cycle time vs. time slotted approach

4.2.3 Fixed cycle time back-pressure algorithm

A traffic signal control approach with fixed cycle times is based on a predetermined order of phases that each have a share in the total cycle time. The control period can be equal to the cycle time, if the phase lengths are updated after each cycle. It can also be a multitude of cycle times, then updates take place less frequently. As back-pressure works best with frequent updates, the first option is adopted.

In equation 4.5 the function $f(\gamma_p(t))$ is put as a general representation. One approach is to use the exponential (logit) function (Le et al., 2013):

$$f(\gamma_p(t)) = e^{\theta\gamma_p(t)} \tag{4.10}$$

$$P_p(t) = \frac{e^{\theta\gamma_p(t)}}{\sum_{i\in\mathcal{P}} e^{\theta\gamma_p(t)}} \tag{4.11}$$

The exponential function is common to use in choice problems. In this application it is an advantage that the BP-value for each phase can be positive or negative, the outcome is always a positive share (other than the proportional function of Eq. 4.2). The function can be tweaked by adjusting the parameter θ, which can also be the disadvantage of this method, determining the right value.

There is one issue with this method, that would make it less useful. The back-pressure values for each phase include the service rate value. A higher service rate, would clear the queue faster, therefore is preferred, with the aim of high throughput. The described method of proportioning would use the higher pressure to give this phase more green time. The higher service rate, however, would require less green time.

4.2.4 Time slotted back-pressure algorithm

An approach quite different from using fixed cycles is traffic signal control using time slots. The term 'time slot' is used here, as a reference to the back-pressure methods for communication networks in literature. More than the method in Subsection 4.2.3 it is like the 'original' back-pressure procedure.

This approach splits time into fixed steps, such as 10 seconds. Each time step represents a slot that can be assigned to one activated phase. This phase has the highest back-pressure value, as defined by Equation 4.6.

In case two (or more) phases have exactly the same BP-value, the next slot will be assigned randomly among the two. Or, another possibility is to choose the phase with the longest time of inactivity.

A possible drawback, especially from the road user point of view, is that traffic streams with a low BP value will wait 'forever' if they are dominated by others. And on the other hand, a domination traffic stream should give way to others once in a while. To facilitate these requirements a *maximum waiting time* and a *maximum green time* are introduced. For example, whatever the conditions, after 90 seconds of facing red light, an approach should receive green. And after 45 seconds of green, the approach is excluded from receiving green in the next control step.

The back-pressure procedure can result in negative BP values for all phases. In that case, the *all red* phase can be implemented. It is considered an option, however, to pick the phase with the least negative value. This raises the (at least local) throughput of the intersection, besides the problem of user acceptance.

A possible drawback of this approach is that it is difficult to implement with currently available systems. The high frequency of control decisions requires the availability of reliable and real-time measurements. The same goes for the first method, but to a bit less extent since the typical control period is several times longer.

Algorithm 1 Traffic signal control update

procedure BACK-PRESSURE UPDATE
  for $i \leftarrow 1, n$ **do**
    $turnprob \leftarrow$ getCurrentTurnProb(i)
    $phaseBPs \leftarrow$ backpressureTSC(i,$turnprob$)
    $phaseTimes \leftarrow$ determinePhasetimes($phaseBPs$,i)
    $TSC(i).phasetimes \leftarrow phaseTimes$
  end for
end procedure
function BACKPRESSURETSC(i,TB)
  for all $a \in SendingLinks(i)$ **do** ▷ $SendingLinks$ is defined in the data structure for TSC i
    $LinkPressure$ = CalculatePressure(a,i)
    $InQ_a \leftarrow LinkPressure$ ▷ pos is the position of a in $SendingLinks$.
  end for
  for all $b \in ReceivingLinksLinks(i)$ **do**
    $LinkPressure$ = CalculatePressure(b,i)
    $InQ_b \leftarrow LinkPressure$ ▷ pos is the position of b in $ReceivingLinks$.
  end for
  for $p \leftarrow 1, |\mathcal{P}|$ **do**
    $pressure_{a,b} = \mu_{a,b,p}\,(InQ_a - InQ_b TB_b)$
    $phasePressures_p = \sum_a \sum_b pressure_{a,b}$
  end for
  return $PhaseBPs$
end function
function DETERMINEPHASETIMES($phaseBPs$,i)
  if $allowallred$&max($phaseBPs$) <= 0 **then** ▷ $allowallred$ is a global boolean
    $PT_p = 0, \forall p \in \mathcal{P}$
    return PT
  else if $splitcycle$ **then**
    $proportions$ =LOGITASSIGNMENT($phaseBPs$,η) ▷ Instead of the logit assignment, a proportional assignment could be opted
    $timeproportions = proportions \cdot cycletime_i$
    add later
  else
    add later
  end if
end function

4.2.5 Comparison of the two methods

In Figure 4.4 the traffic signal control model is sketched, with the two methods 'cycle time' and 'slot time'. The two methods are compared briefly.

- On the aspect of control periods, the fixed cycle time method typically has a longer update time than the slotted time method.
- The fixed cycle time method can assign precise phase times. The slotted time method assigns each slot to one phase only, phase times can only be precise if the time slots are short enough. For practical applications, the update frequency for the slotted time method may be too high.
- The slotted time method better fits the 'original' back-pressure control examples, where each time step a new transmission is assigned. It provides the phase with the highest pressure at any time, which can prevent overestimation of the needed green time.
- The slotted time method has the advantage that phases are automatically put into the right order, whether the fixed cycle time method has a fixed order of phases.
- The fixed cycle time method includes a parameter θ that should be calibrated.
- The slotted time method could lead to unfair waiting times if one phase dominates the others, even if a maximum waiting time is considered.

Both methods have their strengths and weaknesses, therefore both are evaluated in the next chapters. Algorithm 1 illustrates the total traffic signal control model to be put into the simulation model.

4.3 Route guidance based on back-pressure control

Back-pressure hasn't been applied to route guidance yet. In the research approach and literature review (2.2.3) some starting points were formulated. This section develops a first attempt of a route guidance procedure based on back-pressure.

4.3.1 Translating back-pressure to route guidance

First, a short motivation on why the concept of back-pressure is thought to be promising also for route guidance.

Route guidance, from a DTM point of view, is about making travellers take those routes that lead to the best performance for the road network as a whole. One way to improve the performance is to make better use of the available road capacity, diverting traffic from busy roads or parts of the network to parts of the network with more 'space' available. Back-pressure fits this idea, since the principle favours low pressured links above high pressured links. At the same time, maximum throughput and stability are characteristics of back-pressure controlled networks, these are indicators of performance.

In the back-pressure applications in the field of communication technology route guidance is an implicit characteristic, as a data unit is sent to the node with the shortest queue. For the traffic control application route guidance and intersection control are pulled apart. They are difficult to be integrated, as intersection layouts usually have approaching lanes dedicated to specific continuation links. Once a vehicle is on one lane, it can't switch back to change its route or link choice.

Other than most examples from the field of communication technology, for route guidance of traffic it is important to take into account not only the next first links, but the possible routes as a whole.

One task will be to provide the routes by representative pressure values.

Like for traffic signal control in Eq. 4.4, the 'weight' (or attractiveness) for a route could be defined by:

$$A_r(t) = \mu_r BP_{\text{route},r}(t). \tag{4.12}$$

$BP_{\text{route},r}$ represents the back-pressure value of the route, and μ_r the service rate to the route. The service rate to a route is hard to define. For an intersection it is the saturation flow from one link to the next. For routes, it can be the capacity of the first link of the route. However, multiple routes are assigned at the same time, all taking a share of the capacity. Therefore, taking the capacity as a 'service rate' is probably not correct, but a first step, an assumption that might work. If the first link has a double capacity, compared to the alternative route, it is able to handle a higher throughput.

In the following subsections a number of aspects is discussed. First, a general algorithm is presented for route guidance based on back-pressure. Then, the overall utility function is formulated, followed by paragraphs that discuss specific aspects of the algorithm and utility function.

4.3.2 Route guidance back-pressure algorithm

The general algorithm for route guidance with back-pressure can be summarized as follows:

- For every link l with multiple next links (other needn't be considered):
  - Determine the destinations that can be reached through this link, d_l:
  - For every $d \in d_l$:
    - For every partial route $r_{l,d}$ starting from l with destination d:
      - Determine the service rate $\mu_{r_{l,d}}$, as the capacity of the first link.
      - Determine the 'utility' $U_{r_{l,d}}$ (incl. the use of pressures).
    - Determine the proportions for each partial route.
    - Determine the new split vector, based on the route proportions.

The utility function is described next, followed by specific parts of the utility function. In Subsection 4.3.7 the procedure for route proportions and split vectors is explained. The last subsection gives some further remarks on the use of service rate in this context.

4.3.3 An overall utility function

A general function for the utility of a route, that incorporates the factors that are taken into account:

$$\text{BP}_r = \alpha_1 P_{\text{route},r} + \alpha_2 P_{\text{firstlink},r} + \alpha_3 P_{\text{user},r} \tag{4.13}$$

Here, BP_r stands for the utility of the route. $P_{\text{route},r}$ is a measure for the pressure of a route, and $P_{\text{firstlink},r}$ a measure for the pressure of only the first link of the route. These are explained in Subsection 4.3.5. The factor $P_{\text{user},r}$ is used to take user preference into account, as described in Subsection 4.3.6. The α parameters can be tweaked to the degree a factor should contribute to the utility.

4.3.4 Route definitions

Route guidance in this thesis relates to en-route route guidance. The general algorithm states that for every destination every relevant route should be evaluated for its utility, in order to determine proportions for each route. This subsection discusses the types of routes.

A route is a set course from a starting point to a destination. In a model, routes are used to direct traffic along a specific path of sequential links, and to serve as options for route choice or route guidance. There are several ways routes can be defined in a model, such as the following.

A *complete route* consists of a list of links that lead from an origin to a destination.

Instead of using complete routes, *partial routes* can be defined for every intermediate decision point. These routes contain a list of links and a destination. Partial routes are generalized regarding to origins, so they could apply to vehicles going from origin 1 or from origin 2, etc.

In this thesis partial routes are used, which include the complete routes. The main reason is that this approach is easy to fit en-route route guidance.

The available (partial) routes for one destination overlap each other, if they share the same links. This needs to be taken into account with route guidance.

The number of routes in a network grows fast, when the network gets bigger and complex. In this thesis the networks are relatively simple, for bigger networks another kind of route representation might be better. The method of *recursive logit (RL)* (Fosgerau et al., 2013) could be applied here, which also has potential to be used in a distributed manner.

A simplified way of modelling routes, is to use only the *next link* after a decision point. This means that each next link represents a route to a destination d, if there is a path possible through that link to destination d.

4.3.5 Link and route pressures

Representative route pressure

A representative value for the 'route pressure' (P_{route}) is not obviously found. Route choice (and guidance) are usually based on travel times, link additive characteristics. If a route consists of links 1, 2 and 3, its travel time is a summation of their individual travel times. For pressures, this can't be done, a summation of (relative) densities doesn't say much about the utility of a route. The challenge is to find a representative value of all individual link pressures.

The individual link pressures can be defined as in Section 4.2.2.

Alternative approaches to route pressures

There are many ways imaginable to combine the pressures of individual links into one representative value, $P_{\text{route}}*$.

- Use the *weighted average* value of all link pressures related to their link length.
- Use the *maximum* value of all pressures. If one link is 'full', the whole route has a high pressure, disregarding the rest of the route.
- Use the weighted average plus a weighted standard deviation. This way, the average is represented, as well as a value for deviations. The underlying assumption is that deviations generally worsen the state of a route, especially the deviations of higher pressures.

- Use the *mean of the weighted average and the maximum*. The idea is similar to the previous one, but explicitly considers the maximum deviation and not the others.

Positive values for route pressures

So far, route pressures result in a value between 0 en 1, where 0 stands for an empty route and 1 for a full route. Comparing this with traffic signal control, the values should actually be negative. The least negative pressure has the highest capacity for additional traffic.

However, the function of Eq. 4.12 indicates that both the pressure and the service rate add to the total 'attractiveness' of a route. If service rate and pressure were to be multiplied as total measure of route attractiveness, the outcome would be counterproductive. Take a pressure of −0.5, for a service rate of 10 it would lead to a value of -5, for a service rate of 20, a value of -10. With a higher service rate, the final value should be higher instead of lower. When service rates and pressures were both taken as positive values, it would be more natural to combine the two.

To prevent confusion in further development of the algorithm, a choice is made to keep the route pressures as positive values, but turn them around, where the lowest value represents a full link (no room) and a value of 1 stands for an empty link.

$$P_{\text{route}} = 1 - P^*_{\text{route}}. \tag{4.14}$$

Only pressures for routes, not for current links

For traffic signal control a pressure difference is defined to combine the demand upstream and the supply downstream. For route guidance that is not applicable. A traffic stream on a link that faces a decision point could be given a pressure, but it can't be compared to the route pressures, and it would make no sense, since for every route (for a certain destination) this pressure would be the same.

Pressure of the first link

The value $P_{\text{firstlink},r}$ in Equation 4.13 stands for the pressure of only the first link of the route. This pressure can be defined in the same way as the route pressure, but it is based on only one link.

4.3.6 User preference

Route guidance based on only route pressures doesn't take delays into account, but with some modifications this is possible. User satisfaction is equivalent to the desire of road users to travel along the shortest route (expressed in distance, or preferably in time). Especially at low traffic demands, when congestion is unlikely to occur, the system is expected to perform best when travellers use the shortest routes, instead of the routes with the lowest pressures (pressure is low enough everywhere). More travellers will reach their destination within a certain time period (high throughput and output), while the total time spent and distance travelled is minimum.

Other than the method of (Zhang, 2012) the approach of (Neely et al., 2005; Ying, Shakkottai, et al., 2011) is adopted, and user satisfaction is put into the utility formula of Equation 4.13.

The question now is how to define $P_{\text{user},r}$. For the cases where all links of a route are considered (as in variant 2 and 3 of the previous subsection) a travel time can be determined. This travel time is an estimate for the instantaneous travel time, based on the current speeds on the links of the route. For variant 1, where only the first link is considered, a the (static) minimum necessary distance is taken as a measure for user preference (length of the link plus a straight line to the destination).

The higher the travel time (or distance), the less favourable the route should be. Therefore $P_{\text{user},r}$ should be negative. Furthermore, the travel times are scaled to a value between 0 and 1, to keep all three terms in Eq. 4.13 of comparable values. The scaling is done by dividing the travel times of the choice set by the maximum travel time.

$$P_{\text{user},r} = -\frac{\text{TT}_r}{\max_{i\in\mathcal{R}}(\text{TT}_i)}$$

This scaling method contains a weakness. Suppose there are three routes to choose from, their travel times are 1, 2, and 3 minutes. $P_{\text{user},r}$ values are $\frac{1}{3}$, $\frac{2}{3}$ and 1. Now suppose the third route's travel time increases to 10 minutes. Now $P_{\text{user},r}$ values are $\frac{1}{10}$, $\frac{2}{10}$ and 1. The difference between the first two routes has become much smaller, which means that (after applying the exponential function) the difference between their route shares will be smaller. One would expect their relative difference to be unchanged. It would make more sense to use the absolute time differences and find an appropriate value for α_3, or to scale the travel times to a (static) minimum value instead of the maximum.

4.3.7 Methods for route proportions

This subsection describes the methods that can be used to determine route proportions and eventually the splitting rate between the choice for the next link.

Multinomial logit function

A function is needed to determine the probability for each route. A first step is the logit choice function:

$$\text{Pr}_{\text{MNL},r} = \frac{e^{\theta U_r}}{\sum_{i\in\mathcal{R}_{l,d}} e^{\theta U_i}} \quad \forall r \in \mathcal{R}_{l,d}$$

This choice model is fairly simple, but assumes that the alternative routes are completely independent. In most networks, routes shares each other links for a considerate part, and this model is not likely to be applicable.

Path size logit function

Contrary to the Multinomial Logit model, the Path size logit function does consider route overlap, by implementing the Path Size Logit (PSL) model, instead of the for independent choice alternatives. This method uses a path size (PS) factor for each route to account for the overlap. A route that is almost similar to another route has a low path size factor, and will have a lower choice probability than if each route would have been treated independently.

There are several versions of the PS factor, here the original version is used (Frejinger, 2008):

$$\text{PS}_r = \sum_{a\in\Gamma_r} \frac{L_a}{L_r} \frac{1}{\sum_{s\in\mathcal{R}} \delta_{as}} \tag{4.15}$$

The value δ_{as} is the number of times link a occurs in a choice set of routes $\mathcal{R}$ which contains partial routes starting from the current link with the same destination, the choice set. The PS value is static, so it needs to be calculated only once, after the routes have been defined.

The route probabilities now become:

$$\text{Pr}_{\text{PSL},r} = \frac{\text{PS}_r e^{\theta U_r}}{\sum_{i\in\mathcal{R}_{l,d}} \text{PS}_i e^{\theta U_i}} \quad \forall r \in \mathcal{R}_{l,d} \tag{4.16}$$

Route proportions and service rate

The functions above define a degree of probability for the routes to be used. To determine the actual route proportions, the service rate is to be incorporated. Two approaches can be adopted to do this.

The first way is to include the service rate into the value for U.

$$U_{\mathrm{v1},r} = \mu_r \mathrm{BP}_r \tag{4.17}$$

Then the route proportion can be calculated directly from the logit or path size logit function above.

$$\mathrm{Proportion}_{\mathrm{V1},r} = \mathrm{Pr}_r \tag{4.18}$$

The second option is to use the utility function BP_r as a value for U and value the outcome of the logit function as a kind of 'attractiveness'. Then, scale these values with the service rate. This approach seems to be logical: first a division is made based on the route utilities, then the capacity of each route is taken into account separately.

$$U_{\mathrm{v2},r} = \mathrm{BP}_r \tag{4.19}$$

$$\mathrm{Proportion}_{\mathrm{V2},r} = \frac{\mathrm{Pr}_r \mu_r}{\sum_{i \in \mathcal{R}_{l,d}} \mu_r} \tag{4.20}$$

The design for the route guidance model is not a 'finished product' up to this point. The way a route pressure can be defined has various options, but especially the right way to incorporate the service rate is not determined, rather a few approaches that could work have been stated. Probably there is still another way of assigning available capacity and routes for each destination group. The challenge is to use the capacity of the next links as much as possible, while taking account of the route utilities.

From route proportions to split values

Calculating the split values when the route proportions are known is straightforward. Each route has a first link that corresponds to one of the outgoing links from the current position. For each outgoing link, the corresponding route proportions are added up to the split values.

4.3.8 Variants of the route guidance approach

The algorithm and functionalities described so far, can be used in various combinations, particularly by altering the α values of the utility function. A few examples:

- Looking only to the first link, ignore the rest of the routes.
- Only look at complete routes, in this approach a PSL choice model is applicable.
- Only take user preference into account; although this wouldn't make use of the 'back-pressure' capabilities of the model.
- A combination of all three factors.

The first option is a special case. In the special approach that only the first link is taken into account, and not the route as a whole, the choice becomes much simpler. The number of outgoing links (that can lead to the destination) determines the number of choices. A multinomial logit model can be used. If user preference is taken into account, a representative value is needed to replace travel times; one could use the distance from the next node, or the fastest current travel time via that link.

4.3.9 Use of the service rate

As mentioned in section 4.3.1, using the service rate for route guidance back-pressure is not a straightforward application.

Assuming that a service rate can be used as proposed, how to define it? A first attempt is to use the ingoing capacity of the first link of the route. A link with higher capacity has a higher potential to facilitate traffic flow. Combined with the pressure of the route, this could give a reasonable estimation for the route attractiveness.

A second thought is to define the service rate by a representative capacity for the whole route. This thought is not further developed. A first problem is how would you define this representative capacity; you can think of a minimum capacity of all links on the route. The second reason is that the state of the route is already represented by the pressure, just like the intersection, the service rate should focus on handling traffic at the point of route guidance.

A third idea is to use the actual supply of the first link of the route, rather than its maximum capacity. This means that if spill back happens towards the entrance of the link, the service rate would drop to a value (much) lower than the link capacity. This approach has been tested, however the sudden drop of service rate made the routes via that link suddenly that much less attractive, that the effect was not realistic.

A fourth idea is to iteratively assign the service rate to the traffic demand on the link, only if the demand is higher than one of the to be chosen the next links. For example, in the first step half of the demand is handled, translated into flows that are subtracted from the service rate. In the next step, the idea is that the link(s) with a low remaining service rate become less attractive. Just like the third idea, test results showed that the effect was 'unrealistically' high. This probably has to do with the way the service rate has been used in the algorithm, having too much influence (see next paragraph).

For now the approach is taken, that the service rates are independent of the actual link state, and the variable part of the 'route attractiveness' is taken by the route pressures alone.

4.4 A combined system

The combined system is applied in a network that consists of traffic signals and where route guidance is applied.

There is a link between the two modules of traffic signal control and route guidance. As shown in Figure 4.1 the traffic signal controller uses the outcome of the route guidance controller. The traffic signal controller uses route guidance information to determine the split ratio for each approach, so that the downstream pressures can be weighed at the right ratio (see Equation 4.3).

4.5 Summary of control modelling

A general back-pressure algorithm for traffic signal control has been designed, with two main variants: one with a fixed cycle time that assigns phase durations for one or more cycles, and one with

short time slots that repeatedly activates the dominant phase. Pressure values are based on representative link densities, are normalized to the jam density and can be extended by a power function that increases the relative weight of higher pressures. The necessary turning probabilities are based on measurements or on route guidance settings, which integrates traffic signal control and route guidance.

For route guidance a general algorithm has been proposed as well. As a first attempt of service rate, the link capacity of the first outgoing link of the route is used. Furthermore a number of variants for the algorithm have been considered. The first variant is simple but myopic and observes only the first link of each route, which also limits the number of routes that need to be considered. An important limitation of this variant is that congestion on links further downstream is not taken into account. Other variants observe complete routes (from the en-route position). A path size logit choice model is used to determine the proportions for each route. A pressure value that represents the whole route needs to be determined. This can be done by taking the (weighted) average of the link pressures, or by a method that takes outliers specifically into account. Travel time as a measure of user satisfaction, allowing road users to use the shortest routes, can be incorporated into the method. The overall utility function of a route can be written as: $\mathrm{BP}_r = \alpha_1 P_{\mathrm{route},r} + \alpha_2 P_{\mathrm{firstlink},r} + \alpha_3 P_{\mathrm{user},r}$. The three terms represent the (total) route pressure, the pressure of the first link, and the travel time value.

Chapter 5

Simulation environment

In order to quantify the performance of the proposed traffic control system it will be tested in a simulated environment. This chapter describes the simulation system needed for the experiments. Section 5.1 explores available simulation models and selects one that will serve as a starting point. An overview of the model and its components is described in Section 5.2. Modelling the network, traffic demand, and route choice is topic of Section 5.3. Propagation of traffic flow in the model, along links and nodes, is described in Section 5.4. Traffic control facilities that are added to the model are the topic of s Section 5.5.

5.1 Modelling approach

The development of the simulation environment starts with choosing a simulation model that can be used as a base and starting point. This simulation model needs to meet the requirements, as formulated in Subsection 3.3.2 of the Research Approach. The following three subsections describe the possibilities, the choice and motivation for the chosen model, and the adaptations that are needed.

5.1.1 Simulation model possibilities

There are many ways to simulate traffic flows. Simulation models are often categorized as microscopic or macroscopic simulation models.

In microscopic models each vehicle is used as an individual particle that takes part of traffic and follows its route from origin to destination. Each vehicle particles can have specific properties (such as vehicle type, driver character, connectivity) that influence its driving behaviour (car-following and lane changing). The infrastructure is modelled with high level of detail as well, roads have distinct lanes, intersection properties can be very specific. The same goes for traffic control, which can be applied in detail, targeted at individual vehicles and infrastructure parts. Also new concepts of vehicle and infrastructure communications are implemented in microscopic models. Examples of (widely used) microscopic models are: VISSIM, PARAMICS, AIMSUN, MITSIMlab and SUMO.

Macroscopic simulation models look at the aggregate traffic stream instead of individual vehicles. The underlying fluid-dynamic models are based on the principle of the conservation of flow and the fundamental relation between traffic flow (q), density (ρ) and speed (v).

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{5.1}$$

$$q = \rho v \tag{5.2}$$

The LWR model by (Lighthill & Whitham, 1955; Richards, 1956) is a widely used model, where the relation of Equation 5.2 is defined by an 'equilibrium flow': $q = q_e(\rho) = \rho v_e$. This model is called

a first-order fluid-dynamic model, or kinematic wave model. Second order models can take into account velocity dynamics to describe the driving behaviour more precise, but are harder to solve numerically, an example is METANET, used in (Hegyi, 2004).

The LWR model can be numerically solved by analytical methods, but for simulation models usually discretization and numeric integration methods are used, such as the Godunov scheme. The cell transmission model by (Daganzo, 1994, 1995) uses this method and has been expanded by others to include specific features, such as DSMART (Zuurbier, 2005, 2010) for route guidance and Fastlane (van Wageningen-Kessels, 2013) for multiple user classes. Examples that focus on the efficiency of the model are the link transmission model (Yperman, 2007) and the wave front tracking method (Henn, 2005).

A slightly different approach is to use a dynamic traffic assignment model, such as MARPLE (Taale, 2008). This model is based on the relation between flow and travel time.

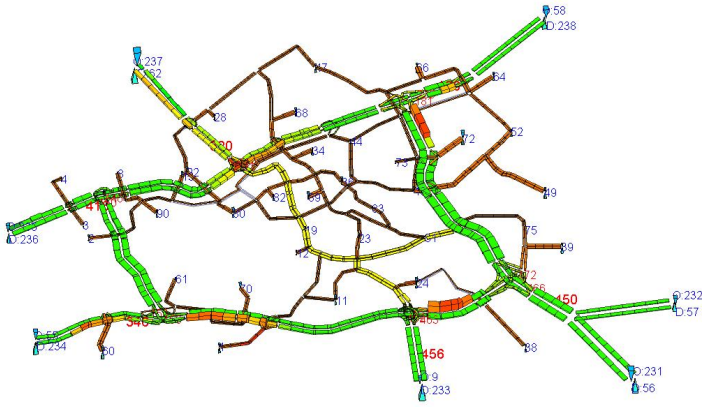


Figure 5.1 – Screenshot of a simulation with DSMART (Rotterdam road network)

5.1.2 Chosen model: DSMART

Based on the simulation requirements a modelling approach has been chosen to use as a core of the simulation environment. Because the model should be fast and efficient, and potentially applied to larger networks, the first decision is to use a first order macroscopic simulation model if possible.

Out of the candidate models DSMART[1] (Figure 5.1) has been adopted as a good choice, based on a number of reasons:

- The model is a cell transmission model, that can simulate traffic dynamics and congestion quite well, and in a traceable way.
- Traffic in the model is specified per destination, as is route choice (instead of having split fractions on the aggregate traffic). Route guidance is built-in and traffic can be rerouted at every link.

[1] DSMART: Dynamic Simple Macroscopic Assignment of Road Traffic

- The model is fast and efficient and can be applied to larger networks. Creating networks and traffic demand profiles is not difficult.
- The model is available and familiar from earlier experience.
- As the model has been programmed in Matlab, it is accessible to make changes, and to add the proposed control algorithms.

A main disadvantage of DSMART is that it doesn't include traffic signal control facilities. This is one of the points where the model should be extended. Another limitation of DSMART is that the lanes of a road are aggregated; taking into account lane selection (for example in case of intersection approaches)is therefore not possible. Therefore in this research only simple intersections will be considered, where complete links are connected rather than specific lanes.

5.1.3 Modifications and extensions to the model

Because the model did not meet all the requirements for this research, it needs to be modified or extended at some points:

- A functionality of *traffic signal control* needs to be added. First, a function needs to be developed for the role of the controller and that defines the right traffic light settings. Secondly, these settings need to be implemented to control the traffic flow at every time step. Thirdly, variable objects can be used to support this functionality.
- The *route guidance* method should be modified to support the proposed algorithms. One of the tasks is to add objects that represent the (pre)defined routes.
- The available model is an older version (Zuurbier, 2005). The newer version (Zuurbier, 2010) contains an approach of 'system class dynamics' (SCD), which improves the precision by separating changes in the composition of traffic flows. This feature, that secures FIFO (first-in-first-out) behaviour and prevents dispersion, is also added to the available base model.
- Traffic flows are not only defined by destination, but also by origin. The SCD model is extended to take the origin data of traffic into account.
- The DMART model already generates results that can be used for evaluation, however, some results are added. An example is the calculation of travel times and values that represent the generalized network fundamental diagram.

The following sections give a condensed description of the DSMART model, although emphasis lies on the parts that require specific attention in this thesis. For a more thorough documentation of DSMART the reader is referred to the original theses (Zuurbier, 2005, 2010).

5.2 Overview of the model

Before getting into some specific aspects of the simulation model, this section gives an overview of the model, its structure and how it works as a process of steps.

5.2.1 Simulation process

The model can be explained as illustrated by Figure 5.2, which draws an outline of the simulation model.

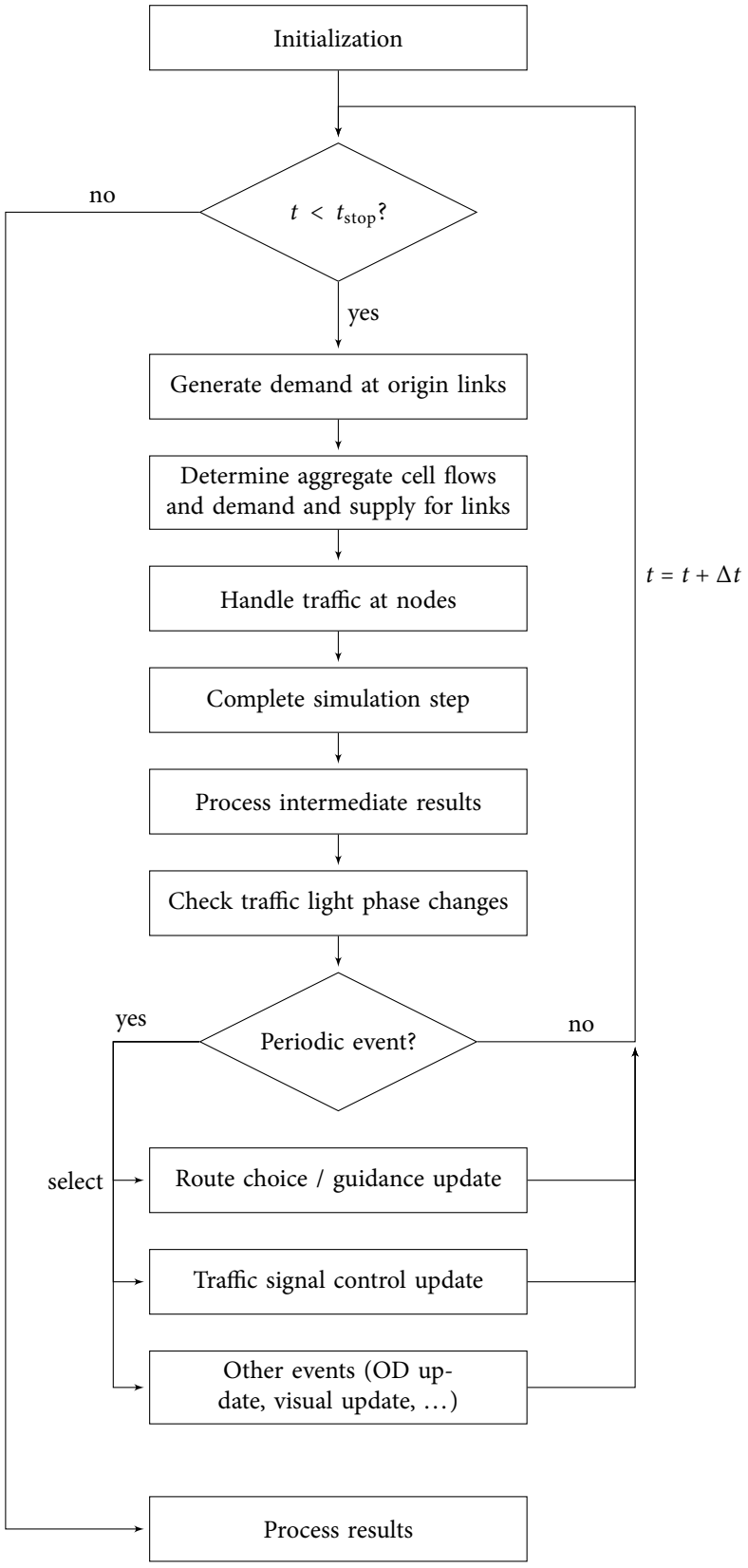


Figure 5.2 – Ouline of the simulation process

The first task is *initialization*. The process of initializing functions prepares all the data structures that are needed to run the simulation, including the network of links and nodes, the traffic demand, routes, and traffic control facilities. To build this data structure input is needed from files (Matlab tables). Simulation parameters, that can be set from the script that runs the simulation, also serve as input.

Then, the main simulation loop starts and is repeated until the simulation time boundary is reached. This loop is the core of the model and calculates the traffic flows and other state variables for each time step.

A simulation step starts with generating traffic demand from the demand profile at each origin, and assigning this to the network entrances (see Subsection 5.3.2). Each link consist of cells. The state of these cells determines the flow that passes between the cells of this link, and for the link as a whole, a supply and demand (5.3 and 5.4.2). The demand of a link represents the potential traffic flow if it could run freely. As an intermediate step, the demand of links before a red traffic light is put to zero. After the links are handled, the next task is to deal with traffic flow at the nodes. Demand and supply of the connected links are matched, taking into account the traffic composition and split ratio of each incoming link. The resulting flow determines the new state of the relevant cells, and the new composition vector of each outgoing link. The main part of the simulation step is completed by executing remaining accounting tasks and processing intermediate results.

After the traffic flows have been updated, traffic signals are checked to change phase for the next simulation time step, and a list of periodic events is checked to be executed before the next time step. These events can be regarded as an outer loop that only need to be performed after a set time period. An example of an event is *route guidance update*, which calls on the controller to update the settings for route guidance. The same goes for *traffic signal update*. Other events provide an update of the current traffic demand or are 'administrative' events, such as 'update the visual network state'. Although the diagram shows them in parallel, the event types are checked in a serial order, and multiple events can be activated in one simulation step.

When all tasks of the current simulation time step have been done, the simulation time is incremented and the simulation continues to run the next time step. If the end of the simulation period is reached, the simulation loop is concluded and the results of the simulation are generated.

5.2.2 Using the model

The original DSMART model can be run from a Matlab script 'main.m' that opens up a GUI, where parameters can be set and and a scenario file (including network and traffic demand data) can be loaded. Then the simulation can be run and visualized on screen, as in Figure 5.1. Results are saved in Matlab variables.

The modified model can be run in the same manner, but an alternative way has been added. For an experiment, it is convenient to run several simulations from one script, changing the relevant parameters and input files. After each simulation the results are stored in a specified file and folder, which makes it easier to review and compare the simulation results.

In order to get meaningful results for simulation experiments a warming up time is applied. This first part of the simulation is used to load the network with traffic until it reaches an acceptable starting point from where results can be recorded. This warming-up facility was added to the DSMART model, by postponing the start of measurements or resetting the intermediate results.

5.3 Network, traffic demand and routes

The network model represents the road network that will be used to carry traffic from origin to destination. The model of traffic demand determines how much traffic is to be put into the simulation at what time. The way routes are defined and traffic can be routed along the network is laid down in the model for routes and routing. These three topics are briefly discussed in this section.

5.3.1 Network model

The network mainly consists of two types of elements: links and nodes. Links represent uniform road stretches that connect one point to the other. These points are called nodes, they provide the exchange of traffic between the connected links.

Link properties

A link is defined by a number of properties. The geometry is characterized by the nodes at the start and the end of the link, the number of lanes, the length, and intermediate vertices that bend the course of a link. Traffic flow characteristics include the capacity of a link (veh/h), the maximum speed and the jam density (veh/km). Each link is divided into a number of cells, that represent road sections with assumed homogeneous conditions.

In Matlab links are represented by objects, which also hold information about the traffic performance and details on the traffic flow (as described in Section 5.4).

Node properties

Nodes have a limited set of properties. They are defined by their location and by the links that enter and exit the node. As a node can represent an intersection, traffic control signals are indexed by the corresponding node. Nodes also represent origins and destinations.

Origins and destinations

Each origin and destination is coupled to a node and a link. The node index is used to identify the origin or destination. Traffic is added to the network at virtual origin links, only connected to their origin nodes. Likewise, traffic leaves the network at destination links, only connected from their destination nodes.

5.3.2 Demand model

The model of dynamic traffic demand is (differently from the original DSMART) based on traffic demand profiles. A three-dimensional array is used to represent traffic demand. Origin and destination nodes represent the first two dimensions, the third dimension is time. This array is accompanied by a vector of corresponding time values. For example, coordinate (2,5,3) represents the traffic demand (expressed in veh/h) from node 2 to node 5 at time point 3, where time point 3 refers to the third entry of time value vector.

The traffic demand profile describes a piecewise linear function between each data point. During the simulation, the traffic demand at the origins is periodically updated. The new value is simply picked by interpolation, see Figure 5.3.

To simplify the creation of demand scenarios, the simulation model contains a load factor parameter, by which the traffic demand as a whole can be scaled.

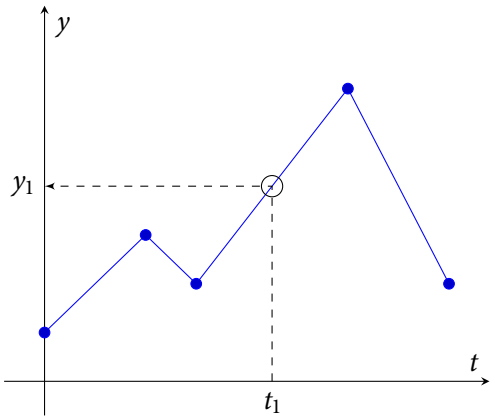


Figure 5.3 – Interpolation of dynamic traffic demand

5.3.3 Routes and route choice model

Routing with turn ratios

If a link is connected to multiple downstream links, traffic on the link needs to be routed, assigned to the downstream links considering the right turn ratios. In the simulation model aggregate traffic flows are specified with composition vectors, so that the traffic volume to each destination is known. For each destination, a specific turn ratio is applied. The turn ratios have been defined during the periodic updating process of route choice or route guidance.

DSMART has been structured in such a way, that a three-dimensional matrix (SPF) contains turn ratio references for each link, destination and time period. The turn ratios themselves are discrete, with 0.05 precision (or as originally, 0.1). Therefore, whenever a new turn ratio is calculated, it is always discretized to some extent.

An example of a turn ratio is [0.1 0.5 0.4], which means that for the current link and destination, 10% goes to the first link, 50% to the second ad 40% to the third link that is downstream of the current link.

Standard route choice model

The standard route choice process is used to simulate the natural route choice behaviour of traffic. In DSMART the standard procedure is to update route choice settings (turn ratios) every period of typically 15 minutes. The route choice is based on (instantaneous) travel time values, of which road users are assumed to be well-informed.

The standard route choice model is a probit choice model. This means that random 'noise' is added to the link travel times as a measure of perceived travel time. Using randomized sets of travel times and the Floyd Warshall (FW) shortest path algorithm, the number of times a route is perceived as the shortest determines its share of route choice.

Explicit route definitions

In the standard route choice model explicit route definitions are not needed, it uses the FW algorithm to find the (perceived) shorest routes. However, in this thesis explicit routes are needed for the application of route guidance, in particular as information for the used Path Size logit model (or multinomial logit model).

As described in Subsection 4.3.4 there are two types of routes. The first group consists of (total) routes that connect a path of links from an origin to a destination. In the model these are mainly used for evaluating the traffic performance from origin to destination, they don't have a function in the main part of the simulation.

The second type is the group of partial routes. Each link with multiple downstream links is a 'choice point' and has a set of partial routes for the destinations that can be reached through this link. These partial routes are deduced from the total routes. Each partial route is defined by its destination, the links that are part of the route, and the length of the route.

The partial routes are used as a basis for route guidance. If the applied choice model is the Path Size logit model, then another property is added to each partial route: the PS factor that is a measure of overlap (where only routes with the same destination are compared).

All (partial) route properties are defined in the initialization phase of the simulation.

Explicit routes have another function as well. They can be used to 'force' traffic in a fixed direction. For example, there is an intersection with two crossing traffic streams, one exclusively going from north to south, the other exclusively going from west to east. The intersection node has two incoming links and two outgoing links. Free route choice would explore both outgoing links, but with a route definition it is possible to direct traffic into one direction only. (N.B. This could also be realized by expanding the node properties with allowed connections.)

5.4 Traffic flow model

This section gives an overview of the traffic flow model that is used in the simulation model. The description focuses on the principles, modelling choices and the process, mathematical details can be found in (Zuurbier, 2005, 2010).

5.4.1 Traffic flow on links

The DSMART model is an extension of the cell transmission model and is based on the LWR model, a kinematic wave model, or first order fluid-dynamic model, as described in Subsection 5.1.1.

Fundamental diagram

The kinematic wave model assumes that all traffic conditions are at equilibrium state and can be defined by a fundamental diagram (Figure 5.4). Flow and speed can be expressed as a function of density.

In DSMART a simple triangular shaped fundamental diagram is used. Each link has a fundamental diagram that is defined by the capacity flow, the maximum speed, and the jam density of the link. The top of the diagram represents the capacity at the point of critical density. The left branch represents free flow conditions where vehicles drive at maximum speed, the right branch indicates the congested conditions.

Advantageous of this traffic flow model is that the main properties of traffic flow theory are covered (congestion formation, shock wave theory) with a simple and efficient model. The drawbacks are related to the same simplicity. The first order model assumes instantaneously reacting traffic (ignores acceleration, hysteresis effects), and the capacity drop phenomenon is neglected.

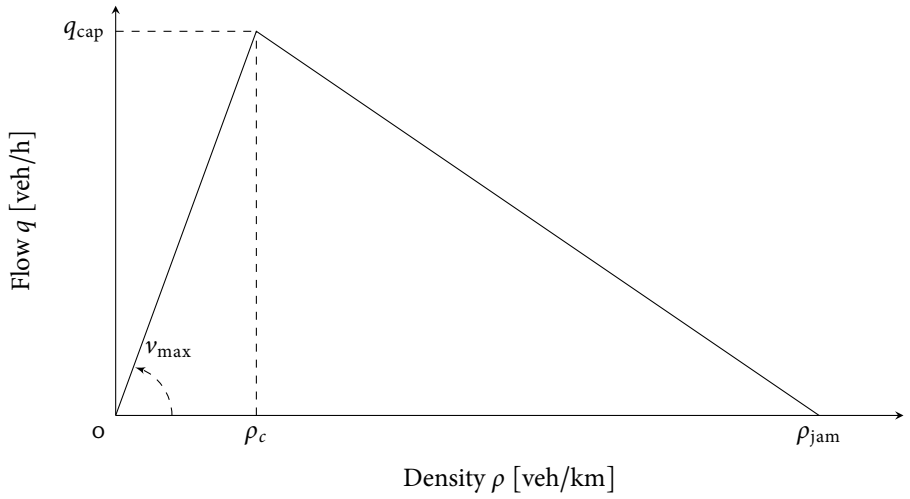


Figure 5.4 – Triangular fundamental diagram

Discretization

To get numeric solutions for the traffic model, time and space are discretizated into simulation time steps and link cells. In DSMART the cells of a link have a length that is based on their maximum speed. The maximum length of a cell is defined by the product of the simulation time step and the maximum speed on the link. This complies to the Courant-Friedrichs-Levy (CFL) condition which prevents traffic flow information from 'skipping' a cell.

The accuracy of modelling the traffic flow depends on the 'resolution' of the simulation. Choosing smaller time steps and smaller cells will improve accuracy, but increase the calculation time on a quadratic scale (decreasing cell length by half, requires twice as many calculation per time step, and twice as many time steps).

The simulation time step also has implications for the traffic control capabilities. Presuming that all relevant phase lengths traffic signals can be simulated, time steps need to be considerately small. A time step of 2 seconds seems a good compromise, although for larger networks this will mean a rather slow simulation.

Determining traffic flow on the link

Traffic flows on a link are defined by the traffic flows (or flux) in between cells, and the flows going in and out at the link borders. In order to determine the flux the state variable density is used. At this stage, the aggregate density can be used, because although the traffic is composed of traffic to multiple destinations the driving behaviour is assumed homogeneous.

The Godunov scheme is a simple way to calculate the progression of flows between cells, and uses the concept of demand and supply.

Each cell has an amount of potential flow that it can deliver to the next cell. In case of free-flow conditions, the demand equals the equilibrium flow from the fundamental diagram. In case of congestion, the demand is maximum and equals the capacity flow of the cell.

Likewise, each cell can receive an amount of flow depending on its state, the supply. In case of free-flow conditions, the supply equals the link capacity. In case of congestion, supply is limited and determined by the equilibrium flow from the fundamental diagram.

The flux from cell 1 to cell 2 is the minimum value of the demand of cell 1 and the supply of cell 2. This is obvious, the flow can't be higher than the demand, or higher than the supply. With the fluxes the new densities can be computed.

For the link boundaries, the demand is equal to the demand of the last cell, and the supply is that of the first cell. Calculating flows from one link to the next is more complex since it involves nodes that can have multiple in- or outgoing links.

System Class Dynamics

In the adopted traffic simulation model it is important to guarantee FIFO (first-in-first-out) behaviour. The original DSMART model used group-specific cell densities which were determined by the cell transmission process described above. This model has the capability to produce satisfying values of aggregated densities and fluxes, however it diffuses the group-specific information (causing this information to travel faster than traffic itself).

Therefore a new methodology in (Zuurbier, 2010) separated the dynamics involved with composition changes from the aggregated traffic flow modelling. This approach ensures FIFO behaviour by keeping track of composition changes. Another advantage is efficiency, the composition of traffic only needs to be dealt with at the beginning and end of the link.

In this approach, called System Class Dynamics (SCD), the composition of traffic is defined by a relative composition vector at the start and end of each link. When the composition of traffic that enters the link changes its composition, the vector at the entrance changes and a moving particle vector is created as a separation between the previous and new composition. Each composition vector is coupled to a z-value, that represents the number of vehicles on the link at the moment it is created. As traffic leaves the link, the composition vector particles are 'progressed' by decreasing their z-value accordingly. When a particle reaches the end of the link ($z = 0$) it overwrites the old composition vector at the end of the link. This last composition vector is used for the dynamics of the link outflow to the next node.

As written earlier, the simulation model used in this thesis is a modified version of the original DSMART model. The SCD approach was added, and instead of composition vectors π that specified the destination of traffic, composition matrices χ are used. This approach also includes the origin of traffic and could be of help for evaluation purposes. Figure 5.5 shows this (extended) SCD concept.

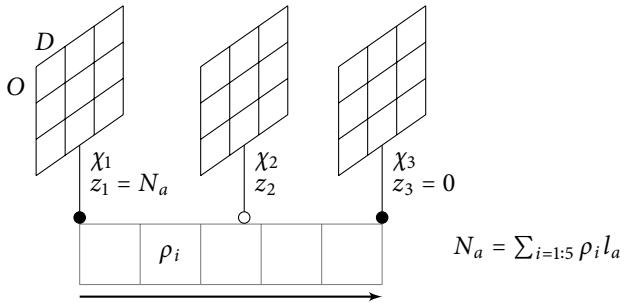


Figure 5.5 – A link with five cells with aggregated densities, and three matrices (two boundaries, one travelling particle) that represent the relative commodity composition of the System Class Dynamics (SCD)

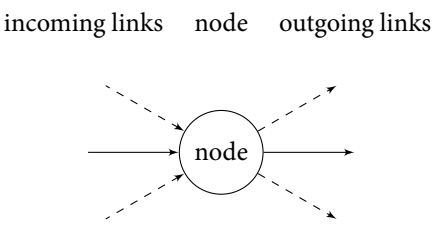


Figure 5.6 – Example of a node. Each node is connected to one or more incoming and outgoing links.

5.4.2 Traffic dynamcis at the nodes

When the link traffic flows have been calculated, the traffic dynamics at the nodes is the next step. The process at each node can be summarized as follows.

Based on the aggregated demand of incoming links, the total demand at the node is determined. Demand of links with a currently red traffic light is set to zero.

Although the dynamics of a fixed simulation time step are considered, the SCD approach requires that this time step is split into smaller parts if needed. The need depends on the change in traffic composition that needs to be considered if composition matrix particles approach the link boundary. As long as the elapsed time is smaller than the simulation time step, the following steps are taken:

- Based on the composition matrices, destination specific split ratios and aggregated demand for each incoming link, the demand for each outgoing link is determined.
- This information and the supply of outgoing links determine the proportion of the demand of each incoming link that can be 'handled' through the node. Here, a simple function is used to assign the supply proportionally to the demand.
- The fluxes between the incoming and outgoing links are determined.
- Based on these fluxes, determine the decisive time increment: the time it will take before one of the composition matrix particles reaches the node, or the remaining time to the end of the simulation time step.
- Using the found time increment, densities of the cells connected to the node can be updated, as well as the z-values of the composition matrices. Also, if one of the composition matrices reached the node, it replaces the one at the border.

After the whole simulation time step has been done, the composition of the outgoing links is determined, and if they differ from the composition matrix at the entrance, a new composition matrix and particle are created.

5.5 Traffic control model

This section outlines the traffic control model that was added to the simulation model in order to model traffic signal control and route guidance. The architecture of the main objects is described, and the way traffic control measures have a direct effect on the traffic flow. An important part is how the controllers themselves are modelled, the processes that are periodically run to adjust the traffic control setting, as shown as the outer loop(s) of Figure 5.2.

5.5.1 Traffic signal control simulation

Object structure for traffic signal control

There are three main object types that are used to model traffic signal control in the simulation.

The first is 'TSC' (traffic signal control). An overview of (most of) its properties:

- *node*: the ID of the node this TSC belongs to;
- *senders* and *receivers* are lists of the incoming and outgoing links;
- *phaseplan* is a three-dimensional matrix that contains the allowed traffic movements from senders to receivers, for each phase;

- *phasetimes* is a list that represents the division of cycle time among the phases;
- *currentphase* is set to the currently active phase;
- *nextphase* represents the time value when the phase should be changed;
- *cptime* records the elapsed time on the current phase, *waitingforgreen* is a list of per phase elapsed times without green (both are used for the time slotted method, to bound maximum green and waiting times).

Ideally, TSC is used as a general traffic signal object. To implement back-pressure as a strategy, a supportive object 'TSCBP' was added. Its main property is *mu* (μ) a three-dimensional matrix that holds the service rate information (fitting the phaseplan of TSC)).

The third object to mention is 'TSCESTIM', which is used to estimate the turn probabilities at the intersection by counting and registering traffic that passes the node.

Traffic signal control in the simulation process

The way traffic signal control is incorporated into the model is quite simple:

- As explained at the first part of Subsection 5.4.2 red traffic lights enforce the related link demand to be put to zero, to make sure no traffic leaves the link at the current time step. The links with red lights are found by taking the links of the intersection that are not green (in the current phase).
- If turn probabilities are estimated (instead of derived from route guidance settings) the TSCESTIM objects count and register the traffic that passes the intersection, as part of the node dynamics (Subsection 5.4.2). For the probabilities only recent measurements are used (latest two minutes).
- At the end of each simulation step, as indicated in Figure 5.2, the traffic control signals are checked for phase changes, using the property *nextphase* of TSC. If the time for the next phase has come, the *currentphase* property is progressed to the next phase. This step is not executed in the time slotted traffic signal control variants (the same phase keeps activated until the controller is updated).

Traffic signal controller

The traffic signal control methodology as discussed in Section 4.2 is translated to the simulation model as a set of scripts and functions. The controller is 'called' after a set period of (simulation) time, based on the variable *TSCtimesteps* (which represents either cycle time or time slot duration).

The main script is called *updateTSC* and performs a few steps. It checks for each signalized intersections (TSC) which strategy should be applied: fixed traffic control, vehicle actuated or back-pressure control. The procedure for fixed and vehicle actuated control is not treated here separately, as they are mainly simple versions of the back-pressure model.

In case of back-pressure, the following steps are taken:

1. The turn probability for all incoming links is retrieved, either from the TSCESTIM object or from current split vectors and traffic compositions (in case of route guidance).
2. The function *backpressureTSC* is used to calculate the link pressures (based on cell density values averaged of the last cycle time. It also uses the service rates of the TSCBP object, and the turn probabilities of the first step, to calculate the weight for each phase.

3. The function *determine_phasetimes* determines the new *phasetimes* property of the TSC object. This function follows one of the two approaches.
    - The first follows the 'fixed cycle time' method of Equation 4.5 to determine the new phase durations. These values are then discretized to fit into the 'grid' of simulation time steps (rounding to the nearest simulation step, considering that phase times add up to the cycle time).
    - The second branch of the function follows the 'slotted time' method of Equation 4.6. It first checks if the current phase can be changed (only if it has reached the minimum phase time parameter). Two functions check if the current phase time exceeds the maximum green time, and if one of more phases should have priority because of outrunning the maximum waiting time. The phase pressure of phases that are hereby excluded is modified to extremely low values. Now , the phase with the highest pressure receives a time slot of phase time.

After the new phase times have been determined, various TSC properties are set. First the current phase is set to the first phase that should be activated, and the *nextphase* and *cptime* time values are updated with the phase time of the first phase. As a last step, the *waitingforgreen* property is updated by resetting the values of the current phase. In case no phase is activated (all signals red) the procedure is slightly different, as the current phase is set to zero.

5.5.2 Route guidance simulation

Object structure for route guidance

The main object types that are relevant for route guidance have already been mentioned in Subsection 5.3.3.

As taken from the original DSMART model, en-route route guidance (or route choice) is enforced to the traffic flow through destination specific turn ratios. For each link and destination a reference to the turn ratio can be found in the SPF matrix, the actual ratios can be looked up in the SPLIT matrix (with the reference row). The SPF and SPLIT matrices are expanded when needed, that is when the route guidance controller (or route choice process) results in a turn ratio that is not yet included.

The explicit route definitions are stored in two ways. The ROUTES object stores all complete routes, but these mainly serve as an input to the model. The actually used partial routes are stored as property *subroutes* inside the link records of the LINKS object (only for links that have multiple outgoing links).

It can be useful to have quick access to the destinations that can be reached from a link, and which of the outgoing links can be routed to those destinations. To this purpose, a property object *routed* is also added to the LINKS. The entries of this property contain a destination and the optional outgoing links.

Route guidance in the simulation process

The effect of route guidance is modelled as part of the node dynamics, see Subsection 5.4.2. The specific traffic demand to each outgoing link is (per destination) represented by the turn ratio found with the SPF and SPLIT object.

Route guidance controller

The route guidance methodology as discussed in Section 4.3 is translated to the simulation model as a set of scripts and functions. The controller is 'called' after a set period of (simulation) time, based

on the variable *RGtimesteps*.

The route guidance model is an extension of the route choice and guidance model of DSMART, the main script is called *pathsupdate*. Depending on the parameter value of *choicemethod* a procedure is followed to implement the standard route choice method (See 5.3.3), or one of the route guidance options.

The route guidance process when the proposed back-pressure based strategy is applied has three options: using the 'first link' approach, of the approach with total routes, or a combination. The first two approaches are briefly described.

When only the first link of the routes is considered the process is relatively compact.

- For every link and for every reachable destination (*routed* property) a check is done to determine the optional outgoing links. If there is only one option, the guidance process can be completed by assigning all traffic to this destination through that link. Otherwise, the process continues.
- For each outgoing link the function *determinepressure_link* is used to determine the link pressure, based on minute averaged density and the set parameters. (Once a link pressure has been defined for one destination, it can be reused for the next.)
- Adding the service rate (based on link capacity) turn proportions are defined using the logit assignment.
- The calculated turn proportions are compared with the existing SPF and SPLIT objects, which are updated if necessary.

When the total routes are considered, the process is as follows:

- Every link with *subroutes* is treated. For each reachable destination the route set for this destination, out of *subroutes* is created.
- For each partial route in the route set the service rate is determined, based on the (outgoing) link capacity.
- The function *determinepressure* calculates the pressure value for the whole route, based on minute average densities and the parameters that define the chosen pressure function.
- If user preference is considered, the instantaneous link travel times are retrieved.
- The overall utility of the partial routes is determined, based on the route pressures and user preference (as in Equation 4.13).
- The function *assignpressures* takes the route utilities and determines the proportions of each route, by the path size logit assignment (or multinomial logit, if wanted).
- The route proportions are converted to turn proportions for each outgoing link.
- The calculated turn proportions are compared with the existing SPF and SPLIT objects, which are updated if necessary.

5.6 Summary of the simulation environment

This chapter gave an overview of the structure and processes in the simulation environment:

- The choice was made for a first-order macroscopic simulation model. The simulation model of DSMART is taken as a starting point, and is to be expanded, particularly to incorporate the traffic control methodology.
- The core of the simulation model is represented by a simulated traffic flow process that is repeated for all simulation time steps. Periodically, as an outer loop, traffic controllers are called to update the activated control strategies.
- The network is represented by links and nodes. Dynamic traffic demand is generated and transported from origin to destination node, using routes.
- The traffic flow model is a kinematic wave model, based on the fundamental diagram of each link. Space and time are discretized (into cells and time steps) and a Godunov scheme based process is used to progress aggregated traffic along each link. System Class Dynamics (SCD) ensure the separation of changes in traffic composition. Aggregated cell densities and SCD information are used to handle the traffic dynamics at the nodes.
- The traffic control model is added to DSMART. Traffic signal control objects directly influence the traffic flow and route guidance is enforced by manipulating turn ratios at intersections and diverges. Periodically, the functions that represent the traffic controllers update their strategy. This goes for traffic signal control and for route guidance.

Chapter 6

Simulation experiments

The previous chapters have defined experiment requirements (Section 3.3.1 of the Research approach), the theoretic framework (Chapter 4), and the simulation model (Chapter 5). The proposed model can now be tested thoroughly in this chapter.

Section 6.1 gives an overview of the experimental set-up. Three case studies are discussed in the next three sections. Section 6.2 describes case 1, which evaluates the signal control model. Then, case 2 does the same for the route guidance model as presented in Section 6.3. The final case, case 3, combines the two control models and is the subject of Section 6.4. Section 6.5 summarizes the experiment.

6.1 Overview of experiment

6.1.1 Three cases

The experiment consists of three parts.

1. In the first part traffic signal control is examined. Back-pressure control variants are compared to a vehicle actuated and fixed control situation. A simple and a little more expanded network are evaluated.
2. The second part focusses on route guidance. Again, several variants are compared. Two networks are used, one with three routes (of which two overlap) and the other is similar but with one route considerably longer than the others.
3. The third case combines the two modules of traffic signal control and route guidance. Several combinations of variants are evaluated and compared. The network is somewhat more complex than that of the two first cases.

6.1.2 Performance indicators

The performance of each simulation can be measured by means of indicators, considering the network and time span of the simulation as a whole, or looking at specific locations, target groups, or time periods.

Basic network indicators

To evaluate the aggregate simulation performance, the following network indicators are calculated:

1. the Total Time Spent (TTS), the time spent by all vehicle in the network during the simulation;
2. the Total Delay (TD), TTS minus the time spent if all vehicles had moved with 'free flow speed'[1];

[1] In the simulation the free flow speed is a link specific maximum speed.

3. the Total Distance Travelled (TDT), the distance covered by all vehicles during the simulation.
4. The average network speed can be derived by: TDT/TTS.

Other performance indicators
Other ways to evaluate the traffic performance:

- The (Generalized) Funadamental Diagram. Back-pressure control aims at high throughput, and comes with (relatively) evenly spread densities. This can be evaluated by means of the GNFD.
- Specific links or routes should be looked separately to explain observations.
- Control variable diagrams, to explain observations from the control settings.
- Travel times, time–space diagrams.

6.2 Case 1: Traffic signal control

In Case 1 traffic signal control is examined. The goal is to evaluate back-pressure control compared to vehicle actuated control, and fixed control.

Two case consists of two parts, case 1A and 1B. The networks have only simple traffic signals, with two phases each. Phase lengths are a multitude of the simulation time step of 2 seconds. In both parts the control variants are the same (see section 4.2.1):

- Fixed: signal control with a cycle time of 1 minute, two phases of 30 seconds.
- Vehicle actuated (VA): cycle time of 1 minute.
- Back-pressure with fixed cycle time (BPcycle): cycle time of 1 minute, $\theta = 4$ (Eq. 4.10).
- Back-pressure with time slots (BPslot): time slots of 6 seconds.

For both back-pressure variants, Equation 4.8 with $m_1 = 2$ is used to determine the pressure.
The simulation runs for a period between 7 and 9 o'clock, with a half hour warming up time before.

6.2.1 Case 1A: two intersections

The network for Case 1A is a series of two simple intersections and unidirectional roads, as in Figure 6.1. All links have one lane, a maximum speed of 60 km/h, and are 2 km long, except for link 2 (0.5 km). Figure 6.2 shows the profiles of traffic demand for each direction, the route from node 1 to 2 has the highest traffic demand. To keep this case as transparent as possible, all traffic goes straight on.

The hypothesis for this case is:

- Fixed control can't handle the higher volume on link 1, a queue will form; other directions profit.
- VA control will give priority to stream 1 to 2. The crossing directions face longer queues.
- Back-pressure variants will perform in between. As link 2 gets congested and causes 'back-pressure' to link 1, the latter will be queued more, leaving more space for traffic from 3 to 4.

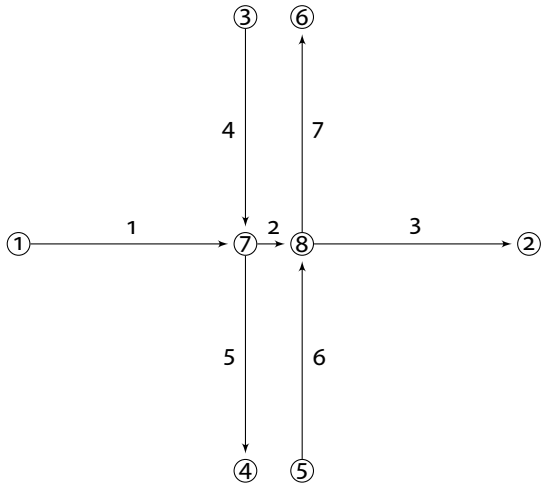


Figure 6.1 – Network Case 1A

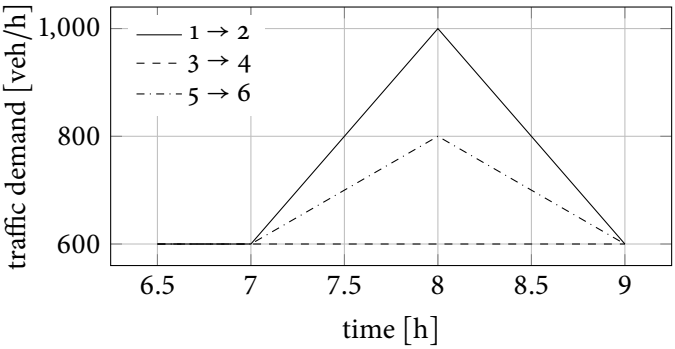


Figure 6.2 – Origin--Demand data Case 1A

The simulation results confirm the hypotheses for the main part. Figure 6.3 shows the delay on the four links that lead to the intersections. The total delay with fixed signal control is low compared to the other variants. Only on link 1 does it have (as expected) the highest delay value. It is noted that (only) in this case at the end of the simulation there was still a queue covering the whole length of link 1, meaning that there were (approximately 70) vehicles still waiting outside of the network. The simulation model doesn't account for vehicles waiting outside of the network when determining the delays, only traffic on the links is measured.

The delays on link 2 and 4 are largest for VA control. Traffic from link 1 to 2 is given a lot of green time, which causes queueing on the crossing link 4, and on link 2. With the back-pressure variants, queues on link 2 impose a decreasing throughput from link 1, enabling link 4 to release its queue. For link 6, the difference is smaller, because back-pressure from the links downstream of the intersection at node 8 play hardly a role. Variant BPcycle has a lower delay on link 6 than BPslot. This is (probably) due to the fact that in BPcycle every phase is served every cycle. In case of BPslot, link 2 is the dominant traffic stream, and link 6 has to built enough queue to get its turn. This mechanism is also the cause of the higher delays on link 1, for BPcycle versus BPslot (more green time to link 6,

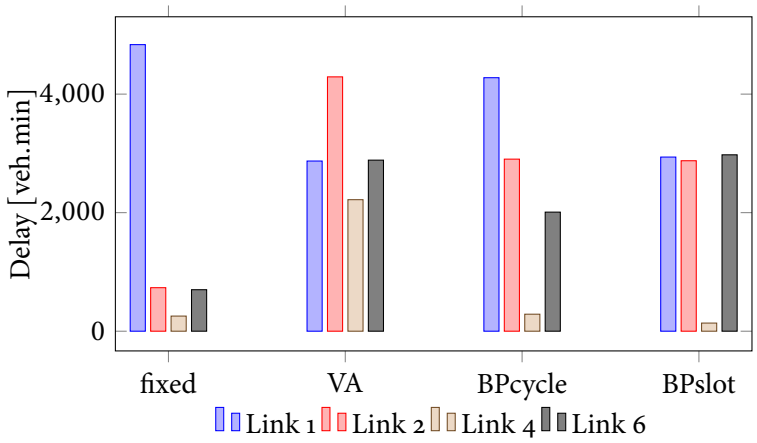


Figure 6.3 – Case 1a, delay per link

less to link 2, and consequently less green time to link 1).

Figure 6.4 shows the network production (based on NFD data) over the simulation time, for each variant. This diagram shows that total throughput is the highest for BPslot, which reaches the highest peak flow. Looking at the total flows that passed the intersections shows similar indications.

Overall, this case works as expected.

6.2.2 Case 1B: small grid network

For Case 1B, the network is a small grid network (Figure 6.5), with two main routes (north–south) and three crossing routes (asymmetrical). The main routes have 2 lanes, the others 1. The links are 1 km or 0.5 km long. The OD profile is in Figure 6.6.

The hypothesis for this case is:

- Fixed control will lead to long queues on some links that aren't adequately served by their intersection.
- The back-pressure signal control variants are expected to perform well, and well-balanced.
- VA control performs less effective than back-pressure control, as concluded in Case 1A.

Unfortunately, the simulation did not deliver all the desired results. A lot of queueing took place outside of the network, and not accounted in the traffic performance (such as delay) of the network (a explained in Case 1A).

From the running simulation images, the observed effects were as expected. A new simulation with longer links, or lower traffic demand, could give more insight in the network performance. The diagram in Figure 6.7 shows that again the highest production in the network is reached with back-pressure control, especially the slotted variant.

6.3 Case 2: Route guidance

In Case 2 route guidance is examined. The goal is to evaluate variants based on back-pressure control and compare them to the standard route choice behaviour (probit route choice, in the simulation model).

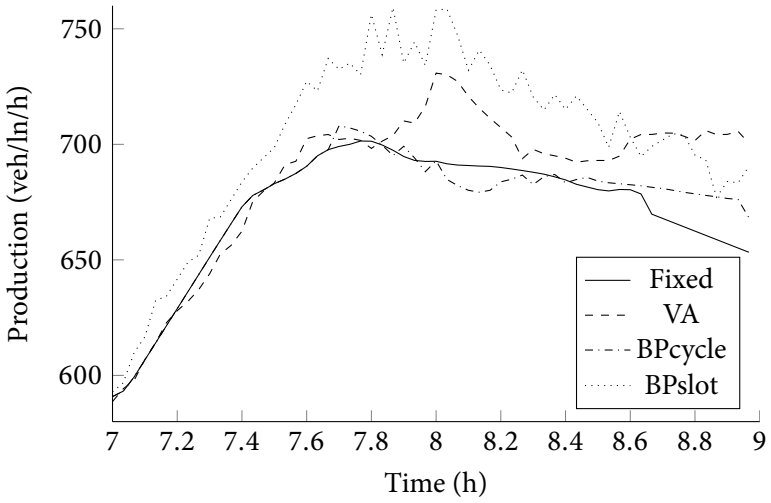


Figure 6.4 – Case 1A, Average network production

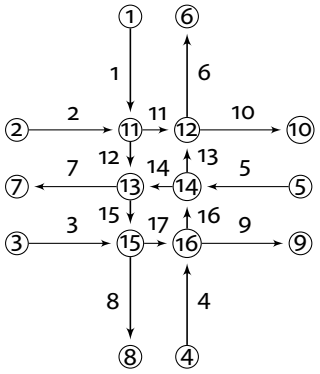


Figure 6.5 – Network Case 1B

Two case consists of two parts, case 2A and 2B. The networks contain a simple link structure with three routes, of which 2 overlap. The capacity of the links has been adjusted to create bottlenecks along the routes.

The variants are:

- Reference variant (Std): probit route choice (9 trials). This stochastic variant is run three times, to determine the average behaviour.
- Back-pressure based on the first link (BP-1st)
- Back-pressure based on multinomial logit (BP-MNL)
- Back-pressure based on path size logit (BP-PSL)
- Back-pressure based on a combination of the first link and PSL (BP-1stP)

In case 2B 'user preference' terms are added:

- Back-pressure based on the first link, and distance (BP-1st*)

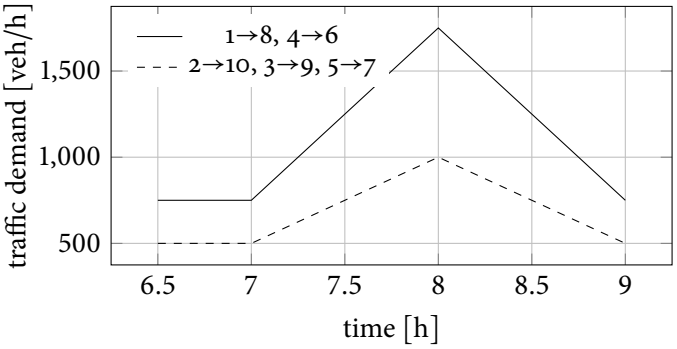


Figure 6.6 – OD Case 1B

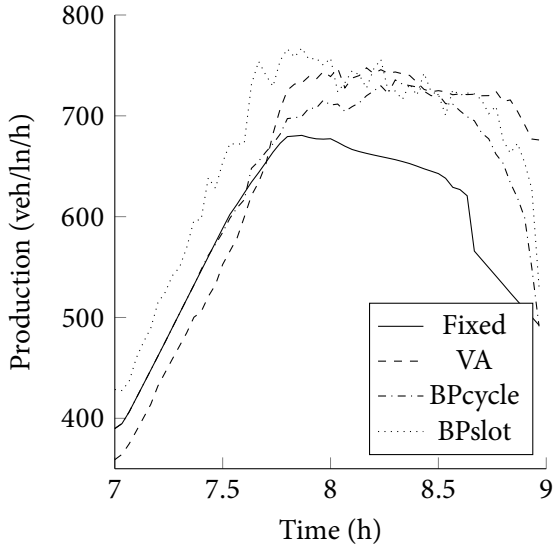


Figure 6.7 – Average network production, Case 1B

- Back-pressure based on path size logit, and travel time (BP-PSL*)
- Back-pressure based on a combination of the first link and PSL, and travel time (BP-1stP*)

In Table 6.1 the alpha values are listed for each variant (Equation 4.13).

For back-pressure variants, link pressures are based on the relative density, and route pressures are based on the weighted average link pressure. The parameter $\theta = 10$, for all (PS) logit functions (after testing this seemed to give well-balanced results). Equation 4.17 is used to incorporate the service rate, but it shouldn't have influence since the capacities are equal for each route decision.

Full compliance is assumed, all vehicles follow the advice given by route guidance.

The simulation runs for a period between 7 and 9 o'clock, with a half hour warming up time before. Time step is 5 sec, and the update frequency for route choice / guidance 30 seconds.

Table 6.1 – Alpha parameters per variant

Variant	α_1	α_2	α_3
BP-1st	0	1	0
BP-MNL	1	0	0
BP-PSL	1	0	0
BP-1stP	0.5	0.5	0
BP-1st*	0	1	0.5
BP-PSL*	1	0	0.5
BP-1stP*	0.5	0.5	0.5

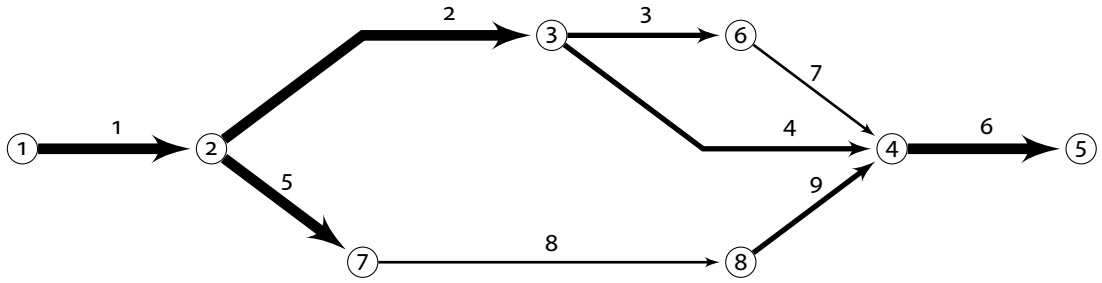


Figure 6.8 – Network Case 2A

6.3.1 Case 2A: three routes

The network for Case 2A is sketched in Figure 6.8. There is one origin and one destination, three routes. Links are 5 or 10 km long, and have 1 to 4 lanes (indicated by line width). The profile of traffic demand is illustrated in Figure 6.9. The network is overloaded for most of the simulation time, as the capacity of link 1 is 8000 veh/h and the demand varies from 7500 up to 15000 veh/h. This is not a problem, it means that the network is evaluated at peak loading.

The hypothesis for this case is:

- Bottlenecks at link 7 and 8 will cause congestion on those routes, to which the algorithms should respond.
- Variant BP-1st will direct traffic via the route(s) that are not congested at the first link.
- Variant BP-PSL will direct more traffic through the route via links 5–8–9 than BP-MNL, due to overlap of the other two routes.

The simulation results leave room for improvement (see Figure 6.10–6.12). The standard (probit) route choice performs best, in terms of total delay, (low) remainder at the end of the simulation, and highest average speed. A plausible explanation is that all back-pressure variants send more traffic through the bottleneck of link 8. This is caused by the definition of the route pressure. In this case, an average of the link pressures is used. When link 8 becomes a bottleneck, only link 5 gets congested. For the route including link 8 and 9 the overall pressure is relatively low. Further study is required to see if another route pressure method (Section 4.3.5) would perform better.

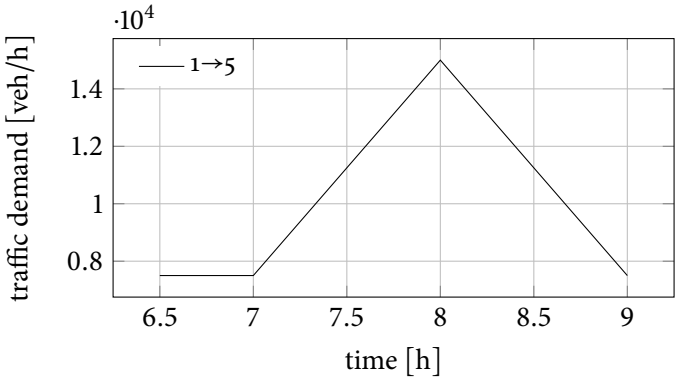


Figure 6.9 – OD Case 2

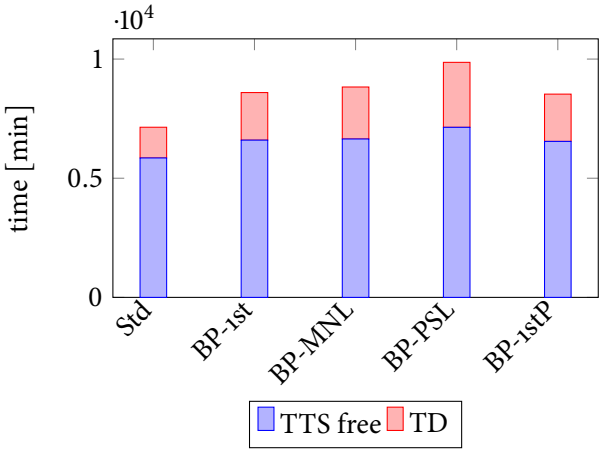


Figure 6.10 – Case 2A, Total traveltime spent including total delay

As expected, BP-PSL does send more traffic through link 8 than BP-MNL, which explains the bad performance of BP-PSL. BP-1st performs relatively well, since it reacts on the congestion on link 5, the same goes for BP-1stP.

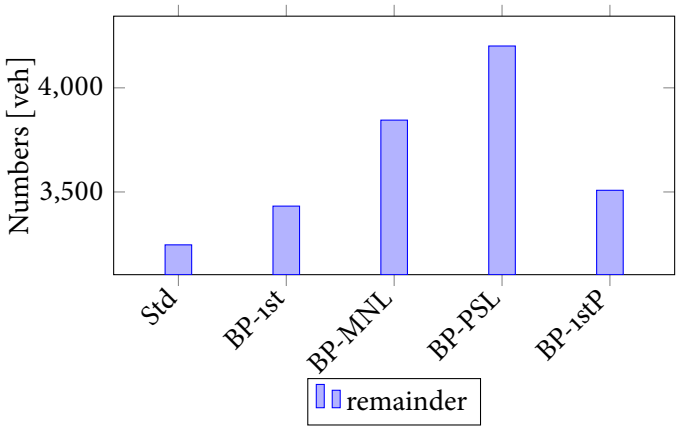


Figure 6.11 – Case 2A: Remaining vehicles in the network after simulation

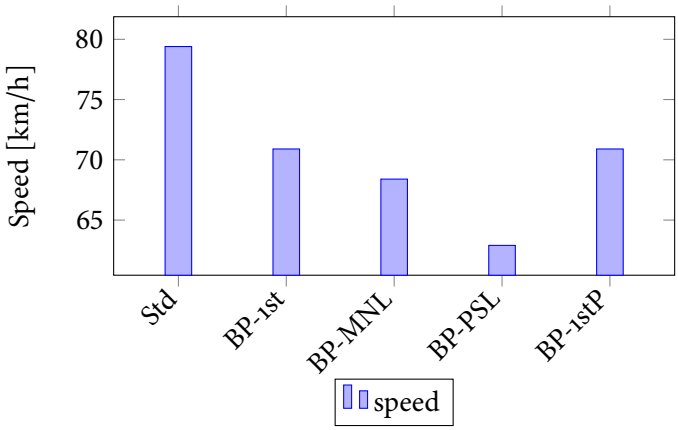


Figure 6.12 – Case 2A: network average speed

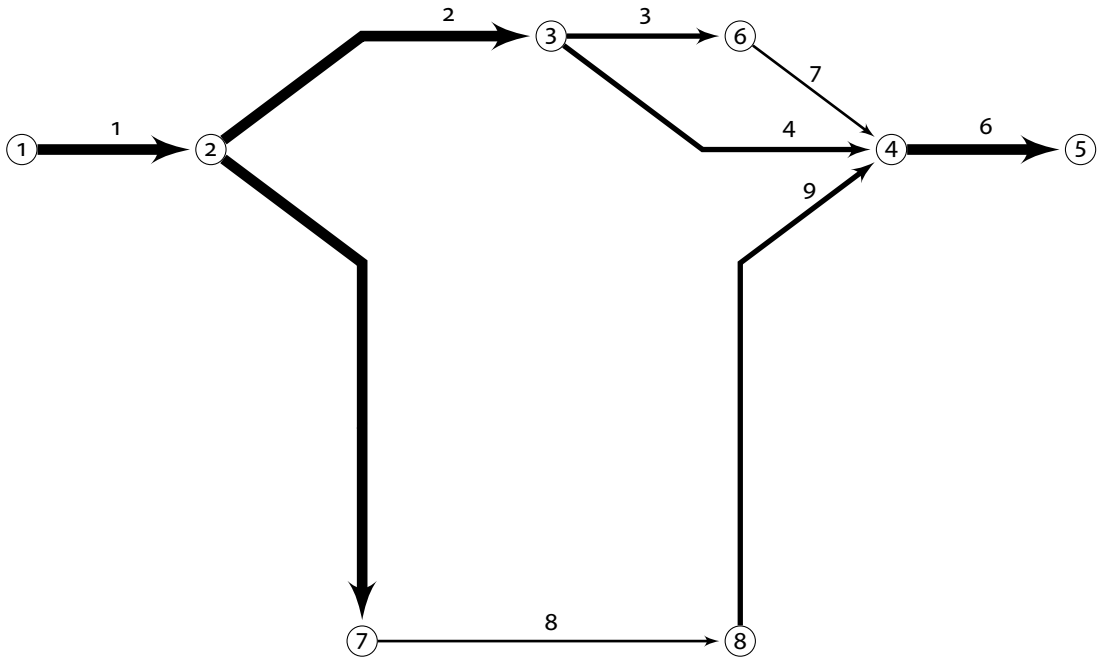


Figure 6.13 – Case 2B network

6.3.2 Case 2B: three routes (one is longer)

The network of Case 2B (Figure 6.13) is similar to Case 2A, except that link 8 is positioned 10 km downwards, making the route twice as long. The traffic demand is the same as Case 2A.

The hypothesis:

- The standard route choice will prefer the two top routes, because of the shorter travel time. The bottom route will only be taken if, due to congestion, the travel time reaches a value similar to the bottom route.
- The (first 3) back-pressure variants guide traffic much like the case 2A, since it doesn't consider travel times.
- The back-pressure variants that take into account travel time will end up somewhere in between.

The simulation results are overall as expected 6.14–6.17. The difference in TTS and TD between standard route choice and the back-pressure variants is now much smaller. That is due to the fact that standard route choice uses the two upper routes, which become highly congested. The variants of back-pressure that include user satisfaction (travel time) have the lowest delays, perform best. They seem to find a balance between short individual travel times and the spread of pressure (density). The number of remaining vehicles and the speed values show the same pattern.

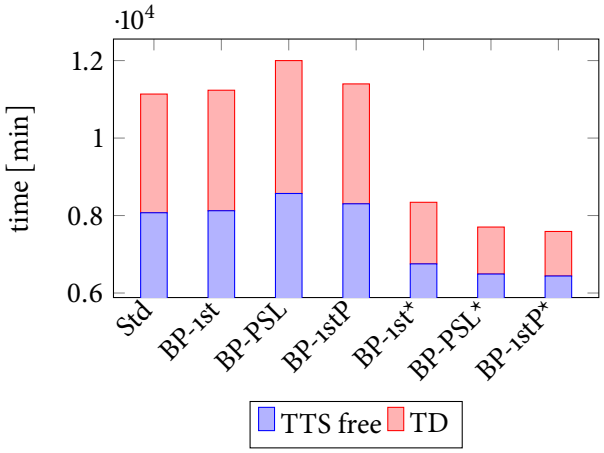


Figure 6.14 – Case 2B, Total traveltime spent including total delay

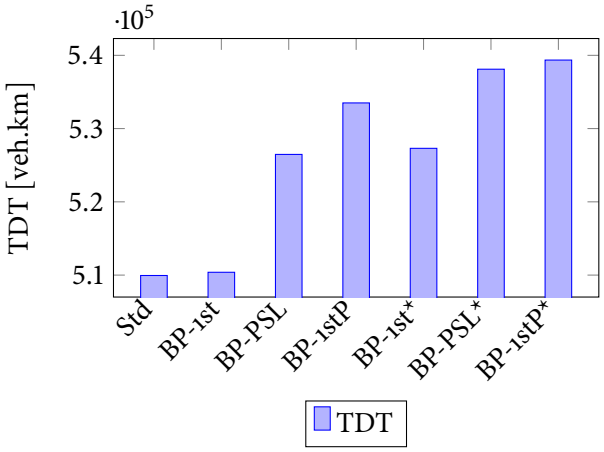


Figure 6.15 – Case 2B, Total distance travelled

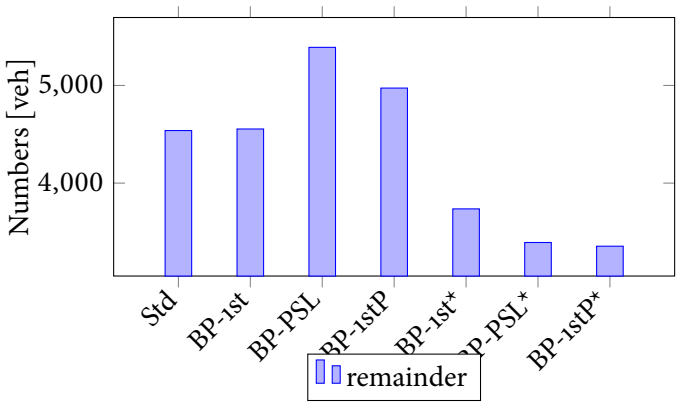


Figure 6.16 – Case 2B, Remaining vehicles in the network after simulation

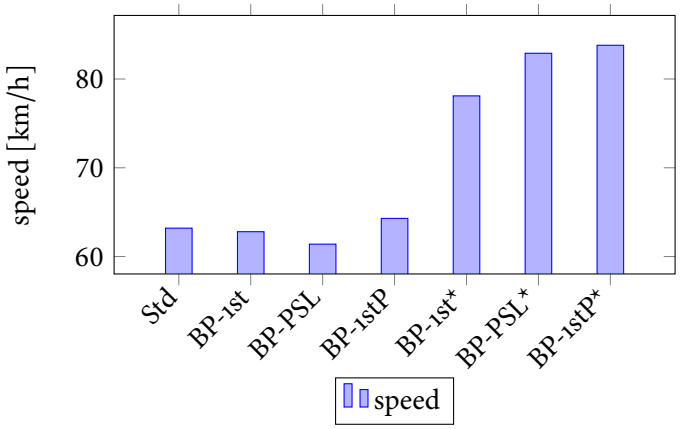


Figure 6.17 – Case 2B, network average speed

6.4 Case 3: traffic signal control and route guidance

General

In case 3 traffic signal control and route guidance are combined. The network is shown in Figure 6.18, and is derived from (unknown). The standard length of links is 1 km, and each has one lane, except the links from the origins (node 1 and 2) and to the destinations (node 3 and 4). The traffic demand (Figure 6.19) is equal for all 4 OD relations. Traffic signal control is installed on nodes 8, 9, 11–15, and 17.

Simulation control scenarios

This case study evaluates an overview of combinations, based on the variants in the previous two cases. Four traffic signal control variants:

- fixed cycle control;
- vehicle actuated control (VA);
- back-pressure with fixed cycle time (BPcycle);
- back-pressure with time slots (BPslot).

Six route guidance variants:

- reference variant (Std): probit route choice (9 trials, 3 simulations to determine the average behaviour);
- back-pressure based on the first link (BP-1st);
- back-pressure based on path size logit (BP-PSL);
- back-pressure based on a combination of the first link and PSL (BP-1stP);
- back-pressure based on a combination of the first link and PSL, and travel time (BP-1stP*);
- back-pressure based only on travel time estimation (BP-TT). This one is introduced in this case as another point of reference. Parameters: $\alpha_1 = \alpha_2 = 0, \alpha_3 = 1$.

Each combination of these 2 sets of variants are evaluated by simulation. The simulation settings (parameters) are taken from Case 1 and 2, except the following. The update frequency for route guidance is now once per minute. The simulation step is 5 seconds, which allows the signal control phases a 'precision' of 5 seconds. The time slot (BPslot) is 10 seconds.

The simulation runs for a period between 7 and 10 o'clock, with a half hour warming-up time in advance.

Hypothesis

The hypothesis for this case is that the combined approach will show a similar performance for traffic signal control and route guidance to the previous cases.

Results

Regarding the general network performance, Figures 6.20–6.23 show the main differences. Some observations:

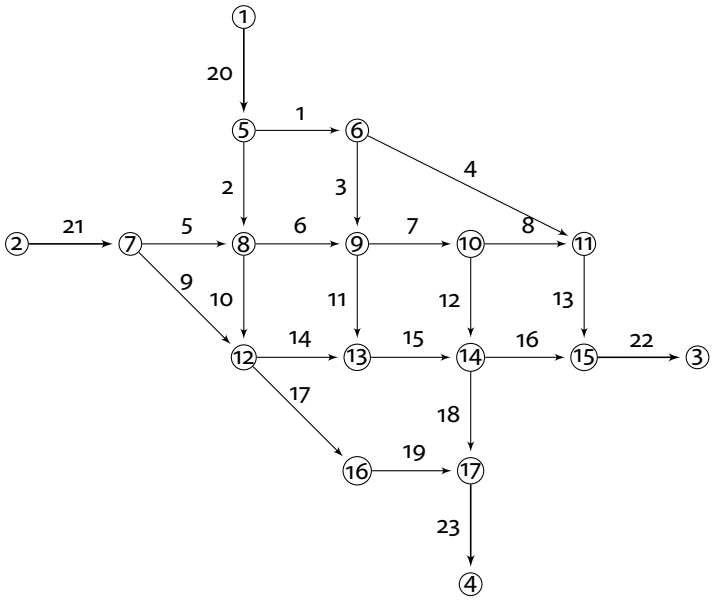


Figure 6.18 – Case 3 network

- Fixed signal control performs worst (high delays, low overall speed, least in- and output). This was expected. In defense of fixed signal control, the control settings (all 50–50 tactics) hadn't been optimized (or coordinated), other settings could have given somewhat better results. Still, fixed control isn't adaptive at all.
- In this network, vehicle actuated control seems to work best. The reason behind this is not clear, a possible explanation is that in this case pressures downstream of intersections are relatively equal and therefore don't play a big role.
- BPslot performs almost as good as VA control. There is a possible explanation for not performing better. If two intersection approaches (or phases) have nearly the same pressure, but one continues to dominate the other direction, then the queue on the other direction that doesn't get served might back-propagate through the network.

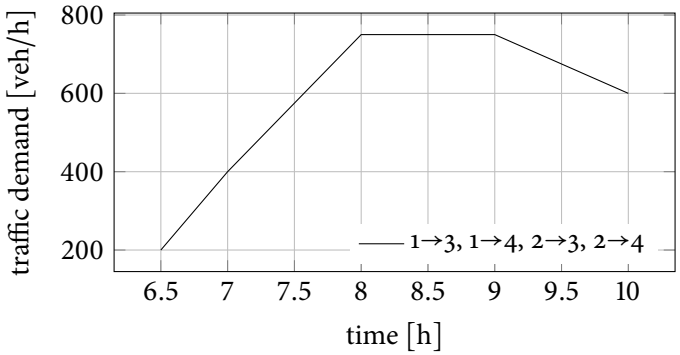


Figure 6.19 – OD Case 3

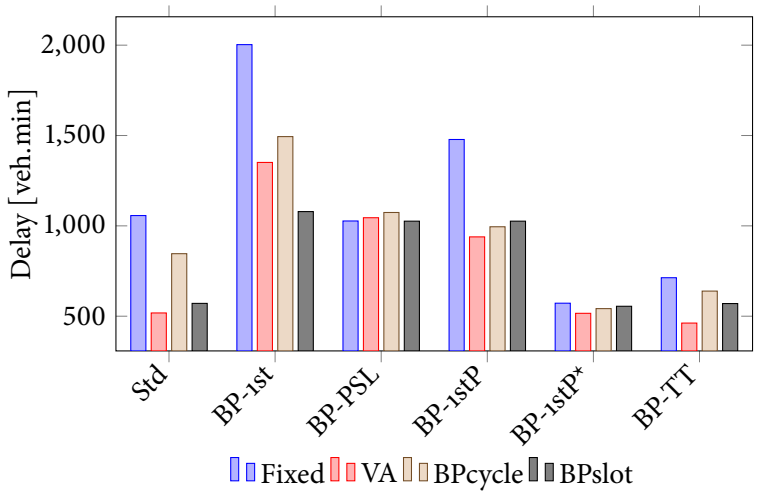


Figure 6.20 – Case 3, Total delay per combination of variants

- BPcycle performs slightly worse.
- Regarding variants of route guidance, the standard route choice performs well. The difference with BP-TT is small, and this is logical, since both are based on travel time (just different algorithm).
- The back-pressure variants of route guidance don't perform that well. This probably has to do with the pressure function not being optimal, as described in case 2.
- The back-pressure combination (BP-1stP*) performs best and is stable among its signal control variants. Probably taking a combined strategy (of search low densities and short travel times) leads to good overall results. Of course this is no full proof, since only one network has been tested, at one traffic demand profile.
- From the simulation images (not -yet- in the report) it becomes clear that BP-1st is a 'myopic' approach of route guidance. Link 15 is the first link to get congested. However traffic from origin node 2 to destination 3 is still sent through this path. Only when congestion spills back to link 14 and 9, more traffic is guided along links 5, 6 and 7.

6.5 Summary of simulation experiments

The simulation experiments are divided into three parts: case 1 examines traffic signal control, case 2 focuses on route guidance, and case 3 combines traffic signal control and route guidance.

For traffic signal control, back-pressure control is a good way to generate high throughput at the intersections, while keeping the queues evenly distributed and within boundaries. Back-pressure signal control based on time slots is more effective than the cycle time based version. Some aspects of the algorithm and its practical use require special attention.

For route guidance an optimal use of back-pressure has not yet been found, as the results are on a most part lacking compared to the standard route choice model. Two main aspects are the definition of a representative route pressure, and the capacity of routes related to the assignment of (destination

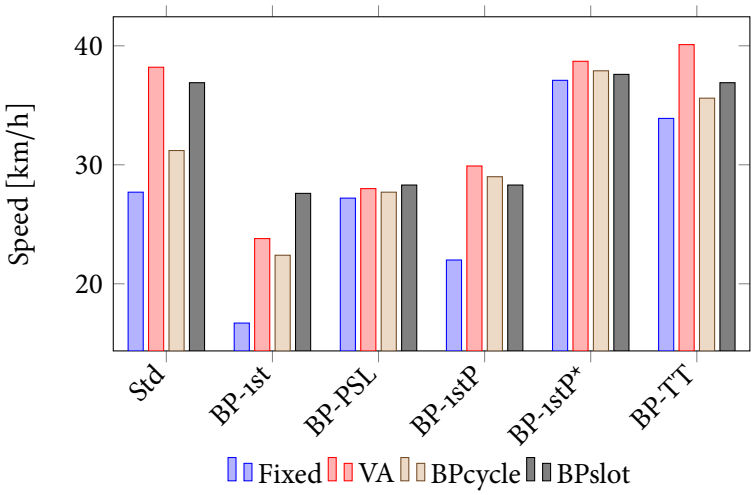


Figure 6.21 – Case 3, Average network speed per combination of variants

specific) traffic to each route. The performance is expected to be better if the pressure were (partly) based on the critical links, instead of on average density. Yet the basic algorithm that was formulated works to some extent, and so far the simulated effects can be understood and are up for improvement. It was also found that a 'pressure' function that includes not only route densities, but also travel time can give good results.

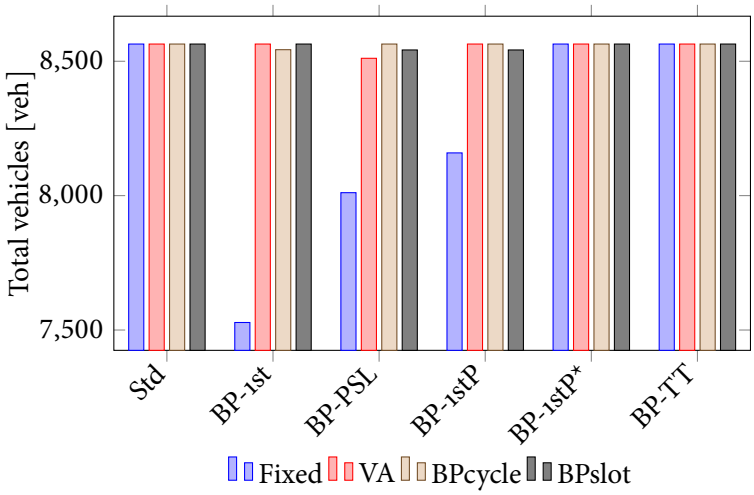


Figure 6.22 – Case 3, Input

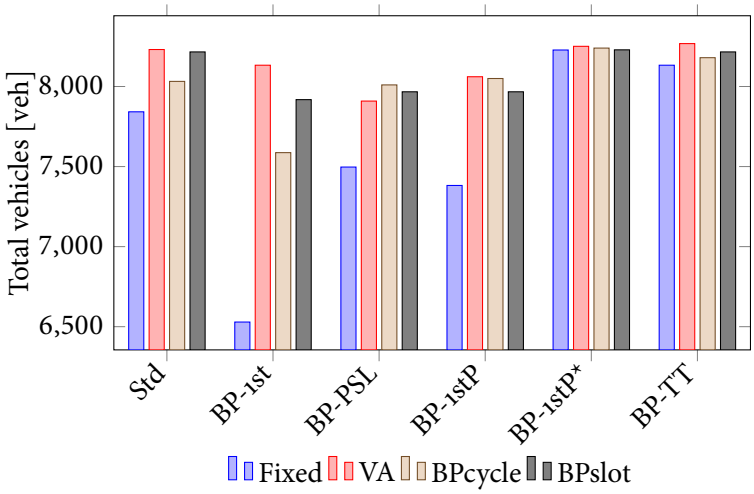


Figure 6.23 – Case 3, Output

Chapter 7

Discussion

The experiment of Chapter 6 has produced results that have been analysed to some extent, but require further discussion. This chapter does two things. In Section 7.1 the results are condensed to a summarizing evaluation. Then, Section 7.2 offers ideas for improving the methodology. These observations will be used in the conclusions and recommendations in the final chapter.

7.1 Evaluation of results

The results from the experiment have can be summarized to a few key points:

- For *traffic signal control*, back-pressure control is a good way to generate high throughput at the intersections, while keeping the queues evenly distributed and within boundaries. Back-pressure signal control based on time slots is more effective than the cycle time based version. Still, some aspects require special attention (see the following sections), related to the algorithm itself and its practical use.
- For *route guidance* an optimal use of back-pressure has not yet been found. Two main aspects are the definition of a representative route pressure, and the service rate for routes related to the assignment of (destination specific) traffic to each route.

  Still, the basic algorithm that was formulated works to some extent, and so far the simulated effects can be understood and are up for improvement. It was also found that a 'pressure' function that includes not only the 'fullness' of the route, but also travel time from a user perspective can give good results (although not tested, this approach is expected to work well during low traffic densities, with higher traffic loads the travel time term will have less influence compared to the vehicle density).

7.2 Ideas for improvement

After having developed the model and tested it in a simulation experiment, a lot of things came up that could have been better or should be improved when the research on this topic is to be continued.

Enhancements to the control model

On the part of traffic signal control:

- The used algorithm could be made applicable to complex intersections with advanced approaches and phase modelling, and taking clearance times into account.

- The back-pressure signal control model with time slots could be enhanced. One proposal is to add 'virtual' pressures related to waiting times at an intersection. This enables 'fair treatment' for every queue, and prevents continuous domination by one phase.
- If back-pressure signal control model based on cycle times is further adopted, each intersection could have an independently optimized cycle time.

On the part of route guidance:

- The formulation of a representative route pressure should be further investigated, for example with the suggestions in the control modelling sections in this thesis. Taking the (weighted) average of relative densities has shown to perform weak.
- As mentioned before as well, the way the 'service rate' is used, or just the assignment of traffic to the routes, should be further researched.
- Related to the previous point, it might be worth to find a way to deal with queues at the location of route guidance. For example, the queues on approaches to a signalized intersection are currently not taken into account, as well as congested lanes before a highway diverge.
- So far, explicitly defined routes are used in the route guidance model. This is hard to maintain at large networks. At the same time, one of the advantages of back-pressure control is that it can be applied in a distributed system. Therefore, it is suggested to investigate distributed types of route guidance, for example by applying the recursive logit route model.

Testing

The control model proposed in this thesis has a lot of variants and room for options. However, the experiment has been too limited to test every idea or calibrate parameters (they have been calibrated roughly through trial-and-error). Also, to proof if an algorithm really works, more networks, more traffic demand scenarios should be evaluated. Not only the new control model, also the reference control systems could be evaluated on the same situations to compare the control model with, for example, optimal control or model predictive control.

Besides the scope of testing, perhaps a more detailed simulation model would fit the control approach better. For example, intersections can only be modelled roughly in the simulation model used in this research.

Chapter 8

Conclusions and recommendations

The possibilities of using the concept of back-pressure control for traffic signal control and route guidance have been described and examined in the previous chapters.

This last chapter concludes the research of this thesis. Section 8.1 reviews the main objective (Chapter 1) in two steps. Subsection 8.1.2 evaluates the components of the study with regard to the research questions. Then, subsection 8.1.1 draws the overall conclusions. As the research contains some loose ends and gives rise to new questions, recommendations for further research are made in section 8.2.

8.1 Conclusions

This section aims to give an overview of the main findings of the research done in the thesis. Most importantly, the main objective as presented in the Introduction (Chapter 1) is to be evaluated.

The main objective of this thesis is:to develop a framework that integrates route guidance and signal control based on the back-pressure principle, and determine its feasibility and potential benefit.

The objective has been divided into a series of research questions, in order to set out the necessary steps to accomplish the objective. Next, first the overall conclusions are stated, followed by the supportive answers to the research questions.

8.1.1 Overall conclusions

The overall conclusions of this research are:

- This thesis demonstrates that it is possible to create a methodology, based on back-pressure control, that integrates traffic signal control and route guidance.
- Traffic signal control based on back-pressure control performs well in the simulations, especially the variant with time slots. Throughput is high and queues remain within reasonable boundaries.
- In the context of road traffic management it is hard to fully integrate traffic signal control and route guidance, which is much more straightforward in the field of wireless communication where the method has its origin. A modest step of integration is to use route guidance settings to determine turning probabilities at the intersection, which are used to determine the influence of queues on outgoing links.
- Using back-pressure for route guidance requires some (artificial) design choices. The challenge is to define a representative function of route pressure or utility, and to combine this with a service rate value, in order to obtain a high throughput and stability with minimum delays. Several ideas have been presented. The average density has turned out not a good measure for route pressure. It is possible to combine factors of pressure based on density and travel time, in

order to use the shortest routes in case of low traffic, and shift to the routes with open capacity if needed.

8.1.2 Evaluation of research questions

1. What are the characteristics of route guidance and what are its strengths and weaknesses?

In chapter 2 route guidance has been described as a means of traffic control that influences the route choice of road users. Depending on the used algorithm route guidance can be used to optimize travel time from a user perspective or for the network as a whole. In operational route guidance routing decisions are often based on instantaneous travel times, but a predictive approach that uses expected travel times is possible as well. The benefit of route guidance is that traffic can be directed along routes that lead to low travel times and delays. The positive effects of route guidance can be influenced negatively by overreaction, oscillation, and delayed decisions. Besides, the compliance of route guidance is an issue.

2. How can traffic signal control complement route guidance into an integrated approach?

Traffic signal control is used on signalized intersections. A control system that contains both traffic signal control and route guidance could be structured in a way that the two methods can cooperate. By sharing information the integrated control parts could be optimized and could even aim for common strategies. A first effective step is to make traffic signal control depending on route guidance. In this architecture route guidance determines the higher level (strategic) control, and traffic signal control is used facilitate the occurring traffic flows with the knowledge of the direction of traffic.

3. How can the concept of back-pressure control be used as an algorithm for traffic signal control?

Back-pressure control has its origins in multi-hop communication networks. It is used to assign server activation to move data packets to their destination. The main strengths of back-pressure control is that the total throughput is maximized, and that the queues in the network are stabilized as much as possible. The basic back-pressure algorithm can been extended to cope with finite queues and delay.

On the application of traffic signal control back-pressure control has been used in literature. The concept aims to activate the signalling phase that has the highest total weight, a summation of the weights of the allowed traffic streams. The weight is the product of pressure and service rate (saturation flow). The pressure of a traffic stream is the difference between the queues at the incoming link and outgoing links (weighted by proportion). There are several variants to this basic model.

4. How can the concept of back-pressure control be used as an algorithm for route guidance?

In multi-hop communication networks the use of routes is a variable and an outcome of the algorithm, whereas in the case of intersection control route choice is used as a fixed input. The back-pressure concept can however also be used as a method for route guidance. Instead of determining the right phase to be activated, the task is to determine the ratio to direct traffic to following links. This is done for each destination group, that each has its collection of routes. For each route a certain pressure can be formulated, that expresses how filled up the route is (or the counterpart, how much capacity it has left). Each route can also be coupled with a service rate. Route pressure and service rate are less straightforward to define than for the intersection control. Then, a choice model is to be used to determine the ratio of routes, based on the product of route pressure and service rate.

5. What is the purpose of the new control approach, and what are the design requirements?

Starting from the objective in the Introduction, in Chapter 3 a system design concept is presented, that defines a control loop and the parts that represent the control system. Also design choices are made for the two 'modules' traffic signal control and route guidance. Lastly, the chapter contains the scope of the proposed experiments and the simulation model to be used.

6. What are the proposed control algorithms and their alternatives?

This question has been mainly covered in Chapter 4. The general back-pressure algorithm for traffic signal control consists of three steps: 1. determine turning probabilities and back-pressure values for each traffic stream; 2. determine the accumulated weight (back-pressure value) for each phase; 3. allocate the right amount of time to the phase(s) to be activated.

Two main variants for back-pressure algorithm were formulated: fixed time back-pressure, and time slotted back-pressure. The first presumes a fixed cycle time and assigns a portion to each phase, following a logit choice function. The second method assigns the whole current time slot to the phase with the highest weight.

Pressure values are based on the density on the links, averaged over a short time period and normalized to the jam density. A variant where this pressure value is further modified by a power function makes that lower pressures are further decreased and higher pressures count heavier.

The turning probabilities of the first step can be based on measurements or on known route guidance settings, which integrates traffic signal control and route guidance.

For route guidance the following general algorithm is proposed. For every link with multiple next links, and for every destination that can be reached through this link: 1. for every (partial) route starting from the current link to the current destination, determine the route pressure and the service rate value; 2. determine the utility for these routes; 3. determine the proportions for these routes; 4. determine the new split vector for the current link and destination.

A first attempt for the service rate is to use the capacity of traffic flow from the current link to the first link of the route.

Many varieties of the general algorithm have been considered. A first simple but myopic approach is to observe only the first link of each route. In this case the number of choice options is limited to the number of outgoing links, and the proportions and split vector are determined with a multinomial logit model, based on the pressure of these links.

If complete routes are observed, a path size logit choice model is used to determine the proportions for each route, and these proportions can be converted to a split vector. Now, instead of the pressure of one link, a pressure value that represents the whole route needs to be determined. This can be done by taking the (weighted) average of the link pressures, or by a method that takes outliers specifically into account. The resulting pressure value is subtracted from 1, to make sure that higher values go with higher utility. Pressure becomes 'available capacity'.

Besides pressures based on traffic load expressed in density, also user satisfaction can be incorporated into the method. This would allow road users to use the shortest routes to some extent, and especially in case of low traffic it would benefit the users. Travel time is a good measure for user satisfaction, represented by instantaneous travel time as a practical value.

The overall utility function of a route can be written as: $BP_r = \alpha_1 P_{\text{route},r} + \alpha_2 P_{\text{firstlink},r} + \alpha_3 P_{\text{user},r}$. The three terms represent the (total) route pressure, the pressure of the first link, and the travel time value, each with a corresponding weight factor.

7. How can simulation be used to evaluate the performance of the control algorithms?

The Matlab-based model DSMART has been chosen to macroscopically simulate traffic flows and the effects of the traffic control algorithms. The original model needed to be modified and expanded, to incorporate traffic signal control. The proposed algorithms for traffic signal control and route guidance were added to the Matlab code.

Other modifications were done to add more detail to the traffic flow itself and to generate the desired results. Lastly, scripts based on input variables were used to partially automate the process of experiment.

8. How should the simulation experiment be conducted, in order to get meaningful results to evaluate?

In the experiments the performance of the proposed traffic control algorithms was tested. First the traffic signal control algorithm and its variants are evaluated, then the algorithm and variants for route guidance, each with a suitable network and traffic demand scenario. Thirdly, a network is simulated that combines traffic signal control and route guidance. All simulations are evaluated with network performance indicators, and detailed results can be used to explain differences.

9. Based on simulation results, what are the differences in performance, between various variants in various scenarios, and compared to the prior expectations?

The simulation experiments are divided into three parts.

Case 1 examines traffic signal control. It shows that the effect of back-pressure is as expected. Green time is assigned based on (waiting) traffic at the intersection approaches, and traffic on the outgoing links has a diminishing effect on the green time of the corresponding phases. The throughput, production, of the (small) networks increases in case of back pressure control with time slots, but the variant with fixed cycle time doesn't perform better than the vehicle actuated variant.

Case 2 focuses on route guidance, without the use of signalized traffic control. The simulation of a network with three similar route lengths (case 2A) showed that the back-pressure algorithm technically worked, but the results were not optimal (worse than the standard route choice model). The performance is expected to be better if the pressure were (partly) based on the critical links, instead of on average density. The variant that only takes into account the first link of the route performed relatively well, since the first links contained a large portion of the total congestion. However, it would fail if congestion would be concentrated at links further downstream.

In the simulation of a second network (case 2B) one of the routes of case 2A is extended to twice its length. Here, the standard (probit) route choice model resulted in high delays, because the shortest routes were overly used, and the back-pressure based variants performed still a bit worse, for the same reason as in case 2A. A pressure function that includes (instantaneous) travel time seems to result in a balance between spread of pressure (density) and user satisfaction (travel time), as expressed by low overall delays.

Case 3 integrates traffic signal control and route guidance, and simulates a large number of combinations of the variants, in a network that is a bit more extensive than the previous two cases. The findings in this case mainly confirm those of the first 2 cases. On the traffic signal control part, time slotted back-pressure performs well, just still slightly worse than vehicle actuated control.

The results from the experiment have can be summarized to a few key points:

- For *traffic signal control*, back-pressure control is a good way to generate high throughput at the intersections, while keeping the queues evenly distributed and within boundaries. Back-pressure signal control based on time slots is more effective than the cycle time based version. Some aspects of the algorithm and its practical use require special attention.
- For *route guidance* an optimal use of back-pressure has not yet been found. Two main aspects are the definition of a representative route pressure, and the service rate for routes related to the assignment of (destination specific) traffic to each route.

  Still, the basic algorithm that was formulated works to some extent, and so far the simulated effects can be understood and are up for improvement. It was also found that a 'pressure' function that includes not only the 'fullness' of the route, but also travel time from a user perspective can give good results.

The last research question is answered as part of the overall conclusions and continues in the recommendations section.

8.2 Recommendations

The recommendations put below relate to things that can be improved or further investigated in future research. They are grouped in a few categories.

Traffic control model in general

Recommendations for the traffic control model as a whole:

- The integration of traffic signal control and route guidance can be intensified by combining the two into one algorithm. The increased complexity might be solved by optimization from iteration.
- In order to further enable network management and strategies, the proposed traffic control model could be integrated in a hierarchical control structure, where strategies are translated to 'virtual' pressure modifications.
- Related to that, coordination and prioritizing of certain traffic flows (also public transport) could be incorporated into the model.
- The choice for control strategy can be responsive to the traffic situation. For example, it could be beneficial to use back-pressure control in heavy traffic, and to use another control type in low traffic situations.

Traffic signal control model

Recommendations for the traffic signal control model:

- The used algorithm could be made applicable to complex intersections with advanced approaches and phase modelling, and taking clearance times into account.
- The back-pressure signal control model with time slots could be enhanced. One could use waiting times to add 'virtual' pressures. This way, queues are treated more 'fair', and continuous domination by one phase is less likely.
- The back-pressure signal control model based on cycle times can be improved by optimizing the cycle time for each intersection independently.

- Back-pressure controlled intersections are automatically coordinated to some degree, but it might be good to add explicit coordination, to be enforced by modifying the pressure values.

Route guidance model

- The formulation of a representative route pressure should be further investigated, for example with the suggestions done in this thesis.
- Also the way the 'service rate' is treated, in relation to the assignment of traffic to the routes, should be a point of further research.
- It might be worth to take into account the queues at the location of route guidance. For example, the queues on approaches to a signalized intersection are currently not taken into account, as well as congested lanes before a highway diverge. Route choice starts with selecting a lane to begin the route, the state of this lane should influence route choice as well.
- In large networks there are too many routes to be predefined and treated individually. At the same time, one of the advantages of back-pressure control is that it can be applied in a distributed system. Therefore, it is suggested to investigate distributed types of route guidance, for example by applying the recursive logit route model.
- As with intersections, coordination can be applied to route guidance.
- The influence of compliance to route guidance has been ignored in this thesis but should be a point of further research.

Simulation and testing

Recommendations regarding simulation and testing of the control model:

- Using a more detailed simulation model (both in network detail and traffic flow accuracy) would enable more detailed applications of the control model to be evaluated (e.g. complicated intersections) and better estimation of the traffic performance.
- To further evaluate the value of the proposed control approach, it is recommended to do tests on other networks and demand scenarios. Scenarios should also include incidents, to evaluate the response of the control model in case of a sudden disturbance of the traffic state.
- There are several variants of the control model that require more testing and calibration, in particular the route guidance pressure functions and the parameters (weight factors, logit function parameters) in the model.
- To better estimate the value of this control approach, it could be compared to other advanced types of traffic control, such as model predictive control.

Outlook for practical use

Before the proposed control approach can be implemented in real traffic networks, it needs further development as described in the recommendations of the previous paragraphs. Furthermore, several developments need to take place to provide the necessary technical requirements, in particular the following.

- Necessary steps need to be taken for vehicles to be equipped with devices that support communication between the vehicles and the traffic control system, where vehicles provide data and the traffic control system provides route guidance and traffic signal control.

- Further research is required to enable reliable and fast state estimation (density profiles).

Other applications

The concept of back-pressure control could have potential in traffic control applications outside the scope of intersection control and route guidance. An example is ramp metering, where the throughput and length of the waiting queue can be controlled according to the traffic state on the highway.

Bibliography

Athanasopoulou, E., Bui, L. X., Ji, T., Srikant, R., & Stolyar, A. (2013). Back-pressure-based packet-by-packet adaptive routing in communication networks. *IEEE/ACM Transactions on Networking*, *21*(1), 24–257. doi:10.1109/tnet.2012.2195503

Ben-Akiva, M. & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. International Series in Operations Research & Management Science, (23), 5–33. doi:10.1007/978-1-4615-5203-1_2

Berg, M. van den, Hegyi, A., Schutter, B. de, & Hellendoorn, H. (2007). Integrated traffic control for mixed urban and freeway networks: a model predictive control approach. *European Journal of Transport and Infrastructure Research*, *7*(3), 223–250.

Bui, L., Srikant, R., & Stolyar, A. (2009). Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing. In *Ieee infocom 2009 - the 28th conference on computer communications*, Apr. 19–25, 2009. Rio de Janeiro, Brazil. IEEE. doi:10.1109/infcom.2009.5062262

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996). A modified logit route choice model overcoover path overlappover problems. specification and some calibration result for interurban networks. In J. B. Lesort (Ed.), *Proceedings of the 13th international symposium on transportation and traffic theory.* July 24–26, 1996 (pp. 697–711). Lyon, France.

Connekt. (2013 Oct.). *Better informed on the road: Roadmap 2013 – 2023* [Summary]. Ministery of Infrastructure and the Environment.

Daganzo, C. F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, *28*(4), 269–287. Working paper title: The Cell Transmission Model. Part I: A Simple Dynamic Representation of Highway Traffic. doi:10.1016/0191-2615(94)90002-7

Daganzo, C. F. (1995). The cell transmission model, part ii: network traffic. *Transportation Research Part B*, *29*(2), 79–93. doi:10.1016/0191-2615(94)00022-R

Feldmann, A., Fischer, S., Kammenhuber, N., & Vöcking, B. (2009). Management of variable data streams in networks. *Lecture Notes in Computer Science*, 266–291. doi:10.1007/978-3-642-02094-0_13

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013 Oct.). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70–80. doi:10.1016/j.trb.2013.07.012

Frejinger, E. (2008). *Route choice analysis: models, algorithms and applications* (PhD thesis, École Polytechnique Fédérale de Lausanne). Verkregen van http://biblion.epfl.ch/EPFL/theses/2008/4009/EPFL_TH4009.pdf

Gao, R., Dovrolis, C., & Zegura, E. W. (2006). Avoiding oscillations due to intelligent route control systems. *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*. doi:10.1109/infocom.2006.124

Giaccone, P., Leonardi, E., & Shah, D. (2005). On the maximal throughput of networks with finite buffers and its application to buffered crossbars. In *Infocom 2005. 24th annual joint conference of the ieee computer and communications societies. proceedings ieee*, Mar. 13–17, 2005 (Vol. 2, pp. 971–980). Miami, FL, USA. doi:10.1109/INFCOM.2005.1498326

Gregoire, J., Frazzoli, E., La Fortelle, A. de, & Wongpiromsarn, T. (2013 Sept.). Capacity-aware back-pressure traffic signal control. eprint: 1309.6484. Verkregen van http://arxiv.org/abs/1309.6484

Gregoire, J., Frazzoli, E., La Fortelle, A. de, & Wongpiromsarn, T. (2014 Jan.). Back-pressure traffic signal control with unknown routing rates. eprint: 1401.3357

Gupta, P. & Javidi, T. (2007). Towards throughput and delay optimal routing for wireless ad-hoc networks. In *2007 conference record of the forty-first asilomar conference on signals, systems and computers*, Nov. 4–7, 2007. Pacific Grove, CA. IEEE. doi:10.1109/acssc.2007.4487207

Hegyi, A. (2004). *Model predictive control for integrating traffic control measures* (PhD Thesis, TRAIL Research School, Delft, Netherlands).

Henn, V. (2005 Oct.). A wave-based resolution scheme for the hydrodynamic lwr traffic flow model. In S. P. Hoogendoorn, S. Luding, P. H. L. Bovy, M. Schreckenberg, & D. E. Wolf (Eds.), *Traffic and granular flow '03* (pp. 105–124). Delft, Netherlands: Springer Berlin Heidelberg. doi:10.1007/3-540-28091-X_10

Hoogendoorn, S. P., Landman, R., & Kooten, J. van. (2013). Integrated network management amsterdam: control approach and test results. In *16th international ieee conference on intelligent transportation systems (itsc 2013)*, Oct. 6–9, 2013. The Hague, The Netherlands. IEEE. doi:10.1109/itsc.2013.6728276

Hoogendoorn, S. P., Taale, H., Wilmink, I., Katwijk, R. van, Immers, B., & Schuurman, H. (2012). *The future of traffic management: State of the art, current trends and perspectives for the future*. TrafficQuest.

Katwijk, R. van & Taale, H. (2012, 29 Aug.). *Coördinatie van maatregelen*. TrafficQuest.

Lämmer, S. & Helbing, D. (2008). Self-control of traffic lights and vehicle flows in urban road networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(4), P04019. doi:10.1088/1742-5468/2008/04/P04019

Landman, R., Hegyi, A., & Hoogendoorn, S. (2012, 31 Dec.). Network wide service level-oriented route guidance in road traffic networks: traffic management in line with the policy objectives of the road authorities. In *Trail beta-congress: mobility and logistics – science meets practice*, Oct. 30–31, 2012. Rotterdam, The Netherlands. Trail Research School. Verkregen 28 Feb. 2014, van http://repository.tudelft.nl/view/ir/uuid:b902773c-50ad-4798-abe5-f29e9af7f3d2/

Le, T., Kovacs, P., Walton, N., Vu, H. L., Andrew, L. L., & Hoogendoorn, S. P. (2013, 3 Oct.). Decentralized signal control for urban road networks. eprint: 1310.0491. Verkregen 7 Mar. 2014, van http://arxiv.org/abs/1310.0491

Lighthill, M. J. & Whitham, G. B. (1955). On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of The Royal Society A*, *229*, 317–345. doi:10.1098/rspa.1955.0089

Lin, S. (2011). *Efficient model predictive control for large-scale urban traffic networks* (PhD thesis, Delft University of Technology). TRAIL Thesis Series T2011/3, the Netherlands TRAIL Research School.

Naghshvar, M. & Javidi, T. (2010). Opportunistic routing with congestion diversity and tunable overhead. In *2010 4th international symposium on communications, control and signal processing (isccsp)*, Mar. 3–5, 2010 (pp. 1–6). Limassol. IEEE. doi:10.1109/isccsp.2010.5463431

Neely, M. J., Modiano, E., & Rohrs, C. (2005 Jan.). Dynamic power allocation and routing for time-varying wireless networks. *IEEE Journal on Selected Areas in Communications*, *23*(1), 89–103. doi:10.1109/JSAC.2004.837349

Neely, M. J. & Urgaonkar, R. (2008). Opportunism, backpressure, and stochastic optimization with the wireless broadcast advantage. In *2008 42nd asilomar conference on signals, systems and computers*, Oct. 26–29, 2008 (pp. 2152–2158). Pacific Grove, CA. IEEE. doi:10.1109/acssc.2008.5074815

Neely, M. J. & Urgaonkar, R. (2009 July). Optimal backpressure routing for wireless networks with multi-receiver diversity. *Ad Hoc Networks*, *7*(5), 862–881. doi:10.1016/j.adhoc.2008.07.009

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., & Wang, Y. (2003). Review of road traffic control strategies. *Proceedings of The IEEE*, *9*(12), 2043–2067. doi:10.1109/JPROC.2003.819610

Richards, P. I. (1956 Feb.). Shock waves on the highway. *Operations Research*, *4*(1), 42–51. doi:10.1287/opre.4.1.42

Schreiter, T. (2013). *Vehicle-class specific control of freeway traffic* (PhD thesis, Delft University of Technology).

Smith, M. (1980). Local traffic control policy which automatically maximises the oveover travel capacity of an urban road network. *Traffic Engineering and Control*, *21*(6), 298–302.

Taale, H. (2008, 5 Nov.). *Integrated anticipatory control of road networks: A game-theoretical approach* (PhD thesis, Technische Universiteit Delft). TRAIL Thesis Series nr. T2008/15.

Tassiulas, L. (1995). Adaptive back-pressure congestion control based on local information. *IEEE Transactions on Automatic Control*, *40*(2), 236–250. doi:10.1109/9.341781

Tassiulas, L. & Ephremides, A. (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, *37*(12), 1936–1948. doi:10.1109/9.182479

Varaiya, P. (2009). A universal feedback control policy for arbitrary networks of signalized intersections. Topl Group. doi:10.1.1.308.5717

Varaiya, P. (2013a Nov.). Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, *36*, 177–195. doi:10.1016/j.trc.2013.08.014

Varaiya, P. (2013b). The max-pressure controller for arbitrary networks of signalized intersections. In S. V. Ukkusuri & K. Ozbay (Eds.), *Advances in dynamic network modeling in complex transportation systems* (Chap. 2, 2, pp. 27–66). Complex Networks and Dynamic Systems. Springer, New York. doi:10.1007/978-1-4614-6243-9_2

Wageningen-Kessels, F. van. (2013). *Multi-class continuum traffic flow models: analysis and simulation methods* (PhD thesis, Delft University of Technology, TRAIL).

Wahle, J., Bazzan, A. L. C., Klügl, F., & Schreckenberg, M. (2000 Dec.). Decision dynamics in a traffic scenario. *Physica A: Statistical Mechanics and its Applications*, *287*(3–4), 669–681. doi:10.1016/s0378-4371(00)00510-0

Wang, Y., Papageorgiou, M., & Messmer, A. (2003 Jan.). Predictive feedback routing control strategy for freeway network traffic. *Transportation Research Record*, *1856*, 62–73. doi:10.3141/1856-07

Wongpiromsarn, T., Uthaicharoenpong, T., Wang, Y., Frazzoli, E., & Wang, D. (2012). Distributed traffic signal control for maximum network throughput. In *2012 15th international ieee conference on intelligent transportation systems*, Sept. 16–19, 2012 (pp. 588–595). Anchorage, AK. IEEE. doi:10.1109/itsc.2012.6338817

Ying, L., Shakkottai, S., Reddy, A., & Liu, S. (2011). On combining shortest-path and back-pressure routing over multihop wireless networks. *IEEE/ACM Transactions on Networking*, *19*(3), 841–854. doi:10.1109/tnet.2010.2094204

Ying, L., Srikant, R., Towsley, D., & Liu, S. (2011). Cluster-based back-pressure routing algorithm. *IEEE/ACM Transactions on Networking*, *19*(6), 1773–1786. doi:10.1109/tnet.2011.2141682

Yperman, I. (2007). *The link transmission model for dynamic network loading* (PhD thesis, Katholieke Universiteit Leuven).

Zhang, R. (2012). Traffic routing guidance algorithm based on backpressure with a trade-off between user satisfaction and traffic load. In *2012 ieee vehicular technology conference (vtc fall)*, Sept. 3–6, 2012 (pp. 1–5). Quebec City, QC. IEEE. doi:10.1109/VTCFall.2012.6399177

Zuurbier, F. S. (2005 May). *A generic approach to a route guidance strategy* (MSc thesis, TU Delft).

Zuurbier, F. S. (2010). *Intelligent route guidance* (PhD thesis, Technische Universiteit Delft). TRAIL Thesis Series nr. T2010/8.