# NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Thesis submitted in fulfilment of the requirements of the
PhD degree, Hamilton Institute, National University of
Ireland, Maynooth

# New Perspectives on Modelling and Control for Next Generation Intelligent Transport Systems

Arieh Christopher Schlote

March 2014

Advisor : Professor Robert Shorten

Department Head: Professor Douglas Leith

Examiners : Professor Harvey J. Miller
Professor Seán McLoone

Für Corinne und Johanna.

# Contents

# Abstract

*This PhD thesis contains 3 major application areas all within an Intelligent Transportation System context.*

*The first problem we discuss considers models that make beneficial use of the large amounts of data generated in the context of traffic systems. We use a Markov chain model to do this, where important data can be taken into account in an aggregate form. The Markovian model is simple and allows for fast computation, even on low end computers, while at the same time allowing meaningful insight into a variety of traffic system related issues. This allows us to both model and enable the control of aggregate, macroscopic features of traffic networks. We then discuss three application areas for this model: the modelling of congestion, emissions, and the dissipation of energy in electric vehicles.*

*The second problem we discuss is the control of pollution emissions in fleets of hybrid vehicles. We consider parallel hybrids that have two power units, an internal combustion engine and an electric motor. We propose a scheme in which we can influence the mix of the two engines in each car based on simple broadcast signals from a central infrastructure. The infrastructure monitors pollution levels and can thus make the vehicles react to its changes. This leads to a context aware system that can be used to avoid pollution peaks, yet does not restrict drivers unnecessarily. In this context we also discuss technical constraints that have to be taken into account in the design of traffic control algorithms that are of a microscopic nature, i.e. they affect the operation of individual vehicles. We also investigate ideas on decentralised trading of emissions. The goal here is to allocate the rights to pollute fairly among the fleet's vehicles.*

*Lastly we discuss the usage of decentralised stochastic assignment strategies in traffic applications. Systems are considered in which reservation schemes can not reliably be provided or enforced and there is a significant delay between decisions and their effect. In particular, our approach facilitates taking into account the feedback induced into traffic systems by providing forecasts to large groups of users. This feedback can invalidate the predictions if not modelled carefully. At the same time our proposed strategies are simple rules that are easy to follow, easy to accept, and significantly improve the performance of the systems under study. We apply this approach to three application areas, the assignment of electric vehicles to charging stations, the assignment of vehicles to parking facilities, and the assignment of customers to bike sharing stations.*

*All discussed approaches are analysed using mathematical tools and validated through extensive simulations.*

# Preface

## Acknowledgements

This thesis is the culmination of 4 years of being a PhD student. Looking back, I notice that I have many fond memories. Many people I met along the way have contributed to this thesis in many different ways.

I spent most of this time at the Hamilton Institute at National University of Ireland, Maynooth, which is a truly great place to work. I also had the opportunity to spend 10 months at Technische Universität Berlin and Fraunhofer-Institut für Offene Kommunikationssysteme.

First and foremost, I would like to thank my supervisor Professor Robert Shorten. Bob has guided me - scientifically and otherwise - all the way and opened many doors for me. Between urgent work calls late on a Sunday evening, 40 cent coffee breaks, work meetings in the deepest snow - deep for Ireland anyway, Christmas lunches, and crazy road trips halfway through Europe, I feel that we have become friends.

In much of my work I have worked together with Professor Emanuele Crisostomi. I have very much enjoyed collaborating with him in person and via heated video chats. He is to some degree my second supervisor and a very dear friend.

My other co-authors and collaborators should also be mentioned. It has been a pleasure and an honour to get to work with all of you. Thanks to Professor Fabian Wirth, Professor Avi Berman, Professor Christopher King, Łukasz Buszisz, Rade Stanojević, Professor Stephen Kirkland, Florian Häusler, Mahsa Faizrahnemoon, Thomas Hecker, Astrid Bergmann, Ilja Radusch, Bei Chen, and Mathieu Sinn.

What I liked in particular about the Hamilton Institute is that you can always find someone to help you with a problem or point you in the right direction. In particular, I would like to mention Professor Stephen Kirkland, David Malone, Oliver Mason, Professor Douglas Leith and Kenneth Duffy. Many thanks also to the administrative staff. Rosemary, Kate, and Ruth have made my life much easier on many occasions.

I thank Professor Jörg Raisch for hosting me along with my supervisor for a long-term visit at TU Berlin in 2011/2012. I also thank Ilja Radusch who hosted me for 3 months at Fraunhofer FOKUS during the same period.

Life consists not only of work but also of fun. There are many people that have made the time in Ireland that much more enjoyable; in particular Andres, Fernando, Alessandro, Karl and Mark. Many thanks to everyone who ever joined a lunch break, a coffee break, an International Lunch, a golf outing, a dinner in the Roost, a city trip, a long walk, a movie night, or a board game evening. You are too numerous to be mentioned here, but you all made life in Ireland far more enjoyable.

Special thanks also go to Mark, Karl, Jule, Markus, and Moritz, who helped me with spell checking and made a number of very useful suggestions to improve readability of the non-technical chapters, and to Bob who gave invaluable feedback to all chapters.

Keeping in contact over a long distance is not easy and gets harder as time goes on. I am grateful for the few very good friends and relatives that manage to keep this precious contact.

I would also like to thank my family. Stefanie, Inge, Klaus, Doris, and Manfred. You all have supported me in the last years and decades in numerous ways. I am deeply grateful to each and every one of you. Without you, this thesis would have never been written.

Finally, I would like to thank the two most important people in the world. Corinne, you have stood by me through these many years of long distance relationship and uncountable hours of waiting for me and travelling to see me. You mean the world to me and I am incredibly glad that I found you. Johanna, you bring so much joy and happiness into my life. Every day with you is something new and a little adventure. I hope that we get to see each other more in future years than in the last year.

## Publications

The following publications report on the work I have done during the course of my PhD.

1. "Markov Chain Based Emissions Models: A Precursor for Green Control", published in Green IT: Technologies and Applications, Springer Verlag, 2011.

   Joint work with Emanuele Crisostomi, Stephen Kirkland, and Robert Shorten.

2. "Traffic modelling framework for electric vehicles", published in the International Journal of Control, vol. 85, no. 7, pp. 880-897, 2012.

   Joint work with Emanuele Crisostomi, Stephen Kirkland, and Robert Shorten.

3. "Cooperative regulation and trading of emissions using plug-in hybrid vehicles", published in IEEE Transactions on Intelligent Transportation Systems, vol.14, no. 4, pp. 1572-1585, 2013.

   Joint work with Florian Häusler, Thomas Hecker, Astrid Bergmann, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten.

4. "Stochastic park-and-charge balancing for fully electric and plug-in hybrid vehicles", to appear in IEEE Transactions on Intelligent Transportation Systems.

   Joint work with Florian Häusler, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten.

5. "Stochastic algorithms for parking space assignment", accepted for publication in IEEE Transactions on Intelligent Transportation Systems.

   Joint work with Christopher King, Emanuele Crisostomi, and Robert Shorten.

6. "On closed loop bike availability prediction", submitted to IEEE Transactions on Intelligent Transportation Systems, 2014.

   Joint work with Bei Chen, and Robert Shorten.

7. "A Big-Data Model for Multi-Modal Public Transportation Optimisation", submitted to the IEEE Transactions on Intelligent Transportation Systems, 2014.

   Joint work with Mahsa Faizrahnemoon, Lorenzo Maggi, Emanuele Crisostomi, and Robert Shorten.

The fourth item has been very well received. We have the honour to have this work mentioned in The Economist [The Economist, 2012].

Complementary videos were produced to showcase some important aspects of two of the above publications. One video was produced for Publication 3, it can be found online at `http://www.hamilton.ie/aschlote/twinLIN.mov`. Another video was produced for Publication 5; it can also be found online at `http://www.hamilton.ie/aschlote/sumo_movie.mov`.

The above work has further been presented at a number of conferences.

8. "A Google-like model of road network dynamics", presented in the 15th Yale Workshop on Adaptive and Learning Systems, Yale University, New Haven, 2011.

   Joint work with Emanuele Crisostomi, Stephen Kirkland, and Robert Shorten.

9. "Cooperative regulation of emissions using plug-in hybrid vehicles", published in the proceedings of the 1st International Conference on Connected Vehicles and Expo, pp. 201-202, Beijing, China, 2012. [Best Short Paper - Runner-Up]

   Joint work with Florian Häusler, Thomas Hecker, Astrid Bergmann, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten.

10. "Balanced routing for EVs", published in the proceedings of the 1st International Conference on Connected Vehicles and Expo, pp. 343-344, Beijing, China, 2012.

    Joint work with Emanuele Crisostomi, and Robert Shorten.

11. "Stochastically balanced parking and charging for fully electric and plug-in hybrid vehicles", published in the proceedings of the 1st International Conference on Connected Vehicles and Expo, pp. 341-342, Beijing, China, 2012.

    Joint work with Florian Häusler, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten.

12. "Stochastic algorithms for parking space assignment", published in the proceedings of the 9th ITS European Congress, Dublin, Ireland, 2013.

    Joint work with Christopher King, Emanuele Crisostomi, and Robert Shorten.

13. "A framework for optimal real-time emissions trading in large-scale vehicle fleets", published in the proceedings of the 9th ITS European Congress, Dublin, Ireland, 2013. [Best Paper Award]

    Joint work with Florian Häusler, Mahsa Faizrahnemoon, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten.

14. "A Google-like Model for Public Transport", presented at the 2nd International Conference on Connected Vehicles and Expo, Las Vegas, USA, 2013.

    Joint work with Mahsa Faizrahnemoon, Emanuele Crisostomi, and Robert Shorten.

15. "The effect of feedback in the assignment problem in shared bicycle systems", presented at the 2nd International Conference on Connected Vehicles and Expo, Las Vegas, USA, 2013.

    Joint work with Bei Chen, Mathieu Sinn, and Robert Shorten.

The following further publications also report on my work during the course of my PhD. However, the research reported in them is in the area of communication networks and they are thus not part of this thesis. The relevant contributions to publication 16 can be found in Appendix A.

16. "On the fair coexistence of loss-and delay-based TCP," published in IEEE/ACM Transactions on Networking, Vol. 19, No 6, pp. 1811-1824, 2011.

    Joint work with Łukasz Budzisz, Rade Stanojević, Fred Baker, and Robert Shorten.

17. "On the higher moments of TCP," published in Linear Algebra and its Applications, Vol. 439, No. 4, pp. 899 - 913, 2013.

    Joint work with Fabian Wirth, Abraham Berman, and Robert Shorten.

# 1. Introduction

***Abstract :*** *This chapter introduces the reader to some basic issues in traffic systems and sets the context for the work reported in this thesis. It also reviews the relevant literature.*

## 1.1. The Effects of the Automobile

In 1886, the German engineer Karl Benz was granted a patent for the first combustion driven vehicle. It was celebrated by many as a great environmental advance as it seemed to be the solution to urban pollution. At the time, with most urban transport being carried out using horses, pollution in the form of horse manure was a major issue and health hazard [Morris, 2007]. Today, cars are everywhere. Delivery trucks play a huge role in supply chains and enable easy door-to-door transport of goods. Cars have enabled vital services to be delivered in very little time. For example, the ambulance services in Germany share the common goal of reaching any patient within 10 minutes of an emergency call. Motorised vehicles can not only be rented as taxis for transport at our leisure, but most notably they are themselves affordable for many people and have given us great personal freedom. Cars allow us to go wherever we want and whenever we want to. However, the number of cars, especially in cities, has grown so large that we have long since started to feel their negative effect. The most noticeable aspect of this certainly is congestion. Queueing in cars and waiting for traffic jams to clear up has become a natural daily routine for people living or working in a city.

However, congestion is not the only negative aspect of high traffic densities in cities. In fact, five major issues can be identified, each of which adversely affects human health and well-being.

1. Congestion is not only unpleasant to experience, but it is also a major inhibitor of economic growth. [Dirks et al., 2010] reports that in the USA congestion causes 4.2 billion hours of wasted time annually and incurs a cost of 87 billion US Dollars, when wasted fuel and lost productivity are taken into account. It is further reported that Dublin holds a record in economic loss, with 4.1% of GDP lost to congestion related causes.

2. Car-related safety issues, both for occupants and other road users, are also a major

issue. Car accidents still account for a disproportionately high percentage of injury-related mortality figures [Peden et al., 2002].

3. The contribution of transportation to greenhouse gas emissions is not sustainable at its current levels [European Environment Agency, 2008]. In the European Union (EU) transportation accounts for 20% of greenhouse gas emissions.

4. Perhaps the most compelling reason to reevaluate our concept of personal mobility is the fact that the internal combustion engine (ICE) is extremely damaging to human health [Friends of the Earth, 1999]. By-products of the combustion process include: $CO$, $NO_x$, $SO$, $Ozone$, $Benzene$, $PM10$, and $PM25$. All of these adversely affect human health and are linked to lung disease, heart disease, and certain forms of cancer [Friends of the Earth, 1999]. In a recent study conducted in the US [Cone, 2008] it is claimed that problems with air-quality lead to three times as many deaths as car accidents. Public discourse focuses mainly on greenhouse gas emissions and on vehicle safety. The general public appears to fail to recognise the damaging effects of the ICE on our health.

5. A special case of pollution is noise pollution. In cities, traffic is a significant contributor to environmental noise. According to [Theakston, 2011] environmental noise can be identified as increasing the risk factors of a variety of ailments, including cardiovascular disease and sleep disturbance, and may be linked to tinnitus and disruptions of child development, leading to cognitive impairment.

Governments and municipal authorities have already started to respond to the air-quality issue. Car manufacturers are under constant legislative pressure to produce ever-cleaner vehicles. Furthermore, cities in some countries ban certain vehicles from densely populated areas ( Umweltzonen) [Federal Republic of Germany, 2014], and sometimes speed limits are adapted to respond to pollution peaks [City of Brussels, 2012]. In particular, the concept of the Umweltzonen is widespread throughout Germany. Limits on particulate matter and other pollutants have been in effect in Germany and the EU for some time. For example, in the EU, exposure to a yearly average of 40 micrograms per cubic metre ($\mu$g/m$^3$) and a daily average of 50 $\mu$g/m$^3$ have been set for particulate matter smaller than 10 micrometres (PM10). In Germany, Umweltzonen, or green zones, have been introduced to ensure that these and other limits are met. Vehicles are issued with permits according to their emission class. And access to low emission zones is restricted to vehicles fitted with the proper permit. Going further in this direction, even stricter measures are being planned; most notably, the ban of ICE driven vehicles from our cities in the near future [BBC, 2011]. These measures are highly invasive and adversely affect many drivers. Also, they may not always be necessary. Whether a polluting vehicle should be allowed into a city should depend on the context, in this case its effect on the air pollution levels. On the other hand these measures do not go far enough and can potentially cause a negative feedback. Being awarded access to a green zone gives drivers a feeling of being very environmentally friendly. This may encourage additional

trips leading to an increase in the number of journeys and thus pollution.

An intuitive approach to solve the congestion problem is to enable the road network to allow higher throughput of cars by building new roads and adding lanes to existing roads. However, increasing the capacities in existing traffic infrastructure is usually extremely difficult and does not actually solve the problem. Using the words of the American philosopher Lewis Mumford: "Adding highway lanes to deal with traffic congestion is like loosening your belt to cure obesity." In fact, adding new roads to an existing road network can, under some circumstances, actually lead to an increase in congestion, see for example the famous Braess' paradox [Braess, 1968; Braess et al., 2005]. Hence, there is a high demand for intelligent solutions.

A growing number of cities are currently engaging in attempts to encourage travellers to use bicycles instead of personal vehicles and public transport. Part of this includes a massive roll out of bike sharing schemes all over the world [Shaheen et al., 2010; Midgley, 2011; Pucher and Buehler, 2012; Fishman et al., 2013], see for example *Vélib'* in Paris, *Bicing* in Barcelona, *Capital Bikeshare* in Washington, D.C. and *Forever Bicycle* in Shanghai. A good overview can be gleaned from [DeMaio, 2009] and the references therein. They are not only seen as an integral component in public transportation networks for alleviating congestion and curbing health problems related to low levels of physical activity [Midgley, 2009; Murphy and Usher, 2013], but also as a fundamental tool for combating urban pollution [Pucher et al., 2011]. They are also being suggested as a step towards solving the last mile problem in intelligent transportation systems [Mackechnie, 2013; Pucher and Buehler, 2008]. The last mile problem is the challenge to close the gap between a traveller's origin or destination and public transport access points. *Dublinbikes* is an example of a small, but fast growing bike sharing scheme. Since its launch in 2009, the Dublin bikes have made over 6 million trips and are currently expanding to 102 stations and 1500 bikes from originally 44 stations and 550 bikes [JCDecaux, 2013].

A further, very prominent and noticeable issue in many cities is parking. Finding a parking space in a densely populated area is a non-trivial challenge. The lack of instantaneous parking causes significant damage, both economically and environmentally. People driving around, looking for free or cheap parking waste not only their own time, but also consume road capacity, burn fuel, and produce toxic emissions, thus contributing significantly to the above mentioned problems, congestion, greenhouse gas emissions and pollution. For example, in a small Los Angeles business district, cars cruising for parking spaces burned 47,000 gallons of gasoline and produced 730 tons of carbon dioxide in the course of one year [Shoup, 2007]. It was also recently claimed that the average car owner in Paris spends four years of his life searching for parking spaces [Elfrink, 2012].

The solutions to the above mentioned problems have to be context aware and collaborative. Often in this context the word "smart" is used. Within the context of smart-cities research [IBM, 2007; Hodgkinson, 2011], collaborative mobility is rapidly becoming one

of the main growth areas in applied and theoretical research. The topic is generating interest in a wide range of areas including computer science, electrical engineering, applied mathematics, operations research, networking, as well as transportation engineering. Many important things happened in the last years. Much progress has been made in technologies related to communication networks, vehicles, and optimisation. At the same time we have seen a significant spreading of smartphones. Most notably, social attitudes to mobility have changed. Road networks in our cities are often hugely inefficient and services and products related to smart mobility are expected to become a multi-billion dollar industry over the next decade [Ezell, 2010]. Additionally, research into traffic issues is strongly incentivised. Since 2008, the EU has allocated 1 billion euro for the European Green Cars Initiative, which funds research projects with a focus on traffic electrification and safety as well as traffic efficiency. Transport is also one of the main research interests in the EU Framework Programme for Research and Innovation, Horizon2020 [European Commission, 2014], in which a planned 80 billion euro are going to be invested in research projects in the years 2014 to 2020.

## 1.2. New Technology

At this point it should be noted that traffic has been an active topic in research for a long time. Recently, research in this area has made much headway. More importantly, the last decade has seen an important convergence of technology. New vehicle classes are available and have reached mass production state. This includes the electric vehicle (EV), hybrid and plug-in hybrid electric vehicles (HEV and PHEV) as well as electric and hybrid buses and cargo delivery trucks. These vehicles carry a large number of sensors, directed both at the vehicle itself and at the outside world. Another great advance is taking place in the area of communication. The Global Positioning System (GPS) enables vehicles to determine where they are with high accuracy. Cellular networks enable vehicles to communicate with infrastructure elements and other vehicles. It is an ongoing effort to enable vehicles to directly talk to all vehicles and infrastructure nearby. In addition, modern vehicles are equipped with computers that have significant processing power. Often these technological advances are jointly referred to as Information and Communications Technology (ICT), or in this context as the connected vehicle. All this combined, puts us in a situation, where we have huge amounts of relevant data at our disposal that can be collected and communicated in close to real time [Biem et al., 2010]. This allows traffic researchers a completely new degree of freedom to use data driven models and control strategies and glean better insight into the complex dynamics of traffic networks.

These technological advances enable a number of new approaches to deal with traffic network inefficiencies. With more accurate knowledge of the state of the traffic network, and each actor in it, it is possible to make concise decisions aimed at optimising the network performance. Further, vehicles can gather information and find solutions in a

decentralised manner collaboratively with all nearby vehicles and supported by roadside infrastructure. These new possibilities also raise several new research questions. New models are needed which can handle the huge amounts of data generated by Intelligent Transportation Systems (ITS). At the same time they need to take the specific characteristics of the new vehicle types into account. For example, being stuck in congestion is highly inefficient for ICE driven vehicles, but not so for EVs. On the other hand, power consumption in EVs depends to a significant degree on the use of auxiliary systems such as heating, while this is not the case for conventional vehicles. Driving uphill consumes much energy in EVs, but some energy can usually be recovered when going downhill or decelerating. To be useful, these models have to be both descriptive and enable control strategies. We also need collaborative control algorithms that make use of the new technology and respect the specific constraints of ITS applications. These constraints will be discussed in detail in Chapter 5.

## 1.3. Related and Recent Work

At this point, many PhD theses in traffic give an extensive review of traffic models, see for example [Zegeye, 2011; Chiguma, 2007]. Exhaustive reviews of traffic models can be found in many places, for example [Hartenstein and Laberteaux, 2010; Hoogendoorn and Bovy, 2001]. We do not repeat this exercise here. Instead we want to focus on the recent advances enabled by the new technologies discussed above.

It should, however, be mentioned that in the past it was almost impossible to obtain real time data about the state of traffic networks. Accordingly, traffic models necessarily had to be very complex in the hope of achieving a high degree of accuracy. In particular, the effect of a policy change on a traffic network is very hard to model. This significantly reduces the predictive power and hence usefulness of models, see [Chatterjee, 1999]. A good overview of traffic models can be found in [Hartenstein and Laberteaux, 2010; Hoogendoorn and Bovy, 2001] and the references therein. [Hartenstein and Laberteaux, 2010] investigates traffic models with vehicular communication networks in mind, but it also includes a comprehensive general discussion. We base the following brief discussion on it.

Traffic models can be classified into three different classes regarding their resolution. Microscopic traffic models explicitly consider the operation of individual vehicles, taking into account things such as maximum speed, acceleration abilities, driver behaviour and car following models, see for example [Krauß, 1998]. These models are very accurate, yet come with a very high computational cost that renders these models unpractical for very large networks or on-board computation. At the other end of the spectrum are macroscopic models that focus on aggregate quantities, such as average speed, traffic densities and so on. These models in general have good scaling behaviour but much information about the precise system dynamics is lost. The third category, mesoscopic models aims at bridging the two other model classes by modelling some aspects in an ag-

gregate macroscopic manner and other with microscopic precision. Depending on which approach is used, a number of technical tools are available for the models including stochastic models such as the constant speed model [Fiore and Härri, 2008] and Markov chain models of intersections [Alvarez et al., 2008], ordinary and partial differential equations, see [Piccoli and Garavello, 2006] for a microscopic approach and [Richards, 1956; Lighthil and Whitham, 1955] for a macroscopic variant, and queueing models for continuous time and continuous space models [Gavron, 1998], difference equations for discrete time and continuous space models [Krauß, 1998], and for discrete time and discrete space models cellular automata [Nagel and Schreckenberg, 1992].

All the above models are either extremely complicated and computationally complex or yield limited insight into aspects of interest. The above described advent and convergence of vehicular technologies, the connected vehicle, enables us to use much simpler data driven models. These models can be recalculated at a high frequency using the latest data measurements.

The advent of the connected vehicle has led to a surge of interest in the traffic area, both in academia and industry. Major industrial initiatives in this area include [IBM, 2007; Hodgkinson, 2011; Elfrink, 2012]. This is further supported by the growing public and governmental awareness [European Commission, 2011] of the negative impact traffic has on our quality of life, especially in urban areas.

Many topics are investigated in academia and industry that make use of the connected vehicle. Many research projects aim at producing individual driver profiles and individual guidance for each driver. The goal is to learn what the individual driver's behaviour and preferences are and to suggest optimal personalised options. Projects in this area include SMART-VEI [SMART-VEI Consortium, 2011], SUPERHUB [SUPERHUB Consortium, 2014], ECONAV [ECONAV Consortium, 2014], ECODRIVER [ECODRIVER Consortium, 2014].

Other projects take a higher level approach and aim at optimising social aspects of transport. The goal is to optimise facets of the traffic network, for example, by encouraging drivers to behave less selfishly. Often this is done in the context of multimodal transport, where individual travellers are guided to use multiple modes of transportation, depending on origin and destination, as well as real time information on each mode of transport, see for example Enhanced WISETRIP [Enhanced WISETRIP Consortium, 2014], TEAM [TEAM Consortium, 2014], Mobility2.0 [Mobility2.0 Consortium, 2014], REDUCTION [REDUCTION Consortium, 2014], see also [Botea et al., 2013].

In the context of inter-modality the bicycle is being promoted as an important and powerful factor. BIKE INTERMODAL [BIKE INTERMODAL Consortium, 2013], for example, aims at developing bikes that can be taken on-board other modes of transport. In this context bike sharing also plays a large role. Although being increasingly rolled out globally these systems suffer from a number of inefficiencies. [Chen et al.,

2013] presents a forecast system to estimate the availability of bicycles at a given station based on historic data. This forecast is intended to improve predictability and reliability for the users. A fundamental problem of bike sharing systems is that their inventory becomes unbalanced. To solve this, relocation trucks are being employed to balance the inventory. Much research is carried out investigating optimal routing strategies for these trucks. [Rainer-Harbach et al., 2013; Schuijbroekm et al., 2013] provide solutions to a stationary rebalancing problem, where the desired and the actual number of available bikes at each station are known and the routes of the trucks are optimised. In an extension to this, [Pfrommer et al., 2013] and [Fricker and Gast, 2013] propose to explicitly take user actions into account to solve the dynamic relocation problem. [Pfrommer et al., 2013] is one of the few publications in this area that consider user incentives in their model. Here users are being paid if they deliver a bike to a particular station. Sharing is a concept of growing popularity and is increasingly considered in ITS not only for bikes. The idea of sharing cars and other vehicles has become common practise in recent years. Car sharing has evolved from car-rental companies and privately organised or geographically very limited sharing of vehicles to a fast growing and successful industry, see for example [car2go Deutschland GmbH, 2014; DriveNow GmbH & Co. KG, 2014]. In these systems users can leave the car anywhere within a set area. The relocation of the vehicles is an on-going research topic, see for example [Weikl and Bogenberger, 2012].

A significant number of research projects concentrate on investigating the new vehicle types. Typical directions in research include dealing with the perceived and real disadvantages of EVs compared to conventional vehicles, see ECOGEM [EcoGem Consortium, 2014], ELVIRE [ELVIRE Consortium, 2014], EMERALD [Softeco Sismat S.r.l, 2014], ICT4EVEU [ICT4EVEU Consortium, 2014], or the connection of EVs and PHEVs to the smart grid, see Mobi.Europe [Mobi.Europe Consortium, 2014], MOBINCITY [MOBINCITY Consortium, 2014]. Recent publications in these areas include [King et al., 2013; de Weerdt et al., 2013] for improving perception of the EVs and [Stüdli et al., 2012b,a; Erol-Kantarci and Mouftah, 2010] for the smart grid integration.

A completely different approach is taken in [Knorn et al., 2011a], where it is proposed to let fleets of PHEVs collaboratively control their aggregate emissions and their allocation in a completely decentralised way. Given large enough proportions of PHEVs, it is possible to control emission levels by influencing the mix of the two motors.

A frequent argument in recent traffic publications is that humans and their tendency to make selfish decisions have a negative impact on transport network performance. Humans make errors, unnecessary detours and often exhibit suboptimal driving behaviour. Hence, another approach is to take the human factor out of the picture and use automation as far as possible, see for example [Göhring et al., 2013] and HAVEit [HAVEit Consortium, 2014]. CARBOTRAF [CARBOTRAF Consortium, 2014] reduces the human factor by active traffic management, where properties of traffic networks, such as speed limits, traffic light sequencing and access rules, can be adapted based on available data.

The growing traffic densities in many cities are problematic not only in terms of congestion but increasingly also in regard to finding parking spaces. As mentioned above, people looking for parking spaces in turn contribute to congestion and pollution. Parking is also an issue for multi-modal transport if it takes personal vehicle use into account. Some commercial tools to guide users to parking spaces are already available, for example ParkNow [BMWi, 2014] and SFpark [SFMTA Municipal Transport Authority, 2014]. ParkatmyHouse [ParkatmyHouse, 2014] (now owned by BMW) is a commercial website that aims at making more parking space available by getting private households to advertise vacancies. [Teodorović and Lucić, 2006; Geng and Cassandras, 2011] propose a central reservation system for parking spots using complicated and computationally expensive online optimisation algorithms. [Anderson and de Palma, 2004; Arnott, 2006; Shoup, 2006] discuss the negative impact of free curbside parking. They claim that it incentivises drivers to cruise for a long time to find a free parking space. Different variants of the usage of ICT in finding parking spaces are numerically investigated in [Kokolaki et al., 2011]; it is indicated that a mix of centralised data usage and decentralised self-organising strategies should be developed, as neither strategy is strictly superior to the other.[Caliskan et al., 2006; Klappenecker et al., 2012] present and investigate a prediction model for parking space availability in a single car park.

Electric vehicles require access to charging and parking. Finding a charging station at their destination or along the route is made difficult by the relatively small number of stations at the present time. Due to the limited range of these vehicles, the marginal cost of expending energy to search for spaces may, in some cities, be prohibitively high. The problem of finding charging stations is addressed, for example, in [Mariyasagayam and Kobayashi, 2013]. In an extension to parking services, commercial tools are also available that help with finding and reserving charging spots, for example ChargeNow [BMWi, 2014]. There is an important converse problem associated to this, namely the finding of good locations for charging stations, given certain routes and demand for electric vehicles. This can also be addressed as a joint optimisation problem, see for example [Worley et al., 2012]. Here this problem is formulated for a fleet of vehicles that needs to rent a minimal subset of all available charging stations in order to fulfil their obligations.

## 1.4. The Future of the Connected Vehicle and Contributions of this Thesis

The availability of detailed and up-to-date data to traffic planners and fleet managers enables many new and promising applications. The development of new accurate and data driven models is the first step. Intelligent transport systems provide huge amounts of data, but using this data well is challenging. Data handling has to scale well with the size of the traffic network and the number of travellers. In particular, it would be attractive if the resulting models can be handled by smartphones, to allow individual travellers to investigate specific aspects of the road network taking into account their

personal user profile and to enable distributed computation.

When it comes to the question of how to use ITS models to change and control aspects of the system, many of the approaches discussed in Section 1.3 follow common arguments. They alternate measurements and control decisions. This way, when unfavourable network states start to emerge, measures can be initiated to counteract them. Such first generation feedback control strategies are an important second step after the model development. Feedback control is known to work well and to be easily implemented. However these models ignore a different kind of feedback, which is very important when dealing with systems that involve human decisions. Namely, they do not explicitly take into account the effect ITS services will have on the travellers and their decisions. The next generation of traffic models necessarily needs to integrate this feedback. They need to be able to reliably predict the consequences of control actions. In this way they will be much more efficient and allow for more targeted actions and faster convergence to the desired operating state.

In practice this often depends on the way in which the available data is used. As we shall discuss in greater detail in Chapter 8, giving users direct access to the available data may in some situations actually lead to destructive feedback phenomena and a decrease in performance. This paradoxical fact can be explained by the greedy behaviour humans often exhibit. It is our goal to present a number of approaches to slightly adjust or guide human behaviour to achieve performance improvements without adversely affecting the travellers themselves. In doing this, we present a number of ways in which feedback can be explicitly taken into account in traffic models. Our approaches span a number of traffic related applications, covering assignment strategies for EV charging, parking and bike sharing, cooperative emission control, and data driven Markov chain models of urban traffic networks with applications to the modelling and control of congestion, pollution and electric energy consumption.

A further challenge for future ITS applications is that cars have to be context aware. They have to be able to gather information about their surrounding environment and adapt their behaviour accordingly. Examples of desired behaviour include being quiet in a noise-sensitive area, reduce emissions as far as possible if the surrounding pollution level is high, and make full speed and acceleration abilities available only if the resulting emissions can safely be added to the local environment. This has to be done in a smart way, where dirty vehicles and less dirty vehicles agree on an allocation of the pollutants that may be emitted that is both fair and respects the physical constraints of the vehicles. In this context it should also be possible to avoid pollution in some parts of a road network, for example close to schools, kindergartens and hospitals.

In this context, this thesis discusses three application areas. In Part I we discuss a recently proposed Markov chain modelling approach for road networks. These models treat a road network as a graph and deal with large amounts of data by aggregating relevant information into transition probabilities. The arising Markov chain is computa-

tionally easy to handle and many of its characteristic quantities give meaningful insight into the road network dynamics. The fast computation makes this chain a powerful tool for control applications as it makes it possible to predict the effect of regulatory changes to the network.

In Part II we consider a collaborative control application, in which the switching between operating modes is influenced in a fleet of hybrid vehicles to control the aggregate emissions. As we shall show, this can be implemented with simple broadcast signals from a central infrastructure, without direct communication between vehicles or from the vehicles to the infrastructure, and with minimal computational requirements. This makes the implementation extremely simple and may also allow a certain degree of backward compatibility with older technology. We further show how the additional use of direct communication between vehicles allows us not only to control the emission level but also to influence the share of emissions allocated to each vehicle. Vehicle-to-vehicle communication enables us to implement notions of fairness into the framework.

In Part III we consider the allocation of cars to car parks. Parking guidance systems provide information to users in order to facilitate smart decisions. We are interested in studying the feedback introduced to this system by people changing their behaviour in response to the provided information. We show how this feedback can lead to oscillations in the systems that decrease the performance. We present stochastic assignment strategies that make better use of the available data and lead to higher efficiency in the system. As we show, the same approach works also for the assignment of EVs to charging stations and the assignment of customers to bike sharing stations.

## 1.5. Organisational Aspects

### 1.5.1. Joint Authorship

The work in Chapters 5 and 7 was conducted jointly with Florian Häusler and Mahsa Faizrahnemoon. My contribution was the initial formulation of the optimisation problem with individual utility functions, which was further developed in a joint effort.

Chapter 6 reports work conducted jointly with Florian Häusler. While much of the initial work and the main ideas were developed together, I concentrated on the development, analysis and validation of the different control strategies and proof-of-concept simulations.

Further, the balancing strategy for electric vehicles over available charging stations reported in Chapters 8 and 9 was conducted jointly with Florian Häusler. In this work my contribution was the system performance analysis both mathematically and using simulations.

### 1.5.2. Validation

The models that we developed and investigated in this thesis are of a macroscopic character, in the sense that they concentrate on specific aspects of traffic that we want to study and control. For validation we thus use a complementary microscopic traffic simulator, namely the Simulator of Urban Mobility [Krajzewicz et al., 2006] (Sumo). Sumo is an open source tool and is licensed under the GNU General Public License. Sumo is microscopic and takes into account many different phenomena such as maximum speed, acceleration abilities, lane changing, traffic lights, right of way rules and car following. The traffic model underlying Sumo is described in detail in [Krauß, 1998]. The main features that we use in Sumo are the following: Road networks can be created by hand or automatically and imported from other sources, such as OpenStreetMaps [OSM Contributors, 2014]. Traffic can be specified by assigning routes to individual vehicles, by choosing turning probabilities at each junction, or by using origins and destinations and shortest path routing.

### 1.5.3. Thesis Structure

In Part I we discuss a Markov chain model of an urban road network, that allows to study a number of interesting properties of the network and further to control some aspects by understanding the feedback mechanisms of potential decisions. In particular, in Chapter 2 we discuss the state of the art of this model. We extend this work in Chapter 3 and remove some of its assumptions. We then present a number of new application areas to the Markov chain model in Chapter 4. In Part II we discuss an emission control strategy for fleets of hybrid vehicles. To this end, in Chapter 5 we discuss typical constraints arising from ITS applications and their implementation. In Chapter 6 we present the control strategy. This control strategy is extended in Chapter 7, where also take into account fairness in the allocation of emissions to individual vehicles in the fleet. In Part III, we study the effect of feedback in systems that provide information to users in order to facilitate making smart decisions. We discuss a set of tools to study such systems in Chapter 8 and a number of applications is given in Chapter 9. Finally we conclude the thesis in Chapter 10, where we also discuss future research directions.

# Part I.

# Macroscopic Large Scale Modelling for Intelligent Transport System Applications

# 2. A Recently Proposed Markov Chain Model for Road Network Dynamics

***Abstract :*** *In this chapter we discuss a recently proposed Markov chain based approach to modelling urban traffic networks published in [Crisostomi et al., 2011b]. The construction of the model and its quantities of interest are discussed in detail as this is the base for the original work presented in Chapters 3 and 4. We also discuss a number of the model's shortcomings that will be alleviated in Chapter 3.*

## 2.1. Introduction

The recent advances in traffic related technologies have lead to huge amounts of data being available in close to real time. This enables new ways of modelling traffic using data driven approaches. This also constitutes a challenge as new classes of models need to be developed that can handle the potentially huge amount of data. In order to avail of the advantages of data driven models, these models need to allow fast recomputation of significant quantities. For decentralised and distributed applications in vehicles or roadside infrastructure, these models possibly need to be able to be handled on relatively small computers.

Among the many sources of data for ITS purposes are the following.

1. An increasing number of vehicles is equipped with devices capable of reliably estimating their exact location using, for example, the Global Positioning System. Location data can also be gathered from a fleet of vehicles, such as buses and other public transport vehicles. For example the 1000 buses in Dublin transmit their GPS locations every 20 seconds [IBM, 2013].

2. Smartphones have reached a significant penetration level in the population. For example [Smith, 2013] reports that 56% of all adults in the US have a smartphone according to a 2013 poll. Smartphones can gather and communicate information about traffic flows and much more as recent revelations about secret government activities show.

3. Many cities use induction loops and traffic cameras to count vehicles passing important points in the network. For example, in Dublin 800 road junctions are being observed [Dublin City Council, 2011].

4. A lot of the data volume is caused by the scale of urban networks. Beijing for example is expected to have 6 million cars registered by 2015 [Aizhu, 2013].

Handling these huge amounts of data is made even harder by the complicated dynamics road networks exhibit. Driver behaviour, junction handling, and nonlinear relationships between traffic densities and average travel speed are only some of factors that need to be taken into account.

A possible solution to the data management problem that we believe to have great potential is in aggregating data and using these aggregates to describe Markov chain models. Markov chain models are a key technology in this context. Characteristics of Markov chains are simple to compute and, if they are set up in the right way, they can provide a high level of insight. The use of Markov chains to model a large-scale and highly complicated dynamic system may at a first glance seem strange and overly optimistic. However, Markov chain models of highly complex dynamic systems have been used extensively in the Mathematics and Engineering communities. To give the reader a flavour of this prior work we now briefly describe two applications.

As a first example, we consider a problem from astrophysics; namely, that of a planar circular restricted three-body problem. This is a special case of the three body problem which has inspired many advances in dynamic systems. Here two of the three bodies are assumed to have significantly more mass than the third and of these two bodies - the lighter one - is orbiting the heavier one on a circular trajectory. The third body is free to move around the state space and it is assumed to be without mass. The movement of all three bodies is assumed to be restricted to a plane. We can imagine this as a planet that has a moon on a circular orbit and the third body is an asteroid whose motion is restricted to the plane in which the moons orbit lies. In [Koon et al., 2000, 2002] the planar circular restricted three-body problem is analysed by partitioning the state space into three regions: an interior and an exterior region with respect to the moon, and a region close to the moon, where the asteroids motion is governed by the moons gravitational force. The authors show that for any sequence of these regions they can find a possible orbit of the asteroid that will move through the regions in the given sequence. This is done by identifying a number of important system equilibria and corresponding invariant manifolds. The existence of homoclinic and heteroclinic orbits is discussed. Sequences of these orbits can be constructed such that solutions close to these sequences visit the regions in the given order. This requires advanced methods from dynamic systems theory. In [Dellnitz et al., 2005] it is shown how the results can be refined using a much simpler approach using Markov chains.

Another example is what is known as the transfer operator approach to modelling dynamic systems. Given a Markov process on some state space and some probability distri-

bution on it, the Frobenius-Perron operator determines the evolution of the probability measure in time. Invariant measures of the system can be identified with eigenfunctions of the Frobenius-Perron operator to the eigenvalue 1. In [Dellnitz and Junge, 1999] the spectrum of the Frobenius-Perron operator is numerically approximated by analysing the spectrum of discrete approximations to the Frobenius-Perron operator. This is successfully applied to model the dynamics of complex bio molecules, see [Huisinga, 2001] and [Schütte and Huisinga, 2003]. The idea here is to cut the state space into a finite number of regions. The importance of these regions for the system dynamics can be assessed by creating a Markov chain with the regions as states and to determine the transition probabilities one uses a transfer operator. This operator does not depend on the choice of the partition. Moreover analysing the operators spectrum and the eigenspaces for a certain range of eigenvalues yields information about almost invariant sets and allows one to find a beneficial way of partitioning the state space. Almost invariant sets are subsets of the state space where solutions stay for a long time and can often be identified with important properties of the modelled system. Insight can be gained by using a Markov chain that models these almost invariant sets and the transition rates between them.

A very good and compact introduction to modelling complex dynamic systems as Markov chains can be found in [Froyland, 2001]. Much information about ergodic systems is contained in the distribution of a typical long trajectory, called the systems natural measure. It is shown how this natural measure can be approximated using Markov chains.

In our specific case, traffic in an urban network is a highly complicated dynamical system that can, more often than not, be accurately modelled only by using equally complicated models. However, very often we are interested in aggregation effects (pollution, congestion, travel times etc.). We believe that Markov chain models allow us to compute important properties of traffic networks network with high accuracy and very high speed.

In this chapter we recapitulate the Markov chain based approach to modelling urban traffic networks proposed in [Crisostomi et al., 2011b]. We discuss its abilities and defects to set up the original work presented in Chapters 3 and 4. To this end it is helpful to first review the definition and needed properties of time homogeneous Markov chains on a finite state space and their connection to directed graphs. The definitions are standard and can be found in classic references like [Meyn and Tweedie, 2009] or, in a short summary, given in [Langville and Meyer, 2006, Chapter 15]. More specific references are given where necessary. Then we outline in detail the described modelling approach, which we will use as a basis for the presentation of our own results and we briefly cover the applications discussed in [Crisostomi et al., 2011b].

This chapter is organised as follows. In Section 2.1.1 some basic definitions considering Markov chains are given and related quantities and their computation is discussed. In Sections 2.2 and 2.3 the proposed model and its construction is described in detail and applications that were suggest in [Crisostomi et al., 2011b] are discussed in Section 2.4.

### 2.1.1. Basic Definitions

A Markov chain is a discrete time stochastic process, in which the transition probabilities depend only on the state of the chain at the previous time step and not on the past history of the process. In a time homogeneous Markov chain these probabilities are further independent of time. Let us consider a Markov chain on $n \in \mathbb{N}$ states. For each pair of states, $i, j = 1, \ldots, n$, we denote by $p_{ij}$ the probability of going from $i$ to $j$ in one time step. As we consider only finite state Markov chains, the matrix $P \in [0, 1]^{n \times n}$ of elements $p_{ij}$ together with the initial distribution vector fully describe the evolution of the Markov chain for all times.

A graph is a set of states (or nodes) that can be connected by edges (or arcs). We will only consider directed arcs here. Let us again consider a finite set of $n$ nodes, then for each pair $i, j = 1, \ldots, n$ an arc from node $i$ to node $j$ indicates that it is possible to make a direct transition from state $i$ to state $j$. The case $i = j$ is not excluded; we call such an arc a self-loop. It is possible to give a weight to each arc that corresponds to the cost of using that edge. If the aggregate cost of all outgoing arcs of each node is normed to 1 then the costs can be interpreted as the probabilities of using the corresponding arcs to leave the node.

There is a strong link between Markov chains with finite state space and graphs. States of the chain can be associated with nodes in the graph and non-zero probabilities of transition between two states in the chain can be associated with directed edges between the corresponding nodes with the given probability as a weight. Graphs can thus be analysed using methods for Markov chains. An important property of a graph is connectivity. A directed graph is called strongly connected if starting from any node it is possible to reach any other node by following the arcs. Strong connectivity holds if and only if the transition matrix of the corresponding Markov chain is irreducible. Throughout this part of the thesis we shall assume that all considered Markov chain transition matrices are irreducible. This allows us to apply the Perron-Frobenius theorem, see for example [Horn and Johnson, 1990, Theorem 8.8.4], to ensure the existence of an invariant measure $\pi$ with $\pi^\top P = \pi^\top$. In this situation, $\pi$ is entry-wise positive and its entries sum to 1. We call $\pi$ the chain's stationary distribution.

We now introduce some quantities related to Markov chains that will be useful for our study, namely the mean first passage time and the Kemeny constant. These quantities are related to random walks on a graph. A random walk on a graph is a discrete time stochastic process $S(k)$, $k \geq 0$. Its state space is the nodes of the graph. Starting in an initial node, the random walk in every time step follows one of the arcs leading out of the node it is currently in, where each arc is picked randomly according to that arcs probability. If $S(0) = i$, i.e. the walk starts in node $i$ then the first time that the walk reaches node $j$,

$$\min_{k \geq 0} \{k : S(k) = j\}, \tag{2.1}$$

is called the first passage time from $i$ to $j$. Its expected value is called the mean first passage time. If we denote by $m_{ij}$ this mean first passage time from $i$ to $j$, it has to satisfy the following relationship. With probability $P_{ij}$ it takes exactly one step to move from $i$ to $j$ and for $k \neq j$ the first step takes the random walker from node $i$ to node $k$, from which it takes an expected time of $m_{kj}$ to reach node $j$. This can be formally written as

$$m_{ij} = P_{ij} + \sum_{k \neq j} P_{ik}(1 + m_{kj}) \tag{2.2}$$

$$= 1 + \sum_{k \neq j} P_{ik} m_{kj}. \tag{2.3}$$

and we will be using this equation later to compute mean first passage times.

A way to compute the mean first passage times arises from the study of the group inverse of a matrix. A transition matrix $P$ with 1 as a simple eigenvalue gives rise to a singular matrix $T = I - P$ (where the identity matrix $I$ has appropriate dimensions) which is known to have a group inverse $T^{\#}$. The group inverse is the unique matrix such that $TT^{\#} = T^{\#}T$, $TT^{\#}T = T^{\#}$ and $T^{\#}TT^{\#} = T$. More properties of group inverses and their applications to Markov chains can be found in [Meyer, 1975]. The group inverse $T^{\#}$ contains important information on the Markov chain. For example, if we denote $t_{ij}^{\#}$ as the $ij$ entry of the matrix $T^{\#}$, then the mean first passage times can be computed easily according to

$$m_{ij} = \frac{t_{jj}^{\#} - t_{ij}^{\#}}{\pi_j}, i \neq j, \tag{2.4}$$

where it is intended that $m_{ii} = 0$, $i = 1 \ldots, n$. See for example [Choa and Meyer, 2001].

If the random walk starts in node $i$ and randomly picks a destination node $j$ with probability according to the stationary distribution of the Markov chain, $\pi$, we can compute the expected number of steps to reach the destination according to

$$K_i = \sum_{j=1}^{n} \pi_j m_{ij}. \tag{2.5}$$

A well known, yet remarkable, result [Kemeny and Snell, 1960] is that the right hand side of Equation (2.5) is in fact independent of $i$ and thus $K = K_i$, the Kemeny constant, is a global measure for the Markov chain. It can be interpreted as the expected duration of an average trip. As such the Kemeny constant can be viewed as an efficiency indicator for a network. For example, imagine two graphs with the same vertex set and that differ
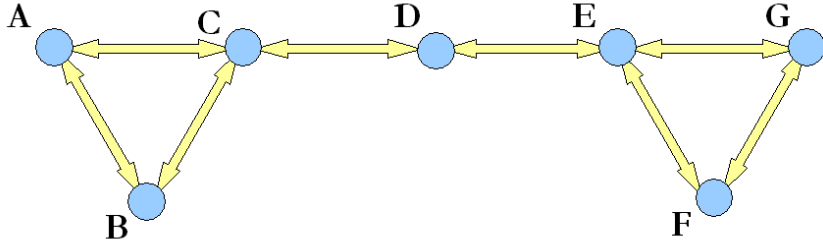
Figure 2.1.: Example of a primal graph of an urban traffic network.

only in their arcs, or even only in the weights of their arcs. If one of them has a smaller Kemeny constant, it means that the average trip is shorter, which indicates that the arc structure is more efficient. We will use the Kemeny constant several times to compare similar networks and order them according to efficiency. If the transition matrix $P$ has eigenvalues $\lambda_1 = 1, \lambda_2, \ldots, \lambda_n$ then another way of computing $K$ is (see [Levene and Loizou, 2002]).

$$K = \sum_{j=2}^{n} \frac{1}{1 - \lambda_j}. \tag{2.6}$$

Equation (2.6) emphasises the fact that $K$ is only related to the particular matrix $P$ and that it increases if one or more eigenvalues of $P$ is real and close to 1. The Kemeny constant is closely related to other well known quantities used in the study of graphs. For example $K_i + 1$ is identical to the expected time to mixing, as defined in [Hunter, 2006].

## 2.2. Base Model

We now describe how a Markov chain can be used to model an urban road traffic network. Graphs and Markov chains can be used quite naturally to do this. This approach was first proposed in [Crisostomi et al., 2011b] and further developed in [Crisostomi et al., 2011c] and [Crisostomi et al., 2011a]. The base chain, which we describe first, is the simplest version of the chain and describes the transitions of vehicles between road segments, i.e., the probability with which a car driving along a particular road will drive onto each of the possible next roads.

Before deriving the Markov chain a directed graph associated with the road network is described. The Markov chain model describes a random walk on this graph. It can be argued that this model gives interesting insight into properties of the road network. Real cars, clearly, do not perform random walks and the model must be understood as a tool to derive network properties, not as a representation of actual traffic dynamics.
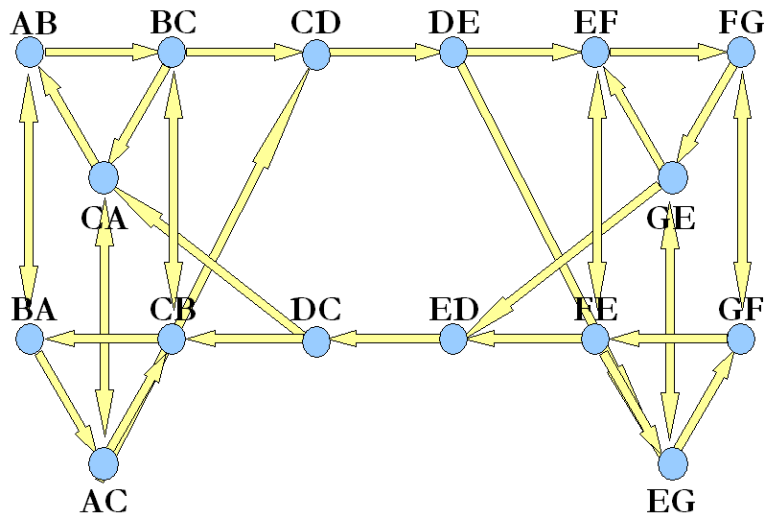
Figure 2.2.: The dual graph corresponding to the graph shown in Figure 2.1.

The graph corresponding to the road network is constructed in the following way. Each intersection in the road network is a node in our graph and there is an edge between two nodes if there is a road segment that connects the two corresponding intersections. An example of such a graph can be found in Figure 2.1, where junctions A,B,...,G are connected by road segments. For instance, the road segment AB is the road that allows a car to go from A to B, and is different from BA which goes from B to A. We will call this the primal graph and we now describe the dual graph. In the dual graph the nodes are the edges from the primal graph, (i.e., the road segments of the traffic network), and there is an edge between two nodes if it is possible to make a direct transition between the corresponding road segments. The dual graph corresponding to Figure 2.1 can be found in Figure 2.2. It can be noted that dual graphs carry more information than primal graphs. For instance, we can see from Figure 2.2 that cars are not allowed to perform u-turns at junction D, while the same information can not be recovered from the primal graph. The weights for the edges in the dual graph are given by the turning probabilities. In the following, we are interested in the Markov chain that corresponds to the dual graph.

The following argument motivates the use of a Markov chain in this context. It is clear that cars entering a road segment have to leave it again eventually. Consequently, we have asymptotic flow conservation (i.e., the long-run fraction of cars entering the road has to equal the fraction of cars leaving again). If we use $\pi \in \mathbb{R}_+^n$ to denote the long-run distribution of cars within the road network, and $p_{ij} \geq 0$ to denote the probability of going from road $i$ to road $j$, then

$$\underbrace{\sum_{j=1}^{n} \pi_j p_{ji}}_{\text{fraction of cars entering road } i} = \underbrace{\sum_{k=1}^{n} \pi_i p_{ik}}_{\text{fraction of cars leaving road } i} = \pi_i \sum_{k=1}^{n} p_{ik} = \underbrace{\pi_i}_{\text{fraction of cars in road } i}, \quad (2.7)$$

where we consider the cars remaining in road $i$ after one time step as cars leaving and re-entering road $i$. In matrix form Equation (2.7) translates into $\pi^\top P = \pi^\top$, where $P = [p_{ij}]$. This is exactly the relation describing the stationary distribution that we also get from the Markov chain as described above. The approach is further justified experimentally using Sumo in [Crisostomi et al., 2011b]. The base model presented in this section describes the redistribution of a fixed number of items between nodes over time. This is a very useful tool. However, the necessary assumption of local (per street) flow conservation is a weakness in the construction of this model. It is a reasonable assumption if the model considers traffic over the course of a full day, however it does not necessarily hold on smaller time scales. An intuitive example of this is commuter traffic, which constitutes a significant part of the total traffic in cities. Clearly, in the mornings more cars leave residential parts of the city than arrive while the opposite is true in the evenings. The traffic flow and the performance of the network may be quite different at different times of the day making it interesting to consider periods that are shorter than the full day. As we shall see, this behaviour can be incorporated into the model. This issue was not discussed in [Crisostomi et al., 2011b] and we will relax the assumption of local flow conservation in Chapter 3.

## 2.3. Multi-Variate and Derivative Models

The base chain as defined above describes the probability of making a transition from one road segment to another in one step. It contains no concept of time that is required to travel, or pollutants that are released, or energy that an electric vehicle requires to travel along its path. In this section we describe how to derive meaningful models from this base chain. In [Crisostomi et al., 2011b] this was proposed in the context of travel time modelling, however following the same approach it is possible to modify the base chain to take into account other features. For example it is possible to describe the probability of transition per unit of time (giving a congestion chain), per unit of pollution (giving a pollution chain), and per unit of energy consumption (giving an energy consumption chain).

The base chain is modified by inserting diagonal entries to the transition matrix $P$ introduced in the previous section, and by scaling the off-diagonal entries appropriately. The diagonal entries are used in our model to account for several factors such as the length of the road, speed limits, road inclination, priority rules, traffic lights, traffic congestion and other non-trivial factors. We use some or all of these factors to compute a cost that has to be paid in order to traverse each road segment. In particular, assume that a cost of $w_i > 0$ is required on average to pass through state $i$, where it is assumed

Table 2.1.: Interpretation of some Markov chain quantities of interest in the congestion application

| Quantity / Markov chain | Congestion chain |
|---|---|
| **Perron Eigenvector (dual)** | Vehicular density in the network |
| **Mean First Passage Times** | Average travel times for a pair of origin/destination |
| **Kemeny constant** | Average travel time for a random trip |
| **Perron Eigenvector (primal)** | Congested junctions in the network |

that $\min_i w_i = 1$ otherwise all weights can be scaled by dividing by $\min_i w_i$. Then the diagonal entries of the transition matrix of the derivative chain are

$$Q_{i,i} = \frac{w_i - 1}{w_i}, \qquad (2.8)$$

where in the originally proposed framework $w_i$ was choosen to be the average travel time of vehicles along road segment $i$. The off-diagonal entries, $Q_{ij}, j \neq i$ are computed via

$$Q_{i,j} = (1 - Q_{i,i})P_{i,j}. \qquad (2.9)$$

According to the way the diagonal entries are chosen, the expected time the random walk spends in each road segment coincides with the average travel time $w_i, i = 1, \ldots, n$. See for instance [Crisostomi et al., 2011b, Appendix] for a motivation for this construction and a description how this was derived. The matrix $Q$ is again row-stochastic and the ratios between the off-diagonal terms in each row are the same as in the initial transition matrix $P$.

The computation of the diagonal entries (with a possible scaling beforehand) in the derivative transition matrix suggest a special meaning to the number 1 in Equation (2.8). However it is not specified how time is measured exactly. Time could be measured in minutes or seconds or in multiples of 7 seconds. It was an oversight of this work that the effect of choosing this base unit, or step size of the chain was not investigated. We rectify this in Chapter 3.

Perhaps the most attractive feature of the proposed paradigm is that typical Markov chain quantities of interest have a straightforward interpretation in the applications. For the congestion chain this is summarised in Table 2.1.

As will be further discussed in Chapter 4, the base chain, built only with information about the turning probabilities, can be used as a starting point for different applications. The distinguishing feature is the choice of diagonal entries in the matrix $Q$, i.e. the choice of the weights $w_1, \ldots, w_n$. It is important to note that the construction of matrix $Q$ is well defined only for strictly positive values $w_i, i = 1, \ldots, n$. In some applications this assumption is not true. For example, in the context of energy consumption in electric

vehicles negative costs occur quite naturally. This will be further discussed in Chapter 4. It is possible to generalise the construction of the Markov chain transition matrix $Q$ to allow that some weights are negative. This generalisation is presented in Chapter 3.

## 2.4. Congestion Chain Applications

A number of interesting applications of the Markov chain framework in the context of travel time modelling have been proposed in [Crisostomi et al., 2011b]. In this section we briefly mention them. Some of these applications have inspired approaches in the context of pollution and energy consumption modelling.

As part of verifying the utility of the Markov chain model using simulations it is first shown that the stationary distribution of the Markov chain corresponds to the density of vehicles measured from simulations. Later it is suggested that the stationary distribution can be used to find traffic light sequences that achieve certain goals, such as the reduction of congestion on individual road segments. To this end the authors further derive formulas to compute the changes in the stationary distribution caused by changes in the value of a given weight $w_i$. One particularly useful fact is that the eigenvector of $Q$ that corresponds to the eigenvalue with the second largest absolute value makes it possible to partition the road network into sub-communities with most of their traffic remaining inside the community and comparably little traffic that transitions between sub-communities. This knowledge is valuable for many applications. Using this eigenvector in this way is a well known technique in spectral graph clustering, see for example [von Luxburg, 2007]. However, most of the available results and techniques in this context hold only for undirected graphs. Information about travelling clusters in traffic networks are not easily obtained from other traffic models or even simulations. Further the intuition that mean first passage times obtained from this model are a good measure for average travel times is validated using simulations. And it is suggested that the mean first passage times can be used to develop new routing algorithms that minimise mean first passage times instead of the more conventional approach to minimise travel time. In this context it is also mentioned that a simple transformation of the weights $w_i, i = 1, \ldots, n$ can be used to change the model from a congestion chain to an emissions chain, yielding a routing strategy that minimises expected pollution outputs. This is the starting point for parts of the work presented in Chapter 4. The last Markov chain related quantity discussed is the Kemeny constant. It is shown that the Kemeny constant is to some degree a measure for the efficiency of a road network. The authors discuss and show via an example that removing different road segments and computing the Kemeny constant of the remaining network enables identification of road segments that have a significant negative impact on the network. It is further described how for a given road network turning probabilities can be found so that the network has the minimal possible Kemeny constant. Even though it may not be possible to influence all vehicles to follow prescribed turning possibilities, it yields an important benchmark for all network optimisation applications. To this end a method is described to immediately

see if the reduction of any given entry of the transition matrix $Q$ has a positive effect on the Kemeny constant.

# 3. Extensions to the Markov Chain Model

***Abstract :*** *The work presented in this chapter builds on the work published in [Crisostomi et al., 2011b].We extend their work in several important and interesting aspects. The main contribution of this chapter is to remove a number of strong assumptions and to further develop the analytic tools to study Markov chain models. The work reported here was conducted jointly with Emanuele Crisostomi, Stephen Kirkland, and Robert Shorten and was published in [Schlote et al., 2011; Crisostomi et al., 2011a,c; Schlote et al., 2012a,b].*

In this chapter we present a number of theoretical extensions to the modelling framework suggested in [Crisostomi et al., 2011b]. These extensions are in three areas. At first we discuss the role of the step size in the model and show that changes in the step size do not affect the results of the model. Then we discuss the flow conservation assumptions made in [Crisostomi et al., 2011b]. The authors assume that vehicles that enter a road segment must also leave the road segment again, which is not necessarily true on time scales of interest. We show that they can be replaced with more reasonable assumptions. In a third step we extend the framework to be able to handle the existence of negative costs for some road segments, which arises quite naturally in the study of electric vehicles.

## 3.1. Step Size

Discrete time models describe the evolution of a system at distinct points in time. When they are applied to continuous time systems, a choice needs to be made as to what these distinct points in time are. In the Markov chain model discussed in this chapter we consider the time instants to be equidistant and we call the time between two consecutive instants the step size. Depending on the specific discrete time model used it is possible that the step size choice has an influence on characteristics of the model. As noted in Chapter 2 there was no discussion of the effect of the choice of step size on the Markov model. We now rectify this. To this end, let costs of $w_i > 0$ be required on average to pass through state $i$ be given. We can generalise the construction of the diagonal entry of the Markov chain transition matrix defined in Equation (2.8) in the following way.

Let the $i'$th diagonal entry be equal to $\tilde{w}_i$, with

$$\tilde{w}_i = \frac{w_i - \alpha}{w_i}, \tag{3.1}$$

where we now choose $0 < \alpha \le \min_i w_i$. This coincides with the originally proposed construction in the case that $\alpha = \min_i w_i = 1$. The parameter $\alpha$ corresponds to the cost associated with one step of the Markov chain; we call $\alpha$ our step size. The step size cannot be larger than the minimum cost associated with a road segment as it would be possible to make a transition between states that are connected via a low cost street but not directly in one time step. This would in turn cause additional edges in the network graph that do not correspond to physically possible connections.

Let us denote by $W$ and $D$ the diagonal matrices $W = \operatorname{diag}(w_1, \ldots, w_n)$ and $D = \operatorname{diag}(\tilde{w}_1, \ldots, \tilde{w}_n)$ respectively. The derivative Markov chain transition matrix $Q$ as defined in Equations (2.8) and (2.9) in its extended version with arbitrary step size can equivalently be obtained by

$$Q = (I - D)P + D \tag{3.2}$$

where $I$ is the identity matrix of appropriate dimensions. The following identities hold.

$$D = I - \alpha W^{-1} \tag{3.3}$$

$$Q = I + \alpha W^{-1}(P - I). \tag{3.4}$$

We now state a lemma describing the relationship between the left Perron eigenvectors of $P$ and $Q$. It yields an alternative way of computing the left Perron eigenvector of $Q$ and is a rigorous presentation of an observation from [Crisostomi et al., 2011b].

**Lemma 3.1.** *Let $P$ be irreducible and let $\pi^\top$ be the left Perron eigenvector of $P$, then $(W\pi)^\top$ is the left Perron eigenvector of $Q$.*

*Proof.* $Q$ has a left Perron eigenvector as it is non-negative and has the same sign structure as $P$ off the diagonal and is thus irreducible. The claim is then proved using Equations (3.3) and (3.4) and the following.

$$\begin{aligned}
(W\pi)^\top Q &= (W\pi)^\top (I + \alpha W^{-1}(P - I)) \\
&= \pi^\top W^\top + \alpha \pi^\top W^\top W^{-1} P - \alpha \pi^\top W^\top W^{-1} \\
&= (W\pi)^\top.
\end{aligned}$$

$\square$

Next, we give a result describing the dependence of the eigenvalues and eigenvectors of $Q$ on the step size $\alpha$ of the chain. To ease exposition, we give $\alpha$ as an argument whenever a quantity depends on $\alpha$, for example we treat $Q$ as a a function on $\alpha$

$$Q(\alpha) = I + \alpha W^{-1}(P - I) \tag{3.5}$$

We show that all generalised eigenvectors of $Q$ are independent of the step size $\alpha$, but the corresponding eigenvalues are not.

**Lemma 3.2.** *Let $\alpha_1, \alpha_2$ be positive step sizes. Then any right (or left) (generalised) eigenvector $x$ of $Q(\alpha_1)$ is also a right (or left) (generalised) eigenvector of $Q(\alpha_2)$, and vice versa. Further, for a common right (or left) (generalised) eigenvector the corresponding eigenvalues, $\lambda(Q(\alpha_1))$ and $\lambda(Q(\alpha_2))$, are related as*

$$\frac{\lambda(Q(\alpha_1)) - 1}{\alpha_1} = \frac{\lambda(Q(\alpha_2)) - 1}{\alpha_2}. \tag{3.6}$$

*Proof.* The proof is based on the special form of $Q(\alpha)$, namely $Q(\alpha) = \alpha W^{-1}(P - I) + I$. Each of $Q(\alpha), Q(\alpha) - I$ and $\frac{1}{\alpha}(Q(\alpha) - I)$ has a common (right or left) Jordan basis. As $\frac{1}{\alpha}(Q(\alpha) - I) = W^{-1}(P - I)$ is independent of $\alpha$, we see that any right or left Jordan basis for $Q(\alpha_1)$ is also a right or left Jordan basis for $Q(\alpha_2)$. This proves the claim on the eigenvectors. Further, note that if $x$ is a right eigenvector for $Q(\alpha_1)$ with eigenvalue $\lambda(Q(\alpha_1))$, then $x$ is also a right eigenvector for $\frac{1}{\alpha_1}(Q(\alpha_1) - I)$ with eigenvalue $\frac{\lambda(Q(\alpha_1)) - 1}{\alpha_1}$; it now follows that x is also a right eigenvector for $\frac{1}{\alpha_2}(Q(\alpha_2) - I)$ with eigenvalue $\frac{\lambda(Q(\alpha_2)) - 1}{\alpha_2}$, and hence that $\frac{\lambda(Q(\alpha_1)) - 1}{\alpha_1} = \frac{\lambda(Q(\alpha_2)) - 1}{\alpha_2}$. $\qquad \square$

So far we analysed how eigenvalues and eigenvectors behave as we change the chain's step size $\alpha > 0$. A natural question to ask is whether the Markov chain itself is well behaved when the step size approaches 0 as a limit. By this we mean that one step at a given step size or two steps at half the step size and so on should yield the same result. For completeness we now end this section by giving a result that answers this question. To this end let $A = W^{-1}(P - (I))$, with this definition we obtain $Q(\alpha) = I + \alpha A$. In the following theorem we denote by $\exp(\cdot)$ the matrix exponential function.

**Theorem 3.3.** *The limit $\lim_{k \to \infty} Q(\frac{\alpha}{k})^k$, where $k \in \mathbb{N}$, exists and is equal to $\exp(\alpha A)$.*

*Proof.* We know from Lemma 3.2 that the generalised eigenvectors of $Q(\alpha)$ are independent of $\alpha > 0$. Thus all matrices $Q(\alpha)$ and $A$ are simultaneously similar to their respective Jordan canonical forms. Let $J = S^{-1}AS$ be the Jordan form of $A$. Then

$$\left(Q\left(\frac{\alpha}{k}\right)\right)^k = \left(I + \frac{\alpha}{k}A\right)^k = \left(I + \frac{\alpha}{k}SJS^{-1}\right)^k = S\left(I + \frac{\alpha}{k}J\right)^k S^{-1} \tag{3.7}$$

By considering each individual Jordan block, it can be seen that

$$\lim_{k \to \infty} \left(I + \frac{\alpha}{k}J\right)^k = \exp(\alpha J). \tag{3.8}$$

And hence $\lim_{k \to \infty} \left(Q\left(\frac{\alpha}{k}\right)\right)^k = S \exp(\alpha J)S^{-1} = \exp(\alpha A)$. $\qquad \square$

**Comment :** This result is important to show consistency of our approach. We now also have a way of deriving a continuous time Markov process. Further its density matrix is $\alpha A$, which is readily computable from the turning probabilities and the weight matrix $W$.

## 3.2. Alternative Flow Conservation Assumptions

As we have already mentioned, a basic problem in the model [Crisostomi et al., 2011b] is that the flow conservation assumptions made to derive the model are not realistic. While it is relatively straightforward to overcome the limitations imposed by these assumptions, it is an oversight in the work described in [Crisostomi et al., 2011b], that these limitations are neither discussed nor resolved there. However these assumptions are widely used in traffic modelling. This is especially true for flow based methods and these assumptions underlie many fluid models [Hartenstein and Laberteaux, 2010]. It should be noted that this assumption is realistic in highway networks [Baskar et al., 2011] or single junctions [Piccoli and Garavello, 2006; Moya and Poznyak, 2009], but can not be generalised to an urban network level. We now rectify this matter by describing a basic modification to the Markov chain that resolves these issues.

The assumption of local flow conservation is too restrictive in real traffic networks. For example, in the morning there will be much traffic from residential areas to commercial areas, so that from a network point of view residential areas function as sources of vehicles while commercial areas are sinks for the vehicles. Other conservation assumptions are more reasonable. For example, it is reasonable to consider the total amount of cars in a city as being constant, or to assume that as one car enters the extra state, another car ("somewhere") starts a journey. Consequently, there are at least two methods of overcoming the limitations of the model in [Crisostomi et al., 2011b]. We now present these two modifications. Afterwards, we show that both approaches give equivalent results in terms of the stationary distribution.

Motivated by the work underpinning the PageRank [Page et al., 1999] algorithm, for the first approach we require that when a car reaches its destination, somewhere else a new car starts its journey. This gives our model two additional degrees of freedom. For each road segment we have a probability that a car will leave or enter the road network at this segment. This allows to model a large number of scenarios. We will refer to this approach as the teleportation approach, following the convention from [Page et al., 1999].

The second approach is inspired by the theory of semi-open Jackson networks [Chen and Yao, 2001]. In closed Jackson networks the number of customers is constant. This class of networks yields comprehensive product form stationary distributions and their behaviour is thus easy to describe. In semi-open Jackson networks the number of customers is bounded from above. To recover the results from the closed case, an extra queue is added to which all customers go that previously left the network and all newly arriving customers come from this queue. This extra state represents users that have departed to the outside world. While the number of customers in the extended network is constant, the number of customers which is in the original network is variable but bounded. We apply the same strategy to our road network model. We add an extra state that is connected to all other road segments. We consider a car that finishes its journey to transition to the extra state. With the probabilities of going to the extra

state from each road segment and of going to each road segment from the extra state we have here the same parameters as in the teleportation approach, but at the cost of having an extra state we gain the possibility of influencing the time cars spend outside the network. Note that we can also include cars physically leaving the considered road network (for example by leaving the city) into this framework.

We now present the technical details of how transition matrices for our Markov chain corresponding to the two approaches can be obtained. We will then show that the respective stationary distributions are equivalent and independent of the weight of the self-loop of the extra state.

Let $R = [r_{ij}] \in \mathbb{R}^{n \times n}$ be the matrix counting transitions between streets, $p \in \mathbb{R}^n$ the vector that counts the number of times each road segment is the origin of a route, and $q \in \mathbb{R}^n$ the vector that counts the number of times each road segment is the destination of a route. We will throughout this section assume that neither $p$ nor $q$ is the zero vector and that $R$ is an irreducible matrix. Let $\tilde{p} = \frac{p}{\|p\|_1}$.

For the teleportation approach let

$$\tilde{U} = R + q\tilde{p}^\top. \tag{3.9}$$

We now make this matrix row stochastic by scaling it in the following way

$$U = F\tilde{U}, \tag{3.10}$$

where $F = \operatorname{diag}(f_1, \ldots, f_n)$ with

$$f_i = \left( \sum_{j=1}^{n} r_{ij} + q_i \left( \sum_{j=1}^{n} \tilde{p}_j \right) \right)^{-1} = \left( \sum_{j=1}^{n} r_{ij} + q_i \right)^{-1}. \tag{3.11}$$

$U$ is row stochastic and describes the transitions in the chain without the extra state.

Now we describe the transition matrix for the extra state Markov chain. Let $c$ be a positive number that we may later use to influence the dwell time in the extra state and let

$$\tilde{V} = \begin{bmatrix} R & q \\ p^\top & c \end{bmatrix}. \tag{3.12}$$

We now make this row stochastic by

$$V = G\tilde{V}, \tag{3.13}$$

where $G = \operatorname{diag}(g_1, \ldots, g_{n+1})$ with

$$g_i = \left( \sum_{j=1}^{n} r_{ij} + q_i \right)^{-1} \tag{3.14}$$

for $i = 1, \ldots, n$ and $g_{n+1} = (c + \sum_{j=1}^{n} p_j)^{-1}$ and thus

$$G = \begin{bmatrix} F & 0 \\ 0 & (\sum_{j=1}^{n} p_j)^{-1} \end{bmatrix}. \tag{3.15}$$

$V$ is row stochastic and describes the transitions in our chain with the extra node.

The following lemma explains in which sense the two approaches yield equivalent stationary distributions; it is a consequence of a well known result presented in [Meyer, 1989].

**Lemma 3.4.** *Let $x^T$ be the stationary distribution of $V$, with $x = \begin{pmatrix} \tilde{x} \\ x_{n+1} \end{pmatrix}$. Then $\frac{1}{1-x_{n+1}} \tilde{x}^T$ is the stationary distribution of $U$.*

*Proof.* We have $x^\top = x^\top V$ and from this we obtain the two equations

$$x_{n+1} = \tilde{x}^\top F q + \frac{c x_{n+1}}{c + \sum_{j=1}^{n} p_j} \tag{3.16}$$

and

$$\tilde{x}^\top = \tilde{x}^\top F R + x_{n+1} \frac{p^\top}{c + \sum_{j=1}^{n} p_j}. \tag{3.17}$$

Solving Equation (3.16) for $x_{n+1}$ and substituting in Equation (3.17) yields

$$\tilde{x}^\top = \tilde{x}^\top F R + \left( \frac{\tilde{x}^\top F q}{1 - \frac{c}{c + \sum_{j=1}^{n} p_j}} \right) \frac{p^\top}{c + \sum_{j=1}^{n} p_j} = \tilde{x}^\top \left( F R + F q \tilde{p}^\top \right) = \tilde{x}^\top U.$$

Normalising with respect to the one-norm we obtain that $\frac{1}{1-x_{n+1}} \tilde{x}^\top$ is the left Perron vector of $U$.

$\square$

The Markov chain with transition matrix $U$ is useful for modelling mobility patterns in the network. In particular, the vectors $p$ and $q$ can be used to model time-dependent directionality. For example, in the morning more cars leave the network in the city centre whereas in the evening the opposite is true.

The Markov chain with transition matrix $V$ is useful when in addition one wishes to regulate the number of cars in the extra state. Encouraging people to use their cars instead of other modes of transportation could be sensible in the following two situations for example. Either when the public transportation network is operating above its capacity, or - if a significant part of the vehicles are electric vehicles or plug-in hybrid electric vehicles - the load on the public power grid is too high.

## 3.3. Networks with Negative and Positive Weights

In Section 2.3 a diagonal matrix $D$ was used to transform the basic Markov chain $P$ into a new Markov chain $Q$, whose step unit can be different from the unit of time (e.g., can be a unit of energy). This was done in Equation (3.2), which we recall here for convenience:

$$Q = (I - D)P + D. \tag{3.18}$$

One of the new application areas of the Markov chain model discussed in this thesis is the modelling of energy consumption in electric vehicles, see Section 4.2. In the case of electric vehicles, the mechanism of regenerative braking can cause non-positive entries in the diagonal matrix $W$. The matrix $D$ is constructed using the elements of $W$, see Equation (3.1). However, for the construction it is assumed that $W$ is a positive diagonal matrix. If one were to follow the construction anyway, negative entries would appear on the diagonal of $W$. This would imply that $Q$ is not a transition matrix of a Markov chain. $Q$ might not be non-negative and row sums could be different from 1, so standard methods to analyse Markov chains would not apply. We note that Lemma 3.1 would still holds in the sense that even if we allow negative entries in the diagonal matrix $W$ then for all step sizes $\alpha > 0$, such that $D$ is non-singular, the vector $(Wx)^\top$ is still a left eigenvector of the matrix $Q$ as defined in Equation (3.18) to the eigenvalue 1. However the matrix $Q$ and its eigenvectors do not seem to have a straightforward interpretation as in the case of all positive weights. In order to apply our previously developed theory and recover some of our analytic tools, we will generalise the construction of the derivative chain. To do this we make only the assumption that the energy required to travel along a road $i$ can not be exactly zero.

To this end, we first define an intermediate Markov chain whose step unit is a unit of energy exchanged between the vehicle and the road network, regardless of whether the unit of energy was spent or gained (thanks to regenerative braking). To do this we define $\widetilde{W}$ to be the diagonal matrix that contains the absolute values of all weights, $\widetilde{W} = \operatorname{diag}(|w_1|, \ldots, |w_n|)$. Accordingly we define $\widetilde{D} = I - \alpha \widetilde{W}^{-1}$, where now $0 < \alpha \leq \min_i |w_i|$. We then obtain the transition matrix $\widetilde{Q} = [\tilde{q}_{ij}]$ of the intermediate chain from

$$\widetilde{Q} = (I - \widetilde{D})P + \widetilde{D}. \tag{3.19}$$

We are interested in storing memory of the sign of energy exchange (i.e., to compute the actual energy required to travel a given route). We assume that transitions between streets occur at the end of the energy step. Hence, the gain or loss of energy while driving along road segment $i$ is independent of the choice of the next road segment $j$. With this assumption in place, we introduce the notation $\sigma_i$ to indicate the sign of the change in energy transferred from the vehicle to the network. That is, $\sigma_i = 1$ if the vehicle loses energy driving along road $i$, and $\sigma_i = -1$ if the vehicle gains energy driving along road $i$.

As the quantity of interest is energy instead of time, we use the term mean first passage energy (MFPE) instead of mean first passage time in this context. We use the intermediate Markov chain to calculate a generalised version of MFPE (generalised to include possible negative values), following and extending the approach of [Grinstead and Snell, 2003]: For $j \neq i$, to calculate $m_{ij}$, the MFPE from $i$ to $j$, we observe that in going from $i$ to $j$ we make a direct transition with probability $\tilde{q}_{ij}$ and spend $\sigma_i$ units of energy. With probability $\tilde{q}_{ik}$ we make a transition to $k \neq j$ where we again spend $\sigma_i$ units of energy to get to $k$, in addition the expected energy required to get from $k$ to $j$ is equal to $m_{kj}$. Thus for $j \neq i$

$$
\begin{aligned}
m_{ij} &= \tilde{q}_{ij}\sigma_i + \sum_{k \neq j} \tilde{q}_{ik}(m_{kj} + \sigma_i) \\
&= \sum_{k \neq j} \tilde{q}_{ik}m_{kj} + \sum_{k=1}^{n} \tilde{q}_{ik}\sigma_i \\
&= \sum_{k \neq j} \tilde{q}_{ik}m_{kj} + \sigma_i.
\end{aligned}
\tag{3.20}
$$

For any fixed $j = 1 \ldots, n$ we can write this in vector form

$$
m_{(j)} = \widetilde{Q}_{(j)}m_{(j)} + \sigma_{(j)},
\tag{3.21}
$$

where $m_{(j)} = \left(m_{1j}, m_{2j}, \ldots, m_{(j-1)j}, m_{(j+1)j}, \ldots, m_{nj}\right)^{\top}$ and equivalently $\sigma_{(j)}$ is the vector of all $\sigma_i$ for $i \neq j$ and $\widetilde{Q}_{(j)}$ is the matrix $\widetilde{Q}$ where we have eliminated the $j'$th row and column. Now we can calculate $m_{ij}$ for all $i \neq j$ using the formula

$$
m_{(j)} = (I - \widetilde{Q}_{(j)})^{-1}\sigma_{(j)},
\tag{3.22}
$$

where we use the fact that $(I - \widetilde{Q})$ is a singular, irreducible M-matrix and according to [Berman and Plemmons, 1994, Chapter 6, Theorem 4.16] each of its principal submatrices is invertible.

**Comment :** Note that Equation (3.22) is identical to a standard formula for calculating mean first passage times in the case of positive weights only. As we have shown it still works if we have negative weights in the chain.

As in the case of positive weights we are interested in the Kemeny constant as a global efficiency measure, but if we introduce negative weights in our graph, the Kemeny constant as defined in Equation (2.5) is no longer independent of the starting node $i$. However we can generalise the notion of the Kemeny constant in the following way. Let $\pi$ be the left Perron eigenvector of $\widetilde{Q}$ and let

$$
K = \sum_{i=1,\ldots,n} \pi_i K_i = \sum_{i=1,\ldots,n} \pi_i \sum_{j \neq i} \pi_j m_{ij}.
\tag{3.23}
$$

Then $K$ coincides with the Kemeny constant as defined in Equation (2.5) in the case where all weights are positive. The generalised Kemeny constant can be interpreted as the expected number of steps of a random walk between a random origin and a random destination, where both origin and destination are picked with probability according to the stationary distribution.

In the spirit of Lemma 3.2, the following lemma shows that generalised mean first passage energies and the generalised Kemeny constant scale linearly with the step size.

**Lemma 3.5.** *For given step sizes $\alpha_1, \alpha_2 > 0$ and for all $i, j = 1, \ldots, n$ the following equalities hold.*

$$\alpha_1 m_{ij}(\alpha_1) = \alpha_2 m_{ij}(\alpha_2) \tag{3.24}$$

$$\alpha_1 K(\alpha_1) = \alpha_2 K(\alpha_2). \tag{3.25}$$

*Proof.* From Equation (3.22) we have

$$
m_{(j)}(\alpha_1) = \left( I - \widetilde{Q}_{(j)}(\alpha_1) \right)^{-1} \sigma_{(j)} = \left( -\alpha_1 \widetilde{W}^{-1}(P - I) \right)_{(j)}^{-1} \sigma_{(j)}
$$

$$
= \frac{\alpha_2}{\alpha_1} \left( -\alpha_2 \widetilde{W}^{-1}(P - I) \right)_{(j)}^{-1} \sigma_{(j)} = \frac{\alpha_2}{\alpha_1} m_{(j)}(\alpha_2), \tag{3.26}
$$

where again by $(-\alpha_1 \widetilde{W}^{-1}(P - I))_{(j)}$ we denote the matrix $(-\alpha_1 \widetilde{W}^{-1}(P - I))$ where we have eliminated the $j'$th row and column. The claim about the generalised Kemeny constant now follows directly from its definition, Equation (3.23). $\qquad\square$

As further explained in the remainder of this section, $W$ represents a diagonal matrix of appropriate weights that accounts for the average quantity of interest that is spent or released along each road segment. We use $W$ to transform the basic Markov chain described by the transition matrix $P$ to a new chain with transition matrix $Q$. This has previously been done in [Crisostomi et al., 2011b] in the context of a model taking travel time into account, yielding a congestion chain.

# 4. Applications for the Markov Chain Traffic Model

***Abstract :*** *In this chapter we discuss a number of novel application areas to the traffic modelling framework suggested in [Crisostomi et al., 2011b]. It is joint work with Emanuele Crisostomi, Stephen Kirkland, and Robert Shorten and was published in [Schlote et al., 2011; Crisostomi et al., 2011a,c; Schlote et al., 2012a,b]. In particular, we apply the framework to model traffic induced pollution and to model the consumption of battery power in electric vehicles.*

## 4.1. Pollution Chain

In this section we change the unit of the Markov chain model to represent the emission of pollutants. The resulting congestion chain can be interpreted to describe the probability of a vehicle going on to the next road segment along its route after spending one unit of pollutant on its current road segment. Thus, the chain does not take into account the dynamics of vehicles over time. In this context, it is very important to bear in mind that the Markov chain model in general is not a representation of actual dynamics of vehicles but rather a tool that allows to extract certain useful informations from a road network. In order to construct the chain, we first need to describe our emissions model. Different pollutants depend on different characteristics of the vehicle and its operation. We concentrate on the modelling of the emissions of a single pollutant. It is possible to take several pollutants into account for example by considering weighted sums or more complicated functions of all pollutants of interest.

### 4.1.1. Pollution Modelling in Urban Road Networks

Following the construction described in Section 2.3, two pieces of information are required to build the pollution chain, the junction turning probabilities to build the matrix $P$, and the average amount of pollution emitted when using a given road segment to build matrix $W$. For an individual car the amount of pollutant emitted along a road segment depends on a number of factors such as the type, age and wear of the engine

and the operation of the vehicle, i.e. speed, acceleration and deceleration, the road geometry, as well as engine temperature, tires, the drive train and environmental factors. Hence, the computation of exact emission values is complicated. We will not take all these factors into account. More accurate models using more or all of the above factors can be integrated into our framework if needed. To compute the average emissions per road segment from this it is further necessary to know the mix of vehicles along the segment.

There are two possible ways of obtaining the information required for the Markov chain:

1. All cars are instrumented to behave as mobile sensors and store the data relative to their travelling history. At regular times, they communicate their data to a central database that collects all the important information.

2. Urban networks are equipped with loop detectors that are positioned at each junction, and they measure the required information to build the Markov transition matrix.

We emphasise here that most of the required information can be easily collected (when it is not already available) by the cars themselves, and we only need to collect and integrate this information. Vehicle positioning systems (GPS) have reached widespread penetration. It is easy to imagine geospatial tagging of vehicle route information, and storing this information locally in vehicle memory. It is also possible to obtain information of the type required from special classes of vehicles (buses, taxis etc.). This may however contain less useful information. Note that such instrumented fleets already exist in cities such as Stockholm [Biem et al., 2010] and Dublin [IBM, 2013]. Furthermore, recent advances in the development of Vehicular Adhoc NETworks (VANET) [Hartenstein and Laberteaux, 2010] are expected to facilitate the proposed emissions model.

It is important to note that we are proposing a new paradigm for pollution modelling that captures large-scale (macroscopic) effects. For the purpose of constructing the model, and for its evaluation, a microscopic model to describe microscopic effects is required. There are several models to evaluate pollution at the microscopic level, with different levels of accuracy and complexity, see [Spence et al., 2009]. A first basic model is the so-called aggregated emission factor model, where a single emission factor is used to represent a particular type of vehicle and a general type of driving, with a usual distinction between urban roads, rural roads and motorways. This model is rather rudimentary and is not realistic for small scale networks, as it omits several phenomena, such as congestion, which are known to significantly affect emissions. In this chapter, we shall make use of more refined average-speed models to build our Markov chain. In this, emission factors are calculated as a function of average speed, see [Barlow et al., 2001; Boulter et al., 2009]. The average-speed approach is described in [UK Department of Transport, 2007] and [Gkatzoflias et al., 2007].

Although the average-speed model has been extensively used for many applications, it suffers from a drawback that very different vehicle operational behaviours (in terms of accelerations, decelerations, maximum speed, gear-change pattern), and therefore different emission levels, can be characterised by the same average speed. A more realistic model is for instance the comprehensive modal emissions model (CMEM) described in [Barth et al., 2000]. According to this model, second-by-second exhaust emissions and fuel consumption are predicted, for a wide range of vehicle categories and ages. For the sake of simplicity, in this work the average-speed model is employed. It is a great advantage of the proposed Markov chain model that any other (more accurate) vehicle emissions model can easily be used for the computation of the weights. However, a high price must be paid in terms of communication between vehicles and infrastructure, as the whole speed profile of individual vehicles is required to be transmitted.

### 4.1.2. Construction of the Markov Chain Transition Matrix

The pollution chain is completely determined by the transition matrix, whose diagonal terms indicate how many units of pollution a fleet of vehicles releases along a road segment, and by off-diagonal terms that indicate the probabilities of turning left, right, etc, when arriving at a junction.

According to the average-speed model, an emission factor $f(t, p)$ is computed for each road segment using

$$f(t,p) = \frac{k\left(a + bv + cv^2 + dv^3 + ev^4 + fv^5 + gv^6\right)}{v}, \tag{4.1}$$

where $t$ denotes the type of vehicle (and depends on fuel, emission standard, category of vehicle, engine power), p denotes the particular type of pollution of interest (e.g. CO, $CO_2$, NOx, Benzene), $v$ denotes the average speed of the vehicle, and the parameters $a, b, c, d, e, f, g$ and $k$ depend on both the type of vehicle and the pollutant $p$ under consideration. For the purpose of this work, the values of the parameters are taken from Appendix D, in [Boulter et al., 2009]. In Equation (4.1) it is assumed that speeds are measured in km/h and emission factors in g/km. Therefore, by assuming that for all $i = 1, \ldots, n$ the average speed of a fleet of vehicles along road segment $i$ is $v_i$, and its length is $l_i$, we use the following costs in Equation (3.2) for our pollution chain

$$w_i = f_i(t,p)l_i, \tag{4.2}$$

where we added the subscript $i$ to $f_i(t, p)$ to emphasise its dependence on the road segment $i$.

Once the transition matrix has been constructed, it is very easy to infer several quantities of potential interest to the designer of a road network. These include the Perron eigenvector, the mean first passage time (or rather mean first passage pollution) matrix and the Kemeny constant. The proposed model has the property that it is driven by

| Markov chain quantity | Green interpretation |
|---|---|
| Left-hand Perron Eigenvector | This vector has as many entries as the number of road segments. Each entry represents the long run fraction of emissions that a fleet of vehicles will emit along the corresponding road segment. It can be used as an indicator of pollution peaks. |
| Mean first passage emissions | This is a square matrix with as many rows as the number of the road segments. The entry $ij$ represents the expected quantity of emissions that a vehicle releases to go from $i$ to $j$. The average is with respect to all possible paths from $i$ to $j$. |
| Kemeny constant | This number is the average number of emissions released along a random route. It is an indicator of pollution in the entire network. |

Table 4.1.: Interpretation of Markov chain quantities in the emission framework

a unit of pollution. For ease of exposition, in the present discussion we use a generic unit of pollutant, while in the simulation examples we will compute CO, CO2, NOx and Benzene. We remind the reader that the difference between the models is in the choice of the parameters in Equation (4.1). Table 4.1 summarises the interpretation of the Markov chain quantities in the emission framework.

There are other quantities that can be computed within the proposed framework:

- Density of emissions along each road (g/km): They can be easily computed from Equation (4.1): it is only required to know the average fleet of vehicles and the average speed along the road.

- Emissions along each road (g): They can be easily computed by multiplying the density of emissions along a road by the length of the road.

- Total Emissions (g): It is sufficient to sum the emissions along each road for all the roads inside the area of interest (e.g. the urban network).

### 4.1.3. Simulations

In this section the proposed approach is described in detail through several simulations. The information required to build the Markov chain transition matrix is recovered from simulating traffic within an urban network using Sumo. To this end we consider the road network depicted in Figure 2.1. Unless stated otherwise all road segments have a

length of 500 meters.

**Simulation 1 :**    We create traffic inside this network in two different operating conditions. In the first case the number of cars in the road network is small, and cars are allowed to travel freely at maximum allowed speed; in the second case heavy congestion is artificially simulated by increasing the number of cars, so that actual speeds are slower than the maximum allowed by speed limits. The computation of emissions is performed according to Equation (4.1), using the data corresponding to EURO 4, Engine Capacity < 1400 cc petrol cars and minibuses with weight below 2.5 tonnes (Code R005/U005 from [Boulter et al., 2009]). The EURO emission standards specify allowed exhaust pollution limits for vehicles sold in the EU. Details can be found at `http://ec.europa.eu/environment/air/transport/road.htm` and the references therein.

Figure 4.1 shows that if the emissions are computed using speed-independent emission factors, then their distribution is not affected by different levels of congestion. This is clearly absurd. In fact, the stationary distribution of emissions is the same, both in the figure on the left (no congestion), and in the one on the right (congestion). On the other hand, the stationary distribution of cars along the roads (represented with the solid line) changes significantly in the case of congestion.

In Figure 4.2 the pollution factors are assumed to depend on velocity, and therefore the stationary distribution of emissions changes from the non-congested (left) to the congested scenario (right), and it remains consistent with the distribution of cars. In both Figures 4.1 and 4.2, and in both scenarios, the junction turning probabilities are the same. Clearly, it is not sensible that the distribution of pollutants does not change with traffic load. Therefore, this simple simulation suggests that speed-dependent emission factors should be used to obtain more realistic results. In Figure 4.2, we can also see that the manner in which pollutants are influenced by the traffic volume is not simply proportional to the number of vehicles. In particular, different pollutants have different densities.

**Simulation 2 :**    In a second simulation we aim at establishing optimal speed limits. By this we mean that we attempt to minimise emissions. However speed limits can not be arbitrarily changed. Existing speed limits usually give a number that is a multiple of 10 km/h, we are thus going to investigate these here. The results are shown in Table 4.2. Within each simulation we varied the speed limit uniformly over all streets from 20 to 120 km/h and calculated the corresponding Kemeny constant. As previously described, the Kemeny constant can be interpreted as an efficiency indicator, in terms of emissions, of the overall road network.
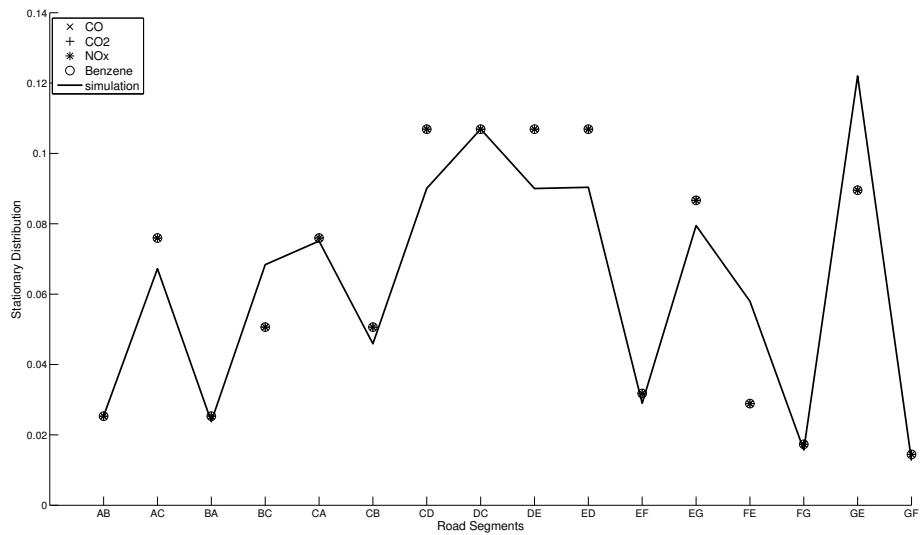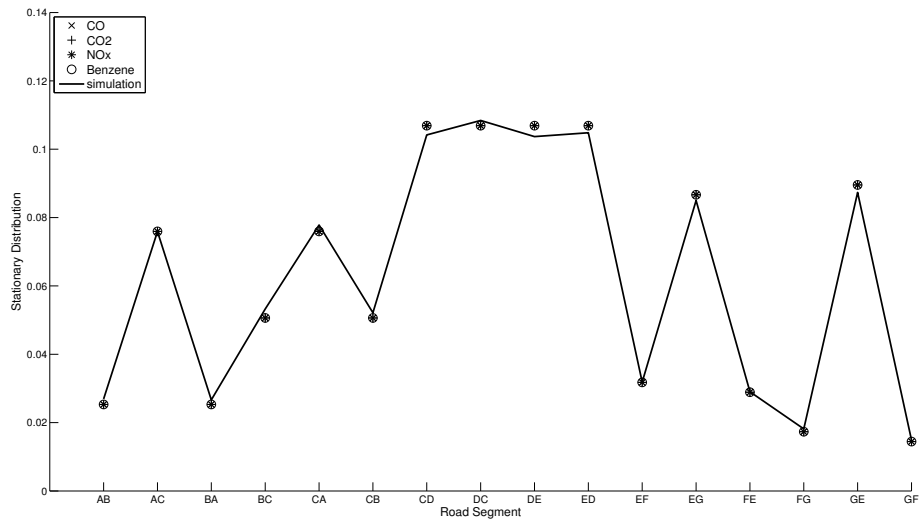
Figure 4.1.: Stationary distribution of cars (solid line) and all pollutants. Pollutant factors are constant with speed. No congestion in the upper, congestion in the lower figure. All emission factors coincide because of density normalisation (their sum is 1).
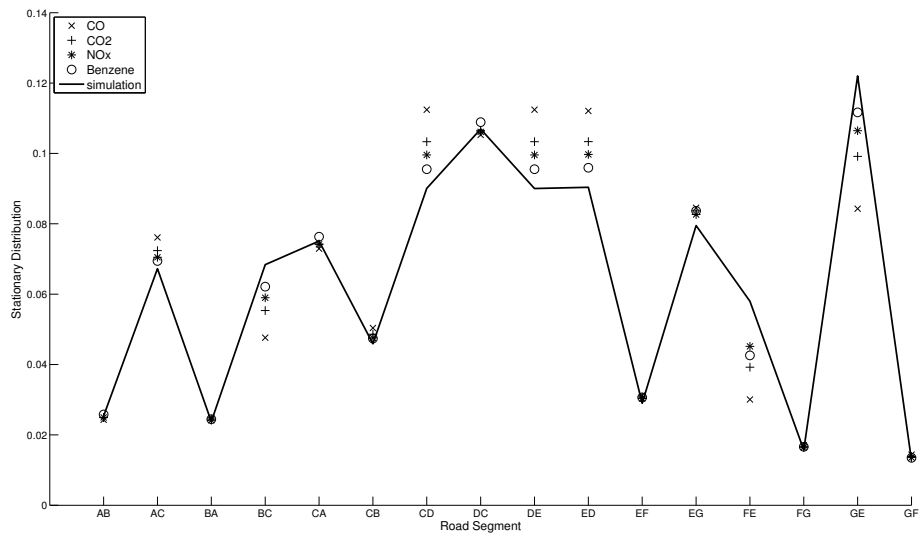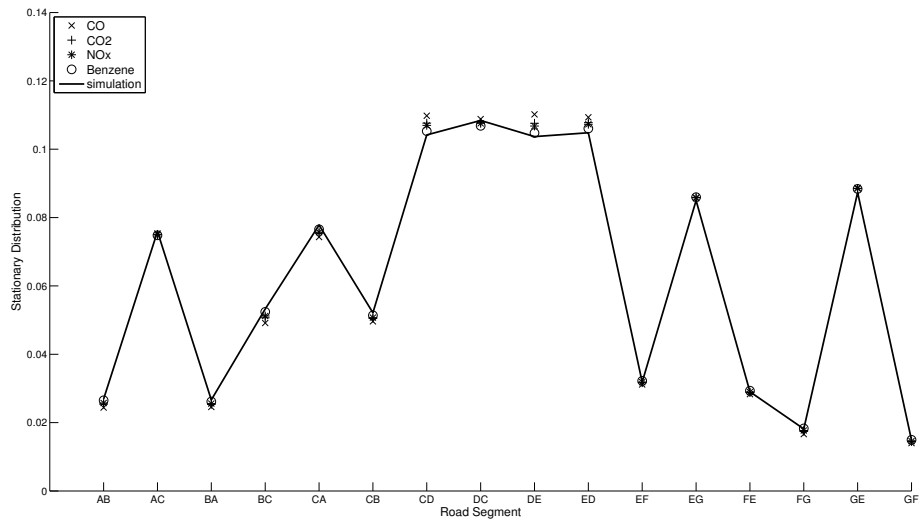
Figure 4.2.: Stationary distribution from simulation and model for time and all pollutants. Pollutant factors are dependent on speed. In this example, the pollutants density follows the car density both in the non-congested scenario (top) and in the congested one (bottom).

| Speed [km/h] | Time [sec] | CO [g] | CO2 [kg] | NOx [g] | Benzene [g] |
|---|---|---|---|---|---|
| 20 | 1304 | 2674 | 1241 | 395.9 | 2.89 |
| 40 | 631 | **2350** | 875 | 243.4 | 2.05 |
| 60 | 437 | 2830 | **812** | 207.8 | 1.44 |
| 80 | 352 | 3758 | 815 | 197.2 | 1.03 |
| 100 | 311 | 4888 | 833 | **194.6** | **0.91** |
| 120 | **291** | 6065 | 853 | 194.9 | 1.07 |

Table 4.2.: Kemeny Constants for Different Global Speed Limits.

Two lessons can be observed from Table 4.2.

1. Too low, and too high speed limits, both lead to high levels of pollution (although travel times are reduced with high speed limits, as expected)

2. Different pollutants have different corresponding optimal speed limits.

To better interpret the results shown in Table 4.2, Figure 4.3a shows how the Benzene and Figure 4.3b the CO emission factors depend on the average speed. The speeds corresponding to minimum Benzene emissions in Figure 4.3 and Table 4.2 do not coincide. This is caused by the fact that the average speed of a car in the simulation is not equal to the speed limit (e.g. due to acceleration and breaking while entering and leaving each road). In this example, high speed limits helped the car to accomplish an average speed closer to the optimal value for minimum emissions.

**Simulation 3 :** The above experiment was repeated with one minor change. Now we vary the speed limit at certain important points in the network, and observe the corresponding global change in pollution. Specifically, we vary the speed limit of the roads CD, DC, DE and ED from Figures 2.1 and 2.2. This corresponds to considering the network to be composed of two clusters of roads, one on the left and one on the right, connected through a bridging road. For this simulation we changed the length of the bridging roads from 500m to 5 km to reduce the gap between the actual average speed from the actual speed limit. We again vary the speed limit on the bridge from 20 to 120 km/h while keeping all other speed limits constant at 50 km/h. The resulting values for he Kemeny constant for all considered pollutants are shown in Table 4.3

Again we can see that both high and low speed limits on the bridge correspond to high overall pollution levels and that optimal speed limits differ for different pollutants.

**Remark :** The last two simulations correspond to a realistic road engineering problem, where the optimal speed limits for a subset of roads must be established to minimise emissions of interest.

**Simulation 4 :** As a further example, we now demonstrate the scalability of our approach by using a significantly larger road network. In particular, we use the network
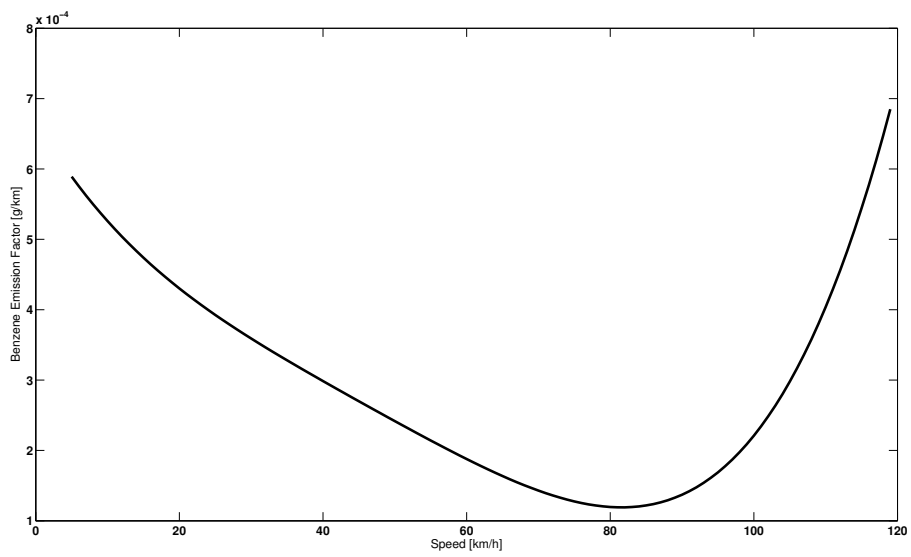
(a) Benzene emissions
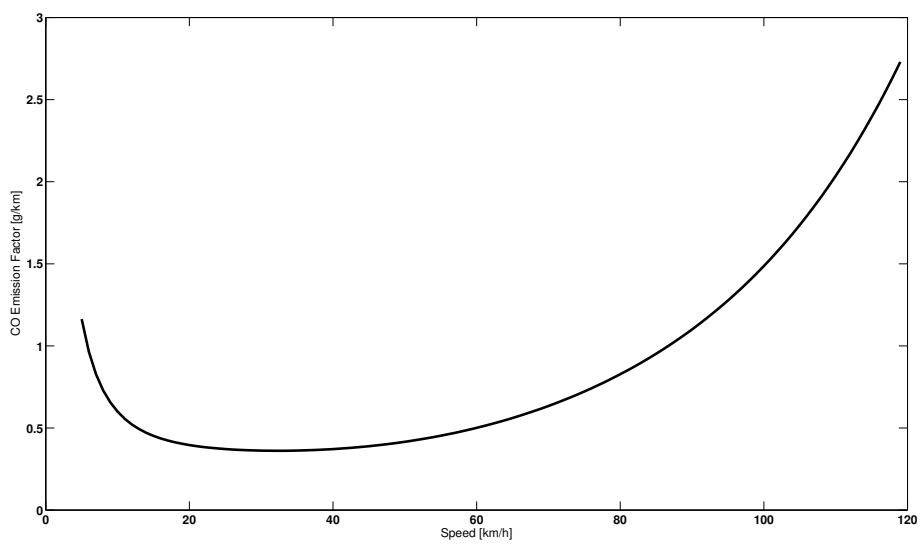


(b) CO emissions

Figure 4.3.: Dependence of emission factor on average speed for Benzene and CO (based on data from [Boulter et al., 2009]).

| Speed [km/h] | Time [sec] | CO [g] | CO2 [kg] | NOx [g] | Benzene [g] |
|---|---|---|---|---|---|
| 20 | 1929 | 5104 | 2109 | 637.4 | 4.76 |
| 40 | 1126 | **4746** | 1673 | 455.7 | 3.73 |
| 60 | 876 | 5606 | **1605** | 412.0 | 2.87 |
| 80 | 756 | 7831 | 1635 | 401.2 | **2.29** |
| 100 | 686 | 12119 | 1710 | **403.0** | 2.77 |
| 120 | **640** | 20045 | 1811 | 411.1 | 5.51 |

Table 4.3.: Kemeny Constants for Different Speed Limits on the Bridge.

depicted in Figure 4.4a, which was generated using the Sumo random network generation facilities.

Figure 4.4b shows the density of pollutants in the road network, together with the distribution of cars resulting from the Sumo simulation. While the road network of Figure 4.4b is composed of 618 road segments, only the subset of roads with ID between 250 and 300 was randomly chosen to be displayed in Figure 4.4b, to improve readability of the result; qualitatively the distribution is the same for any selection of streets. In Figure 4.4b we see again that the distribution of pollutants is heterogeneous, but it is related to the distribution of cars. This is in accordance with common knowledge that congestion should affect emissions.

## 4.2. Electric Vehicle Chain

In this section we show how energy consumption in electric vehicles can be integrated into the framework described above. By energy consumption we mean the charge removed from the battery when traversing a road segment. Following a model presented in [MacKay, 2008], in Section 4.2.1 we calculate this charge using basic formulae from physics and assuming constant losses within the car. These internal losses are to the most part due to the transmission, the motor, and the battery.

Electric vehicles differ from combustion driven vehicles in a number of aspects. Energy is scarce in such vehicles, and significant dissipation is related to road topology, to on-board systems (air conditioning, heating), and to other factors considered not important in ICE based vehicles. A very important feature is regenerative braking. Electric vehicles can recover some kinetic energy back to battery charge during the braking process. This gives rise to the possibility of negative costs on some road segments (energy transferred from the road to the vehicle). To apply our Markov model we shall thus use the extension discussed in Section 3.3. With this, a stationary set of expected energy consumptions may still be calculated (even though, counter-intuitively, some of these have negative values).
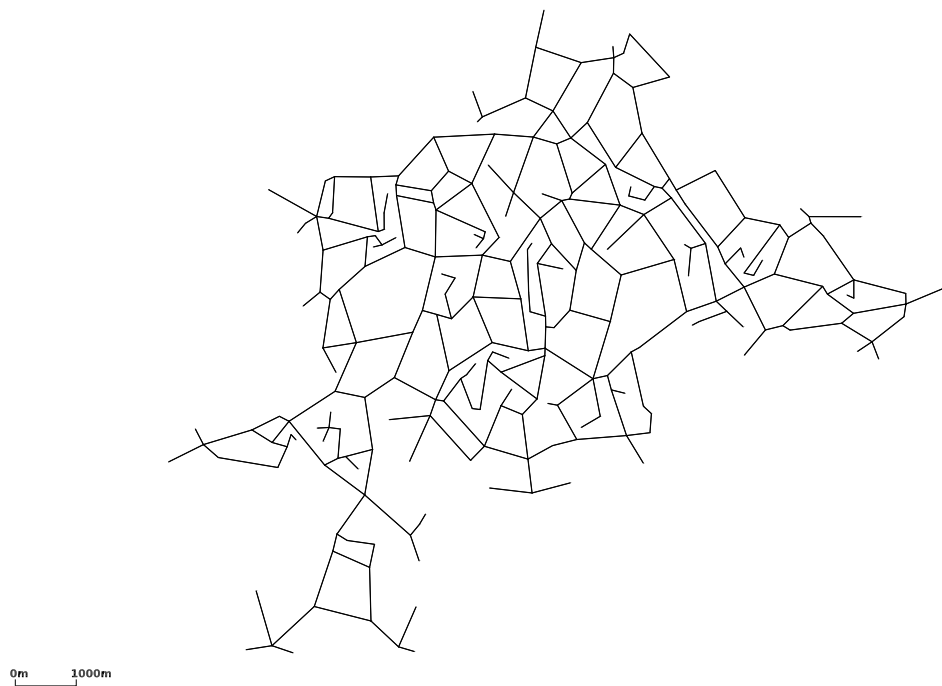
(a)
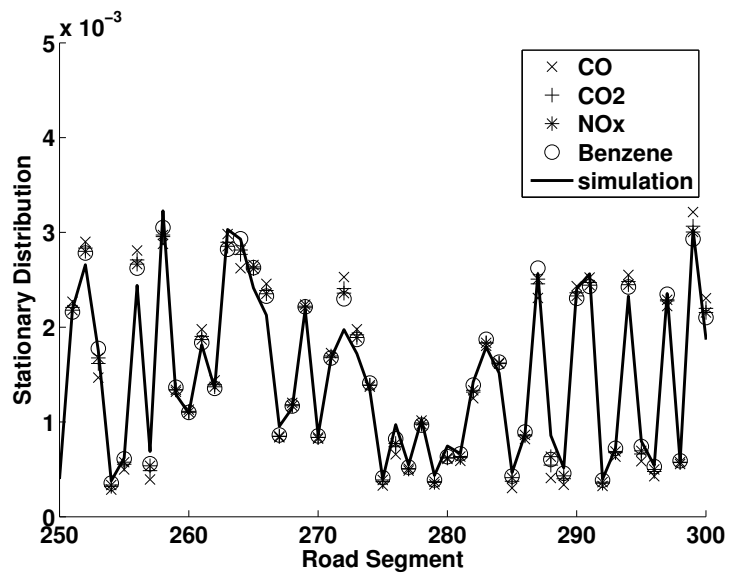


(b)

Figure 4.4.: a: A more complicated realistic road network; b: Stationary distribution of time and pollutants in the more complicated example with heavy congestion. Only a selection of streets is shown to improve readability of the figure.

### 4.2.1. Electric Consumption Model

We now present an elementary model of energy consumption in electric vehicles. Before proceeding, it is important to note that the details of this model are not important; any more accurate model can be embedded in our Markovian framework.

A first approach to modelling the energy consumption is to calculate an expected power needed for a vehicle per unit time, and to multiply this by the expected travel time. The average power (constant speed model) can be obtained as a function of: (1) average speed of the car and (2) the slope of the road. An improvement is to also consider the vehicle acceleration, as suggested in [MacKay, 2008]. This is the approach followed here.

We will assume a driving pattern, where cars are stationary at the beginning of each road segment and accelerate at a constant rate $a_1$ until they reach the cruising speed $v_{cruise}$ and then decelerate at a constant rate $a_2$ to reach the end of the road segment with zero velocity. The energy can be calculated as the sum of the energies in the acceleration phase, the cruising phase and the deceleration phase.

The following forces affect the vehicle:

- $F_{acc} = ma$ is the force needed to accelerate/decelerate the car,

- $F_{rol} = \mu_{rol} mg$ is the force needed to overcome the rolling resistance,

- $F_{ad} = \frac{1}{2}\rho A C_d v^2$ is the aerodynamic drag force,

- $F_{slope} = mg\sin(\phi)$ is the hill climbing force,

where $a$ is the vehicle's acceleration, $A$ is its frontal area, $m$ its mass and $v$ its speed, $\mu_{rol}, \rho, C_d$ and $g$ are constants and $\phi$ is the inclination of the road segment. For our calculations we assume a car weight of 1235 kg, gravitational acceleration of $g = 9.81$ m/s$^2$. Further reasonable parameter choices for a medium size electric vehicle are: $\rho = 1.2$ kg m$^3$, $\mu_{rol} = 0.01$, $C_d = 0.35$, $A = 1.6$ m$^2$ [MacKay, 2008].

We will assume the slope along a road segment is constant. In general (given an accurate velocity profile) we can calculate the energy expended as the integral of the force over the path. As discussed above we split this in three parts.

$$W = \underbrace{\int_0^{x_1} (F_{acc} + F_{rol} + F_{ad} + F_{slope}) \, \mathrm{dx}}_{W_1} + \underbrace{\int_{x_1}^{x_2} (F_{rol} + F_{ad} + F_{slope}) \, \mathrm{dx}}_{W_2} \qquad (4.3)$$

$$+ \underbrace{\int_{x_2}^{x_3} (F_{acc} + F_{rol} + F_{ad} + F_{slope}) \, \mathrm{dx}}_{W_3}, \qquad (4.4)$$

where $x_1$ is the distance after which the cruising speed is reached and $x_2$ is the distance after which the deceleration process is started and $x_3$ marks the end of the road segment. If we consider speed and distance as functions of time $t$ and denote them $v(t)$ and $s(t)$ respectively, then speed, time, distance are related by

$$v(t) = v(0) + at \tag{4.5}$$

$$s(t) = s(0) + v(0)t + \frac{1}{2}at^2. \tag{4.6}$$

We also assume a constant acceleration and deceleration rate of $a_1 = -a_2 = 3\frac{\mathrm{m}}{\mathrm{s}^2}$. If the length of the road, $x_3$, is large enough, we obtain

$$x_1 = \frac{1}{2}a_1 t_{acc} = \frac{1}{2}a_1 \left(\frac{v_{cruise}}{a_1}\right)^2 = \frac{1}{2}\frac{v_{cruise}^2}{a_1} \tag{4.7}$$

$$x_2 = x_3 + \frac{1}{2}\frac{v_{cruise}^2}{a_2}, \tag{4.8}$$

where $t_{acc}$ is the time it takes to accelerate the vehicle to cruising speed at rate $a_1$ and $v_{cruise}$ is the cruising speed. Thus

$$W_1 = \int_0^{x_1} \left(ma_1 + \mu_{rol}mg + \frac{1}{2}\rho AC_d v^2 + mg\sin(\phi)\right) \mathrm{dx} \tag{4.9}$$

$$= \frac{1}{2}mv_{cruise}^2 + \frac{1}{2}\frac{v_{cruise}^2}{a_1}mg\left(\mu_{rol} + \sin(\phi)\right) + \frac{1}{4}\rho AC_d \frac{v_{cruise}^4}{a_1}.$$

For the second integral we obtain

$$W_2 = \int_{x_1}^{x_2} \left(\mu_{rol}mg + \frac{1}{2}\rho AC_d v^2 + mg\sin(\phi)\right) \mathrm{dx} \tag{4.10}$$

$$= \left(x_3 + \frac{1}{2}\frac{v_{cruise}^2}{a_2} - \frac{1}{2}\frac{v_{cruise}^2}{a_1}\right)\left(mg\left(\mu_{rol} + \sin(\phi)\right) + \frac{1}{2}\rho AC_d v_{cruise}^2\right)$$

and for the third integral

$$W_3 = \int_{x_2}^{x_3} \left(ma_2 + \mu_{rol}mg + \frac{1}{2}\rho AC_d v^2 + mg\sin(\phi)\right) \mathrm{dx} \tag{4.11}$$

$$= (ma_2 + mg\left(\mu_{rol} + \sin(\phi)\right))\left(-\frac{1}{2}\frac{v_{cruise}^2}{a_2}\right) - \frac{1}{4}\rho AC_d \frac{v_{cruise}^4}{a_2}.$$

We further assume that the losses along the drive train are a constant 15% and during deceleration 50% of the energy can be saved by regenerative braking. An additional and highly important factor in the consumption of battery load is on-board equipment such as heating, light, air-conditioner, radio and many others that draw power at constant

rate over time. Their demand is much harder to model as it depends on the individual driver. However, their aggregate effect can not be neglected as will be exemplified in Section 4.2.4.

We now have a way to approximate energy requirements for traversing given road segments. The method is of course crude, but as in the case of the pollution chain it is possible to substitute any arbitrarily precise model.

### 4.2.2. Applications

We now give a number of applications to illustrate the utility of our model. The first application illustrates the effect of relaxing the flow conservation assumptions from [Crisostomi et al., 2011b] as done in Section 3.2. The second application concerns using the model for routing. This idea is already explored in the context of conventional vehicles in [Crisostomi et al., 2011c] and we now suggest how this basic idea can be used for routing of electric vehicles. In the third example we investigate how the Markov chain can be controlled both in an idealised context (using Lemma 3.1), and in a practical context using decentralised control.

**Comment :** To give some intuition that Lemma 3.1 still holds in the sense that $(Wx)^\top$ is still an eigenvector of $Q$ corresponding to the eigenvalue 1, consider again the network depicted in Figure 2.1. Simulating with Sumo and using the Markov chain model, where we use the expected energy consumption as weights, we obtain Figure 4.5. It depicts the vehicular density as measured from Sumo (solid line) and $(Wx)^\top$ (red dots), the extended notion of the stationary distribution of battery charge consumption/gain, as described above. For the simulation we lift junction B to be at the top of a small hill. Note that energy consumption increases drastically on the upward slopes.

### 4.2.3. Realistic Traffic Model: Relaxed Flow Conservation Assumptions

In Section 3.2 we discussed an extended Markov chain model with relaxed flow conservation assumptions. We used two approaches, one based on teleportation and one on an extra state. We now give an example of the approaches utility. For this purpose, we simulate the traffic network depicted in Figure 2.1 using Sumo by randomly choosing pairs of source/destination roads. To make sure that the local flow conservation assumption is not valid, sources are located mainly in the left triangle, while destination roads can mainly be found in the right triangle. We let cars leave the network as soon as they reach their destination; similarly new cars appear in the network starting from the source road at random times. Figure 4.6 shows that the teleportation method (whose stationary vector is, as previously stated, a scalar multiple of the appropriate subvector of the stationary distribution obtained by the extra state method) is in close accordance with the measurement of the simulations in Sumo. At the same time, it can be seen from Figure 4.6 that the previous method completely fails to describe the stationary distribution in this example where flows of cars are not conserved.
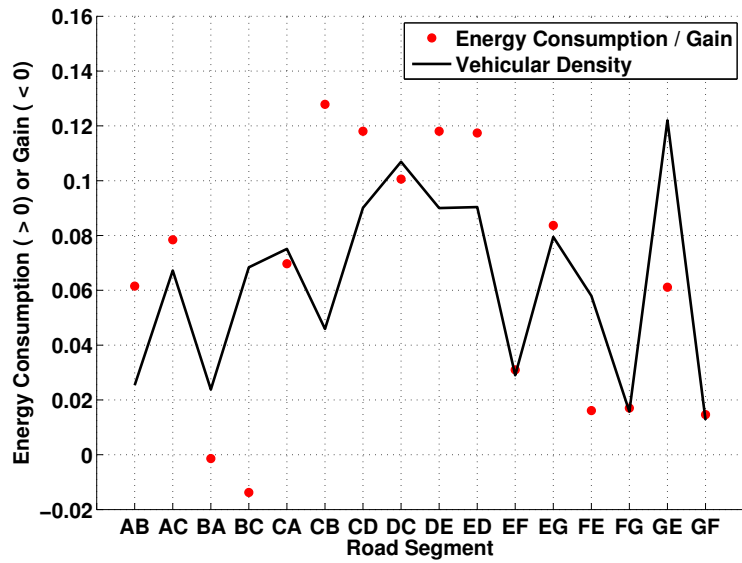
Figure 4.5.: Comparison of distribution of cars (line) and the average amount of energy used or gained on each road segment (dots). The energy consumption distribution is positive on road segments where on average energy is expended and negative where energy is gained.
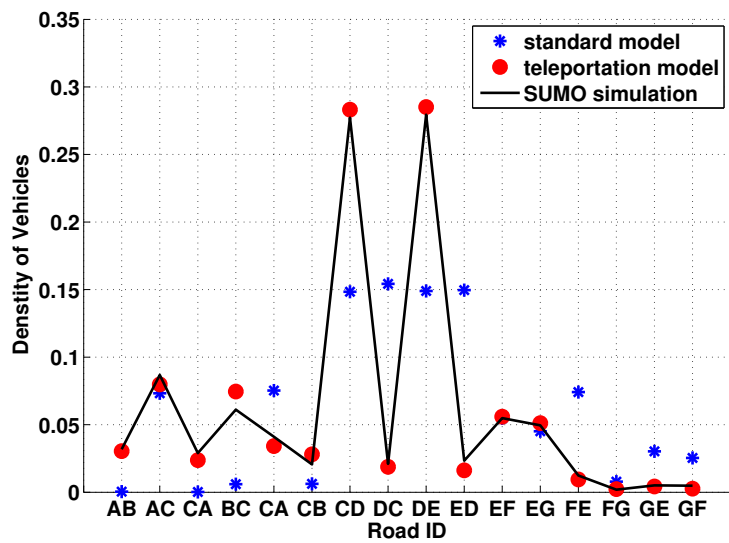


Figure 4.6.: Comparison of simulation results with the models with and without teleportation. The simulated traffic was generated such that the local flow conservation assumption is not valid.
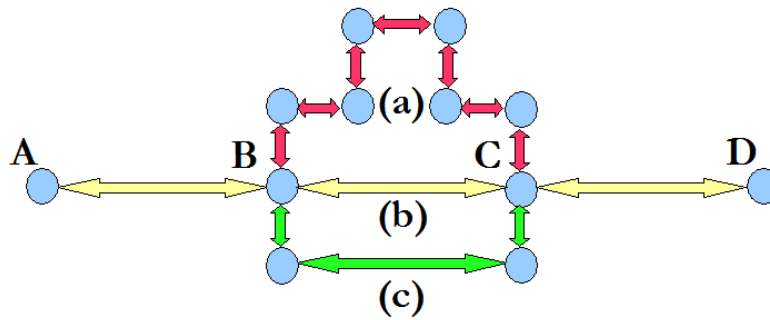
Figure 4.7.: There are three possible paths to go from B to C. Different routing strategies based on minimum distance, minimum energy consumption suggest different paths, also depending on the different environmental conditions (e.g., air conditioning switched on/off).

### 4.2.4. Optimal Routing

Even for a combustion driven vehicle it is usually hard to find an optimal route, as there are always trade-offs between travel time, travel distance, cost, pollution, etc. Taking the limited range and special requirements of electric vehicles into account certainly adds yet another dimension to this problem. However, at the beginning of our research navigational software and path finding algorithms were not taking this new dimension into account and were thus of limited use for the coming generation of vehicles. This has changed over the last couple of years, see for example [Artmeier et al., 2010; de Weerdt et al., 2013] and [Wade, 2012]. Our Markov chain model has proven a reliable first step to modelling electric vehicles. In addition, from this model it is possible to derive a number of routing strategies more tailored to the requirements of electric vehicles. We outline some of them in this section. We give details to two of the approaches. These routing strategies may help to promote the cause of electric and hybrid electric vehicles. All of the following routing strategies address distinct issues related to electric vehicles and traffic in general. Each strategy is associated with a different kind of route optimality.

**Minimum Energy Routing :** A very simple yet important routing strategy for electric vehicles can be obtained by minimising the distance to the destination not in terms of travel time or actual distance, but instead in terms of energy charge needed to finish the journey. Using the weights calculated in Section 4.2.1 combined with a classic graph search algorithm like [Dijkstra, 1959] it is possible to find minimum energy paths for each origin-destination pair.

Let us consider the road network depicted in Figure 4.7 as a case study for minimum energy routing. There are three routes connecting B to C. Assume that Route (a) is 1.8 km long and is travelled at an average speed of 50 km/h, Route (b) is 1 km long and

Table 4.4.: Required energy (in kWs) to travel from B to C in the road network depicted in Figure 4.7

| Auxiliary Power Demand | 500 W | 3500 W |
|---|---|---|
| Route (a) | 535 | 924 |
| Route (b) | 915 | 1050 |
| Route (c) | 695 | 884 |

travelled at 80 km/h and Route (c) is 1.4 km long and travelled at 80 km/h. Further assume that Route (b) is not flat, but rises to a small hill at its centre at a constant slope of 5 % in both directions. Table 4.4 reports the energy required to travel each route calculated with two different powers needed by auxiliary systems in the vehicle. According to [Farrington and Rugh, 2000] the power demand of accessory systems in an electric vehicle is largely affected by the usage. For instance, the power may vary from a minimum of 500 W to a maximum of 3500 W, for instance in winter, when the heating is running at full power.

In travelling from A to D, the driver has a choice to take one of the three routes (a),(b) or (c). The shortest route between B and C is Route (b) but we can see in Table 4.4 that both the longer routes perform better in minimising the required energy for travelling. This is due to the non-flatness of Route (b) and the high energy costs for electric vehicles on upward slopes. Additionally, we can see that Route (a) is more energy efficient then Route (c) if the auxiliary power demand is low, while this relationship is reversed if the auxiliary power demand is high. This implies that, especially in more complex routing tasks, an estimate of the power demand and its changes over time is necessary to achieve efficient and close to optimal solutions.

We believe that it is possible to assist users in making energy conscious route choices by providing energy road maps, i.e. road maps in which the displayed distance corresponds to energy consumption instead of travel distance.

**Minimum Popularity Routing :** In order to minimise the risk of traffic accidents and congestion along the vehicles path, in some situations it may be advisable to avoid the roads that most people use. Such popular roads could for example be close to shopping areas or train stations. The base Markov chain presented in Section 2.2 describes how often roads are taken, while ignoring travel times or energy consumption. Thus, it carries information about popularity and its stationary distribution can be used to find the minimum popularity route, again using conventional graph search algorithms.

**Mixed Minimum Energy/Popularity Routing :** Both minimum energy and minimum popularity routing may be too focused on one aspect of traveling. To circumvent this it is possible to use a weighted sum of the above two weights to find an optimal

path. In this sum we have an additional tuning parameter that can be set according to how important energy and popularity are considered to be. Similarly it is possible to use the derived electric Markov chain presented in Section 3.3, which contains information about both energy consumption and popularity and use its stationary distribution as weights for the graph search.

**Energy Optimal Risk-Averse Routing :** The limited range of electric vehicles can be managed by choosing appropriate routes, as discussed above, or switching to other modes of transportation for long distances. However, this might only mitigate but not completely eliminate "range anxiety" (i.e., the continual concern and fear of becoming stranded with a discharged battery in a limited range vehicle, [Tate et al., 2008]). For instance, unexpected events might require a longer path than expected, and the depletion of the battery before arrival. To avoid this, we propose the following scheme. For each road segment we calculate the impact on the network performance in the case that it is closed (as measured by the Kemeny constant) and use this as a weight for the graph search in order to obtain a route along which incidents have the least impact. This way of routing was first proposed in [Crisostomi et al., 2011b] in the context of conventional combustion driven vehicles, but its impact on electric vehicle routing is much more significant.

**Socially-Aware Optimal Routing :** All conventional routing algorithms and also the ones described above have a main characteristic in common: they are greedy. That is, they try to achieve the best result for an individual in a given situation. It is well known that this can lead to a suboptimal behaviour of the whole system and as exemplified by [Braess, 1968; Braess et al., 2005], this can even cause suboptimal performance for each individual driver. This problem can be solved by centralised routing, that is a central server computes routes for each vehicle and communicates them to the drivers. This approach is of course associated with a number of problems including the required infrastructure and cooperation of all drivers. It should be noted, however, that there is evidence in the literature, see [Avineri, 2009], that if drivers take the impact of their route choices on the overall network into account this can have a significant positive impact on the network performance, even if the weight put on the global effect in the decision making is small.

We believe that the ideal approach lies in randomised routing, where drivers with coinciding origins and destinations may take different routes. One way of achieving this is to affect the drivers route choices by giving incentives, such as different travel speeds or tax advantages. Our Markov chain framework is very close to this way of thinking and can thus play an important role in this routing strategy. We further claim that in our proposed setup the Kemeny constant can be utilised to measure the global efficiency of a given road network configuration and can thus be used to rate different routing strategies.

**Energy Optimal Minimum-Error Routing :** Not only external events such as accidents can cause unplanned route changes. This can also happen by human error. We

Table 4.5.: Expected required energy (in 100 kWs) to travel from OA to TO in the
network in Figure 4.8.

| Error Probability | 1% | 2% | 3% | 4% | 5% |
|---|---|---|---|---|---|
| Route 1 | 1.956 | 2.036 | 2.115 | 2.194 | 2.273 |
| Route 2 | 2.076 | 2.117 | 2.157 | 2.197 | 2.237 |
| **Error Probability** | **6%** | **7%** | **8%** | **9%** | **10%** |
| Route 1 | 2.351 | 2.429 | 2.506 | 2.582 | 2.659 |
| Route 2 | 2.277 | 2.317 | 2.357 | 2.397 | 2.437 |

propose that routes can be chosen in order to minimise the impact that a driving mistake
can have on energy demand. We now explore this routing strategy in more detail.

The Markov chain can be used to evaluate possible routes for drivers that are not confi-
dent or unfamiliar with the road network and are thus prone to make errors, for example
they might take a wrong turn. The following application of the Markovian framework
was first proposed in [Crisostomi et al., 2011c] and more details can be found there.
However, the ideas in [Crisostomi et al., 2011c] assume an even greater importance
when applied to electric vehicles because of their limited range. In such vehicles the
cost of a wrong turn may be high (especially if it leads to a one-way system) as sig-
nificant amounts battery power may be consumed in correcting the mistake. The basic
idea, therefore, is to find routing strategies where the cost of making a wrong turn is low.

To see how this may be achieved, consider the road network depicted in Figure 4.8.
There are two reasonable routes a vehicle might take from O to T, Route 1: OA-AB-
BC-CT and Route 2: OA-AT. Assume that the driver has a probability of making a
wrong turn at each junction of $\epsilon > 0$. The routes are evaluated by constructing an indi-
vidual transition matrix $Q$ for each route, where the probability of making the correct
turn at each junction is set to $1 - \epsilon$ and all other choices are given the same probability.
As a performance indicator we use the MFPE between origin and destination.

In Table 4.5 we show the MFPE from road segment OA to road segment TO of both
routes as a function of the error probability. We can clearly see that Route 1 performs
better when the probability of making an error is less than 4%, otherwise Route 2 is
superior. For this example we used a constant cruising speed of 50 km/h on each road
segment and travel times proportional to the length of each road.

As can be observed, the robust routing algorithm is particularly appealing in the case
of electric vehicles, as a driving mistake might lead to a longer path and an unexpected
higher energy consumption. The situation gets even worse if the driver ends up in a
zone without charging stations.

**Comment :** Note that routing for electric vehicles is particularly challenging. This fol-
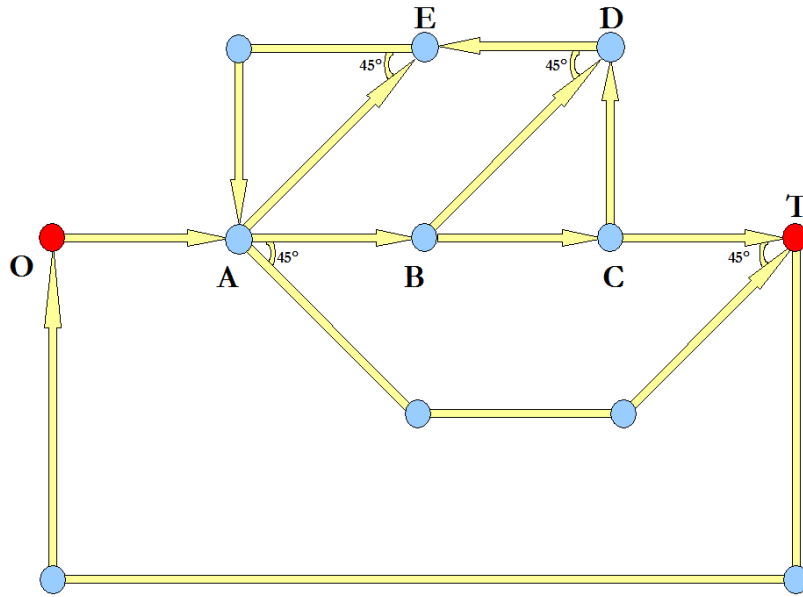
Figure 4.8.: Example of a road network with two possible routes. The route performance depends on the individual driver.

lows from the fact that energy is scarce in electric vehicles, and that significant amounts of energy can be consumed by on-board systems. This means that the time taken for a journey is important if the energy consumption per unit time is high. For example, in winter when the vehicle is being heated, or in summer when the air-conditioner is used, the duration of the journey becomes very important. In contrast, when the energy consumption is low, optimal routes depend mostly on road network topology and cruising speed. One fundamental observation about travelling is: The slower the cruising speed is, the less energy is consumed per unit of distance. Thus, given a choice between a low speed route and a high speed route of similar length, it is apparently better to take the low speed route. However, given the time dependence discussed above, this is not entirely true.

To illustrate this basic point, consider the traffic network depicted in Figure 4.9. We determine the stationary distribution of energy dissipation for two scenarios, corresponding to the air-conditioner being switched on and off. To achieve this, we assume a constant base load of 500 W in the first scenario and a load of 3500 W in the second scenario [Farrington and Rugh, 2000]. There are two routes to travel from A to D: the northern route (AE-ED) and the southern route (AB-BC-CD). Let speed limits of 50 km/h on the northern route and 30 km/h on the southern route be given. In this situation one notices two facts:
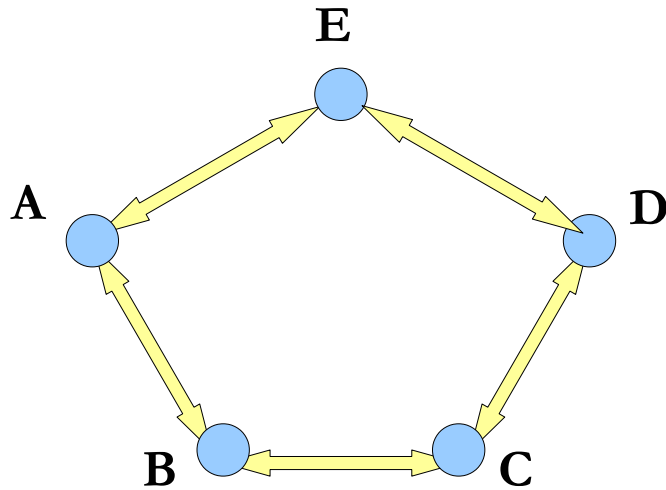
Figure 4.9.: An example road network.

- In travelling from A to D the minimum energy route with air-conditioner on is the northern route, while with air-conditioner off it is the southern route. This means that optimal route calculation is dependent on driver preferences. Issues like the weather have to be taken into account.

- The stationary distributions corresponding to the two scenarios are given in Figure 4.10. Note that the stationary properties of the network are very different in the two cases.

## 4.2.5. Traffic Load Control

Being able to control the stationary distribution of a road traffic network opens up a wide range of possibilities and different applications. For many goals corresponding optimal stationary distributions can be calculated and the system can then be driven towards them. For example, using the congestion chain it may be possible to equalise the travelling time on parallel routes to distribute the load within the network. Alternatively road network designers may be able to unburden some road segments in order to facilitate maintenance activities. In the context of the pollution chain, it may be interesting to be able to dissolve pollution peaks or even to minimise pollution along some roads or sub-networks in critical areas, for example near hospitals or schools. Here, we consider the Perron vector control in the context of the electric consumption modelling. There are many applications which benefit from the ability to control the stationary distribution. From a network designing point of view it may be interesting to be able to match high energy consumption with free charging point capacity. This can be used to equalise the
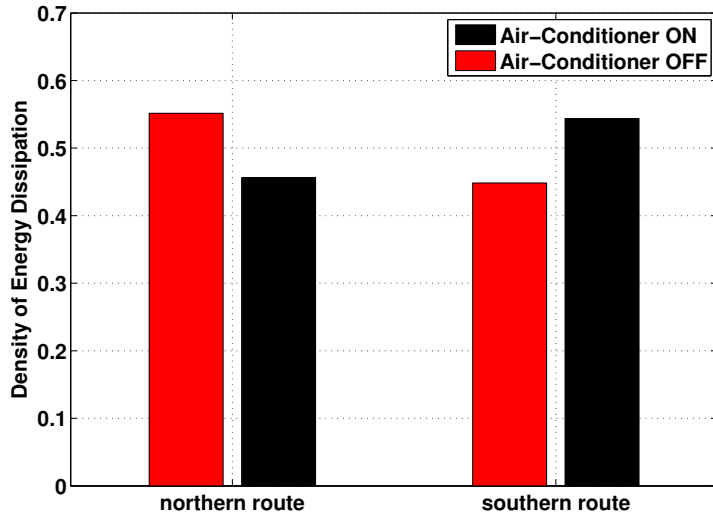
Figure 4.10.: Stationary distribution of energy dissipation in two cases.

demand on each charging stations. We now investigate in more detail how the Perron vector of our chain can be controlled.

### Theoretical Approach

Lemma 3.1 suggests a simple strategy to regulate the Perron eigenvector of the chain. Note that the Perron vector is determined by the Perron vector of the basic chain, and the diagonal scaling. The scaling in turn is determined by a host of factors, some of which can be controlled by the network designer. One of these is the speed limit. Recall that if the left Perron vector of $P$ is $x^\top$ and we wish through feedback to achieve a target left eigenvector $z^\top$ of $Q$ for some positive vector $z = (z_1, \ldots, z_n)^\top$ we set $w_i = \frac{z_i}{x_i}$ according to Lemma 3.1.

A useful application of the Perron vector control is now illustrated through an example which exploits again the road network of Figure 2.2. The dashed line in Figure 4.11 depicts the nominal density of cars in the case of uniform speed limits set to 50 $km/h$. Let us assume that the road engineer is interested in manipulating speed limits in order to achieve a uniform density of cars along all road segments (traffic balancing). Lemma 3.1 can be used to compute the "optimal" weights, and predict the "optimal" speed limits accordingly. New simulations in Sumo show that the desired objective is indeed obtained, as shown with a dotted line in Figure 4.11, although such "optimal" speed limits are unrealistic, as shown in Table 4.6, second line. A good trade-off is shown with solid line in Figure 4.11, where reasonable speed limits are used instead (e.g., they are all multiples of 10), as reported in the third line of Table 4.6.

Table 4.6.: Speed limits (in km/h) in the uncontrolled case, in the unrealistic optimal solution, and in the realistic balanced case

| Road Segment | AB | AC | BA | BC | CA | CB | CD | DC |
|---|---|---|---|---|---|---|---|---|
| Uncontrolled Case | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| Optimal (Unrealistic) Case | 15 | 44 | 15 | 29 | 44 | 29 | 117 | 122 |
| Realistic Solution | 30 | 50 | 30 | 40 | 50 | 40 | 100 | 100 |
| Road Segment | DE | ED | EF | EG | FE | FG | GE | GF |
| Uncontrolled Case | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| Optimal (Unrealistic) Case | 117 | 117 | 45 | 21 | 42 | 10 | 23 | 8 |
| Realistic Solution | 100 | 100 | 50 | 40 | 50 | 30 | 40 | 30 |

**Comment :** The proposed application assumes that drivers will not change their routes as a reaction to the new imposed speed limits; we also varied speed limits under the implicit assumption that this would correspond to proportionally adjusting average travel speeds. Despite these assumptions, Figure 4.11 shows that the road network traffic, as observed from Sumo simulations, is in accordance with the theory.

**Decentralised Traffic Load Control**

In the section above we controlled the Perron vector by choosing appropriate weights in the graph. This weight is however not necessarily something we can access directly. In a realistic scenario we may be able to change speed limits or traffic light sequencing to affect the weight, but the relation to the weight is not explicitly given. We use an algorithm from [Stanojević and Shorten, 2009a], which presents a decentralised way to reach an implicit consensus. We present two simulations using this algorithm. In the first example we take a large road network and one junction with three incoming streets and one outgoing street, as depicted in Figure 4.12. We adjust the speed limits on the three incoming streets such that the amount of energy required to traverse these streets is equalised. In the second example the large road network is considered as consisting of two main components that are connected by 4 bridging streets as depicted in Figure 4.13. In order for the algorithm to function, each road segment is assumed to be able to measure the density of cars travelling on the road, and to communicate this information to neighbouring roads.

As proposed in a different context in [Stanojević and Shorten, 2009a], our objective here is to determine speed limits that equalise the load across certain road segments. To this end let $\mathcal{I}$ be the set of road segments of interest, let us discretise time in steps $k = 1, 2, \ldots$ and let speed limits $v_i(0)$ at time $k = 0$ be given for all $i \in \mathcal{I}$. Let us denote by $\theta_i(k)$ the density of energy dissipated on each road segment at time $k$. We use the following iterative equation to update the speed limits
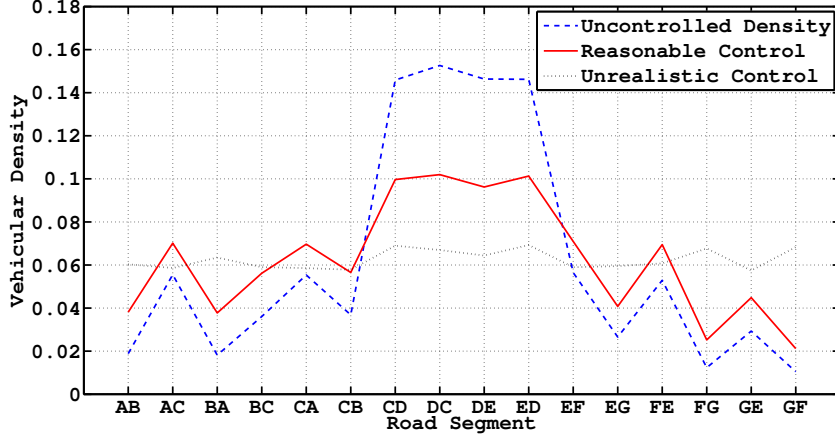
Figure 4.11.: Comparison of traffic density as a function of speed limits. Uniform speed limits lead to the unbalanced solution shown with a dashed line. A very good balance, in dotted line, is achieved using the unrealistic speed limits reported in Table 4.6, second line. A trade-off solution shown with solid line is obtained by using realistic speed limits as shown in Table 4.6, third line.

$$v_i(k+1) = v_i(k) + \eta \sum_j \left( \theta_j(k) - \theta_i(k) \right), \tag{4.12}$$

where $\eta$ is a positive parameter and the sum is taken over all road segments $j = 1, \ldots, n$ that road $i$ can communicate with.

The implicit consensus is conducted by alternating the following two steps:

1. Determine densities of interest (e.g., energy dissipation) at time $k$.

2. Update speed limits according to Equation (4.12).

For our two examples we use Sumo simulation runs and calculate the stationary distribution of the energy Markov chain. As the density of each road segment we use the corresponding entries of the stationary distribution. We perform an update of the speed limits according to Equation (4.12) and continue with the next instance of the simulation. In Figures 4.14a and 4.14b we give the relevant entries of the stationary distribution as a function of the number of simulation runs. Full details of the algorithm are given in [Stanojević and Shorten, 2009a]. At this point we just use the implicit consensus algorithm without proving that it achieves the balancing of the stationary vector.
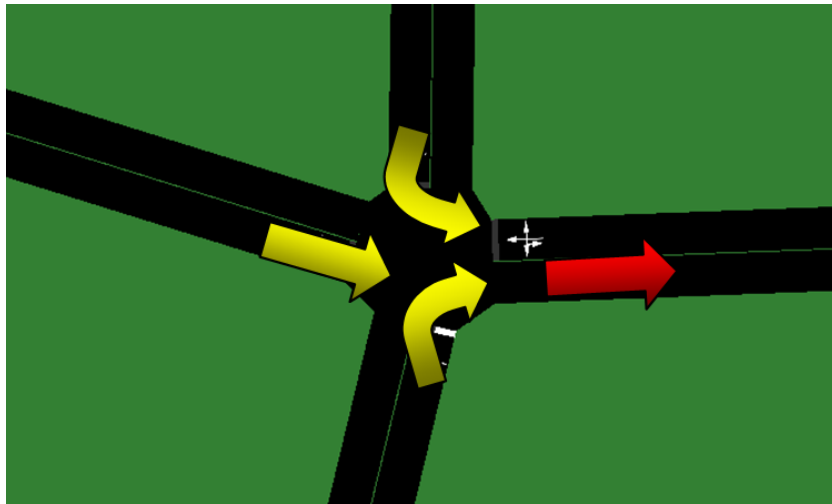
Figure 4.12.: Scenario with three streets with a common continuation, where we try to equalise the amount of energy expended on each street.



0m          1000m
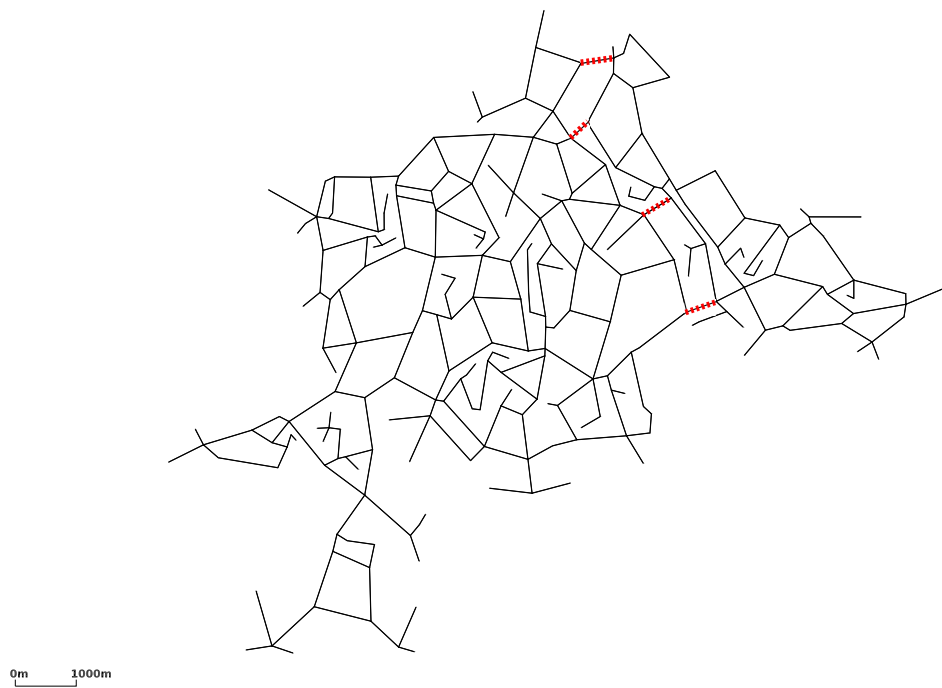
Figure 4.13.: Scenario of a big city and a suburb with 4 connecting streets, where we try to equalise the amount of energy expended on each street out of the suburb.
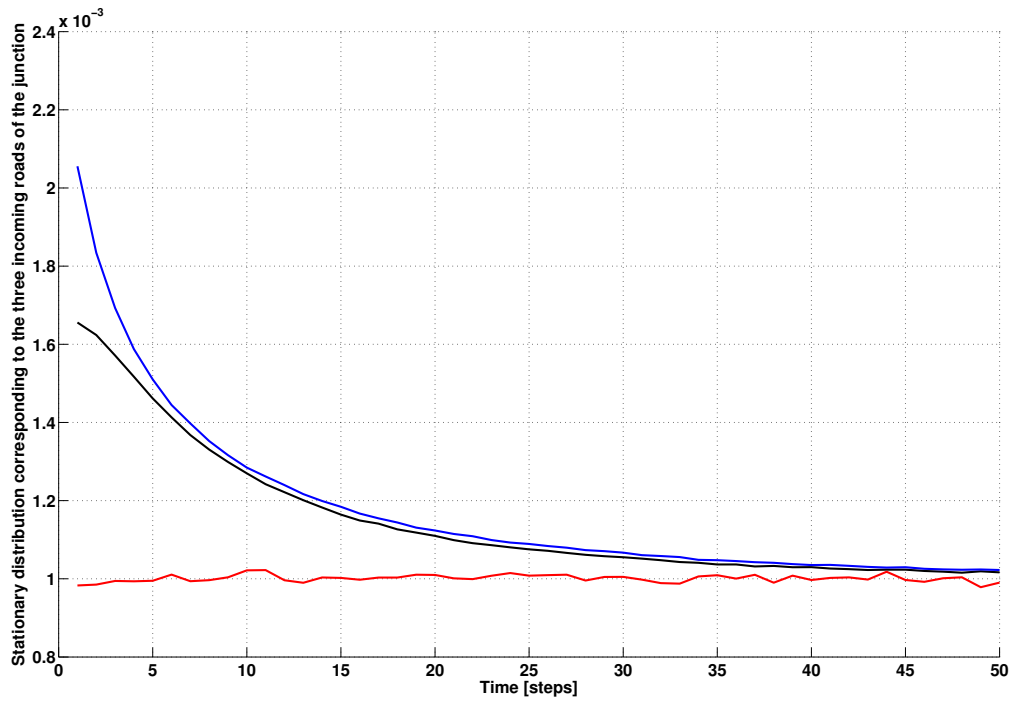
(a)
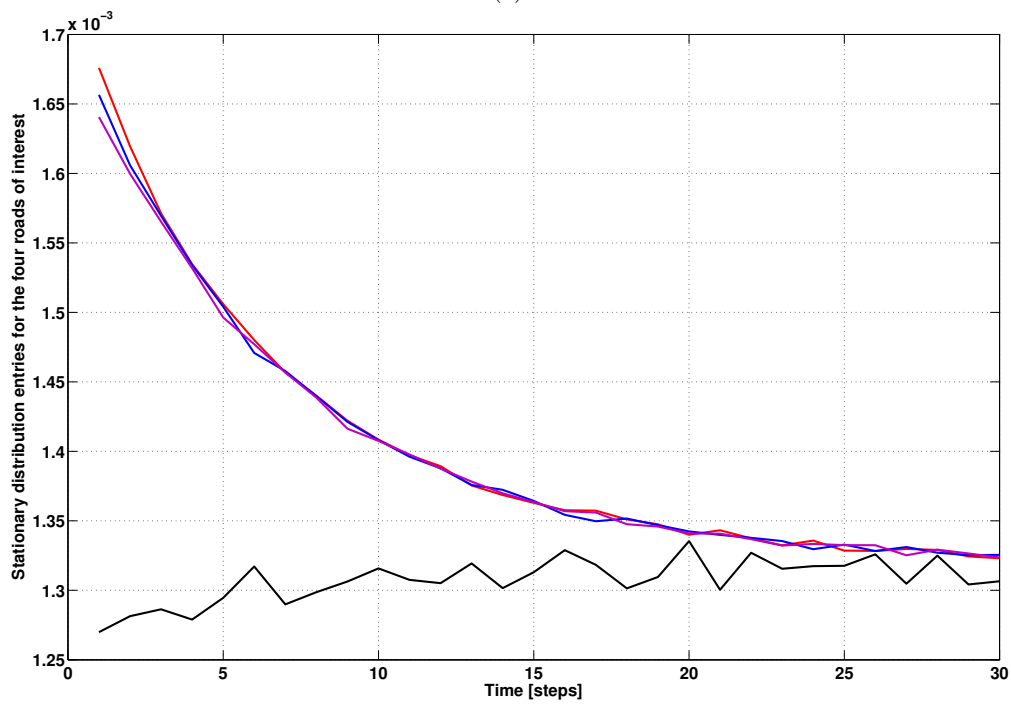


(b)

Figure 4.14.: The vertical axis shows the entries of the Perron vector. Each time step corresponds to one simulation instance. a: Convergence results for the single junction scenario; b: Convergence results for the suburb scenario.

72

# Part II.

# Microscopic Interventions and Cooperative Control

# 5. Cooperative Control of Pollution in Urban Scenarios - Setting

***Abstract :*** *In this chapter we discuss the typical constraints arising from control application in Intelligent Transportation systems. The material presented in this chapter is joint work with Florian Häusler, Mahsa Faizrahnemoon, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten and was published in [Häusler et al., 2013b].*

In this chapter we discuss the typical constraints arising in Intelligent Transport System (ITS) applications and discuss different levels of involvement of infrastructure units in the algorithm design.

In a number of problems in traffic systems, an available budget $C > 0$ has to be allocated to $n$ users. User $i$ obtains an allocation $D_i \in [0, C]$ such that $\sum_{i=1}^{n} D_i \leq C$. Each user may have a benefit, or utility, from the allocated share. This can often be described by a utility function $f_i : \mathbb{R}_+ \to \mathbb{R}$. Hence user $i$ obtains a benefit $f_i(D_i)$. Often in this context one wishes to allocate the available budget in a way that optimises the aggregate utility. This problem can be written as

$$\max_{\{D_1, \ldots, D_n\}} \sum_{i=1}^{n} f_i(D_i) \tag{5.1}$$

subject to the constraint that

$$\sum_{i=1}^{n} D_i \leq C. \tag{5.2}$$

Many problems in an ITS context map to this scenario. For example, we shall investigate a scenario where $C$ is a pollution budget assigned to a group of vehicles and $D_i$ is the pollution budget assigned to vehicle $i$. $f_i(D_i)$ is the benefit of the $i$th vehicle using the allocation $D_i$. A benefit expressed by $f_i$ in this context could for example come from higher speeds that are possible using more fuel. Other examples that map to this problem formulation include the following. $C$ could be the energy output of a charging station that is serving $n$ electric vehicles (EV) and $f_i(D_i)$ could be a function of the remaining charging time for the $i'$th vehicle. Alternatively, in the context of a

multi-modal transport application, the task could be to assign $C$ travellers to $n$ modes of transportation in a way that maximises traffic network efficiency.

It should be noted that it is also possible to cast much more complicated problems in this framework. For example, $C$ could be a vector. This corresponds to several seperate budgets allocated to the same group of users. For example $C$ can be a budget for $m \in \mathbb{N}$ pollutants. In this case $f_i$ maps from $\mathbb{R}_+^m$ to $\mathbb{R}$ and Equation (5.2) has to be understood component-wise.

A rich repertoire of techniques exist to deal with such problems. This is, in particular, true if the problem can be posed in a way such that the functions $f_i$ are concave, see for example [Boyd and Vandenberghe, 2004] for a general introduction. Many of these techniques have been successfully applied in other application domains [Stanojević and Shorten, 2009b]. In particular, in this context it should be mentioned that there is often strong similarity between traffic related challenges and problems in communication networks; streets are similar to data links, cars resemble data packets, and so on. Often opportunities arise to use well established tools from communication in the traffic context and many of the approaches discussed in this thesis are inspired by algorithms used in the Internet or the mobile cellular network.

Given reasonable assumptions on the functions $f_i$ a centralised optimisation is always possible, but may be challenging in the ITS context. It is possible to characterise algorithms that solve the above problem (and many others) by the level at which a central infrastructure is involved in them. Roughly speaking, three variants can be identified.

1. Fully centralised solutions: involve solving the optimisation problem defined in Equations (5.1) and (5.2) centrally. Clearly, the requirements to implement such a solution involve perfect knowledge of the $f_i(D_i)$, bi-directional communication between infrastructure and all vehicles, and significant computational ability of the centralised node. Advantages of the centralised approach include speed of convergence (instantaneous) and that there is no requirement for direct communication between neighbouring vehicles (vehicle-to-vehicle communication,V2V). Disadvantages of this approach include sensitivity to node failure, sensitivity of solutions to noise and uncertainty, and the feasibility of centralised approaches in large-scale environments.

2. Fully decentralised solutions: involve solving the optimisation problem defined in Equations (5.1) and (5.2) using local computation and communication only. A decentralised approach requires V2V communication capability. Advantages include that there is no need for a centralised optimisation step, the ability to deal with uncertainty in the $fi(Di)$, and the robustness with respect to node failures. Disadvantages include the very slow convergence rates especially in large-scale applications.

3. Hybrid solutions: involve using a small amount of centralised computation and

communication to speed up the convergence of decentralised solutions and to mitigate some of the constraints of ITS environments. In this part of the thesis we shall mainly investigate such hybrid approaches.

Solving problems that arise in an ITS context are very challenging due to a number of factors. We now summarise some of the basic issues.

1. Scale: ITS problems tend to be characterised by a large-scale level of participation. Optimisation problems over very large graphs are considered to be some of the most challenging in the theory of Operations Research [Hager et al., 1994].

2. Non-homogeneous levels of vehicle participation: ITS problems are also characterised by varying levels of vehicle participation as vehicles leave and join the system. Deployed algorithms should function satisfactorily irrespective of level of vehicle participation.

Algorithms deployed to solve problems like the one given in Equations (5.1) and (5.2) must also respect constraints imposed by the communication capabilities of the ITS components.

3. I2V broadcast: Here vehicles listen to a broadcast from a centralised node. This scenario, the infrastructure-to-vehicle broadcast, is probably the most common and well-established communication link in our context. The required technologies to implement I2V broadcasts are already available in most vehicles and a feature of broadcasting is, that the total communication effort is scalable and does not depend on the number of vehicles in the fleet.

4. I2V unicast: Direct communication from the central infrastructure (e.g., a fleet management back-office) to individual users using cellular networks is a feature, which could not be implemented easily today, due to network address translation (NAT) traversal problems. Available technologies for NAT traversal require extra infrastructure and communication protocols or control over network infrastructure.

5. V2I unicast: In contrast to unicast I2V, communicating into the other direction is technically easy to implement. A standard cellular communication network is required and the communication effort increases linearly with the number of vehicles.

6. Asymmetric and symmetric V2V: Vehicle-to-vehicle communication using IEEE 802.11p technology is currently under standardisation but not yet deployed. V2V communication may be classified as being symmetric and asymmetric, where symmetry refers to a situation where vehicle $i$ communicates with vehicle $j$ if and only if $j$ can communicate with $i$.

A further issue that arises in ITS applications is the ability to solve large-scale optimisation algorithms. In some situations it may not be possible or desirable to solve optimisation problems in a centralised manner.

7. Computation: A basic question concerns as to whether the central infrastructure is capable of solving large-scale optimisation problems and communicating bidirectionally to mobile vehicles at relatively high sampling rates.

Finally, a number of very important performance constraints also arise in ITS applications. These include the following:

8. Robustness to failure and uncertainty: A basic requirement is that the optimisation algorithm should be robust with respect to the effect of failures in the system. Also, in many situations, utility functions are not known exactly.

9. Speed of convergence: A further requirement, in the context of decentralised algorithms, is that algorithms should converge to the optimal solution reasonably quickly.

In the following two chapters we introduce some novel approaches to solve a problem in ITS that can be formulated in the form of Equations (5.1) and (5.2). Namely, we want to allocate emission rights in a fleet of hybrid vehicles. At first, in Chapter 6, we shall present a simple algorithm with minimal communication and computation requirements, for example no V2V or V2I is required. This algorithm achieves the constraint (5.2) and is thus to be seen as a control algorithm. In Chapter 7, we shall then propose two algorithms that use V2V and are able to solve the optimisation problem.

# 6. Cooperative Control of Pollution in Urban Scenarios - Control

***Abstract :*** *In this chapter we present a new approach to regulate traffic related pollution in urban environments by utilising hybrid vehicles. We assume that these vehicles have two operating modes, a fully electric mode (EV mode) and a mode that allows usage of the internal combustion engine (ICE mode). To do this we orchestrate the way each vehicle in a large fleet combines its two engines based on simple communication signals from a central infrastructure. The results presented in this chapter are joint work with Robert Shorten, Emanuele Crisostomi, Astrid Bergman, Florian Häusler, Thomas Hecker, and Ilja Radusch and were published in [Schlote et al., 2012c, 2013b].*

In this chapter we present a new approach to regulate traffic related pollution in urban environments by utilising hybrid vehicles. Inspired by well established strategies in communication network congestion control, most notably the Random Early Detection active queue management scheme [Floyd and Jacobson, 1993], we orchestrate the way each vehicle in a large fleet of hybrid electric vehicles combines its two engines based on simple communication signals from a central infrastructure. The primary goal is to regulate emissions and we discuss a number of control strategies to achieve this. The efficacy of our approach is exemplified both by the construction of a proof-of-concept vehicle, and by extensive simulations, and verified by mathematical analysis.

This chapter is organised as follows: In Section 6.1 we present some important notions from networking research and describe how these translate in a seamless fashion to fleets of hybrid vehicles. We then propose our control approach in Section 6.2. In Section 6.3 we describe our hybrid car and the modifications we made to it as well as the smartphone application we used. The pollution model we used is given in Section 6.4. Section 6.5 outlines a number of employed control algorithms. Simulation results are presented in Section 6.6.

## 6.1. An Analogy with Internet Congestion Control

The principal objective of the work presented in this chapter is to use the opportunity afforded by new vehicle communication technology, and by new vehicle types, to develop effective techniques to regulate pollution generated by road vehicles. Essentially we wish to place a feedback loop around a group of vehicles, and use this loop to control the group's emissions. Our basic requirements are as follows.

(i) *Aggregate pollution levels should not exceed pre-described levels :* Basically the objective here is to decouple pollution in a certain area (or group of vehicles) from the number of vehicles in that area as depicted in Figure 6.1.

(ii) *Best effort behaviour :* The network traffic should exhibit a best effort type behaviour. Vehicles in a geographic area should adjust their behaviour so as to share their the allowed pollution level irrespective of the number of participating vehicles. They should also respond to non-vehicle pollution generation by becoming cleaner if pollution levels rise.

(iii) *Fairness :* Ideally, we would like to see a notion of fairness in the network. As is always the case in dealing with fairness, there are several notions of fairness, potentially leading to very different situations.

Each participating vehicle should strive to achieve these objectives by using as little battery power as possible. Electric power is valuable as the on-board storage capacity is limited, and it takes a long time to recharge the batteries. Further, zero-emission power may give these vehicles access to certain otherwise restricted areas of the city, for example there may be zero-emission zones close to particularly sensitive areas, such as hospitals or kindergartens. Another important point is that we do not require zero pollution in a geographic area on a city scale. This does not seem to be necessary as low levels of pollution seem to be tolerable for humans. The validity of this claim as well as the precise value of the safe levels is not part of our study, however this is reflected by many countries specifying legal safe limits for several pollutants. Also a zero emission city imposes a high burden on participating vehicles, as they in effect have to function as electric vehicles or cannot be used inside the city. Rather, we impose a pollution budget and insist, that the aggregate level of pollution, among all participating vehicles, never exceeds this level.

Items (i)-(iii) closely resemble the requirements of resource allocation algorithms found in networking applications (such as the Internet). The notion of aggregate behaviour, best effort behaviour, and fairness, resemble in particular elastic traffic control in the Internet as depicted in Figure 6.2, for example the congestion avoidance algorithm of the Transmission Control Protocol (TCP) has these characteristics [Jacobson, 1988]. Essentially, we are viewing the pollution control problem as a resource allocation problem, where a certain amount, or budget, of pollution is shared among competing vehicles. In our situation the objective of each vehicle is to be as polluting as possible (use as little
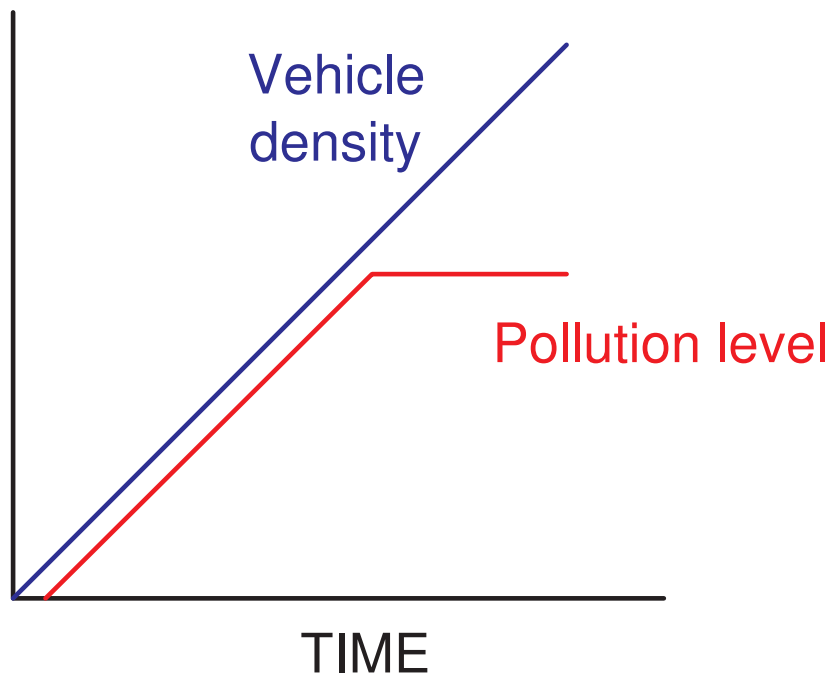
Figure 6.1.: Typically pollution levels follow the number of vehicles. We wish to decouple pollution from vehicle density.
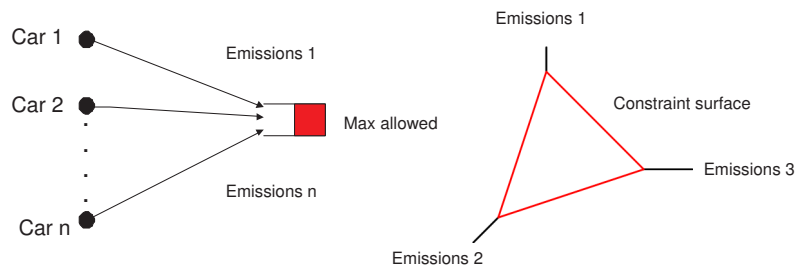


Figure 6.2.: Pollution as a resource allocation problem

battery power as possible) while at the same time being fair to other vehicles, and such that the aggregate level is close to the allowed threshold. To do this, we require vehicles to change their behaviour to respond to other vehicles joining and leaving the network (best effort behaviour).

The realisation that the pollution control problem can be recast in a resource allocation framework is fortunate. Network resource allocation problems are at a mature stage, and typically involve a large-scale decentralised optimisation problem. This latter feature makes algorithms from this community attractive in a transportation context. Relevant ideas in this direction include the Kelly framework [Kelly, 1999], the Random Early Detection active queue management (RED) [Floyd and Jacobson, 1993; Srikant, 2003], and the additive increase multiplicative decrease (AIMD) congestion control algorithm implemented in TCP [Jacobson, 1988], to name but a few. In the next section we illustrate how some of these ideas, together with new vehicle types, can be used to great effect to manage aggregate vehicle emissions.

## 6.2. Pollution Control

We wish to use inter-vehicle cooperation to regulate pollution in a non-invasive manner, i.e. we want vehicles to work together without inconveniencing the individual driver. Our approach to achieve this is to use hybrid vehicles. The basic idea is to orchestrate and coordinate switching between drive modes in a fleet of hybrid vehicles so as to achieve regulated pollution levels. Feedback is used to adjust the level of coordination to achieve the desired level of regulation. The control loop is depicted in Figure 6.3. In our work, the infrastructure uses algorithms of the form of Algorithm 6.1, where $g$ is a function of present and past emission levels and depends on the chosen control algorithm.

---
**Algorithm 6.1** Central probability control
---
$E(k) \leftarrow$ Measurement or estimate of current emission levels
$p \leftarrow g(E(k), E(k-1), ...)$
Broadcast $p$ to all vehicles.

---

---
**Algorithm 6.2** Vehicle Mode Choice
---
$p \leftarrow$ broadcasted probability value
Use ICE with probability $p$ and EV mode otherwise.

---

To realise this objective we assume the availability of a context aware hybrid vehicle, whose switching into fully electric mode can be made dependent on the location of the vehicle as well as in response to an external signal. Our approach here is to allow vehicles to randomly select their mode of operation in response to these signals (in contrast to the approach in [Knorn et al., 2011b,a]) according to Algorithm 6.2. Here $p$ is the

probability that a hybrid electric vehicle engages its combustion engine. Another important point that has to be considered is the availability of accurate and real time pollution estimates or measurements. These can be obtained in a number of ways: by roadside infrastructure measurements, by communication of on-board measurements of emissions of the vehicles themselves, or using geosensor networks. This is not the focus of our work, but the technology is available, see for example [Norris and Chormaic, 2002; Jeong et al., 2014]. Throughout this chapter we will assume that the required pollution data is available to us.

**Comment :** Before proceeding, it is worth noting that the objective in this chapter is to illustrate what is possible when vehicles cooperate with each other. As we shall see, our approach allows regulatory authorities to enforce different desired policy objectives depending on context. For example, road usage based on carbon emissions could easily be implemented. Contrast this to the situation today where all road users have equal access to roads no matter how polluting the vehicle is. To enforce ideas such as ours, regulatory actions would be necessary or some form of incentivisation through pricing. However, such regulation already exists in many countries. For example, in Germany, green zones [Sadler Consultants Ltd., 2013] are only accessible to certain vehicle types, and London already has extensive congestion charging [Transport for London, 2013]. However both these approaches are rigid, put many constraints on the drivers, ignore the actual levels of pollution and may encourage additional trips by giving drivers with access to the road network a misleading sense of being environmentally friendly. This behaviour may counteract the positive effect of banning dirty vehicles. We present an approach that is flexible, puts very little constraint on the drivers, and regulates the pollution. We do this by automating vehicle mode choice based on entry to the green zone, with an additional feedback loop to ensure an elastic and best-effort type behaviour of the vehicles.

Clearly, our approach does not require fleets of hybrid vehicles. Many modern vehicles allow the driver to choose different operational settings depending on whether the driver wants to be environmentally friendly or likes to have the full acceleration and maximum speed of his vehicle available. If the switching between the modes is made available as a control variable, the same ideas work. However hybrid vehicles and their ability to switch off the combustion engine completely will make the approach work better.

We now describe one such vehicle, which we have constructed for demonstration purposes, and then proceed to outline our basic modelling assumptions, and control strategies.

## 6.3. The Demonstration Vehicle

To demonstrate the application of our approach we equipped a 2008 model Toyota Prius with additional technology. The car is depicted in Figure 6.4.
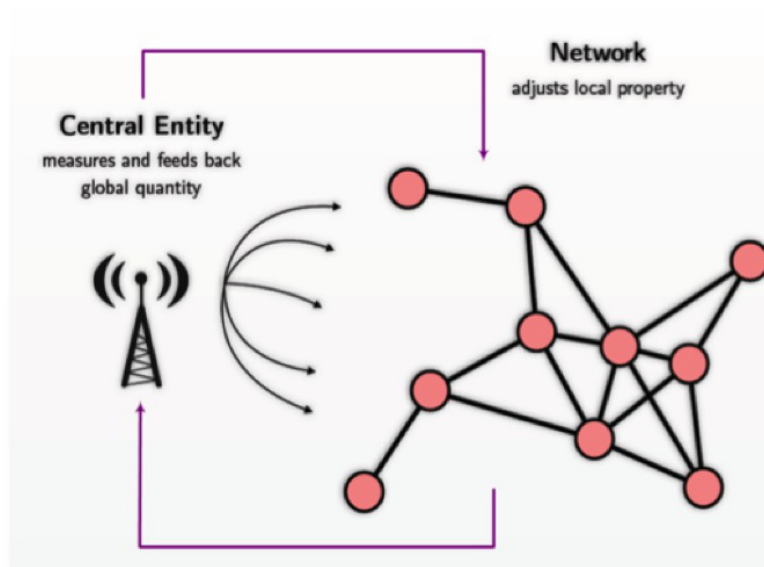
Figure 6.3.: Feedback loop



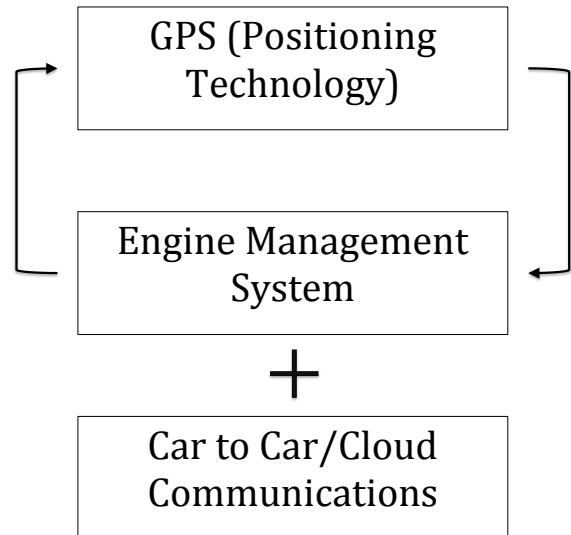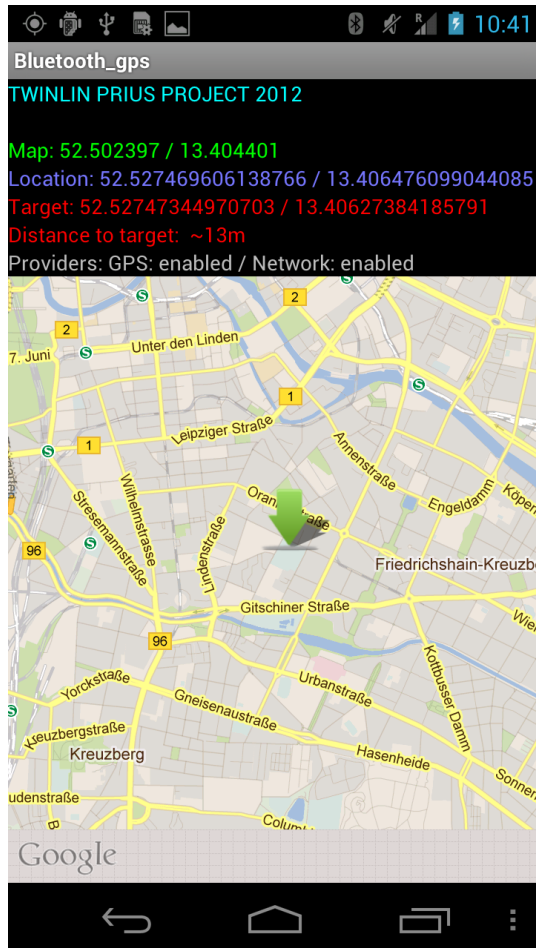Figure 6.4.: The modified Prius.

Figure 6.5.: Left: Android based App; Right: Communication flow chart. Information from top to centre block includes: Position; Topological information; Speed constraints; Environmental zones; Time constrained access and Road surface information. Information from centre to top block includes: Fuel levels; Type of engine; Wishes of driver and Congestion charging information.

The Toyota Prius is a power split hybrid vehicle [Ehsani et al., 2010]. The transaxle of the Prius allows the car to be powered by the ICE engine alone, the battery, or using a combination of both. It is this degree of freedom that we exploit to control aggregate emissions using a fleet of these vehicles. In principle by controlling the mix of electric and ICE engine torques, one can control the vehicle consumption and emissions levels more accurately as was done in [Knorn et al., 2011b,a]. For the purpose of this work we use a simpler approach that is significantly easier to implement. In particular, we only use the ability to switch between EV mode and ICE mode. Importantly, in this context, the Prius is equipped with an "EV-mode button" allowing the vehicle to be manually switched to pure electric driving mode if the battery level permits. This mode is automatically turned off if the button is pushed again; when the vehicle speed rises above 45 km/h; or when more torque is required than the electric motor can deliver (or when the battery level is too low). Hence an implementation of our proposed approach can be achieved without interfering with the vehicles on-board communication and computation system. This also allows a platform independent implementation using different vehicle models from different manufacturers given that they provide a similar EV mode button. Our basic modification to the vehicle is to automate this mode change by building appropriate hardware to override the manual function.

Thus, for the purpose of this project we have made some important modifications to the basic vehicle to make it behave as a context aware vehicle. First, we automated the switching of the vehicle from ICE to EV mode by adapting the "EV-mode button" hardware in the vehicle. For this purpose a dedicated Bluetooth controlled mechanical interface was constructed to override the manual EV button based on signals from a smartphone. The switching is based on GPS location, external context information, and on-board signals such as speed and battery level. Second, special purpose hardware was constructed to permit communication between a smartphone and the CAN-bus. However, there are also commercial tools available to do this. The Prius provides a CAN (Controller Area Network) access on the vehicles diagnostics OBDII-interface (On-Board Diagnostics II). This CAN interface carries much in-vehicle data such as energy consumption, rotational speed, engine mode and pollution. Our special hardware module CAN-Gateway acts as a gateway between this CAN interface and the smartphone. The module is directly connected to the CAN and to the Smartphone via Bluetooth. Using special commands on the Bluetooth channel, the smartphone subscribes to specific in-vehicle data. The CAN-Gateway reads this data from the vehicles CAN interface, filters the requested data and delivers them to the smartphone. Due to safety reasons, the module works in a read-only mode as CAN networks are very sensitive to adaptions. Communication to other vehicles, to GPS, and to a cloud server is also realised using a smartphone device. For our work a SAMSUNG Galaxy NEXUS device and HTC EVO-3D were used, both running Android 4.0. To control the driving mode, the software connects via Bluetooth to a mechanical switch to switch driving mode between the EV mode and the ICE mode. Thus, the smartphone has access to in-vehicle bus data such as driving mode, battery level, and pollution levels via a central server. A screenshot of our smartphone application is depicted in Figure 6.5. The Android application thus

Table 6.1.: CO Emission factors for petrol cars with engine capacity between 1400 and 2000 cc (from [Boulter et al., 2009], pag. 97 ). For all reported emission classes we have $d = e = f = g = 0$ and $k = 1$.

| Emission standard | Coefficients | | | Valid speed range (Km/h) | |
|---|---|---|---|---|---|
| | a | b | c | Minimum | Maximum |
| Euro 1 | 28.26 | 1.54 | $9.77 \cdot 10^{-4}$ | 5 | 120 |
| Euro 2 | 56.75 | -2.26 | $2.67 \cdot 10^{-2}$ | 5 | 140 |
| Euro 3 | 35.04 | -1.8 | $2.49 \cdot 10^{-2}$ | 5 | 140 |
| Euro 4 | 22.63 | -0.69 | $1.44 \cdot 10^{-2}$ | 5 | 140 |

allows the vehicle to interact with its environment in a very smart manner. Importantly, it allows controlled delivery of emissions and pollution into the city environment (be it noise pollution, or more directly harmful pollutants), allowing us to control where and when these emissions are delivered into the environment. A movie demonstrating operation of the vehicle can be found at *http://www.hamilton.ie/aschlote/twinLIN.mov*.

## 6.4. Pollution Modelling and Simulation

To investigate the cooperative behaviour of a fleet of hybrid vehicles, we use a modified version of the traffic simulator Sumo. Sumo is used to simulate traffic for a fleet of hybrid vehicles traversing a specified geographic location. As before in Section 4.1.2, average speed emission models are used to estimate specific pollutants of each vehicle [Gkatzoflias et al., 2007; Boulter et al., 2009]. The average speed approach is described in detail in [Gkatzoflias et al., 2007]. To ease exposition of this chapter, we repeat how the emission factor $f(t, p)$ is computed according to this model.

$$f(t,p) = \frac{k}{v} \left( a + bv + cv^3 + dv^3 + ev^4 + fv^5 + gv^6 \right), \tag{6.1}$$

where $t$ denotes the type of vehicle (and depends on fuel, emission standard, category of vehicle, engine power), $p$ denotes the pollutant of interest (e.g. $CO$, $CO_2$, $NO_x$, Benzene), $v$ denotes the average speed of the vehicle, and the parameters $a, b, c, d, e, f, g$ and $k$ depend on both the type of vehicle and the pollutant $p$ under consideration. For the purpose of this work, the values of the parameters are taken from Appendix D, in [Boulter et al., 2009]. For quick reference, we report in Table 6.1 the emission factors for $CO$ which have been used later in Section 6.6.

In Equation (6.1) it is assumed that speeds are measured in $km/h$ and emission factors in $g/km$. As before, we use the average speed model for simplicity. More realistic and accurate models can easily be embedded into the simulation environment without changing the qualitative features of the simulation (or the analysis).

## 6.5. Control Algorithms

We investigate three distinct control methods for regulating pollution in a geographic area by orchestrating the switching of vehicles into EV mode. We describe the traffic system with a discrete time model in which time is measured in steps $k = 0, 1, \ldots$ of a given length. Denote by $E(k)$ the aggregate amount of pollution produced by all vehicles in a geographic area at time instant $k$. At the beginning of each time step a control variable, $p(k)$, may be computed according to Algorithm 6.1 and broadcasted from the central infrastructure and each vehicle may choose its mode of operation by tossing a coin as described in Algorithm 6.2. These choices can not be changed until the beginning of the next time interval. This benefits the driver as high frequencies of changes between the modes and the associated starting and stopping of the ICE can be very bothersome.

All investigated control strategies are based on the following assumptions.

(i) Pollution levels in a geographic area can be measured or estimated without explicit communication from the vehicles.

(ii) There is no explicit car to car communication.

(iii) There is no explicit car to infrastructure information, but participating vehicles can listen and respond to broadcast signals.

(iv) At each time step, the decisions the vehicles make are independent of each other.

The first controller we investigate for the central infrastructure is a classical proportional-integral-differential (PID) regulator.

**PID control :** With these assumptions we can set up the regulation problem. To this end let $E^*(k)$ denote the desired level of pollution at time instant $k$. The first control strategy uses a discrete implementation of the classical PID control:

$$p(t) = K_p e(t) + K_i \int_0^t e(\theta) d\theta + K_d \frac{de(t)}{dt}, \qquad (6.2)$$

where $e(t) = E^*(t) - E(t)$. We do not choose to do it here, but with this controller systems like ours are often modelled as a sector-bounded time-varying nonlinearity with additive noise term (to account for the randomness), followed by an averaging filter. Note that the nonlinear term depends on traffic densities and is thus time-varying. Given this background, the problem can be viewed as a classical Lure problem and the control design carried out in this framework [Kellet et al., 2006]. Consequently, stability of the closed loop can be guaranteed by selecting the control gains in accordance with the circle criterion. A detailed description of a design using the circle criterion for a similar problem is given in [Kellet et al., 2006]. In this paper, a similar model is derived in the context of a buffer length regulation problem for a network router and the circle criterion is employed to give a stable controller. We shall not repeat this description here.

Rather, it is worth noting that even though the Lure framework allows filter dynamics to be incorporated into the feedback loop (this is an advantage over other approaches to nonlinear control designs such as multiple-increase, multiple-decrease algorithms [Stanojević et al., 2006]), the modelling of the process is not satisfactory. The system to be controlled is a stochastic one; the Lure approach models this as a deterministic system with an additive noise term and in some sense ignores the stochastic nature of the problem. In what follows we give an exact analysis in a stochastic framework for a simple integral control. A similar approach can be taken to obtain the same results if we use a PID controller.

To this end we further assume that the desired pollution level $E^*(k) = E^*$ is the same at each time step $k$. For a fixed number of cars, let the emissions of car $i$ at time $k$ be given by

$$E_i(k) = \begin{cases} C_i & \text{, with probability } p(k), \\ 0 & \text{, with probability } 1 - p(k), \end{cases} \tag{6.3}$$

for some $C_i \geq 0$. The overall emissions at time $k$ are given by $E(k) = \sum_i E_i(k)$. We now analyse this system for our integral controller, where we use a filtered value for the emission measurements given by $\bar{E}(k+1) = \lambda E(k+1) + (1-\lambda)\bar{E}(k)$ for some $\lambda \in (0,1)$ and update the probability according to

$$p(k+1) = Q\left(p(k) + K\left(E^* - \bar{E}(k)\right)\right), \tag{6.4}$$

where $K > 0$ is the gain and $E^*$ is the desired pollution level which satisfies $E^* \in [0, C]$, where $C = \sum_i C_i$ is the maximum possible pollution level. To ensure that $p(k) \in [0,1]$ we use $Q : \mathbb{R} \to [0,1]$, a projection to the unit interval, i.e.

$$Q(p) = \begin{cases} 0 & \text{if} \quad p < 0, \\ p & \text{if} \quad 0 \leq p \leq 1, \\ 1 & \text{if} \quad 1 < p. \end{cases} \tag{6.5}$$

We now focus our attention on the expected values of $\bar{E}$ and $p$ conditioned on the initial values $\bar{E}(0)$ and $p(0)$. If we let $X(k) = E[\bar{E}(k)|\bar{E}(0), p(0)]$ and $Y(k) = E[p(k)|\bar{E}(0), p(0)]$, then if $Y(k)$ is far enough from the boundary of the unit interval for all $k \in \mathbb{N}$ we obtain the relationships

$$X(k+1) = \lambda C Y(k) + (1-\lambda)X(k) \tag{6.6}$$

$$Y(k+1) = Y(k) + K(E^* - X(k)). \tag{6.7}$$

**Theorem 6.1.** *The system described by (6.6) and (6.7) has a unique fixed point, which is independent of the initial conditions. If further $K < \frac{1}{C}$ then the fixed point is locally asymptotically stable.*

*Proof.* We can describe the system in matrix notation as

$$\begin{pmatrix} X(k+1) \\ Y(k+1) \end{pmatrix} = P \begin{pmatrix} X(k) \\ Y(k) \end{pmatrix} + W, \tag{6.8}$$

with $P = \begin{pmatrix} 1-\lambda & \lambda C \\ -K & 1 \end{pmatrix}$ and $W = \begin{pmatrix} 0 \\ KE^* \end{pmatrix}$. The fixed point is given by $\begin{pmatrix} E^* \\ \frac{E^*}{C} \end{pmatrix}$. It can be seen that under the assumption that $K < \frac{1}{C}$ the spectral radius of $P$ is less than 1. And hence the system described by (6.8) converges locally exponentially fast to its asymptotic value.

$\square$

**Random Early Detection: RED** As an alternative to the PID controller we also investigate a feedback strategy that emulates the RED algorithm used in active queue management in Internet congestion control. More details of RED can be found in [Floyd and Jacobson, 1993]. The basic idea in RED is to adjust the probability of a car switching into fully electric mode as a non-linear function of the average pollution levels $\bar{E}$. The control objective in RED is to maintain pollution levels between pre described thresholds: $E_{min}$ and $E_{max}$. To do this, one describes a probability curve

$$p(\bar{E}(k)) = \begin{cases} 0 & \text{if} \quad \bar{E}(k) < E_{min}, \\ 1 & \text{if} \quad \bar{E}(k) > E_{max}, \\ p_{max} \frac{\bar{E}(k) - E_{min}}{E_{max} - E_{min}} & \text{else.} \end{cases} \tag{6.9}$$

Basically, the hope is that for a range of traffic densities there exists an equilibrium probability $p^*$ corresponding to $E^*$ such that $E_{min} \leq E^* \leq E_{max}$, and that this equilibrium is stable and that the system converges to it.

The stability of RED has been widely studied in the context of internet congestion control [Srikant, 2003]. Some work in this direction was done in the course of my degree. We studied the stability of a distributed implementation of an emulated RED system. Parts of the relevant publication [Budzisz et al., 2011] can be found in Appendix A. We will not repeat the stability discussions here. Rather we concentrate on the merits of RED, or otherwise, in the present context. Clearly, RED is much simpler than the PID control described above. It requires only a simple piece-wise linear function for implementation. However, it is also clear that the probability curve is predicated on a range of vehicle densities. Not all densities can be accommodated using a single curve, and to do this requires adjustment of $f$ based on feedback. A further concern is the performance of the algorithm in the presence of rapidly varying traffic densities.

**Multiplicative Increase Multiplicative Decrease: MIMD** As a further alternative control algorithm we investigated. Here again the goal is to keep the filtered emissions $\bar{E}(k), k = 0, 1, \dots$ within some interval $[E_{min}, E_{max}]$. To this end, two factors $m$ and $M$ satisfying $0 < m < 1 < M$ are chosen and the probability that a vehicle will choose the ICE mode is updated according to

$$p(k+1) = \begin{cases} mp(k) & \text{if} & \bar{E}(k) > E_{max} \\ Mp(k) & \text{if} & \bar{E}(k) < E_{min} \\ p(k) & \text{else.} \end{cases} \qquad (6.10)$$

This algorithm is designed for quick convergence and fast reaction to changes in the system and disturbances. However its scaling behaviour is not great. The parameters $m$ and $M$ have to be adapted to the specific system. If they are too close to 1 then convergence will be too slow. If they are too far apart from 1 then we will see oscillations. It is also possible to influence this behaviour by choosing a good filter for the emissions. Algorithms like this are being used in the control of computer networks, see for example [Stanojević, 2007].

## 6.6. Simulations

This section illustrates the simulation results obtained by implementing the pollution control mechanisms previously described. We use Sumo to simulate traffic on a grid-like network. We assume that the aggregate levels of pollution can be either measured or estimated. In reality the estimated pollution levels will be a complicated function of both the aggregate pollution levels and the mixing that takes place in a geographical zone. It seems reasonable that this relationship is an increasing function of the aggregate pollution levels, and all our results and algorithms hold in this case as well.

We first describe the grid-based simulation setup in detail. Then we present the results obtained with the different control approaches and for different test scenarios.

### 6.6.1. Grid Simulation Setup

We simulate traffic in a grid-like transportation network, i.e., where streets intersect orthogonally. The investigated grid is depicted in Figure 6.6. Although it does not correspond to a particular city, it still reflects the topology of many planned cities that can be found for example in North America. Traffic is simulated using Sumo and data is sampled every 10 seconds from it. The pollution control strategies are implemented by interfacing Matlab with the Sumo simulator.

For the purpose of the simulations, cars enter at 5 different points in the network and drive along random routes. Over a time interval of 10000 seconds we increase the number of cars in the network until 2000 cars have entered. We then continue the simulation for 6 hours. For the purpose of emissions modelling we assume all cars to be petrol electric hybrid cars with weight below 2.5 Tonnes and with combustion engine capacity between 1400 and 2000 cc, whose emission factors were reported in Table 6.1. We further assume a vehicle mix in terms of their emission standard, in particular 88% of the vehicles are in EURO 4, 9% in EURO 3, 3% in EURO 2, and 0.4 % in EURO 1. The evolution of the number of cars over time in the network is depicted in Figure 6.7. To better understand
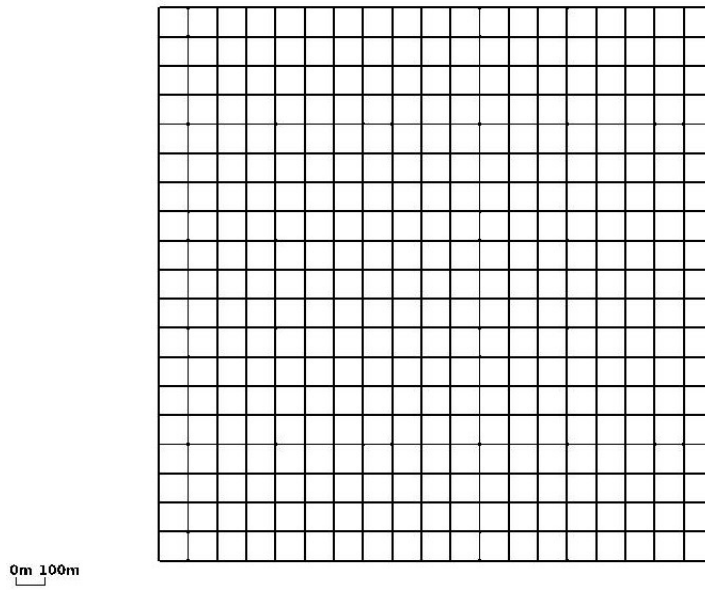
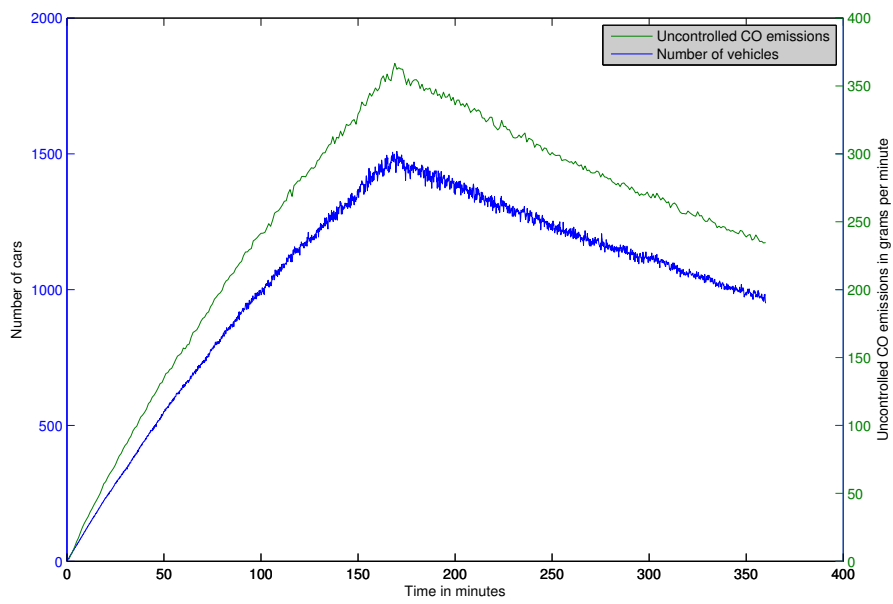Figure 6.6.: Simple grid-like road network for traffic simulations.



Figure 6.7.: Evolution of the number of vehicles and corresponding CO emissions in the network if all cars are combustion driven.

the performance of our algorithms, we simulated what happens if we assume that all cars decide not to use their electric drive at all, or where all cars are just conventional combustion driven vehicles. Figure 6.7 shows the evolution of emissions over time in this uncontrolled scenario. It can clearly be seen that the amount of CO emitted during any time interval is more or less proportional to the number of vehicles present in the network, although it should be mentioned, that it clearly also depends on how polluting each car in the fleet is.

### 6.6.2. Different Pollution Control Strategies

In this simulation we compare three different control strategies for regulating pollution. Recall the plot depicted in Figure 6.7. This figure illustrates the obvious fact that harmful emissions increase as the number of vehicles in the network increases. Our objective now is to show that using our strategies we can decrease the coupling between the number of vehicles and the aggregate emissions. Specifically, using very elementary ideas from Control Theory we can prevent the aggregate emissions level from exceeding some upper limit. We concentrate here on the control of the carbon monoxide emissions, but the same approach can be used for any other pollutant or combinations of pollutants. First we use an MIMD control algorithm, where the aim is to keep the value of the emissions between 140 and 160 grams every minute. Next we use the RED like control algorithm with the same boundary values. Lastly we use an integral control strategy, where we aimed at regulating emissions to the value 150 grams of CO every minute. The evolution of the emissions and the corresponding evolution of the broadcasted probability in all three scenarios are depicted in Figure 6.8. Note that the target values we picked for the pollutants were selected arbitrarily but can easily be adapted to safe levels.

As mentioned before, the approach should also work in situations where pollution is produced by either a number of old conventional vehicles, that we cannot control, or even a source of pollution that is independent of traffic, such as industrial facilities, fossil fuel power plants, or naturally occurring variances in pollution levels. To this end we repeat the above simulations where we also add an external source of CO that is active from minute 180 to 270 and contributes 40 grams of CO per minute during this time. In reality this could for example happen if the wind turns and carries pollution from an industrial area into the city for some time. In this situation the cars treat this source of pollution as a disturbance and adjust their behaviour to compensate for it. Paradoxically, the aggregate emissions from the hybrid vehicles can be much cleaner than the background air quality. Alternatively this temporary exogeneous pollution may also be regarded as a temporary change in the available pollution budget. As we aim at developing best effort strategies, this should not pose a problem to our algorithms. The results are depicted Figure 6.9. It can indeed be seen that all three algorithms cope well with this abrupt change.

It can be seen in Figure 6.7 that if electric drives are not used, the level of pollution follows a curve that mainly depends on the number of vehicles that are currently traveling

in the road network. However, if some of the vehicles travel in EV mode, then the level of emissions does not depend on the number of vehicles anymore, and can be controlled in order to be close to a desired level of pollution, for instance around 150 grams per minute as done here. It can be seen that all three control strategies work well if their respective parameters are tuned to the given scenario. This holds both with and without external pollution.
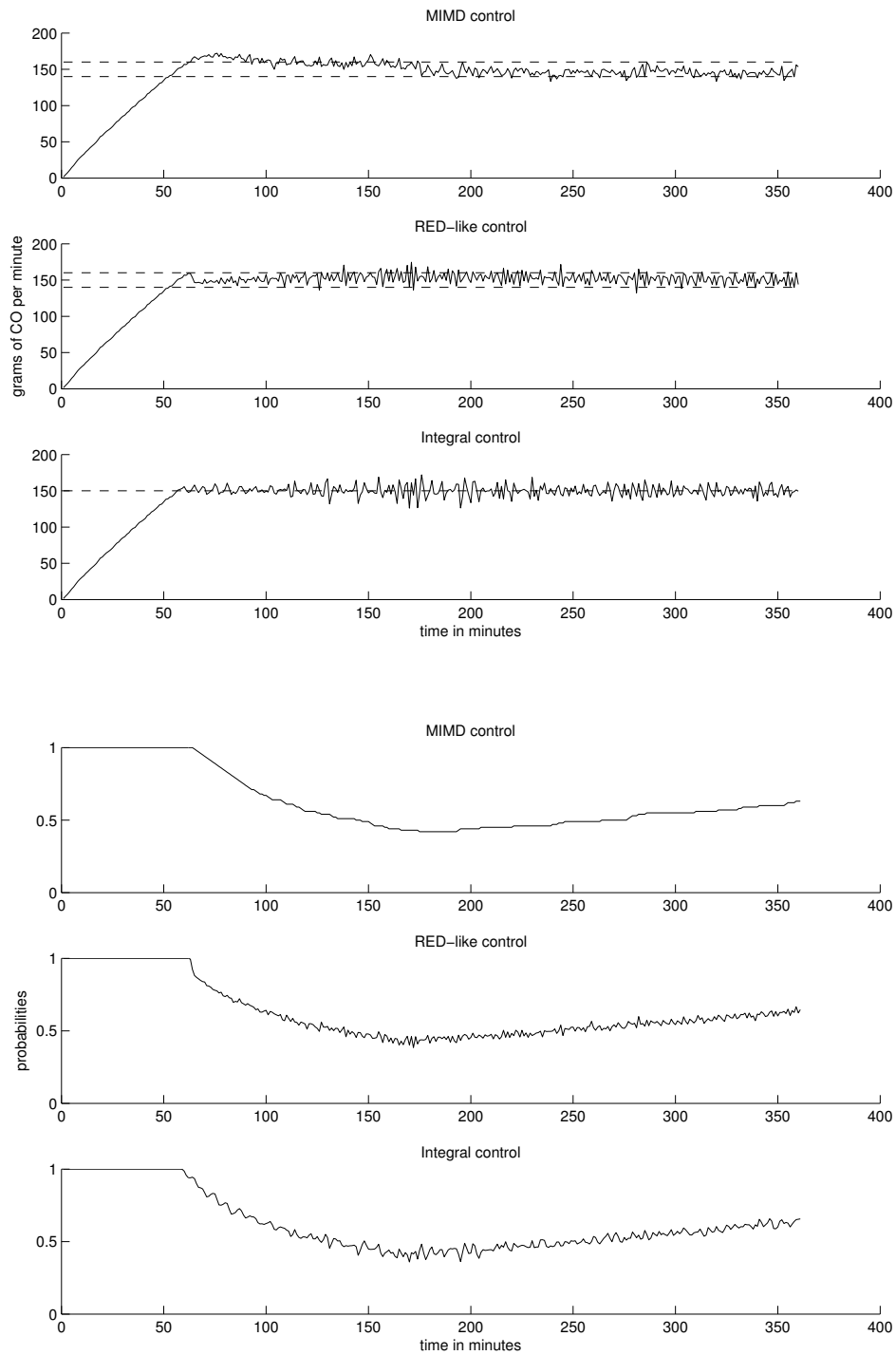
Figure 6.8.: Comparison of three control approaches. Top: Evolution of emissions under the three different controllers. Bottom: Corresponding probabilities.

Figure 6.9.: Comparison of three control approaches in a scenario with an external source of pollution. Top: Evolution of emissions under the three different controllers. The black line depicts the contribution of the vehicles to the pollution and the grey line shows the aggregate pollution. Bottom: Corresponding probabilities.

# 7. Cooperative Control of Pollution in Urban Scenarios - Optimisation

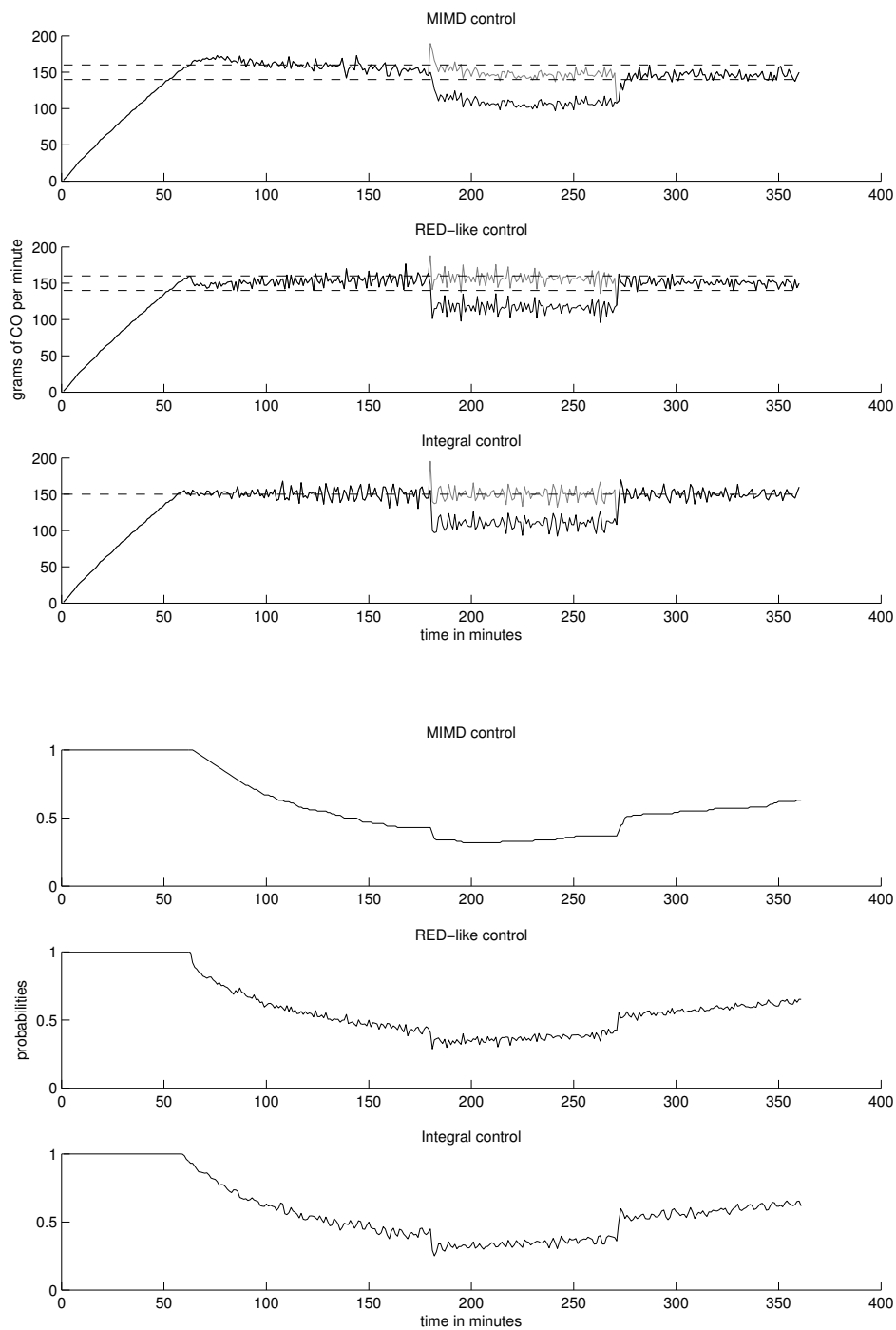***Abstract :*** *In this chapter we extend the approach to regulate traffic related pollution in urban environments by utilising hybrid vehicles presented in Chapter 6. By using ideas from consensus and direct vehicle to vehicle communication, we enable the vehicles to allocate an available pollution budget in a fair way using decentralised trading techniques. This work was conducted jointly with Florian Häusler, Mahsa Faizrahnemoon, Emanuele Crisostomi, Ilja Radusch, and Robert Shorten and was published in [Häusler et al., 2013b].*

In Chapter 6 we proposed a novel pollution control algorithm for fleets of hybrid electric vehicles (HEVs). The idea was to control the fraction of HEVs that was allowed to use their combustion engine at any given time. This was done using only simple broadcast signals from a central infrastructure. Clearly there is a non-unique way in which cars can cooperate to achieve the desired pollution levels. An exciting aspect of our approach is that this non-uniqueness can be exploited to deliver certain notions of fairness to individual drivers. For example, notions of fairness can be introduced that reflect individual driver's battery levels and hence their ability to participate in our pollution control scheme. This can be achieved using simple on-board computations and does not affect the approach's ability to control pollution. In [Schlote et al., 2013b] a very simple approach to do this was proposed and exemplified. We now present a more involved approach, which extends ideas from [Knorn et al., 2011b] on implicit consensus algorithms, that were applied to vehicular emission control problems in [Knorn et al., 2011a]. We follow their idea closely. The idea is to make each vehicle adjust the probability of traveling in ICE mode according to the residual level of its battery, so that vehicles that are close to running out of battery could drive in fuel mode, while at the same time keeping aggregate pollution levels in tolerable limits.

The approach presented in [Schlote et al., 2013b] involved using the broadcasted signal and individual factors, such as current battery level, to compute an individual probability to use the ICE. To achieve this, the $i'$th vehicle adjusts the probability to travel in ICE mode according to

$$p_i = f(p, SOC_i) \qquad (7.1)$$

where $SOC_i$ is the state of charge of the battery in vehicle $i$ and $f : [0,1] \times \mathbb{R}_+ \to [0,1]$ is appropriately computed so that the compliance with the algorithm is strict when the battery is full and $p_i$, the probability of car $i$ of driving in ICE mode approaches 1 when the vehicle is running out of battery. Using the battery fairness approach it was possible to reduce the number of vehicles running out of battery from 106 out of 2000 to 11 in one simulation scenario. Note that no communication between vehicles or direct communication between individual vehicles and the central infrastructure is needed to implement this.

It is possible to include other criteria such as a vehicle's individual emission output, the driver's behaviour, and many more. Fairness considerations are completely standard in the networking community and are captured by the concept of utility fairness, where a utility function for each vehicle is used to encapsulate the desired level of fairness between individual drivers. In the context of [Schlote et al., 2013b] this amounts to tailoring the probability communicated to each vehicle according to the properties of that vehicle (e.g., the type of vehicle or the remaining charge in the battery) with minimal communication requirements. Many techniques exist in optimisation as well as the networking community to achieve utility maximisation in a distributed manner. It is possible to borrow these ideas for our framework and give a formal discussion of strategies to realise utility maximisation that use standard techniques from stochastic optimisation. We now present some preliminary ideas regarding this.

The work presented in this chapter is an extension to the previously discussed pollution control in a geographic area in two senses. First, it applies to a fleet of vehicles belonging to the same organisation, such as a taxi or delivery company or a public transport provider and not all vehicles have to be physically close to each other. However the same techniques can be used to refine the control of emissions in a geographic area. Second, we discuss algorithms that more clearly and explicitly take fairness and utility maximisation into account. In the considered scenario, the fleet vehicles have to share a pollution budget between them in a decentralised fashion using only limited communication.

Formally, we consider a network of $n$ vehicles with individual utility functions $f_i(D_i)$, and a constraint $\sum_{i=1}^{n} D_i \leq C$, where $D_i$ is the amount of the shared resource allocated to vehicle $i$. This leads to an optimisation problem of the form described by Equations (5.1) and (5.2).

We make the following assumptions.

- Utility functions $f_i(D_i)$ are concave. If we are interested in minimising the sum of utility functions (e.g., minimum $CO$), then we assume that the utility functions are convex. In addition, the precise nature of the $f_i(D_i)$ may be unknown, but the

vehicles have access to a measurement or a-posteriori calculation based on observable variables of $f_i(D_i)$ at each time step. In some cases, e.g. the utility fairness case, we will also assume that the utility functions are increasing (decreasing) in the maximisation (minimisation) case.

- We assume that an optimal and feasible solution exists. For instance, this corresponds to assuming that vehicles have the ability to achieve the optimal quality of service or to produce the optimal quantity of $CO$ emissions.

- We assume that all vehicles can communicate directly (or indirectly through a measurement) with a centralised infrastructure, so that reliable pollution levels are always available.

- We assume that the centralised infrastructure may broadcast information to the vehicles, but not communicate directly with each vehicle.

- Cars can communicate directly using V2V. This communication has limited range and it is not necessarily symmetric.

The above assumptions appear to be a reasonable reflection of current and near future state-of-the-art in ITS systems, and suggest a hybrid solution whereby the optimisation is solved using decentralised computation with aid of limited signalling from the centre. Given these constraints we now wish to find efficient algorithms to solve the utility maximisation problem.

## 7.1. Two Decentralised Algorithms to Solve the Optimisation Problem

Equations (5.1) and (5.2) define a classical optimisation problem whose solution may be found for example with the aid of Lagrange multipliers [Boyd and Vandenberghe, 2004]. Let $H(D_1, ..., D_n, \lambda)$ be the Lagrangian associated with the problem, i.e.

$$H(D_1, ..., D_n, \lambda) = \sum_{i=1}^{n} f_i(D_i) - \lambda(\sum_{i=1}^{n} D_i - C). \qquad (7.2)$$

The theory behind convex optimisation is well developed and completely standard. We refer the reader to [Boyd and Vandenberghe, 2004] for an introduction to the relevant notions, which we will use in the following. The solution to the maximisation problem is given by the $n$ primal optimal values $(D_1^*, \ldots, D_n^*)$ and the dual optimal value $\lambda^*$. According to the KKT conditions they need to satisfy

$$\lambda^* = \frac{\partial f_i(D_i^*)}{\partial D_i}, \ i = 1, .., n, \qquad (7.3)$$
$$= g_i(D_i^*),$$

subject to the linear constraint (5.2) being satisfied. Many algorithms that solve convex optimisation problems update primal and or dual variables iteratively and hence give rise to discrete time dynamic systems. We take a similar approach here and assume that in each optimisation step, $k$, the central infrastructure can compute and broadcast a value $\lambda(k)$, which is used by car $i$ to compute a value $D_i(k)$. The goal is to update $\lambda(k)$ and $D_i(k)$ for all $i$, such that $\lambda(k) \to \lambda^*$ and $D_i(k) \to D_i^*$ as $k \to \infty$. The form of Equation (7.3) suggests two (hybrid) approaches to solving the maximisation problem. First, Equation (7.3) suggests that the maximisation problem can be solved using feedback in combination with local on-vehicle computational power. Namely, by regulating the value of $D_i$ so that the outputs of the $i'$th vehicle, $g_i(D_i(k))$, follows the value of $\lambda(k)$ which is broadcast from a centralised infrastructure. The task of the central infrastructure then becomes to find the appropriate value for the dual variable $\lambda(k)$ so that in the limit as $k \to \infty$, or if possible in finite time, the utility maximisation problem can be solved. As we shall see, this can be solved by embedding the multiplier as part of a feedback loop as is done in subgradient methods. This approach requires no V2V or I2V unicast capabilities, can deal with uncertainty in the utility functions, and requires no centralised computation. A disadvantage of the approach is that it is not robust to failure of the centralised node and selecting the control gains is an issue.

A second approach is based on the fact that optimality is also achieved when each of the vehicles reach consensus of the values of the functions $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$ subject to the constraint (5.2). This observation was first exploited in [Stanojević and Shorten, 2009b] in the context of symmetric communication graphs, and later extended to the asymmetric scenario in [Knorn et al., 2011b,a], where the $g_i(D_i)$ are adjusted as part of a feedback loop. Advantages of these approaches are that no I2V unicast capabilities are required, and the techniques can deal with uncertainty in the utility functions. In addition, no centralised computation is required and the technique is robust with respect to node failure.

The first algorithm we propose, Algorithm 7.1, places almost no computational burden on the central infrastructure, and requires no V2V capabilities. To formally present it we need the following additional assumptions.

- The infrastructure and the vehicles have a common clock and update their decision variables in discrete time steps.

- The value of $D_i(k)$ can be directly and arbitrarily adjusted in each of the vehicles. In some situations these variables are not directly accessible and must be adjusted indirectly, for example when $D_i$ is a complicated function of the driving behaviour. For exposition, this case has not been considered. Note however that the framework extends directly to this case, albeit with much more difficult analysis. See [Knorn et al., 2011b,a] for work in this direction.

The basic idea in Algorithm 7.1 is to use an integral control to find the optimal Lagrange

---
**Algorithm 7.1** Feedback control and local computation.

$E(k) \leftarrow$ total emissions produced by the fleet

$\lambda(k+1) = \lambda(k) - \gamma(C - E(k)$

$D_i(k+1) = D_i(k) - \epsilon_i(\lambda(k) - \frac{\partial f_i(D_i(k))}{\partial D_i(k)})$

---

multiplier $\lambda^*$. The centralised infrastructure updates according to:

$$\lambda(k+1) = \lambda(k) - \gamma(C - \sum_{i=1}^{n} D_i(k)), \qquad (7.4)$$

where $\gamma > 0$ and where we have assumed that the centralised infrastructure either receives communication of the $D_i(k)$ from each vehicle, or measures the sum over all $D_i(k)$. Each node (vehicle) receives a broadcast of $\lambda(k)$ and updates its utility (implicitly) according to:

$$D_i(k+1) = D_i(k) - \epsilon_i(\lambda(k) - \frac{\partial f_i(D_i(k))}{\partial D_i(k)}), \qquad (7.5)$$

where $\epsilon_i > 0$.

This corresponds to a subgradient method from optimisation, see [Boyd and Vandenberghe, 2004; Nedic and Ozdaglar, 2009] and the gains can for example be selected using methods from nonlinear control theory. The convergence of the algorithm, and the choice of the gains $\gamma$ and $\epsilon_i$, is beyond what we want to show here. However, it is possible to reuse results and approaches from standard systems theory, see e.g. [Siljak, 2007; Stanojević and Shorten, 2009b; Khalil, 2001], to guide the choice of the gains and prove convergence.

**Comment :** The algorithm in its current form implements a utility maximisation policy. Other policies can be implemented by replacing $\frac{\partial f_i(D_i(k))}{\partial D_i(k)}$ in Equation (7.5) with an appropriate function. For example, if we use $f_i(D_i)$ instead of its partial derivatives, the policy implemented will be a utility fairness policy.

The second algorithm we present here, Algorithm 7.2, uses the above mentioned observation that optimality is ensured when consensus of the values of the functions $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$ constrained to $\sum_i D_i \leq C$ is reached .

---
**Algorithm 7.2** Implicit consensus with input.

$V_i(k) \leftarrow$ average value of utility of all direct neighbours of node $i$

$E(k) \leftarrow$ total emissions produced by the fleet

$D_i(k+1) = D_i(k) + \epsilon_i \left( g_i(D_i(k)) - V_i(k) \right) + \gamma_i \left( C - E(k) \right)$

---

It was observed in [Stanojević and Shorten, 2009b] that utility maximisation problems can be formulated as a consensus problem leading to a completely decentralised solution.

However, consensus based solutions require symmetric communication graphs to guarantee that $\sum_i^n D_i(k) \leq C$ for all $k$, as well as suffering from other convergence related issues. Some of these issues are addressed in [Knorn et al., 2011b,a] in which a basic feedback error signal, $C - \sum_i^n D_i(k)$, from the infrastructure is used to yield a modified integral control of the following form:

$$D_i(k+1) = D_i(k) + \epsilon_i \left( g_i(D_i(k)) - \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} g_j(k) \right) + \gamma_i \left( C - \sum_j^n D_j(k) \right), \quad (7.6)$$

where $\gamma_i > 0$ and $\mathcal{N}_i$ is the set of vehicles that can send information to vehicle $i$ and $n_i$ is its cardinality. Note that we have already used this algorithm in Section 4.2.5. The sign of the parameters $\epsilon_i$ in Equation (7.6) depends on the choice of the functions $g_i$. In particular, the sign has to be positive if functions $g_i$ are decreasing, and negative if functions $g_i$ are increasing. Roughly speaking, the first term of the control achieves a consensus on the utility derivatives, and the second term ensures that the global constraint is satisfied. The stability and convergence properties of Algorithm 7.2, and the choice of $\gamma_i$ and the $\epsilon_i$, can again be determined using standard ideas from systems theory. See [Knorn et al., 2011b,a] for details. While a proper discussion of this issue is again beyond what we want to show here, we note that Algorithm 7.2 is a more complicated control law, requiring V2V communication, and has slow convergence properties [Knorn et al., 2011b,a]. Our contribution here is to present a modification to Algorithm 7.2 to speed up convergence, which we outline now. At each time step the vehicles communicate the maximum and minimum values of $\frac{\partial f_i(D_i(k))}{\partial D_i(k)}$ to the central infrastructure. These values are broadcasted to all nodes who then augment their neighbourhood set with these values.

We now present Algorithm 7.3, a modification to Algorithm 7.2, to speed up convergence. In the communication graph we add edges from the vehicles with minimum and maximum utilities to all other vehicles. To implement this, we propose that at each time step the maximum and minimum values of $g_i(D_i)$ are communicated to the central infrastructure. These values are then broadcasted to all nodes who in turn temporarily augment their neighbourhood information to incorporate the new values, unless the vehicle that has the respective value was already in the neighbourhood.

---

**Algorithm 7.3** Modified implicit consensus with input.

$V_i(k) \leftarrow$ average value of utility of augmented neighbour set of node $i$

$E(k) \leftarrow$ total emissions produced by the fleet

$D_i(k+1) = D_i(k) + \epsilon_i (g_i(D_i(k)) - V_i(k)) + \gamma_i (C - E(k))$

---

**Comment :** As before other policies can be implemented by replacing $\frac{\partial f_i(D_i(k))}{\partial D_i(k)}$ in Equation (7.6) with an appropriate function. For example, if we use $f_i(D_i)$ instead of partial derivatives, the policy implemented will be a utility fairness policy. A proper

discussion on the properties of the candidate functions $g_i$ again is beyond what we want to discuss here, but conditions on the functions $g_i$ and a related discussion can be found in [Knorn et al., 2011b,a] and [Stanojević and Shorten, 2009b].

## 7.2. Acceleration of Convergence

Simulation examples indicate that the implicit consensus algorithm reaches convergence in a faster manner when the maximum and the minimum value of the second term in Equation (7.5) are also revealed to the other nodes, see Section 7.3. This result is not totally unexpected. We give the following plausibility argument. It is well known that in undirected graphs, the second smallest eigenvalue of the graph Laplacian, also called algebraic connectivity, quantifies the speed of convergence of consensus algorithms, see for instance [Olfati-Saber et al., 2007]. Also, it is well known [de Abreu, 2007] that in this case the algebraic connectivity of a new graph $G_2$, $a(G_2)$, obtained from adding a new edge to graph $G_1$ is characterised by:

$$a(G_1) \leq a(G_2) \leq a(G_1) + 2, \tag{7.7}$$

thus establishing that adding a new edge can potentially increase the speed of convergence, or at least does not decrease it. Note that having two more nodes broadcasting their own value, as suggested from the improvement to Equation (7.5), corresponds in practice to adding up to $2(n-1)$ new edges.

The case of directed graphs is notoriously more difficult to handle, and there is not even a unique definition of the graph Laplacian. As a consequence, there are no theorems similar in spirit to Equation (7.7), at least to our knowledge. However the following result from [Wu, 2005] suggests that the out-degree of all nodes, i.e. the number of directed edges leaving from a given node, plays an important role in this. They define the algebraic connectivity of a directed graph $G$ with adjacency matrix $A$ as $\min_{x \perp e, \|x\|=1} x^\top L x$, where $e$ is the all-ones vector of appropriate size and $L = D - A$ is the Laplacian matrix associated with the directed graph and $D$ is the diagonal matrix of out-degrees of the nodes. The author then shows that

$$a(G) \leq \min_{v \neq w} \{d_o(v) + d_o(w)\}, \tag{7.8}$$

where $v$ and $w$ are nodes of the graph $G$. This can be used to state that increasing the out-degree $d_o$ of some nodes of the graph (as in practice occurs when the algorithm is modified to circulate the minimum and the maximum values) also increases (or at least, does not decrease) an upper bound of the algebraic connectivity of the graph.

Finally, we also mention that similar strategies to improve the convergence of network consensus algorithms in directed graphs were also devised in [Cao and Wu, 2007], where the authors claim that having as many as possible vertices with the maximum out-degree

Table 7.1.: Relation between amounts of $CO$ and $CO_2$ produced for vehicle $i$.

| Emission class | $CO = f_i(CO_2)$ |
|---|---|
| EURO 1 | $2.3073 CO_2^2 + 12.25 CO_2 - 2.0531$ |
| EURO 2 | $0.87556 CO_2^2 + 2.1907 CO_2 - 0.24558$ |
| EURO 3 | $3.3209 CO_2^2 + 2.5657 CO_2 - 0.7812$ |
| EURO 4 | $0.53745 CO_2^2 + 1.3241 CO_2 - 0.1064$ |

of $n - 1$, and having the in-degree of each vertex around $m/n$, where $m$ is the number of edges and $n$ that of nodes, can improve the convergence speed.

## 7.3. Simulations

We now present a number of simple simulations to illustrate the efficacy of our decentralised emission trading algorithms. In the following two simulations, we assume that a fleet of 20 hybrid vehicles has a total budget of $CO_2$ they can emit, and the objective is to minimise the total quantity of emitted $CO$. Furthermore, as in [Knorn et al., 2011b,a] we assume that vehicles can regulate the quantity of emissions by deciding which fraction of the desired speed is supplied by the ICE or by the electric motor. This is consistent with a power split hybrid vehicle.

Here we again use the average speed emission model given in Section 4.1.2. According to this, for light duty vehicles the amount of $CO$ produced by vehicle $i$ is related to the amount of $CO_2$ according to the functions $f_i : \mathbb{R}_+ \to \mathbb{R}_+$ given in Table 7.1.

Our objective is

$$\min \sum_{i=1}^{N} f_i(D_i) \tag{7.9}$$

$$\text{subject to } \sum_{i=1}^{N} D_i \leq C, \tag{7.10}$$

where $C$ is the overall $CO_2$ budget allocated to the fleet of vehicles, and $D_i$ is the amount of $CO_2$ allocated to the $i'$th vehicle. Note that this minimisation problem is equivalent to the maximisation problem described by Equations (5.1) and (5.2) which has already been discussed in the previous sections.

### Utility Minimisation within a Single Class of Vehicles

In the first simulation, we consider a fleet of 20 vehicles all belonging to the EURO 4 class, with slightly different parameters that are characteristic of the particular vehicle (e.g., brand, age).

Figure 7.1 shows how the $CO_2$ budget is shared among the fleet of vehicles in order to minimise $CO$ emissions. In particular, note that the solution obtained with the proposed distributed approach converges to the optimal solution computed directly by solving the optimisation problem with Lagrange multipliers, and shown with horizontal blue bars. Note however that the second approach is not viable in large-scale optimisation problem as it would require full-communication and massive centralised computational power.

**Utility Minimisation among Different Classes of Vehicles**

In the second simulation, we consider a fleet of 20 vehicles with ICEs belonging to EURO 1, 2, 3 and 4 classes. The objective is again to share a $CO_2$ budget among the vehicles, while minimising the overall emission of $CO$. As the emission of $CO$ is a very different from vehicles belonging to different classes, the optimal solution computed as in the previous simulation is not feasible any more, as it might allocate unrealistic quantities of $CO_2$ to some vehicles. For this reason, we solve again the optimisation problem with the proposed distributed algorithm, but we also force the unknown quantities of interest (i.e., $D_i$) to belong to the feasible sets for each vehicle. Note that the range of feasible $D_i$ strongly depends on the specific EURO class of the vehicle.

Figure 7.2a illustrates how the same $CO_2$ budget is now shared among the different classes of vehicles, while Figure 7.2b both shows that the total level of $CO_2$ matches the desired one, and that at the same time the amount of $CO$ is minimised.

**Utility Fairness**

The objective of the previous simulations was to minimise the production of $CO$ while keeping the total amount of $CO_2$ equal to a desired threshold. However, a key aspect regards the fairness of the previous solution. Note that while utility is maximised in these simulations, the solution suggested by the optimisation may be very unfair. In some situations it may be of interest to make all vehicles produce an equal amount of pollution. This is easily accommodated in our framework.

In the next simulation we are therefore interested in making the vehicles produce the same quantity of $CO$, while still respecting the total $CO_2$ budget. While such a solution is clearly not overall optimal for the environment (for example more $CO$ than before is produced), it provides a solution that is fairer for the hybrid vehicle owners. In the following, we compare the three algorithms discussed in Section 7.1, and for simplicity, we assume that the utility function of node $i$ is given by a function of the form $f_i(D_i) = \alpha_i \log D_i$, which is increasing and concave. The choice of this function is motivated by dispatch type problems where economic utility is an increasing function of allocated emissions. Recall, for such applications $g_i(D_i(k)) = f_i(D_i(k))$. Figure 7.3a depicts the performance of Algorithm 7.1 , Figure 7.3b the unmodified Algorithm 7.2, and Figure 7.3c the performance of the modified Algorithm 7.3.
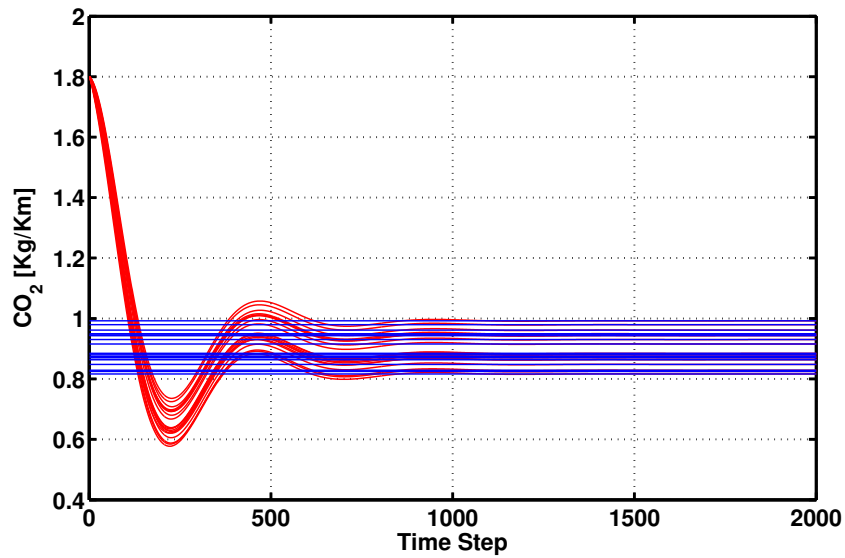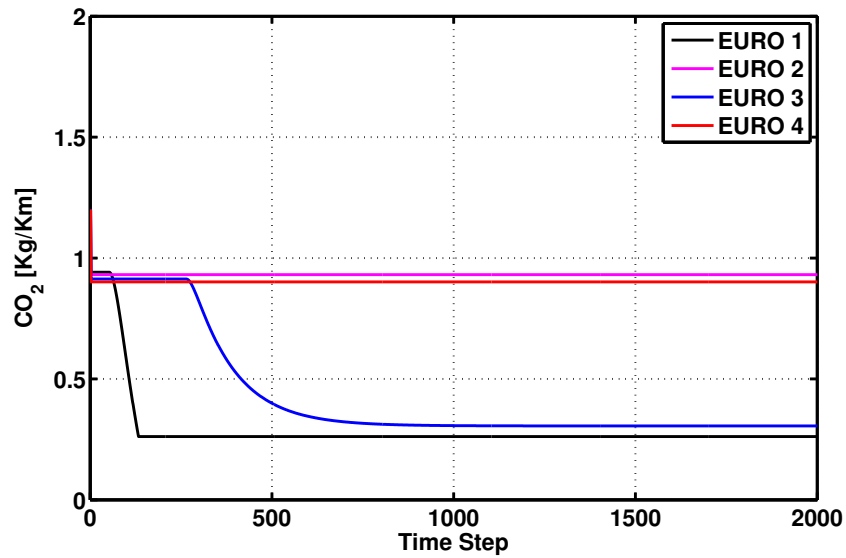
(a)
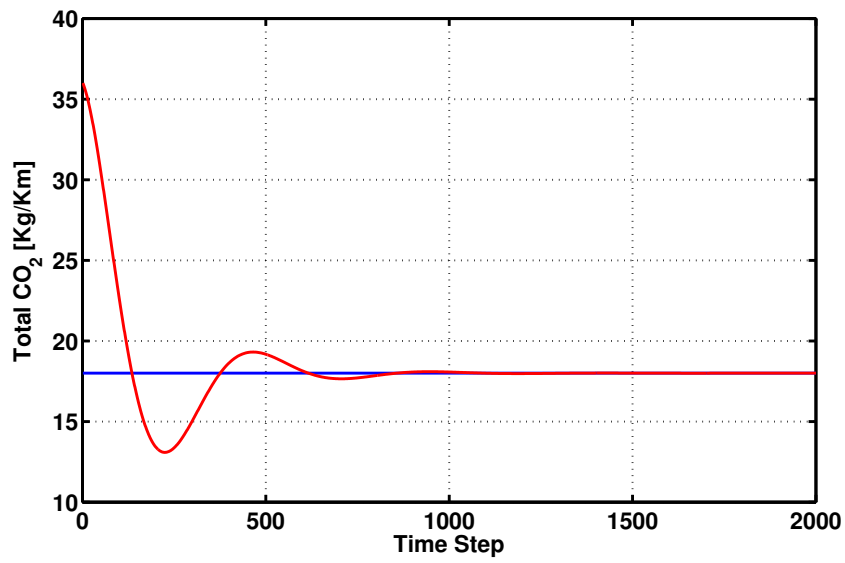


(b)

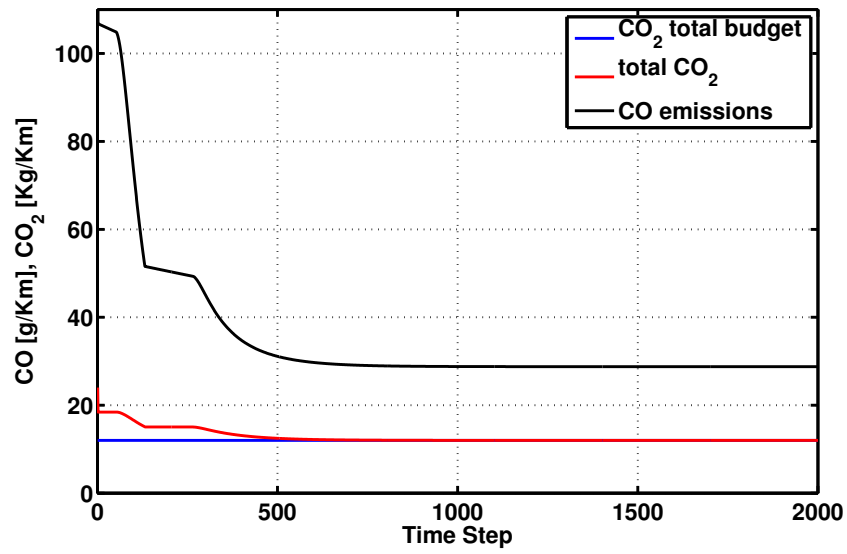Figure 7.1.: Figure a shows how the $CO_2$ budget is correctly shared among the fleet of vehicles in order to minimise $CO$ emissions. At the same time, the $CO_2$ budget is respected, as shown in Figure b.

(a)



(b)

Figure 7.2.: Figure a shows how the $CO_2$ budget is shared among the vehicles belonging to different EURO classes. While the $CO_2$ budget is respected, $CO$ is minimised as shown in Figure b.
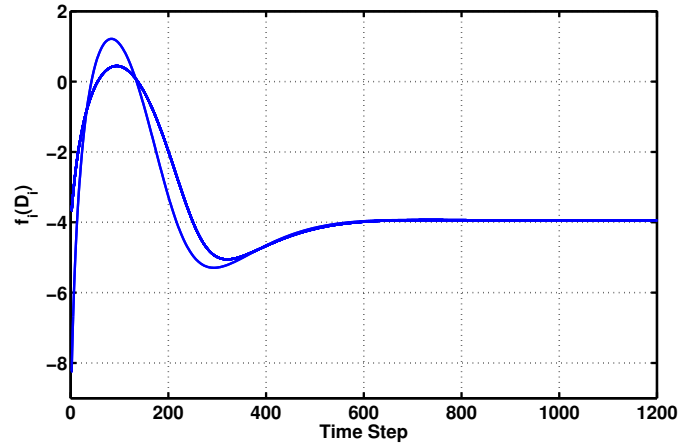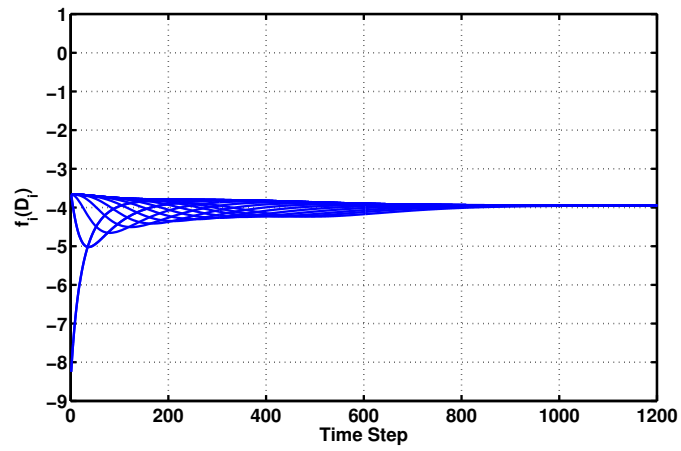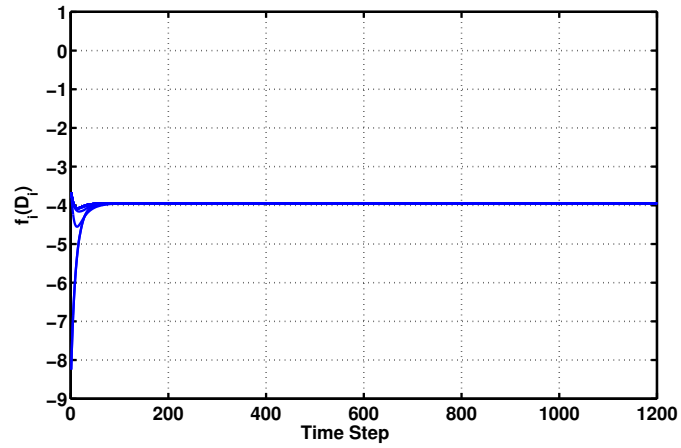
Figure 7.3.: The $CO_2$ budget is shared in such a way to equalise the production of $CO$. In (a) the vehicles do not communicate; in (b) they can communicate only with the "next" vehicle; in (c) they can communicate with the next vehicles and two other randomly chosen vehicles.

Figure 7.4.: Road network imported from OpenStreetMap [OSM Contributors, 2014], area around Rosenthaler Platz in Berlin, Germany

**Utility Fairness: a Case Study**

In this simulation we further investigate the utility fairness strategy from a geographic point of view, and test our ideas using the mobility simulator Sumo, with its HBEFA-emission model and TRACI interface [Krajzewicz et al., 2006]. We simulate traffic in an arbitrary area of Berlin, shown in Figure 7.4, which we divided into nine equivalent rectangular zones of the same size (in a $3 \times 3$ grid). The goal is to regulate emission in each zone separately. To this end, we let the infrastructure broadcast the probability of traveling in EV mode in order to achieve the $CO_2$ budget *in each of* the 9 zones of Berlin. This result is shown in Figure 7.5. Then, we allow single vehicles to locally adapt their probabilities to further achieve $CO$ fairness, as discussed in the previous section. Figure 7.6 shows the normalised $CO$ produced by vehicles belonging to 5 different classes of vehicles (with ICEs equivalent to EURO 1 to 5 vehicles).
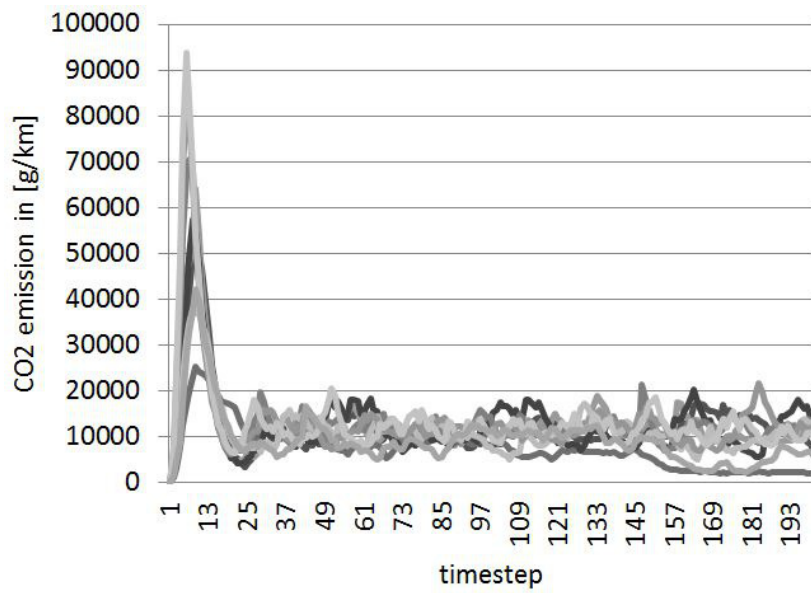
Figure 7.5.: The level of $CO_2$ is maintained equal to the $CO_2$ budget in each subzone of the Berlin area of interest.
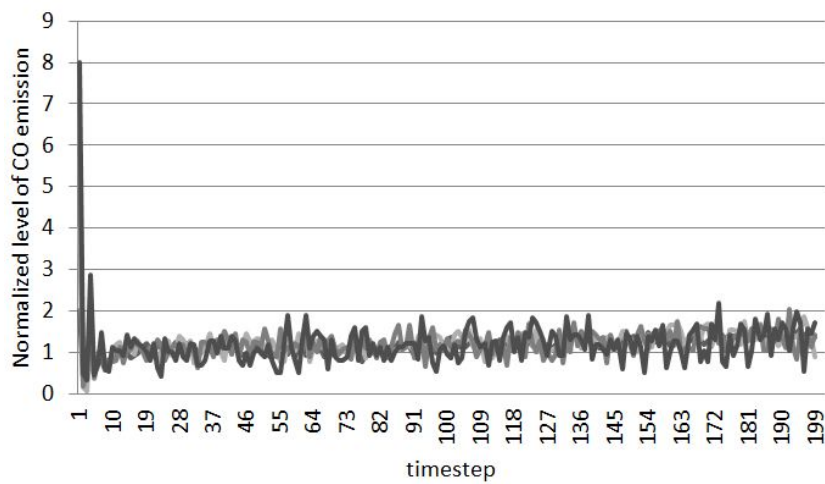


Figure 7.6.: Vehicles belonging to different classes produce the same quantity of $CO$.

109

# Part III.

# Feedback and Resource Allocation

# 8. Analytical Tools to Study Feedback Effects in Resource Allocation Strategies

***Abstract :*** *Inspired by a number of problems in Intelligent Transportation Systems, in this chapter we discuss assignment strategies to servers with queues with the application to traffic related problems in mind. The work in this chapter was conducted jointly with Florian Häusler, Emanuele Crisostomi, Christopher King, Robert Shorten and also with Bei Chen and Mathieu Sinn. It was published as [Schlote et al., 2014, 2013c,a; Häusler et al., 2012, 2013a].*

In many Intelligent Transport System applications, we try to make better use of scarce resources such as parking spaces, charging spots for electric vehicles, or shared transport vehicles such as bicycles. In general terms, we wish to improve the access of users to some commodity. Given the new technologies that enable smart mobility, there are three strategies to do this.

- The first strategy is to use a forecast system that informs the user of the available commodities at the time the user requires them. Such predictions are hard if there is uncertainty in the system, especially as this prediction influences the customer decisions and thus introduces feedback in the system that may invalidate predictions if they do not take this effect into account.

- The second is a reservation system, where customers communicate their choice in advance so that the availability data can be enhanced with this information and resource availability can be guaranteed. In real world implementations such systems are complex. They require much communication and possibly complicated optimisation algorithms. Furthermore, inefficiencies arise as customers' choices are based on virtual information - for example a server may be idle and capable of quickly satisfying the customer's demands yet is booked by other customers and thus discourages the customer from using it. Additional inefficiencies are caused if the duration of the delay fluctuates or customers do not meet their reservations for some reason.

- The third approach is to use stochastic decentralised balancing strategies. Users

make random decisions guided by the availability data. It is possible to implement such systems with minimal communication and computation requirements and they can be very successful in balancing customers and evenly distributing commodities. Decentralised stochastic strategies are also known to be very robust against disturbances.
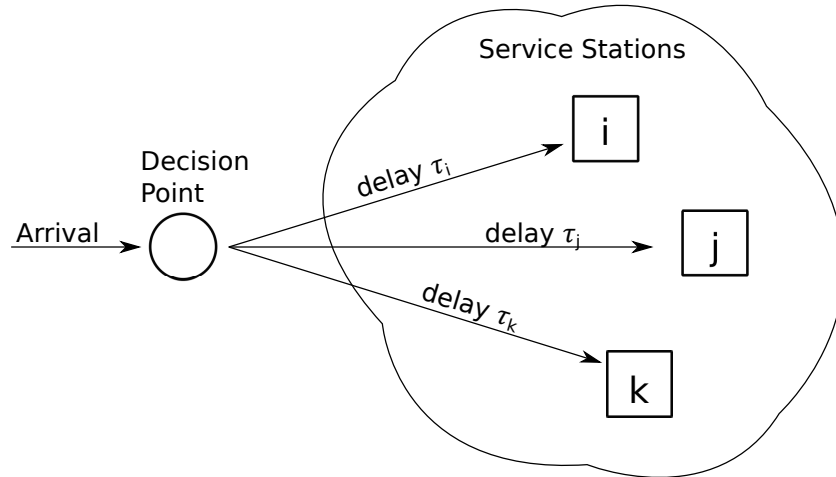


Figure 8.1.: Assignment Setup.

It is the third approach that we adopt here. Hence, in this chapter we consider the problem of assigning customers to service stations. The general setup is depicted in Figure 8.1. We consider systems where customers require some service or commodity and get to make a choice between different providers. We assume that data on the availability of the resource of interest from each provider is available to the customer at the time he makes a decision. Our systems of interest involve several customers competing for those resources with the added difficulty that there is a delay between choice and resource consumption. We provide a number of analytic tools to investigate assignment strategies to servers with queues with the application to traffic related problems in mind. These tools can also be used to analyse investigate systems with prediction type ITS applications. We shall apply all tools in Chapter 9.

In this chapter we consider two scenarios.

1. In the first scenario there is a single server with a queue. Customers require service and decide whether to use this server based on information about the queue occupancy. In this, we concentrate on what happens at this server and disregard customers that choose other options. The goal is to control the queue length at the server to keep waiting times short. There is a delay between the customer choice and their joining of the queue. As we shall see, this makes simple deterministic choice strategies ineffective. Instead, this setting favours stochastic decisions. A

well known example of such a system is RED [Floyd and Jacobson, 1993], already mentioned in Chapter 6.1.

2. In the second scenario we consider a system with more than one server each with a distinct queue. Customers that require service can choose one of these servers based on queue occupancy information. Again there is a delay between decision and joining of the queue. This delay can introduce oscillations and inefficiency to the system under some choice strategies. This model is relevant for a large number of applications, most notably for load balancing in communication networks [Stanojević and Shorten, 2009b], but also for many applications involving user choices such as route choices in journey planning.

## 8.1. Taking Feedback into Account

As discussed previously, the great advances in automotive, computation and communication technology are opening up great potential for improving safety and efficiency of transport. The availability of exact and extensive data regarding shared resources naturally inspires prediction type ideas as a first step. Basically the question asked by many researchers is the following. Given real-time and historic demand data regarding a certain resource, what is the best way of predicting the availability of this resource at a given point in the future. Approaches like this are for example being proposed for predicting the availability of parking spots in a car park [Caliskan et al., 2006; Klappenecker et al., 2012], the availability of bicycles in a particular station in a bike sharing scheme [Chen et al., 2013], the availability of charging spots in charging stations for electric vehicles [Mariyasagayam and Kobayashi, 2013], and the price of charging electric vehicles in smart grids with dynamic pricing [Erol-Kantarci and Mouftah, 2010].

The goal of these prediction models is to keep the number of customers that find no available resource when they arrive at the service station to a minimum. Prediction models, of course, have a long history and many computational tools are available including for example time series analysis and neural networks. In the context of load forecasting for electric grids, highly accurate predictions are necessary for electricity companies to plan the production of power and complicated and computationally expensive tools can be used as the prediction is done centrally. If a similar prediction would be made by an individual user's smartphone, or smart charging device, the same tools cannot be used. Instead, the tools being used in this context have to be as computationally light as possible. Markov chains are often used in this context as they are computationally easy to handle. This is done for example in [Caliskan et al., 2006; Klappenecker et al., 2012; Chen et al., 2013].

An additional factor that is ignored by all the above mentioned papers is that the availability of predictions changes user behaviour. Strangely, all those papers propose prediction models to guide customer choices, yet none of them investigates the consequences of this. For example, in [Caliskan et al., 2006; Klappenecker et al., 2012] an

occupancy prediction algorithm for individual car parks is proposed, but the effect to the system caused by customers adapting their behaviour based on these predictions is ignored. If a prediction is used by only one customer or a small group of customers, it may be valid to assume that the choices of the small group have a negligible effect on the dynamics of the system. However, if the percentage of customers to which the prediction is available is high, the change in their behaviour impacts the dynamics of the whole system and may thus invalidate the prediction. An example of this effect, based on real data, is shown in Section 9.4.3. In this simulation customers want to rent a bicycle and have to choose between two possible bike stations. We show that ignoring the feedback can lead to sub-optimal system performance.

Feedback is a well studied phenomenon in the area of control and many tools are available. Many aspects of such systems with feedback can be investigated and controlled if it is possible to influence the system. A first step in this direction is usually to consider the mean or the long run performance and how long it takes for the system to reach this state from a given initial condition. The second step then is to inspect the variation of the system. Often, especially in complicated or stochastic systems, variations of the state can not be avoided and it is thus of interest to study the effect of disturbances to the system. If one uses state information to control the system, this has to be done carefully. Poor implementations of this can lead to persistent oscillations that can leave the system alternately over- and under-utilised with potentially fatal effect on the system's performance. This can be very frustrating to the customers and be a main impediment for acceptance of the system in the population. We aim to raise awareness of this these effects in ITS and provide a number of tools with which to study traffic related applications.

In what follows, we present three approaches to analyse a distributed stochastic balancing system. Each approach has unique advantages and drawbacks. The first approach is very simple Markov chain, that does not take delays between assignment and arrival to the chosen queue into account and serves as a starting point for the study of such systems. Next, we present an advanced Markov chain that takes these effects into account, but works only for systems with one service station in its present form. Lastly, we present a system of delay differential equations that model a fluid limit of a system with several service stations and delays.

## 8.2. A Simple Markov Chain that Ignores Delays

Consider a system with $n \in \mathbb{N}$ service stations at which customers are being served. Customers that arrive to the system wait for availability signals from service stations and join them only after having received such a signal. There are discrete time steps in which each service station may broadcast a simple signal that indicates that it has free capacity and it decides to do so randomly with a probability that is a decreasing function of its occupancy. In particular if station $j$ has a buffer size of $C_j \in \mathbb{N}$ and a

queue length of $N_j \in \mathbb{N}$, then it uses an algorithm of the form of Algorithm 8.1, where $F : \mathbb{N} \to [0, 1]$ is a decreasing function with $F(0) = 1$ and $F(C_j) = 0$. In addition, we assume that customers do not listen for availability signals at all times. Instead, at each time step they decide to listen for signals randomly according to Algorithm 8.2, where $G : \mathbb{R}_+ \to [0, 1]$ is a decreasing function with $G(0) = 1$. $e_i$ is a commodity associated with the user which vanishes over time and as it does the $i'$th customer's probability of listening for signals increases. For example, in Chapter 9 $e_i$ will be the electric charge left in the battery of an electric vehicle.

---

**Algorithm 8.1** Free capacity indication algrithm

---

$p \leftarrow F(N_i)$
$r \leftarrow \text{Rand}$
**if** $r < p$ **then**
    Broadcast indication signal.
**end if**

---

---

**Algorithm 8.2** Electric Vehicle

---

$q \leftarrow G(e_i)$
Listen for indication signals with probability $q$.

---

## 8.2.1. QoS Analysis: Balancing Behaviour

We first discuss a simple Markov chain approach that models this system with decentralised and stochastic customer assignment. This chain does not take the delay between assignment and arrival at the queue into account and shall only be considered as a first step in modelling such systems. To do this we assume the following.

1. Customers arrive to the system according to a Poisson process with rate $\lambda > 0$.

2. After receiving a signal the customer is immediately added to the queue of the sending station.

3. We assume that each customer that is waiting at a station is being serviced with a rate independent of the total number of customers waiting for service. The service times for each customer are described by i.i.d exponentially distributed random variables with rate $\mu > 0$.

With these assumptions in place, the above model is a Markov chain, where the state is a $n$ dimensional vector that reports the occupancy of each station. The transition probabilities between the states can be readily computed from the rates of the arrival process and the service process and the used assignment rule for each station. For example, the probability of a given number, $m \in \mathbb{N}$ of new customer arrivals in a single time-step of length $T > 0$ is given by $e^{-\lambda T} \frac{(\lambda T)^m}{m!}$. If the state of the system at time step

$k \in \mathbb{N}$ is given by $(N_1(k), \ldots, N_n(k))$ and we know that $m$ new arrival occur, the state in the next time step satisfies

$$\sum_{j=1}^{n} N_j(k+1) \leq \sum_{j=1}^{n} N_j(k) + m \tag{8.1}$$

and for each state with this property we can compute its probability by taking into account the number of necessary departures from each station and the distribution of the $m$ new customers. Should there be several ways of reaching the new state we use the sum of these probabilities. For Algorithm 8.1 we know that each possible state can be reached from each other state in a number of steps of the chain with a positive probability and hence the chain is irreducible. Further as there is a positive probability that the state does not change during a step of the chain it immediately follows that the Markov chain is aperiodic. See [Norris, 1998] for an introduction to these notions. Irreducibility and Aperiodicity together assure uniqueness of the stationary distribution and convergence of the system dynamics to it. If the possible number of customers in each station is bounded from above, then the number of states is finite and we can use the transition matrix to compute the stationary distribution, which yields information about what fraction of time the chain spends in each state.

### 8.2.2. QoS Analysis: Assignment Delay

We now investigate the delay between a customer arrival and assignment to a charging station. According to Algorithm 8.1 customers are not immediately assigned to servers. Rather, they have to wait until they receive communication from a free server. We now give some complementary results to indicate how quickly customers are assigned to a place in a queue at a server. We call this time the assignment delay. Obviously, the system can be sped up by reducing the length of the time steps. However, we are interested in the number of time steps until assignment. To facilitate exposition, we make the following assumptions.

1. If a customer detects several availability signals, then he is assigned to only one server. A server is chosen according to some priority rule or randomly.

2. We assume that the system has reached steady-state.

**Theorem 8.1.** *Consider a network of $n$ servers with $N_j$ customers in the queue of server $j$. In this situation the $i'$th customer arrives to the system. Then the assignment delay, $\tau_i$, before being assigned to a server is on average given by*

$$\tau_i = \left( G(e_i) \left( 1 - p_{red} \right) \right)^{-1} \tag{8.2}$$

*where*

$$p_{red} = \prod_{j=1}^{n_m} \left( 1 - F(N_j) \right). \tag{8.3}$$

*Proof.* At any one time step the $i$'th customer gets assigned if it receives at least one green signal and it is currently listening for signals. These events occur with probability $1 - p_{red}$ and $G(e_i)$ respectively. As the two events are independent, the joint probability is $G(e_i)(1 - p_{red})$. From this it is possible to compute the expected waiting time, which can be seen to be equal to (8.2). $\square$

According to Theorem 8.1 the expected assignment delay is short if $p_{red}$ is small, which occurs if the number of servers is large or if the queues are small. Theorem 8.1 can be understood as a quality of service measure for the customers. One may equally be concerned about the quality of service for the servers. For them it is clearly undesirable to be idle. Obviously this situation will occur at times with low demand. However, it should be avoided that one or several servers are idle, while other servers have a high workload. The following theorem quantifies the probability that our assignment rule assigns customers in a non-optimal way.

**Theorem 8.2.** *Consider a network of $n$ identical servers. At a given time, let $\mathcal{B}$ be the set of all servers that are busy serving one or more customers, and let $n_I$ be the number of servers that are idle. Assume that at this time instant a customer is assigned to one of the servers. Then the probability that this customer is assigned to a busy server instead of an idle one is given by*

$$1 - \frac{n_I}{n_I + \sum_{i \in \mathcal{B}} F(N_i)}. \tag{8.4}$$

*Proof.* According to our assumptions the new customer is assigned to one of the busy server with probability

$$\frac{\sum_{i \in \mathcal{B}} F(N_i)}{\sum_{j=1}^{n_m} F(N_j)}. \tag{8.5}$$

Further $F(N_j) = 1$ for any idle server. Using this in Equation (8.5) and rearranging it yields Equation (8.4). $\square$

**Comment :** Theorem 8.2 addresses the undesirable situation where a customer is assigned to a busy server even though an idle one is available. It is possible to obtain more comprehensive quality of service measures along the same lines as in the theorem. For example it is possible to compute the probability that a newly arriving customer is assigned to a specific server. Such problems can be addressed as a queueing theory problem in an almost identical manner to Theorem 8.2. Given the exact shape of the function $F$ it is possible to further refine this result. For example, if one uses a function of the form $F(N_i) = 10^{-N_i}$ the probability of a new customer being assigned to a particular server depends only on the difference of the queues' lengths and not on the absolute values of the lengths.

In Theorem 8.2 the probability that a new customer is assigned to a busy server instead of an idle station depends strongly on the shape of the function $F$. Obviously a rapidly

decreasing of $F$ is preferable. For example, consider a system with 10 servers, of which 2 are idle. Assume that all other servers each have 2 customers in their queue. If we use for example $F(N_i) = 10^{-N_i}$, then the probability given by Equation (8.4) is less than 0.04.

## 8.3. Improved Markovian Model for a Single Server Scenario

We now present a second analytic tool that makes it possible to take the delay between assignment and arrival into account in a simplified scenario with only one station. To this end we consider a single server with an infinite queue. The first $C$ customers in the queue are being served simultaneously at a constant rate independent of the number of customers in the queue. The remaining customers are not being served until such times that enough customers have left the system and they are among the first $C$ customers in the queue. Alternatively this can be seen as a service station with $C$ parallel servers that can each serve at most one customer and a common queue. Customers realise their need for service according to a Poisson process with rate $\lambda$. When they do this at time $t$, they query the queue's occupancy $N(t)$ and may decide to either use the server or leave the system based on this information. To this end drivers make a random decision and decide to go to the server with a probability $p(N(t))$. The information that the customers can access is updated by the server in discrete time steps of length $T > 0$. If the $i'$th customer decides to join the queue he does so after a delay $\tau_i > 0$. We will model this delay as a random variable and assume that $T$ is chosen such that $\tau_i \leq T$ for all customers $i$. If the load on the system is high, there are two competing goals. On the one hand as many customers as possible are to be served, on the other hand as few customers as possible should arrive to a queue of length $C$ or longer, so that they do not immediately receive service.

We will now describe the evolution of the queue length process in discrete time steps. As a step size we use $T$ such that for all $k \geq 0$ the $k'$th time interval is given by $[kT, (k+1)T]$. We denote by $N(k)$ the queue length at time $kT$. The evolution of $N(k)$ can be described as a difference equation of the form

$$N(k + 1) = N(k) + A(k) - D(k), \tag{8.6}$$

where $A(k)$ is the number of customers that arrive to the queue during the interval $[kT, (k+1)T]$ and $D(k)$ is the number of customers that have completed their service in the same interval. $D(k)$ takes values in $\{0, 1, \ldots, N(k)\}$ and we model it as a random variable with distribution depending only on $N(k)$. In particular we assume that it follows a Poisson process that terminates after generating $N(k)$ events. Further this Poisson process has rate $\mu Q(N(k))$, where $Q(N(k)) = \min\{N(k), C\}$. Accordingly, the distribution of $D(k)$ is described by

$$P(D(k) = t | N(k) = n) = e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!} \tag{8.7}$$

for all $n \in \mathbb{N}$ and all $t = 0, 1, \ldots, Q(n) - 1$ and

$$P(D(k) = Q(n)|N(k) = n) \tag{8.8}$$

$$= \sum_{t=Q(n)}^{\infty} e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!}$$

$$= 1 - e^{-Q(n)\mu T} \sum_{t=0}^{Q(n)-1} \frac{(Q(n)\mu T)^t}{t!}.$$

We assume that $\tau_i$ is described by independent identically distributed random variables with support $[0, T]$. In this setting customers that take a decision to join the queue in $[(k-1)T, kT]$ will reach it either in $[(k-1)T, kT]$ or in $[kT, (k+1)T]$. Accordingly, we can rewrite Equation (8.6) as

$$N(k+1) = N(k) + A_1(k) + A_2(k) - D(k), \tag{8.9}$$

where $A_1(k)$ and $A_2(k)$ are arrival processes of customers to the queue in $[kT, (k+1)T]$. $A_1(k)$ is the arrival process of customers that make their decision to join the queue in $[(k-1)T, kT]$ and $A_2(k)$ is the arrival process of customers that make their decision in $[kT, (k+1)T]$. It is important to note that the arrival process of customers to the queue is not a homogeneous Poisson process. It is however piecewise homogeneous, in the sense that for all $k \geq 1$ the arrival process of customers to the queue in the interval $[kT, (k+1)T]$ is homogeneous. This is formally proved in the following lemma.

**Lemma 8.3.** *Given $N(k-1) = m$ and $N(k) = n$ for some $n, m \in \mathbb{N}$, for any given distribution of the i.i.d random variables $\tau_i, i \in \mathbb{N}$ the process $A_1(k) + A_2(k)$ is Poisson with rate $\lambda \left( \alpha p(m) + (1-\alpha)p(n) \right)$, where $\alpha = \frac{T - E[\tau]}{T} \in [0, 1]$.*

*Proof.* It is enough to prove that $A_1(k)$ is Poisson on the first update interval, $[0, T]$, with rate $\lambda \alpha p(N(0))$. Customers query the queue occupancy and decide to use the server with rate $\tilde{\lambda} = \lambda p(N(0))$. This process generates an infinite number of vehicles, but we are only interested in $A_1$, the number of vehicles that arrive to the queue before $T$. For $i \in \mathbb{N}$ let

$$a_i = \begin{cases} 1 & \text{, if customer } i \text{ reaches the queue before } T, \\ 0 & \text{, else.} \end{cases} \tag{8.10}$$

The $i'$th customer makes his decision at time $t_i$ and arrives at the queue after a delay of $\tau_i$ if he chooses to use the server. Let the cumulative probability density function of $\tau_i$ be given by $F_\tau : [0, T] \to [0, 1]$. $t_i$ is generated from a Poisson process with rate $\tilde{\lambda}$, and hence it is distributed according to an Erlang distribution with parameters $(i, \tilde{\lambda})$ and its probability density function $f_{t_i}$ is given by

$$f_{t_i}(x) = \frac{\tilde{\lambda}^i x^{i-1} e^{-\tilde{\lambda} x}}{(i-1)!}. \tag{8.11}$$

Accordingly, the probability that $a_i = 1$ is given by, $P(a_i = 1) = F_{t_i + \tau_i}(T)$, where $F_{t_i + \tau_i}$ is the joint cumulative probability density function of $t_i$ and $\tau_i$, which can be computed according to

$$F_{t_i + \tau_i}(T) = \int_{-\infty}^{\infty} F_\tau(T - x) f_{t_i}(x) \, dx \qquad (8.12)$$

$$= \int_0^T F_\tau(T - x) \frac{\tilde{\lambda}^i x^{i-1} e^{-\tilde{\lambda}x}}{(i-1)!} \, dx, \qquad (8.13)$$

where we used Equation (8.11) and the fact that $\tau_i$ and $t_i$ are positive random variables to change the limits of integration. Using Equation (8.10), we obtain $A_1 = \sum_{i=1}^{\infty} a_i$ and using the linearity of the expectation operator, we can rewrite the expected number of customers that arrive to the queue before $T$ as

$$E[A_1] = E[\sum_{i=1}^{\infty} a_i] = \sum_{i=1}^{\infty} E[a_i] = \sum_{i=1}^{\infty} P(a_i = 1) \qquad (8.14)$$

$$= \sum_{i=1}^{\infty} F_{t_i + \tau_i}(T) \qquad (8.15)$$

$$= \sum_{i=1}^{\infty} \int_0^T F_\tau(T - x) \frac{\tilde{\lambda}^i x^{i-1} e^{-\tilde{\lambda}x}}{(i-1)!} \, dx. \qquad (8.16)$$

According to Lebesgue's monotone convergence theorem, we may exchange summation and integration and obtain

$$E[A_1] = \int_0^T F_\tau(T - x) \tilde{\lambda} e^{-\tilde{\lambda}x} \underbrace{\sum_{i=0}^{\infty} \frac{(\tilde{\lambda}x)^i}{(i)!}}_{=e^{\tilde{\lambda}x}} \, dx \qquad (8.17)$$

$$= \tilde{\lambda} \int_0^T F_\tau(T - x) \, dx \qquad (8.18)$$

The remaining integral is independent of the Poisson arrival process. Using integration by parts we obtain

$$E[\tau] = \int_0^T x \frac{d}{dx} F_\tau(x) dx \qquad (8.19)$$

$$= [x F_\tau(x)]_{x=0}^T - \int_0^T F_\tau(x) dx \qquad (8.20)$$

$$= T - \int_0^T F_\tau(T - x) dx. \qquad (8.21)$$

And hence

$$E[A_1] = \tilde{\lambda}(T - E[\tau]). \tag{8.22}$$

The total number of customers expected to make a decision in $[0, T]$ is $\tilde{\lambda}T$, hence a fraction of $\alpha = \frac{T - E[\tau]}{T}$ arrives to the queue in $[0, T]$ and the rest, i.e. a fraction of $\frac{E[\tau]}{T}$ arrives in the interval $[T, 2T]$. $\qquad\square$

The system that we have described allows for the undesirable situation where customers arrive to a fully occupied server and have to wait to receive service. The following theorems guide the choice of functions $p$ to ensure that this undesirable situation is a rare event. To this end let $U(k)$ be the number of customers in the queue that have not started to receive service at time $(k+1)T$. It can be described by

$$U(k) = \max\{N(k) + A_1(k) + A_2(k) - D(k) - C, 0\}. \tag{8.23}$$

The following theorem quantifies the probability that $U(k)$ takes positive values in a given situation, i.e. the probability that the server is fully occupied at the end of the interval $[kT, (k+1)T]$.

**Theorem 8.4.** *Given $N(k-1) = m$ and $N(k) = n$ for some $n, m \in \mathbb{N}$ the probability that the number of customers waiting at time $(k+1)T$ is positive is given by*

$$1 - \sum_{l=0}^{C} e^{-\nu T} \frac{(\nu T)^l}{l!} \tag{8.24}$$

$$+ \sum_{t=0}^{Q(n)-1} e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!} \sum_{l=C-Q(n)+t+1}^{C} e^{-\nu T} \frac{(\nu T)^l}{l!},$$

*where $\nu = \lambda(\alpha p(m) + (1-\alpha)p(n))$ and $\alpha = \frac{T - E[\tau_j]}{T}$*

*Proof.* As $U(k)$ is a non-negative random variable

$$P(U(k) \geq 0 | N(k-1) = m, N(k) = n)$$
$$= 1 - P(U(k) = 0 | N(k-1) = m, N(k) = n). \tag{8.25}$$

From Lemma 8.3 we know that $A(k) = A_1(k) + A_2(k)$ is Poisson with rate $\nu = \lambda(\alpha p(m) + (1-\alpha)p(n))$. Note that $A(k)$ and $D(k)$ are independent when conditioned on $N(k-1) = m$ and $N(k) = n$. Let us use the following shorthand notation

$$P_{U,0} = P\left(U(k) = 0 | N(k-1) = m, N(k) = n\right) \tag{8.26}$$

Hence for all $n, m \in \mathbb{N}$ according to Equation (8.23)

$$P_{U,0} = P(A(k) - D(k) \leq C - n | N(k-1) = m, N(k) = n),$$

where we have rearranged the terms in the inequality. As $A(k)$ and $D(k)$ are independent conditioned on $N(k-1) = m$ and $N(k) = n$ we further obtain

$$\begin{aligned}
P_{U,0} &= \sum_{t=0}^{Q(n)} P(D(k) = t | N(k) = n) \cdot \\
&\quad \cdot P(A(k) \leq C - Q(n) + t | N(k-1) = m \text{ and } D(k) = t) \\
&= \sum_{t=0}^{Q(n)} P(D(k) = t | N(k) = n) \cdot \\
&\quad \cdot P(A(k) \leq C - Q(n) + t | N(k-1) = m).
\end{aligned}$$

We now use that $A(k)$ is Poisson with rate $\nu$ and thus the probability of $l$ customers arriving in $[0, T]$ is given by $P(A(k) = l) = e^{-\nu T} \frac{(\nu T)^l}{l!}$ for all $l \in \mathbb{N}$. This together with Equations (8.7) and (8.8) then yields

$$\begin{aligned}
P_{U,0} &= \sum_{t=0}^{Q(n)-1} e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!} \sum_{l=0}^{C-Q(n)+t} e^{-\nu T} \frac{(\nu T)^l}{l!} \\
&\quad + \left(1 - \sum_{t=0}^{Q(n)-1} e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!}\right) \sum_{l=0}^{C} e^{-\nu T} \frac{(\nu T)^l}{l!},
\end{aligned}$$

where we separated the case $t = Q(n)$ from the rest of the sum over $t$. Rearranging yields the claim.

$\square$

Theorem 8.4 gives a formula for calculating the probability of an overflow occurring at the server. Thus, it provides a tool to evaluate the performance of the choice of probability function $p$ in a given scenario. However, this QoS measure gives the probability of an overflow occurring at discrete time steps of length $T$ and disregards the probability that an overflow occurs and vanishes between the time steps. It thus underestimates the overflow probability. We now give a complementary result that gives an upper bound for the overflow probability.

**Theorem 8.5.** *Given $N(k-1) = m$ and $N(k) = n$ for some $n, m \in \mathbb{N}$ the probability that at least one vehicle is rejected during the time interval $[kT, (k+1)T]$ is given by the last entry of the vector $\hat{\pi}_{k+1}$ computed according to*

$$\hat{\pi}_{k+1}^\top = \pi_k^\top \exp(Q_k T), \tag{8.27}$$

where $\pi_k$ is a column vector with a 1 in the $Q(n) + 1$ entry and 0 everywhere else, exp denotes the matrix exponential and $Q_k$ is the $(C+1) \times (C+1)$ dimensional tri-diagonal matrix given by

$$
\begin{bmatrix}
-\nu & \nu & & & & \\
s & -(s+\nu) & \nu & & 0 & \\
& \ddots & \ddots & \ddots & & \\
& & s & -(s+\nu) & \nu & \\
& 0 & & s & -(s+\nu) & \nu \\
0 & \ldots & 0 & 0 & 0 & 0
\end{bmatrix},
\tag{8.28}
$$

with $\nu = \lambda\left(\alpha p(m) + (1-\alpha)p(n)\right)$ and $s = Q(n)\mu$.

*Proof.* During the update epoch $[kT, (k+1)T]$ we can model the system as a continuous time Markov chain with $C + 2$ states $0, 1, 2, \ldots, C + 1$, in which transitions from states $N$ to $N + 1$ happen with rate $\nu$ and from $N + 1$ to $N$ with rate $Q(n)\mu$ for all $N = 0, 1, \ldots, C - 1$. The state $C + 1$ corresponds to the situation where at least one customer arrives to the queue and cannot immediately be served. As we are interested in whether this state is reached during the considered time interval or not, we may make it an absorbing state. Transitions from state $C$ to $C+1$ thus occur with rate $\nu$ while transitions from states $C+1$ to $C$ occur with rate 0. The rate matrix of this chain is given by $Q_k$. $\pi_k$ is the distribution of the chain at time $kT$, which is concentrated in the state $Q(n)$. $\hat{\pi}_{k+1}^\top$ is the distribution of the states after time $T$ for our model starting in $\pi_k$ and accordingly it is given by Equation (8.27); with the last entry corresponding to the probability of reaching state $C + 1$. $\qquad\square$

## 8.4. Fluid Limit Model for Multiple Servers with Delay

We now present the third tool to investigate our assignment system. To this end we consider a system with $n$ stations. Each station $j$ has a finite buffer with capacity $C_j > 0$. Its buffer level at time $t$ is given by $X_j(t)$. A constant arrival stream with rate $\lambda > 0$ is split between all stations in a continuous way, depending on state information. In particular, each station is allocated a share of the arrival stream, which is proportional to the amount of free capacity at that station, i.e. $C_j - X_j(t)$. The fraction of the arrival stream allocated to station $j$ takes a time $\tau_j$ to actually arrive to the station. The buffer content is processed at a rate proportional to the buffer level such that the rate at which the whole system processes is equal to the arrival rate. This system can be described using the following set of delay differential equations. For $j = 1, \ldots, n$ let $X_j(t)$ be described by

$$
\frac{dX_j(t)}{dt} = -\frac{\lambda}{N} X_j(t - \tau_j) + \frac{\lambda}{C - N}(C_j - X_j(t)),
\tag{8.29}
$$

where $N = \sum_{j=1}^n X_j(t)$, and for all $j = 1, \ldots, n$ we have $\tau_j \geq 0$ and $C_j \in \mathbb{N}$ and $C = \sum_{j=1}^n C_j$. It can easily be seen that $N$ is indeed constant for all $t \geq 0$.

Models of the form described by Equation (8.29) arise when dealing with the fluid limit of assignment systems. One example of such a limit shall be discussed in Chapter 9.3 in the context of the assignment of vehicles to parking facilities.

**Theorem 8.6.** *Under the usual assumptions on initial conditions of the delay differential equations the system of Equations (8.29) has a constant solution, namely $X_j(t) = a_j = \frac{N}{C}C_j$. Further the constant solution is exponentially stable if at least one of the following conditions holds.*

$$N > \frac{C}{2} \tag{8.30}$$

$$\tau_j < \left(\frac{N}{\lambda}\right) \frac{\cos^{-1}\left(-\frac{N}{C-N}\right)}{\sqrt{1 - \left(\frac{N}{C-N}\right)^2}} \tag{8.31}$$

*Proof.* To investigate stability we look for a solution of the form

$$X_j(t) = a_j + b_j e^{zt}, \tag{8.32}$$

where $z$ is a possibly complex parameter. Substituting into (8.29) gives the delay differential equation's characteristic equation

$$z = -\frac{\lambda}{N}e^{-z\tau_j} - \frac{\lambda}{C - N}. \tag{8.33}$$

According to [Driver, 1977, Chapter VII, Section 28, Theorem B], if all solutions to Equation (8.33) have negative real part this assures exponential stability of the constant solution. We now investigate under which conditions on $\tau_j$ this property holds.

When $\tau_j = 0$, the solution is

$$z = -\frac{\lambda C}{N(C - N)}, \tag{8.34}$$

which implies exponential stability of the constant solution.

For $\tau_j$ sufficiently small all solutions of (8.33) lie in the left half of the complex plane, and thus the constant solution (8.32) is still stable. However as $\tau_j$ increases, the solutions of Equation (8.33) may move toward the imaginary axis. Instability occurs when the first solution crosses the imaginary axis. Letting $z = x + iy$, this occurs when $x = 0$. In this case Equation (8.33) can be rewritten as

$$iy = -\frac{\lambda}{N}\left(\cos(y\tau_j) - i\sin(y\tau_j)\right) - \frac{\lambda}{C - N}, \tag{8.35}$$

which is equivalent to the two equations:

$$0 = -\frac{\lambda}{N}\cos(y\tau_j) - \frac{\lambda}{C - N} \tag{8.36}$$

$$y = \frac{\lambda}{N}\sin(y\tau_j). \tag{8.37}$$

The Equations (8.36) and (8.37) have no solution if

$$N > \frac{C}{2} \qquad (8.38)$$

and hence the constant solution is stable for all values of $\tau_j$.

If $N \leq C/2$ then the solution is stable for $\tau$ sufficiently small. We now determine $\tau_{crit}$, the precise threshold value where instability occurs. To this end, we can rearrange Equation (8.36) to

$$y\tau_j = \cos^{-1}\left(\frac{N}{C-N}\right). \qquad (8.39)$$

Substituting this in Equation (8.37) yields

$$y = \frac{\lambda}{N} \sin\left(\cos^{-1}\left(\frac{N}{C-N}\right)\right) \qquad (8.40)$$
$$= \frac{\lambda}{N}\sqrt{1 - \left(\frac{N}{C-N}\right)^2},$$

where we used a standard trigonometric identity. Substituting $y$ from Equation (8.40) into Equation (8.39) and rearranging yields

$$\tau_{crit} = \left(\frac{N}{\lambda}\right)\frac{\cos^{-1}\left(-\frac{N}{C-N}\right)}{\sqrt{1 - \left(\frac{N}{C-N}\right)^2}} \qquad (8.41)$$

This concludes the proof. $\qquad\qquad\square$

We now consider a similar system of equations obtained by replacing the undelayed term in Equation (8.29) by an additional delayed term. We obtain for all $j = 1, \ldots, n$

$$\dot{X}_j(t) = \lambda\frac{C_j - X_j(t - \tau_{j,2})}{C - N} - \lambda\frac{X_j(t - \tau_{j,1})}{N}, \qquad (8.42)$$

where $\tau_{j,1}, \tau_{j,2} > 0$, $C_j \in \mathbb{N}$, $C = \sum_{j=1}^{n} C_j$, and $N = \sum_{j=1}^{n} X_j(t)$, which again is a constant for all times $t \geq 0$. This model will be used to study bike sharing systems in Chapter 9, where customers have to make two choices: At which station should they pick up a bicycle, and at which station should they return it. We obtain the following result along the same lines as Theorem 8.6.

**Corollary 8.7.** *Under the usual assumptions on initial conditions the system of Equations (8.42) has a constant solution, namely $X_j(t) = a_j = \frac{N}{C}C_j$. Further if*

$$\tau_{j,1} < \frac{N}{2\lambda} \qquad (8.43)$$

*and*

$$\tau_{j,2} < \frac{C - N}{2\lambda} \tag{8.44}$$

*the constant solution is exponentially stable.*

*Proof.* We obtain the characteristic equation of Equation (8.42) by considering solutions of the form $X_j(t) = a_j + b_j e^{zt}$, where $z$ is a complex parameter. The characteristic equation is

$$z = -\frac{\lambda}{C - N} e^{-\tau_{j,2} z} - \frac{\lambda}{N} e^{-\tau_{j,1} z} \tag{8.45}$$

For $\tau_{j,1} = \tau_{j,2} = 0$ the constant solution above is exponentially stable and by continuity it must also hold for small values of $\tau_{j,1}$ and $\tau_{j,2}$. We now investigate for which values of $\tau_{j,1}$ and $\tau_{j,2}$ we cannot guarantee stability. Consider $z = x + iy$, for real numbers $x, y$. We are interested in values of $\tau_{j,1}$ and $\tau_{j,2}$ for which $x = 0$. In this case Equation (8.45) can be written as

$$iy = -\frac{\lambda}{C - N} e^{-\tau_{j,2} iy} - \frac{\lambda}{N} e^{-\tau_{j,1} iy}, \tag{8.46}$$

which is equivalent to the two equations

$$0 = \frac{\lambda}{C - N} \cos(\tau_{j,2} y) + \frac{\lambda}{N} \cos(\tau_{j,1} y) \tag{8.47}$$

$$y = \frac{\lambda}{C - N} \sin(\tau_{j,2} y) + \frac{\lambda}{N} \sin(\tau_{j,1} y) \tag{8.48}$$

Equation (8.48) is solved by $y = 0$. By comparing the slope of the left hand side and the right hand side of Equation (8.48), it can be seen that if $\tau_{j,1} < \frac{N}{2\lambda}$ and $\tau_{j,2} < \frac{C-N}{2\lambda}$ then $y = 0$ is the only solution. However $y = 0$ does not solve Equation (8.47) and hence under these conditions no instability can occur. $\qquad\square$

In Chapter 9 we shall investigate several Intelligent Transport Systems with assignment schemes. In these systems for each individual customer a choice needs to be made between several options. In particular we consider systems with a delay between the assignment and its effect, usually the arrival at a service station. These systems are hard to analyse in general. The tools presented in this chapter allow to gain some insight into the system dynamics and help with the design and parameter choices in assignment strategies.

# 9. Assignment Strategies in Intelligent Transportation Systems

*Abstract :*   *In this chapter we consider three assignment problems originating in traffic networks and apply the methods considered in Chapter 8 to them. The work in this chapter was conducted jointly with Florian Häusler, Emanuele Crisostomi, Christopher King, Robert Shorten and also with Bei Chen and Mathieu Sinn. It was published as [Schlote et al., 2014, 2013c,a; Häusler et al., 2012, 2013a].*

In this chapter we investigate three transport related application areas for stochastic load balancing ideas. These applications are:

1. Assignment of electric or plug-in hybrid vehicles to charging stations

2. Assignment of cars to parking facilities

3. Assignment of customers to bicycle sharing stations.

The common task in these is to assign customers to service stations and our approaches follow a common idea. The starting point is an analogy with the mobile cellular network, where a small number of base stations cover a large geographic region and partition it into zones such that each zone is covered by one base station. Customers that require service are associated with the base station in the cell they are located in. Mobile phone users can move freely within this region without disturbance of their service as they are being handed over to a new station whenever they move between cells. We now consider in a similar vein a geographic region of interest - typically on the city scale - to be partitioned into zones. Each zone contains a number of service stations, see the example depicted in Figure 9.1. Customers can move freely between the zones. At some point customers may realise their need for a service station and we assign them to one of the stations in their current zone. The underlying assumption is that within a given zone a customer has no preference for a particular service station.

The division into partition cells can be based on graph clustering techniques (e.g. Voronoi partitioning) that make use of customer density patterns, for example in a way such that
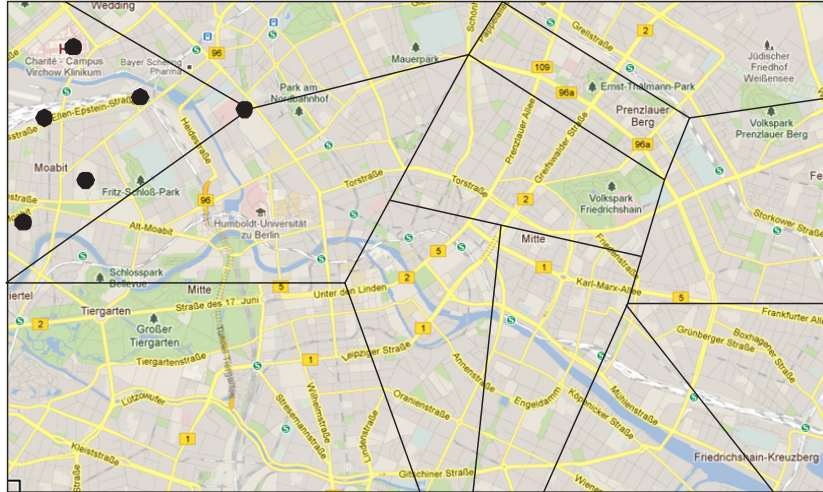
Figure 9.1.: A possible partition of an area of Berlin with hypothetical service stations.

the ratio of customers to service stations is roughly the same in each cell. Note that this partition could be easily adapted based on temporal density patterns making the partition time dependent. The partitioning of the city into cells can be realised in many ways and this is not the focus here. Rather, our focus is on the customer assignment problem within a cell which takes place over faster time-scales than the partitioning problem. To achieve this, we propose stochastic algorithms in which service stations advocate the number of current slots available for customers and customers in turn choose one station randomly with a probability proportional to this number.

We advocate the use of decentralised and stochastic algorithms. They are preferable to centralised and deterministic algorithms for a number of reasons.

1. Low communication overhead

2. No central infrastructure required

3. No complicated optimisation techniques needed

As a benchmark for our algorithms we consider systems in which the same data we use is available to all customers and they themselves make choices. We use our own experience as an indication for how humans decide and assume that people obey simple rules. For example, in the context of choosing one of several servers to receive service, we assume that humans pick the seemingly most favourable option - such as the shortest queue at a check-out till in a supermarket. The approaches presented in this chapter can be seen as an attempt to get people to follow new decision rules. These new rules are only marginally more complicated than joining the shortest queue and in addition have the advantage of having an almost intuitive feeling to them once they have been presented

to the people.

We now discuss the different applications and possible implementation approaches in more detail.

## 9.1. Assignment of Electric Vehicles to Charging Stations

In this section we present the application of stochastic decentralised load balancing techniques to the assignment of electric vehicles to charging stations. In the later sections we will discuss algorithm implementations in which much information is available to the customers, which is in line with the recent developments in ITS. However, in this section we propose the following simple and easy to implement procedure that has minimal communication requirements. Within a partition cell charging stations broadcast a green signal, indicating their ability to accept a new vehicle, with a frequency that is a decreasing function of the current queue length. Hence, charging stations with more spare capacity advertise their service with a higher frequency and thus gain more new vehicles. Similarly, vehicles listen to green signals with a frequency that is a decreasing function of the current level of their battery. In effect we match vehicles with a pressing demand for new energy with the charging stations with the most free capacity. If, at a given time, a vehicle senses a green light from a charging station the vehicle is assigned to the charging station. Note that this happens in a fully decentralised fashion requiring no central infrastructure. This implementation is interesting for customers who are not using dedicated ITS systems as it can be achieved with very little investment into user devices. For example it is possible to use satellite navigation devices to do this. The actual implementation of the protocol is described in detail in the following section.

Throughout this section we use the following assumptions.

1. Charging stations can measure their free capacity and broadcast it to all vehicles in the system.

2. Electric vehicles can measure their battery state of charge and listen to charging station broadcasts.

### 9.1.1. Basic Algorithm

We now describe the basic implementation of the algorithm.

---
**Algorithm 9.1** Charging station
$N_i \leftarrow$ Number of vehicles being charged at station $i$
$p \leftarrow F(N_i)$
Broadcast green light with probability $p$

---

We let all charging stations act at regular points in time. We assume that at every such time step the $i$'th charging station communicates its availability to accept a new vehicle with probability $F(N_i)$. To do this it uses Algorithm 9.1, where $N_i$ is number of vehicles currently queuing for charging at the $i$'th charging station, and $F : \mathbb{N} \to [0, 1]$ is some function mapping the occupancy to a probability. We assume that $F(0) = 1$ and that as the queue grows $F$ monotonically decreases to 0. As charging stations with a shorter queue are more likely to accept a new car than charging stations with a longer queue, this mechanism is able to balance the number of vehicles associated with each charging station.

---
**Algorithm 9.2** Electric Vehicle

$e_i \leftarrow$ State of Charge of vehicle $i$
$q \leftarrow G(e_i)$
Listen for green signals with probability $q$

---

Once the $i$'th electric vehicle has decided to search for a charging station we assume that at each time step it makes a random decision to detect green signals according to Algorithm 9.2, where $G : \mathbb{R} \to [0, 1]$ is a function that maps the vehicle's current state of charge, $e_i$, to a probability. We require that $G$ is a decreasing function and $G(0) = 1$. Throughout this section we use the following simple way to compute $G(e_i)$

$$G(e_i) = 1 - \frac{e_i}{M}, \qquad (9.1)$$

where $M$ is the maximum quantity of energy that an EV battery can hold. By using a decreasing function $G$, we assure that EVs with low state of charge are more likely to sense a green signal than vehicles with plenty of residual charge. This gives EVs priority according to their requirements. At each time step, if a charging station broadcasts a green light and an EV senses it, then the EV is assigned to that charging station. If a vehicle receives green signal from several stations simultaneously, it choses one of those stations at random.

There are many possibilities to refine the basic algorithm. For example, instead of balancing the number of vehicles, it is also possible to balance the required energies requested at each charging station. It is further possible to take other factors into account in the computation of $G(e_i)$, for example the expected energy required to reach a charging station or the distance to it, the expected time to reach the destination in situations where not all stations are equivalent and the user needs, such as the accepted detours, location preferences etc.

### 9.1.2. Protocol Implementation

Note that - contrary to most approaches (e.g. [Geng and Cassandras, 2011], [Kokolaki et al., 2011]) - our approach is fully decentralised. We focus on decentralised solutions here as they can be very attractive for a number of reasons. First, such algorithms

are usually robust to possible failures (including attacks and manipulations) in a way that centralised solutions are not. Moreover, the approach could easily be extended by introducing a reputation system to identify misbehaving vehicles or charging stations. Second, distributed self-organising optimisation algorithms are very suitable to handle scenarios which are highly stochastic in nature; see for example the Transmission Control Protocol (TCP) in communication networks or [Pavone et al., 2011] in the context of vehicular mobility. Third, the decentralised approach we propose has certain privacy and security benefits, as there is no need for an entity, which knows about the locations of all individual vehicles. On the contrary, decisions can be made in a decentralised manner with only limited information. Finally, decentralised solutions facilitate the possibility of implementing plug and play policies, where charging stations may enter and leave the brokerage system. For example private car parks, office blocks, and universities may make their charging stations available during off-peak times. The latter point is very important to emphasise and will be discussed later through an example. In a centralised solution the adding of additional charging stations requires updates to the centralised structure, whereas our solution is completely self-organising leading to a truly plug and play system. However, centralised implementations also have a number of advantages. Many ITS systems that are currently being developed make use of vehicle-to-vehicle and vehicle-to-infrastructure communication and hence it can be expected that in the future the large communication overhead associated with centralised solutions may not be an issue anymore. Such implementations of our algorithms are also possible and we will discuss one later in Section 9.3.

### 9.1.3. Performance Analysis

A system as described above is best analysed in a queueing theory framework. We now describe a possible approach to do this and give results, under simplifying assumptions, regarding our algorithm for two distinct scenarios. For our numerical example, we let the probabilities $F(N_i)$ for Algorithm 9.1 be calculated according to $F(N_i) = 10^{-N_i}$, where $N_i$ is the number of vehicles at the $i'$th charging station. To ease exposition, we make the following assumptions.

1. There is no delay between a car requesting service and it being assigned to a charging station.

2. After being assigned the vehicle is instantly added to the charging stations queue.

3. Each car that is waiting at a charging station is being charged with a rate independent of the total number of vehicles waiting for service.

4. We model the arrival process of cars to the charging stations as a Poisson process and assume that the charging times for each car are exponentially distributed.

**Comment :** The first assumption is justified, as it makes sense to let the assignment happen on a much faster time scale, see for example Theorem 8.2. These delays, if

taken into account, do not change the balancing behaviour of our approach. The second assumption is slightly unsatisfactory and we give some ideas on how it can be removed in Sections 9.2 and 9.3. However, it may be a valid approximation if the partition cell is small and distances to the charging station can be covered much faster than it takes to charge a vehicle. Assumptions 3 and 4 are quite standard in the modelling for ITS, see for example [Caliskan et al., 2006; Klappenecker et al., 2012] for a similar approach.

Under the above assumptions, the system can be modelled as the simple Markov chain proposed in Section 8.2. To indicate the balancing behaviour of our implementation, we now give two examples of the stationary distribution of this chain. The state of the chain is a vector of length $n$ that reports the number of vehicles currently queueing at each station. Given the above assumptions and the assignment probability formula transition probabilities are readily computed. As described in Section 8.2, for finite queues at each charging station, we obtain a transition matrix and can compute its Perron eigenvector, the systems stationary distribution.

The first example we consider is a system with two stations that can accommodate 50 cars each. We choose arrival rate and departure rate such that new cars arrive on average every 13.33 seconds and charge for an average of 1000 seconds or about 16 minutes. Furthermore, to keep the number of states bounded a car will leave the system if it is assigned to a full car park.

In Figure 9.2a we show the distribution of the difference between the queues of the two stations. This is computed from the stationary distribution of the above described chain. It can clearly be seen that states with a high difference in queue lengths (unbalanced situations) occur very rarely.

As a second example we consider a system with three charging stations with capacity for 20 vehicles each. We set the arrival rate so that a new car arrives on average every 25 seconds and the average charging time and assignment rules are the same as above. Figure 9.2b depicts the distribution of the difference between the fullest and the emptiest queue of all three stations. Again states with large differences in queue lengths can be seen to occur very rarely.

**Comment :**   The values of the distribution of differences in the queue occupancy depicted in Figs 9.2a and 9.2b are not strictly decreasing. Note that in many situations it is impossible that the queues are equally long, for instance, this can never occur if the number of vehicles requiring charging is not a multiple of the number of charging stations.
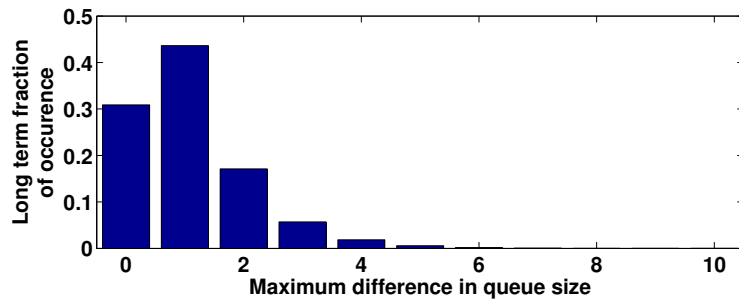
(a)
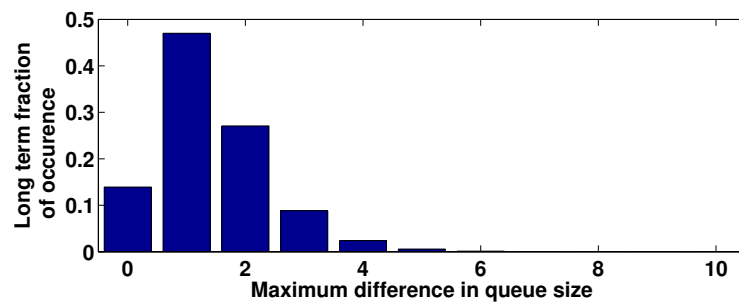


(b)

Figure 9.2.: Stationary distribution of queue length differences for a: two car parks with 50 parking spots each; and b: for three car parks with 20 parking spots each (bottom).

### 9.1.4. Simulations

We now present three simulations to illustrate the efficacy of our approach. Similar to the above discussed analytic tools, in these simulations we do not consider a delay between assignment and arrival of the EV to the charging station.
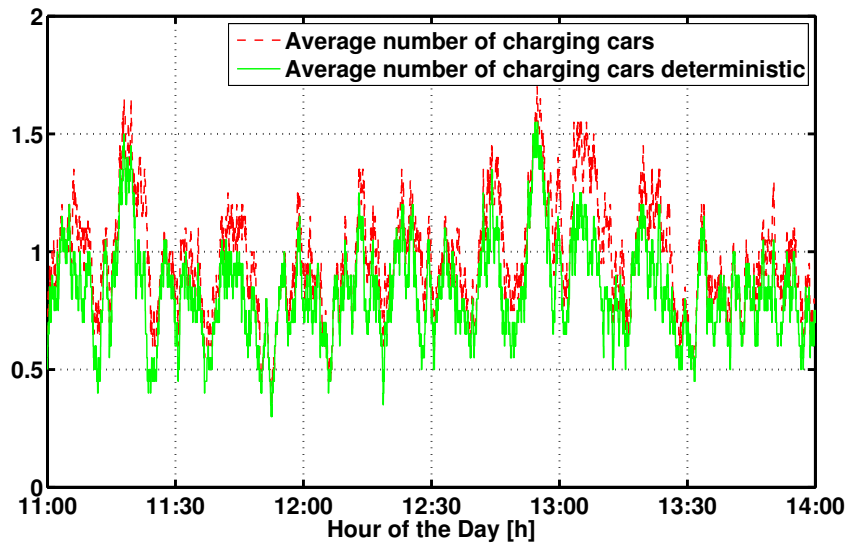
### Deterministic vs. Stochastic

The first simulation compares our stochastic algorithm with a deterministic one. The simulation runs for 8 hours. Moreover, we assume a fixed charge rate of $0.01 kW/s$ per vehicle and a charge request of energy ranging between 1 and $10kWh$, so that the maximum time for recharge (i.e., corresponding to a request of $10kWh$ of energy) is 1000 seconds (or roughly 16 minutes). We also assume that at the beginning of the simulation (8 AM) all charging stations are empty. In the deterministic approach a new car is directly assigned to the charging station having the shortest queue. Figure 9.3 illustrates the results of both approaches.

It can be seen from Figure 9.3 that the deterministic assignment strategy is better than our proposed approach. By its very nature, in the absence of delays the deterministic method is optimal and guarantees that a new car will always be associated with the shortest queue. However, Figure 9.3 clearly indicates that the average number of vehicles waiting at the charging stations using the stochastic algorithm is very close to that obtained from an optimal centralised solution. The two approaches also provide very close results in terms of the variance of the vector of queue lengths (variance equal to zero corresponds to exactly the same number of vehicles queuing at each charging station).

### Assignment Delay

In this simulation we show that the assignment delay, the time between the request for charging and the assignment to a charging station, is within reasonable limits, given that the number of charging stations is high enough to accommodate the number of EVs in the system. The same setup as in the previous Simulation is considered. Figure 9.4 depicts the evolution of the assignment delays and Figure 9.5 depicts the evolution of the average residual charging time of the charging stations. The residual charging time refers to how much time is required until the last car in the queue gets fully charged, and is averaged over all charging stations.

(a)
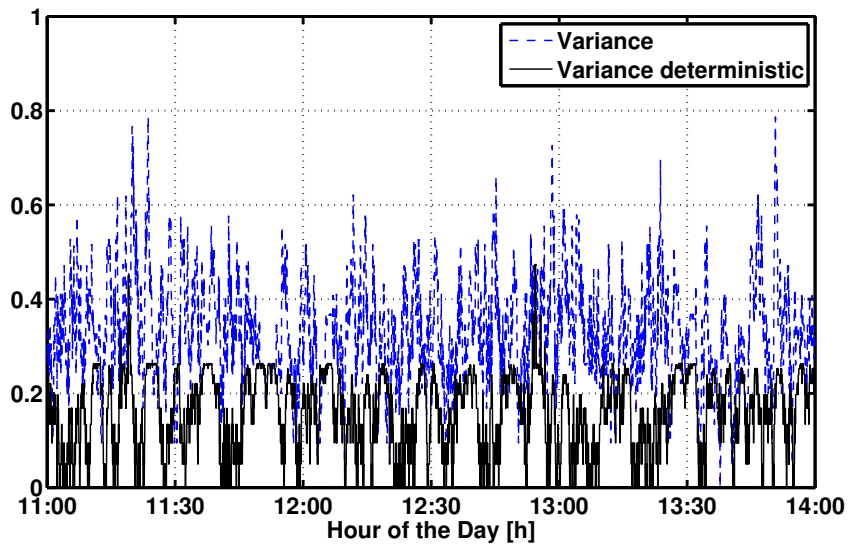


(b)

Figure 9.3.: The decentralised minimum-communication stochastic approach provides results very close to the communication-heavy deterministic centralised approach, both in terms of average number of queuing vehicles at each charging station, above, and in its variance, below.
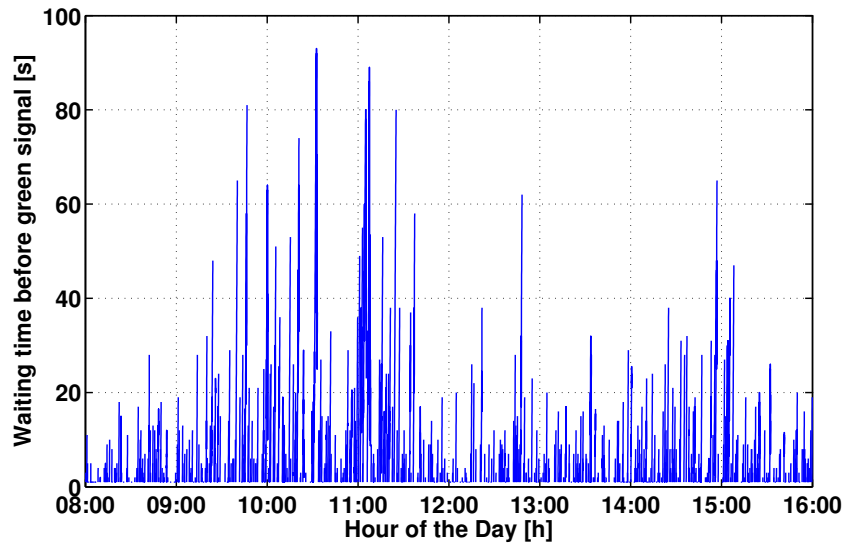
Figure 9.4.: Illustration of the average assignment delay between the moment a vehicle requests charging and the moment it gets associated with a charging station. A probabilistic algorithm facilitates vehicles that have a lower level of battery, but the delay never exceeds two minutes in the simulation example.
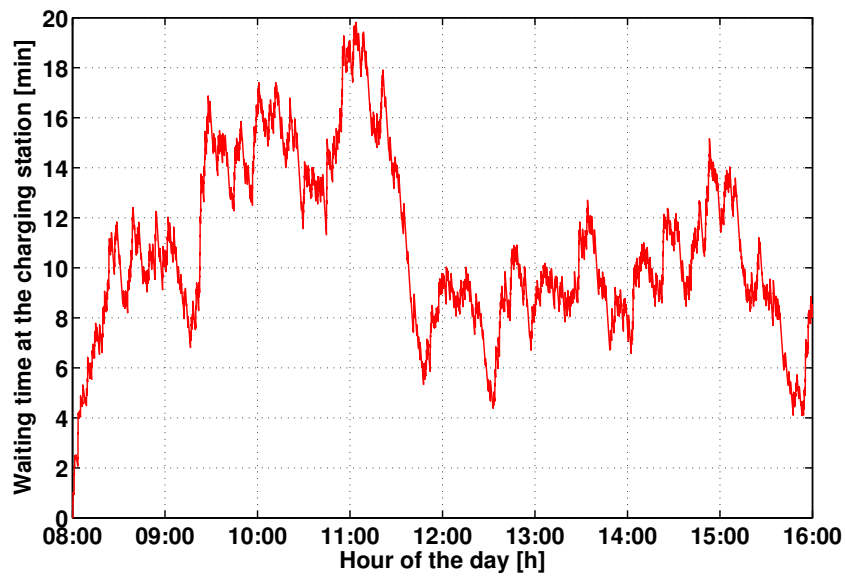


Figure 9.5.: Illustration of the average residual charging time at the charging stations.

**Plug-and-Play Behaviour**

A feature of the proposed approach is that it provides the opportunity of handling a dynamically varying number of available charging stations. In this simulation we give an example of how well our algorithm handles these changes in the system. We assume that we have three available charging stations until 12 PM, at which point two new charging stations join the system and start offering a charging service.

Figure 9.6a illustrates that the maximum difference between the longest queue and the shortest one is usually very small and is particularly large only when the new charging stations become available. As depicted in Figure 9.6b, as the new two charging stations become available, very short waiting times are restored.

## 9.2. Assignment to a Single Car Park

In this section we discuss a second application area for our stochastic assignment approach, namely the choice of a car park in an urban scenario. Motivated by recent publications in the area of car park occupancy prediction [Caliskan et al., 2006; Klappenecker et al., 2012] we first follow their rather artificial approach to consider an isolated car park before developing our algorithm to more realistic scenarios with several car parks in Section 9.3. The choices customers get to make is whether to use this car park or some other possibility which is however not taken into account explicitly in the model. Accordingly, customers simply disappear from the model if they choose not to use the car park. In [Caliskan et al., 2006; Klappenecker et al., 2012] a simple queueing model was proposed and analysed that enables the prediction of the queue length, or car park occupancy, at the estimated time of arrival. This is a very good and useful idea. However the authors have not taken into account the inherent feedback loop that their system exhibits. If a large fraction of customers use the prediction system, the choices they make based on it affect the arrival rate at the car park and thus invalidate the prediction. This section shows how this feedback can be taken into account and how simple decision algorithms can be designed that allow customers to make choices that ensure a high probability of success while keeping the car park occupancy high. To do this we use the following assumptions.

1. The car park is instrumented so that its occupancy can be measured or estimated.

2. This information can be broadcasted to potential customers.

3. We assume that there is an exponentially distributed time between consecutive customer decisions. Also vehicles in the car park stay for an exponentially distributed duration.

These assumptions are standard in studying car park models, see for example [Caliskan et al., 2006; Klappenecker et al., 2012]. The process according to which customers realise
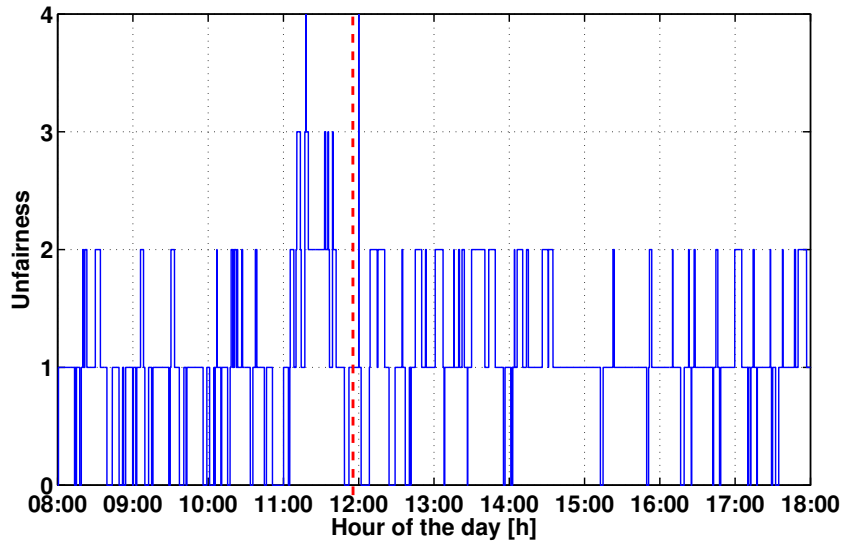
(a)
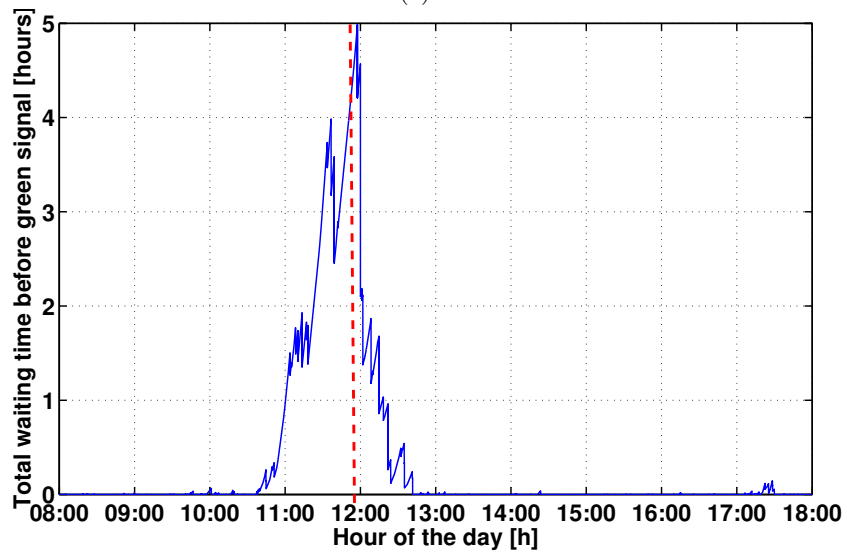


(b)

Figure 9.6.: Figure a measures the fairness of the algorithm in terms of the maximum difference between the busiest and the least busy charging station (larger than 1 implying a sort of unfairness). Figure b measures the aggregate time required before receiving a green signal, which is greatly reduced when the 2 new charging stations are available, as indicated by the red dashed vertical line.
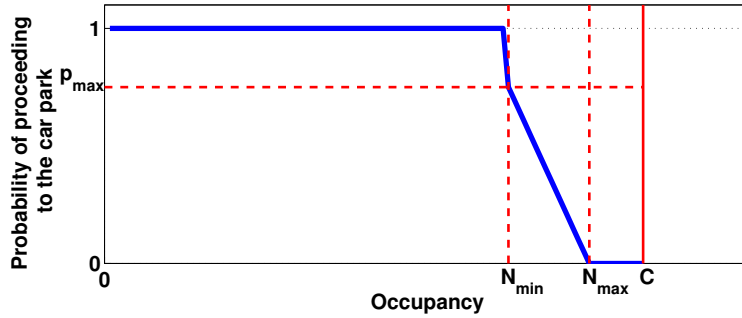
Figure 9.7.: The probability of proceeding to the car park using the RED-approach.

their need for service and start participating in the system is Poisson. The use of Poisson processes to model bursty traffic is well established. Furthermore, the memorylessness of these processes eases analysis.

Our objective here is to develop algorithms which allow vehicle owners to make informed decisions as to whether a car parking space will be available at the car park or not. We model customer decisions in a stochastic fashion with a probability of travelling to the car park that depends on the occupancy of the car park at the time of the query. We assume that customers are more likely to use the car park when the occupancy is low. Thus it seems reasonable to suggest the algorithm given in Algorithm 9.3 for making a decision as to whether or not to go to the car park. This algorithm is based on RED [Floyd and Jacobson, 1993], which we used also in Chapter 6.5.

---
**Algorithm 9.3** Single Car Park
---
$N \leftarrow$ occupancy of car park

$p \leftarrow \begin{cases} 1, & \text{if } N < N_{min}, \\ 0, & \text{if } N > N_{max}, \\ p_{max} \frac{N_{max}-N}{N_{max}-N_{min}}, & \text{otherwise,} \end{cases}$

Go to car park with probability $p$.

---

In Algorithm 9.3 we use the parameters $0 \leq p_{max} \leq 1$, and $N_{min}, N_{max} \in \mathbb{N}$ with $N_{min} < N_{max} \leq C$, where $C$ is the total capacity of the car park. The occupancy of the car park is broadcasted to all participating cars and will be updated in regular time intervals. The probability function defined in the algorithm is shown in Figure 9.7. Note, that the drivers will go to the car park with probability 1 when the occupancy is low and will not go there when the occupancy is high. Note also, that this algorithm can be easily implemented using GPS devices or smartphones.

### 9.2.1. Model

The analysis of this otherwise simple system is made difficult if we allow the following. If customer $i$ decides to go to the car park this takes a time $\tau_i$. We assume that it is possible to identify an upper bound $T > 0$ on $\tau_i$ for all $i$. As we can design the system, we allow for the available occupancy data to be updated in discrete time steps only. In fact, we set the time between updates to be equal to $T$. This enables us to use the tools developed in Chapter 8 but makes the implementation less efficient. However upper bounds regarding the performance are still valid. Updates in this fashion yield discrete time steps $k \in \mathbb{N}$ where the $k'$th time interval is $[kT, (k+1)T]$. We further assume that a car which arrives to the car park during a period when there are no free parking spaces will wait outside the car park until a space becomes available. Denote by $N(k)$ the number of cars parked in the car park plus the number of cars waiting for a parking space at time $kT$. The evolution of $N(k)$ can be described as a difference equation of the form

$$N(k+1) = N(k) + A(k) - D(k), \tag{9.2}$$

where $A(k)$ is the number of cars that arrive to the car park during the interval $[kT, (k+1)T]$ and $D(k)$ is the number of cars leaving from the car park in that same interval. We assume that cars stay parked for a random time described by an exponentially distributed random variable with fixed rate $\mu > 0$. We assume that cars stay, on average, much longer than the time between broadcasts, i.e. $T\mu \ll 1$. Hence the departure process, $D(k)$, changes slowly enough, so that we can approximate it as following a Poisson process that terminates once $N(k)$ cars have left. Further this Poisson process has rate $\mu Q(N(k))$, where $Q(N(k)) = \min\{N(k), C\}$, where $C$ is the car park capacity. Note in particular that this assumption ensures that the evolution of the random variable is independent of all other cars and the occupancy process of the car park. Accordingly, the distribution of $D(k)$ is described by

$$P(D(k) = t | N(k) = n) = e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!} \tag{9.3}$$

for all $n \in \mathbb{N}$ and all $t = 0, 1, \ldots, Q(n) - 1$ and

$$P(D(k) = Q(n) | N(k) = n) \tag{9.4}$$
$$= \sum_{t=Q(n)}^{\infty} e^{-Q(n)\mu T} \frac{(Q(n)\mu T)^t}{t!}$$
$$= 1 - e^{-Q(n)\mu T} \sum_{t=0}^{Q(n)-1} \frac{(Q(n)\mu T)^t}{t!}.$$

These equations coincide with Equations (8.7) and (8.8) and our model has now taken the form of the system described in Section 8.3 and we can apply the results developed there. The system that we have described clearly allows for the undesirable situation

where customers arrive to a full car park and have to wait to gain entrance or leave to find parking at a different location. Theorem 8.4 guides the choice of algorithm parameters to ensure that this undesirable situation is a rare event.

To give a qualitative idea of the order of magnitude of the probability that the number of arriving customers exceeds the available capacity in Theorem 8.4, Figure 9.8 shows this probability for different values of $\lambda$ and $\mu$. In particular, here we assume that the car park has a capacity for 100 vehicles, $m = 80$ and $n = 90$, we choose parameters $N_{max} = 90$, $N_{min} = 75$ and $p_{max} = 0.75$ for Algorithm 9.3, $\tau_i$ is equal to 5 minutes for all vehicles, and the average time between queries (i.e., $1/\lambda$) varies between 10 and 30 seconds, and the average staying time between 0.5 and 1.5 hours (i.e., $1/\mu$). It can be seen that situations with a high probability of overflows occur only when the system is under heavy load, that is when cars arrive with a high frequency and stay for a long period. In Figure 9.9 we study the dependency of the overflow probability on the car parks occupancies at the present and one step back in the past, $m = N(k-1)$ and $n = N(k)$. Here we choose the parameters $C = 100$, $N_{min} = 75$, $N_{max} = 90$, $p_{max} = 0.75$, $\lambda = \frac{1}{20}$, $\mu = \frac{1}{3600}$ and $\tau_i$ is equal to 5 minutes for all vehicles. It can be seen that the probability of an overflow is quite low. It is always 0, when $m \geq N_{max}$. The overflow probability is high only when $n$ is close to $C$ or $n > C$, and at the same time $m < N_{max}$. Note that this figure is slightly misleading: Even though there are some situations that yield a significant probability of an overflow occurring at the next step, these situations themselves occur extremely rarely as they require a large number of cars to arrive during the $k'$th time interval.

Theorem 8.4 gives an upper bound to the car park's overflow probability, a complementary lower bound is obtained from Theorem 8.5. To give a quantitative idea of this bound, we refer to Figures 9.10 and 9.11, which were created in the same setup and with the same parameters as Figures 9.8 and 9.9 for Theorem 8.4. From visual inspection it seems that Figures 9.8 and 9.10, and 9.9 and 9.11 are practically the same. This indicates that the upper and lower bounds computed according to Theorems 8.4 and 8.5 are quite close, and thus they give practical insight into the dynamics of our proposed assignment scheme. However, the figures are not identical, as can be seen in Figure 9.12, where we compare Figures 9.8 and 9.10 for a fixed average staying time of an hour, and in Figure 9.13, where we compare Figures 9.9 and 9.11 for a fixed value of $N(k-1) = 75$. In both cases, the actual probability has to lie between the lower and the upper bounds suggested by the aforementioned theorems.

Figure 9.8.: Probability that the number of arriving customers exceeds the available capacity as a function of average time of arrival and average time of staying.



Figure 9.9.: Probability that the number of arriving customers exceeds the available capacity as a function of the car park's occupancy.
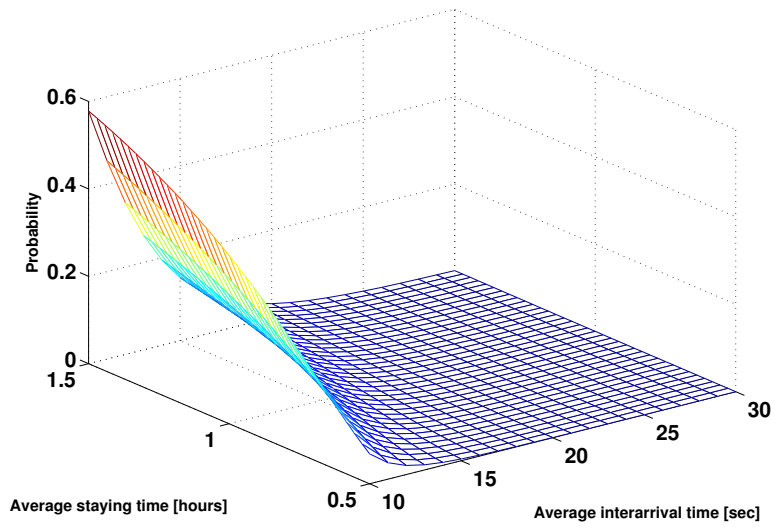
Figure 9.10.: Probability that the number of arriving customers exceeds the available capacity as a function of average time of arrival and average time of staying.
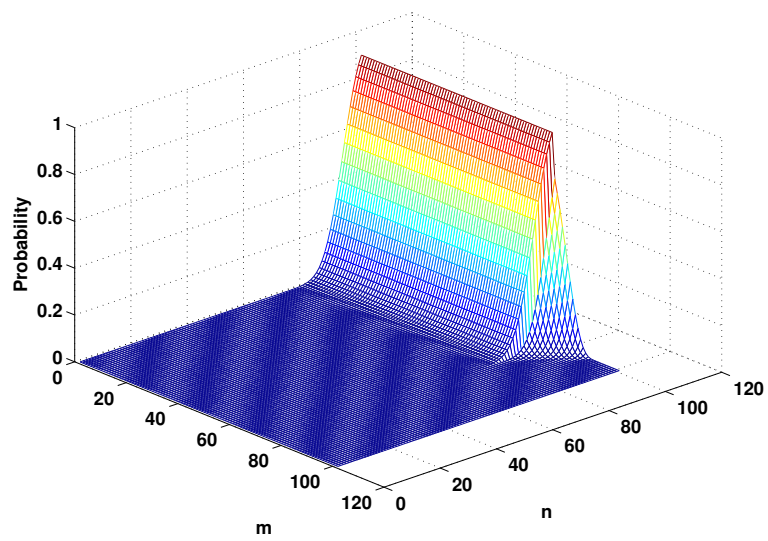


Figure 9.11.: Probability that the number of arriving customers exceeds the available capacity as a function of the car park's occupancy.
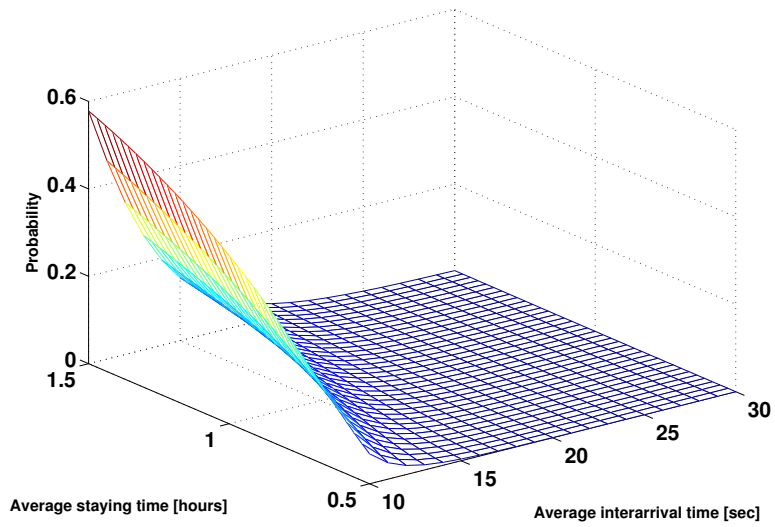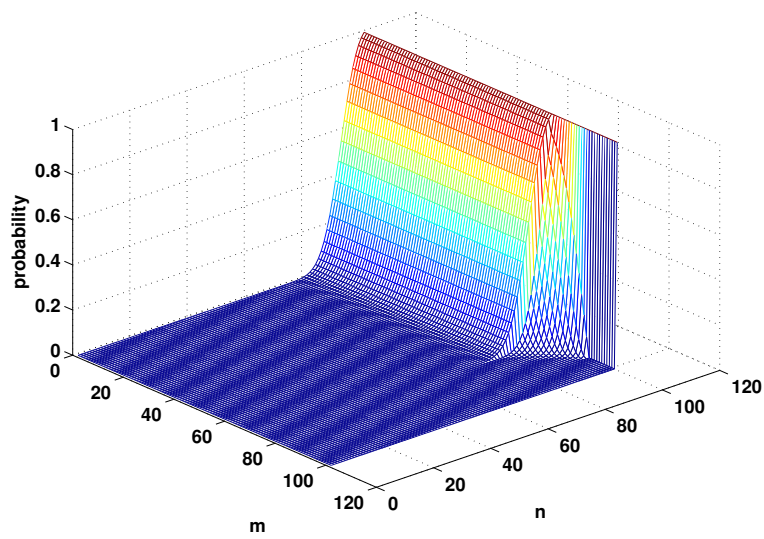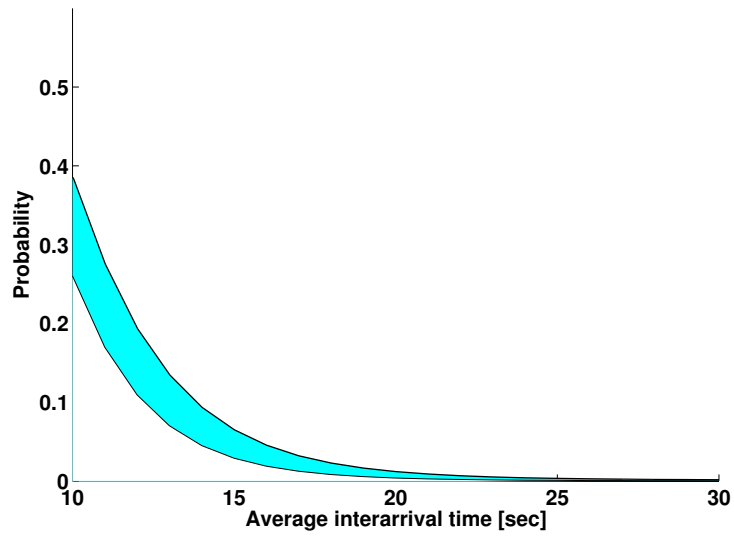
Figure 9.12.: Comparison of Figures 9.10 and 9.8 for an average staying time of one hour as functions of the average time between queries of vehicles.
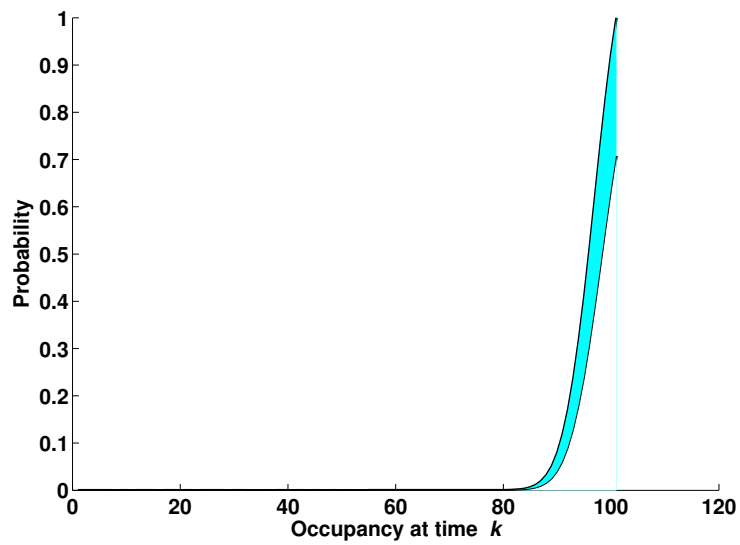


Figure 9.13.: Comparison of Figures 9.11 and 9.9 for $N(k-1) = 75$ as functions of $N(k)$.

### 9.2.2. Simulations

We now show some results obtained from simulations using Sumo. We consider a scenario with a single car park, coloured in red, in the centre of our grid-like city depicted in Figure 9.15. It has capacity for 100 cars and is empty at the beginning of our simulation. Over a time of 3 hours, vehicles appear at random locations throughout the city. We model the arrival of new cars as a Poisson process with expected inter-arrival time of 10 seconds. Cars that decide to drive to the car park and find an available spot stay parked for an exponentially distributed random time with mean 20 minutes and then they disappear. We simulate two instances of this scenarios. In the first scenario occupancy information is not available to the cars and all cars decide to go to the car park. In the second scenario cars have access to the occupancy information. The occupancy information is updated every 100 seconds. So not only do cars experience a delay between making a decision and arriving to the car park, they are also using non-realtime data to make their decisions. In this scenario cars make their decision randomly according to Algorithm 9.3, with $N_{min} = 80$, $N_{max} = 95$ and $p_{max} = 0.75$. In Figure 9.14 the number of cars that are at the car park in both scenarios is depicted. We can see that in the scenario with feedback (green line) the car park's occupancy stays below the capacity, so that no car arrives to a full car park and there are always some spare spaces available. In the scenario without feedback (red line) on the other hand we can clearly see that that the available capacity of the car park is not enough to satisfy the demand. A large number of cars arrives to a car park with no free spaces. The drivers have to wait for someone to leave or have to find an alternative parking facility. This causes an unnecessary waste of time and fuel and contributes to congestion and pollution.

## 9.3. Multiple Car Parks

In the previous section we concentrated on a single car park and cars could only choose between going to the car park or going somewhere else. In this we followed the approach suggested in [Caliskan et al., 2006; Klappenecker et al., 2012]. Clearly "somewhere else" is most likely going to be another parking facility. In this section we investigate how our approach can be extended to the more realistic situation, where the drivers have to choose among several parking facilities. In accordance with the general setup in this chapter, the driver picks a zone in the city in which he wishes to park. In the zone a number of car parks are available and the driver has no preference among those. In particular, customers make a decision to travel to a particular car park based on Algorithm 9.4. This can be viewed as an extension of Algorithm 9.3 to the multiple car park case and returns out focus to the assignment problem discussed at the beginning of this chapter. In this section we shall focus on the effect of the assignment delay on the performance of the algorithm and in this sense the discussion in this section is also an extension to that in Section 9.1.

Put into words, Algorithm 9.4 assigns new customers to the available car parks in a stochastic fashion with a probability of going to a particular car park that is propor-
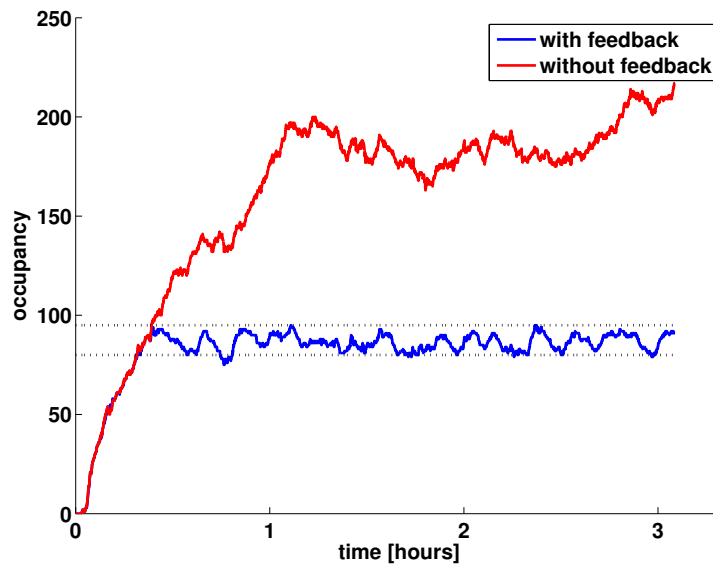
Figure 9.14.: Comparison of occupancy with (red line) and without (blue line) feedback. The dotted lines show the desired lower and upper occupancy of the car park.

---

**Algorithm 9.4** Choose Among Multiple Car Parks

---

**for** $j = 1$ to $n$ **do**
　　$X_j \leftarrow$ number of free spaces in car park $j$
**end for**
**for** $j = 1$ to $n$ **do**
　　$p_j \leftarrow \frac{X_j}{\sum_{i=1}^{n} X_i}$
**end for**
Go to car park $j$ with probability $p_j$.

---

147

tional to the number of available parking spots at this car park. We use the following assumptions throughout this section

1. Car parks can measure or estimate their current occupancy and can make this information available to all customers, for example using a broadcast signal.

2. There is no communication from the vehicles back to the infrastructure and hence no reservation system in place.

3. As customers query the system for a parking space they are immediately assigned to one car park and we also assume that each car proceeds to its assigned facility.

4. There is a delay between assignment of a vehicle to a car park and the vehicle's arrival.

5. Finally, we also assume that cars leave the parking lots in a random fashion, in such a way that the total arrival rate on average is equal to the total departure rate so that the system is in equilibrium.

### 9.3.1. Model

We now derive a model that reflects the dynamics of an assignment system following our approach and using Algorithm 9.4. Assume a cell with $n \in \mathbb{N}$ car parks. The behaviour of the system is determined by the following factors: (1) the statistics of the arrival process for the cars, (2) the statistics of the departure process, (3) the assignment rule, (4) the delays between assignment and parking. For our model we make the following additional assumptions:

1. The process according to which new customers query the system for a parking space is Poisson with rate $\lambda > 0$

2. Each car independently departs after an exponential parking time. Let $C_1, \ldots, C_n$ be the capacities of the parking lots, and let $X_1(t), \ldots, X_n(t)$ be the numbers of free spaces at time $t$. Then the probability that the next departure occurs from parking lot $j$ at time $t$ is

$$q_j(t) = \frac{C_j - X_j(t)}{\sum_{i=1}^n C_i - X_i(t)} \tag{9.5}$$

3. A customer that is assigned to car park $j$ experiences a delay before arriving at the car park which depends on its location and the location of the assigned parking facility, and perhaps also some exogenous factors causing randomness. We thus model these delays as i.i.d random variables and assume that their mean is $\tau_j > 0$, which only depends on the assigned car park.

$\lambda$ refers to the rate at which cars make a decision. Note that, as all cars are assigned to one of the car parks in the zone, this corresponds to the aggregate arrival rate at all car

parks. It is possible to calculate the probability that the number of arriving customers exceeds capacity, in a manner similar to the single car park case. A more pressing issue in this case is whether the protocol balances the load, and whether flapping is avoided. Flapping is a manifestation of instability and occurs when car parks take turns being full. Clearly, this situation should be avoided, and thus, the main question of interest now is to analyse the stability and fairness of the protocol, and to find the dependence on the number of parking lots, the number of available spaces, the arrival rate and the delays.

**Comment :** The assignment rule in Algorithm 9.4 can be adapted to achieve a number of different objectives. For example the individual assignment probabilities can be tuned to divert traffic from certain areas as may be necessary to mitigate congestion or pollution peaks or to reflect a pricing structure.

It is challenging to analyse the stochastic model in full detail, so we begin with the analysis of a simplified deterministic model which describes the so-called fluid limit. This model should apply in the case where the arrival rate and the capacities of the parking lots are very large. In this limit the discrete state of each car park, i.e. its occupancy, is replaced by a continuous variable, and we can view the traffic as a fluid which flows into and out of the car parks. Note that fluid models have often been employed to describe urban traffic, see for example [Garavello and Piccoli, 2009]. The traffic enters and leaves the zone as a steady stream. The entering stream is split into $n$ parts, which proceed to the $n$ parking lots. The amount in each substream varies over time, depending on the available capacity at each car park. There is a delay between assignment and arrival at the car parks. Each lot generates a departing stream, and these combine to form the outgoing stream, which according to our assumptions again has rate $\lambda$. The evolution of this deterministic fluid model is described by a delay differential equation.

Let $C_1, \ldots, C_n$ be the capacities of the parking lots, and let $X_1(t), \ldots, X_n(t)$ be the amount of free space in each lot at time $t$. Note that $0 \leq X_j \leq C_j$, and that $X_j$ is now a continuous random variable. We use the assignment rule according to Algorithm 9.4 and the departure rule (9.5). Thus the variables satisfy

$$\frac{dX_j(t)}{dt} = -\lambda\, \theta(X_j(t)) \frac{X_j(t - \tau_j)}{\sum_i X_i(t - \tau_j)} + \lambda \frac{C_j - X_j(t)}{\sum_i (C_i - X_i(t))}, \tag{9.6}$$

where $\tau_j$ is the delay associated with lot $j$, and where $\theta(X_j(t)) = 1$ for $X_j(t) > 0$ and $0$ else. The factor $\theta(\cdot)$ enforces the condition that the solution to the delay differential equation satisfies $X_j(t) \geq 0$ for all $t$. If we now further assume that $X(t) > 0$, for all $t$, then we obtain

$$\frac{dX_j(t)}{dt} = -\lambda \frac{X_j(t - \tau_j)}{\sum_i X_i(t - \tau_j)} + \lambda \frac{C_j - X_j(t)}{\sum_i (C_i - X_i(t))}. \tag{9.7}$$

It immediately follows that

$$\frac{d}{dt} \sum_{j=1}^{n} X_j(t) = 0 \tag{9.8}$$

and hence the total number of available parking spaces is constant. Define this total to be

$$N = \sum_{j=1}^{n} X_j(t) \tag{9.9}$$

and also define the total capacity of the zone to be

$$C = \sum_{j=1}^{n} C_j. \tag{9.10}$$

Then still assuming that the variables $X_j$ are always positive, we can use Equations (9.9) and (9.10) to obtain the equation

$$\frac{dX_j(t)}{dt} = -\frac{\lambda}{N} X_j(t - \tau_j) + \frac{\lambda}{C - N} (C_j - X_j(t)). \tag{9.11}$$

This equation describing the dynamics of our fluid limit coincides with Equation (8.29), which was discussed in detail in Section 8.4 and we know from Theorem 8.6 that its constant solution is exponentially stable, if the delay satisfies

$$\tau_j < \left(\frac{N}{\lambda}\right) \frac{\cos^{-1}\left(-\frac{N}{C-N}\right)}{\sqrt{1 - \left(\frac{N}{C-N}\right)^2}}. \tag{9.12}$$

Approximately, the condition for stability is

$$\lambda \tau_j \leq \frac{\pi N}{2}. \tag{9.13}$$

Note that $\lambda \tau_j$ is the total number of arrivals during the interval between assignment of the parking lot and arrival at the parking. So condition (9.13) says that the constant solution is stable if this total arrival number during a delay is less than the total number of available parking spaces.

Further, according to Theorem 8.6 the constant solution is given by

$$X_j(t) = \frac{C_j}{C} N. \tag{9.14}$$

This solution describes the situation where the traffic is shared among the lots proportionally to their capacities. This is in accordance with intuition and suggests that our algorithm is fair in this sense.
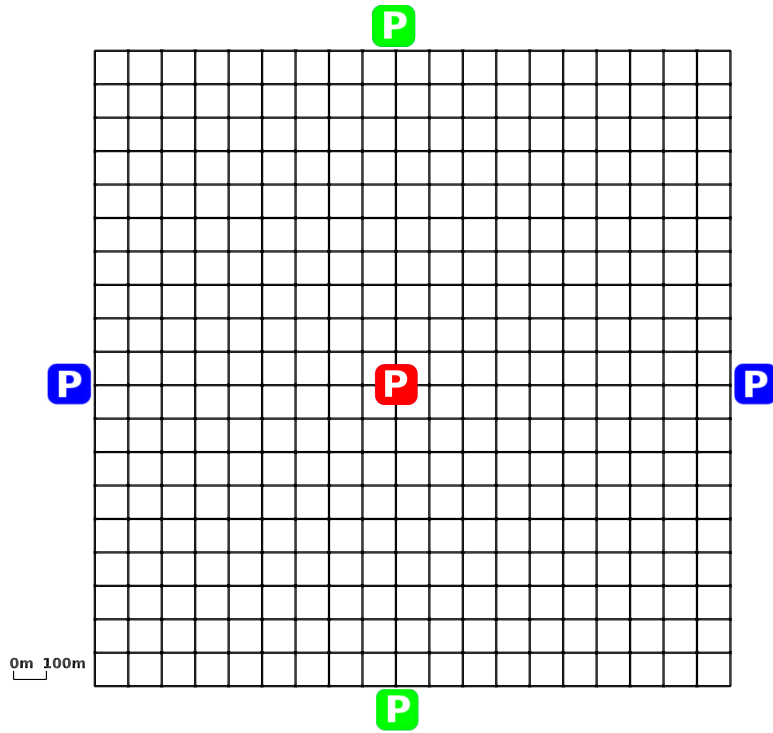
Figure 9.15.: Grid-like road network used for our simulations.

Theorem 8.6 gives an additional condition for stability of the constant solution, namely

$$N > \frac{C}{2} \tag{9.15}$$

So if at least half the parking spaces are empty then the constant solution is stable and in this situation the stability is independent of the delay. This is a remarkable result.

### 9.3.2. Simulations

We now present simulations to illustrate the efficacy of our algorithm.

#### Efficiency of Balancing Strategies from a System Point of View

In this simulation we compare our algorithm to a join-the-shortest-queue like deterministic assignment strategy using the realistic traffic simulator Sumo on a grid-like road network, depicted in Figure 9.15. Vehicles are being routed between an origin and a destination street along the shortest path. To obtain the desired simulation results, we choose roads in the city that are supposed to contain a car park or are an entry or exit

point for cars to the city; these virtual locations are not explicitly taken into account in Sumo, rather we use Matlab to keep track of car park occupancies and origins and destinations of vehicles. In practise we use Sumo to run the simulation until a new car arrives to the city and makes a decision or a previously parked car finishes its service and departs from the car park. Each of these events is generated in Matlab, which adds the new event to Sumo and consecutively starts a new Sumo simulation.

We consider four car parks in our grid-like city coloured in blue and green in Figure 9.15, with capacity for 40 cars each, distributed over the city as well as 4 main access roads located in the corners of the network. Over a time of about three hours 1000 cars arrive to the city according to a Poisson process with average inter-arrival time of 10 seconds and they arrive on each access road with the same probability. Upon arrival they query the occupancies of the car parks and choose one of them according to Algorithm 9.4, then they drive to the chosen car park. Vehicles then stay parked for an exponentially distributed time with mean 20 minutes. The available car park occupancy information is updated only when cars arrive at a car park or leave it, so no communication from the vehicles to the car parks is required. In Figure 9.16a we plot the occupancy of each car park at the time instants at which new cars arrive to the city. As we can see, our approach reasonably balances the occupancies.

In Figure 9.16b we present the same result for a slightly different simulation, where we use a deterministic assignment rule. Namely, we assign cars always to the car park with the lowest occupancy at the time of the vehicle's query. Although this assignment rule intuitively seems reasonable, if it takes a long time for cars to drive to their car park and if many other vehicles arrive in this time, then its performance is poor. By comparing Figures 9.16a and 9.16b it can clearly be seen that our algorithm outperforms the deterministic assignment rule in the sense that cars are better balanced among the available car parks. Specifically, our stochastic approach decreases the time averaged variance of the distribution of parked cars over all car parks from 29.85 to 9.23.

**Effectiveness of Balancing Strategies from a User Point of View**

The objective of the previous simulations was to demonstrate the effectiveness of feedback strategies in terms of efficient usage of the infrastructures. We claim that balancing the vehicles among the available car parks is an efficient way of utilising the available infrastructure. In this simulation we investigate the effect of the delay between between assignment and arrival of a vehicle to a car park on the performance of the two assignment rules discussed in the previous simulation. The objective of this simulation is to show that stochastically balancing the vehicles among the available car parks is in fact a smart strategy that performs better than the strategy of associating vehicles with the emptiest car park. For this purpose, we take the perspective of the users, and as a measure of effectiveness, we consider the percentage of unsatisfied users with respect to the overall number of users. Unsatisfied users are users that arrive to find a full car park. To this end we ran several simulations each simulating a time of three hours, each of them
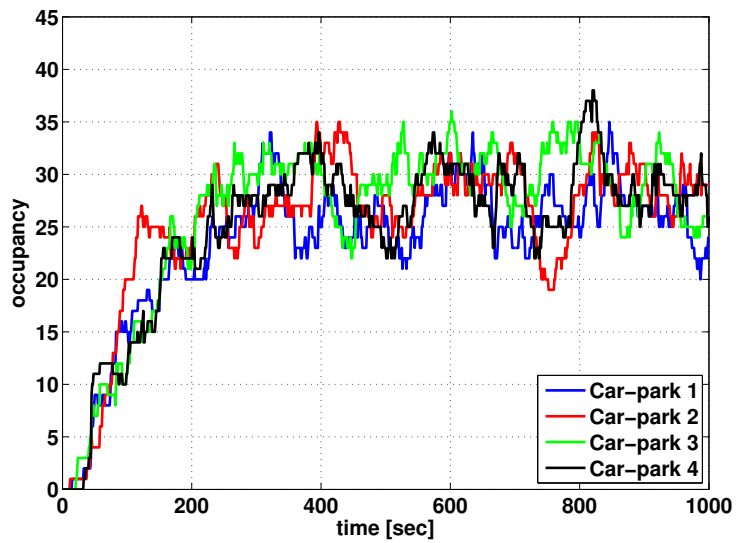
(a) Algorithm 9.4.
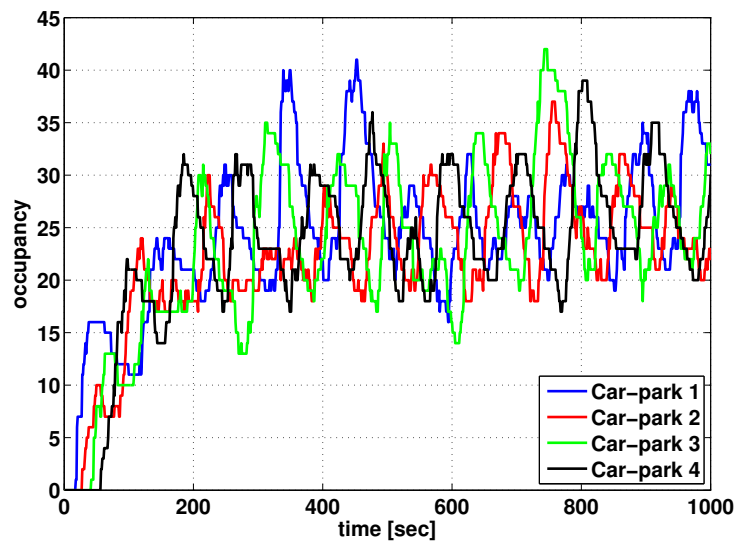


(b) Deterministic assignment.

Figure 9.16.: Number of occupied spaces at each parking lot using Algorithm 9.4 in
Figure 9.16a and using the deterministic assignment rule in Figure 9.16b.
Vehicles are clearly more balanced in 9.16a as the variance is lower.

153
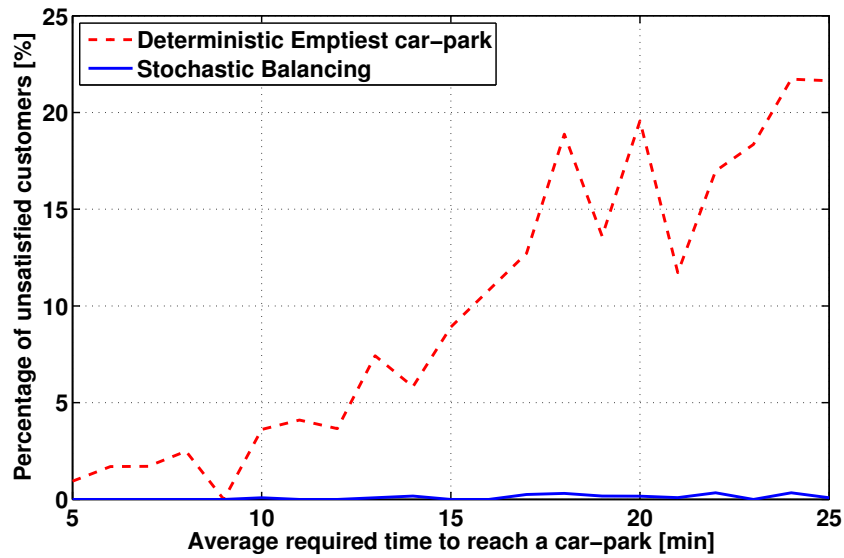
Figure 9.17.: The percentage of unsatisfied users increases if the "emptiest car park" strategy is used, when the distance from the car park increases. On the other hand, the stochastic balancing strategy performs well even in case of long distances, thanks to the intrinsic feedback flavour in the strategy.



Figure 9.18.: Number of occupied spaces at each parking lot. Note that the values are close to the expected average occupied places, shown with the horizontal lines.

for a different value of the average time between the assignment of a driver to a car park and the arrival to the car park. These values range from 5 to 25 minutes. For each car a random number, uniformly distributed bet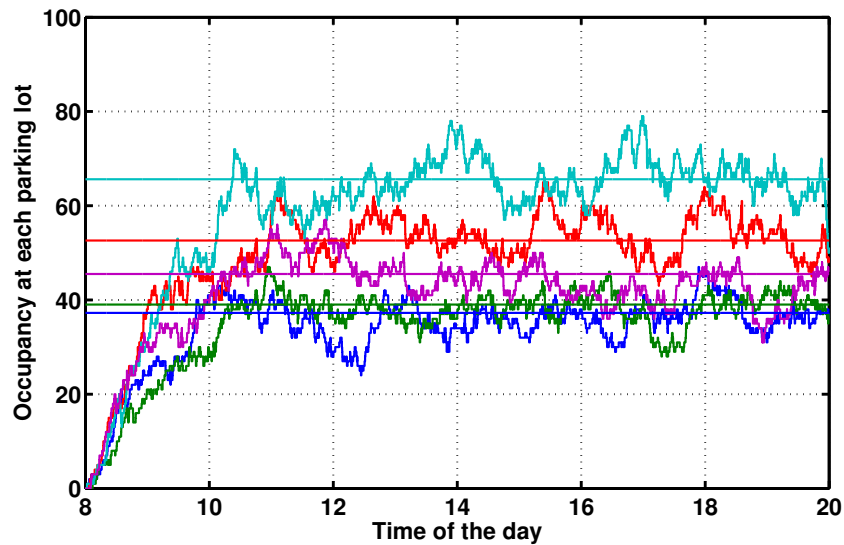ween $-2$ minutes and $+2$ minutes is added to the average delay. Figure 9.17 clearly shows that the percentage of unsatisfied users increases when the average travel time before getting to a car park increases.

**Fairness among Car Parks**

Car parks are often privately owned and will only participate in a scheme such as ours if properly incentivised. At the very least they will want their fair share of customers. From Theorem 8.6 we know this to be the case for the long-run behaviour of the fluid limit. We present the following simulation to validate this kind of fairness for a more realistic system. We show that each car park obtains on average a fraction of all customers that is proportional to its total capacity. For the purpose of the simulation, we assume that there are 5 available car parks with different capacity $C_i$. We also assume that at the beginning of the simulation all parking lots are empty. The arrival process of new cars requiring a parking space is Poisson with average inter-arrival time equal to 15 seconds. We also assume that each car independently departs after an exponential parking time, which on average is one hour long. According to Theorem 8.6, we can predict the average occupancy of each parking lot as $3600/15 * C_i/C$. Figure 9.18 compares the number of occupied parking spaces at each parking lot with the expected average occupancy of each parking lot. Note that the probability of a new car to be associated with a parking lot is proportional to the number of empty spaces. Thus, the average occupancy of each parking lot is proportional to its capacity.

**Comparison to an Occupancy Independent Assignment Strategy**

According to Algorithm 9.4, each vehicle is associated with a parking lot according to the occupancies of all car parks. A simpler and easier to implement algorithm would be to directly assign vehicles to car parks proportionally to each car parks total capacity in an open loop fashion. Obviously, this does not require the monitoring of the parking lots, and if we assume that the capacities are known then no communication is required. On average the car park occupancies should coincide with the ones seen if the system uses Algorithm 9.4, however this open loop approach introduces negative effects in the transient behaviour. In this simulation we compare the two approaches.

For the sake of comparison, we repeat the simulation from the above simulation, assuming that a new car arrives on average every 8 seconds. Arrivals are now so frequent that it becomes likely that a car may not find an available place when it gets to the car park, as there are very few available places on average. We assume that vehicles that do not directly find a parking spot leave the system and are discarded from our model. We compare the Algorithm 9.4 with the simpler open loop strategy. We use the number of vehicles that do not manage to find an available place when they get to the parking

lot as a performance indicator. At the end of the simulated day this happens to 184 cars with Algorithm 9.4, while this number is 289 with the open loop approach. Also, Figure 9.19 compares the number of available places with the two algorithms. As can be seen from the figure, fewer places are available with Algorithm 9.4. This means that cars that would have been discarded in the open loop system have been accommodated, which indicates a more efficient utilisation of the available infrastructure.

**Cooperative vs Non-Cooperative Vehicles**

When introducing a new protocol into a system one has to ensure backwards compatibility. This is a well-known issue in communication networks, where sometimes it is not possible to implement new and superior algorithms as they do not compete well with older implementations and introduce unfairness and performance degradation in the whole system. See for example Appendix A for a related discussion. The same problem may occur here as people join our proposed ITS parking system at different times or maybe not at all. Further people may decide not to comply with the systems recommendations and choose a car park based on personal preference. An algorithm based on reservation policies for example might not be robust enough against such occurrences and the performance might decrease. Figure 9.20 compares the number of empty places at each parking lot in the cases that all vehicles follow the standard procedure of asking the infrastructure where to go, and in the case that about 20% of the vehicle choose the parking lots according to the open loop policy described in the simulation above. To make the figure more readable, we compare the available places at only 3 parking lots. As can be seen from the figure, results do not change drastically if some of the vehicles do not follow the stochastic algorithm; in particular, in none of the two cases vehicles got discarded by the parking lots.

**The Benefit of Load Balancing to the User**

The objective of balancing car park occupancies is to avoid localised congestion, pollution peaks, and to increase the probability of a given driver finding an available space. To do this, drivers are directed to a number of nearby car parks. The cost of this strategy could sometimes be increased driving time for individual drivers and some drivers being further away from their destination. Quantifying these effects in a very detailed manner is beyond the scope of this thesis. However, we give the following simulation to showcase the potential benefits of our approach to the users, and to address in some manner the above concern. We consider again the grid-like network depicted in Figure 9.15. We assume there are the two car parks coloured in blue, located at the left centre and the right centre of the map. Each of these has capacity for 100 cars and is initially empty. Now assume that there is an event happening close to the left car park (car park A), which 200 vehicles, located uniformly distributed over the map, wish to attend. They all start their journey at the same time. We use the mobility simulator Sumo and consider two scenarios: (i) All vehicles drive to car park A and the first 100 to arrive find a space, the rest has to drive to the right car park (car park B) from where the drivers
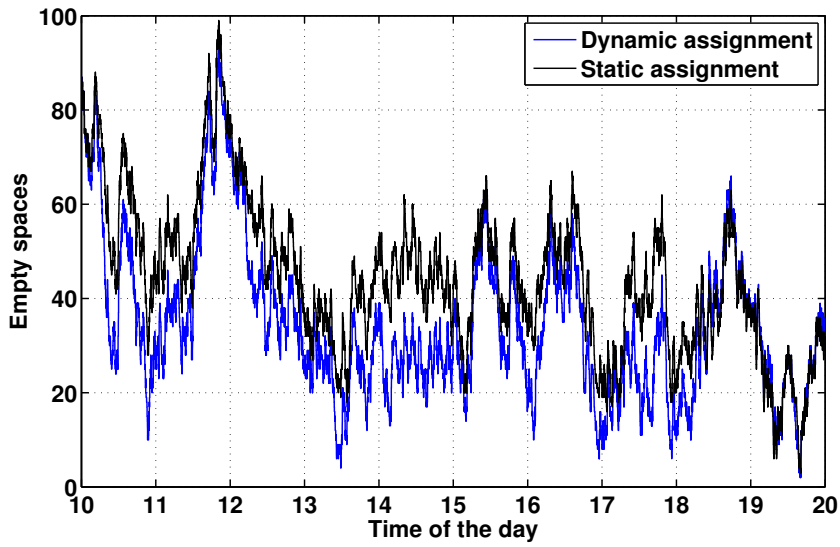
Figure 9.19.: Number of available places using a dynamic strategy that feeds back to the vehicles an information that takes into account the current occupancy of the parking lots, and a static approach. The dynamic strategy makes a more efficient usage of the available infrastructure, as on average less parking places are left empty.
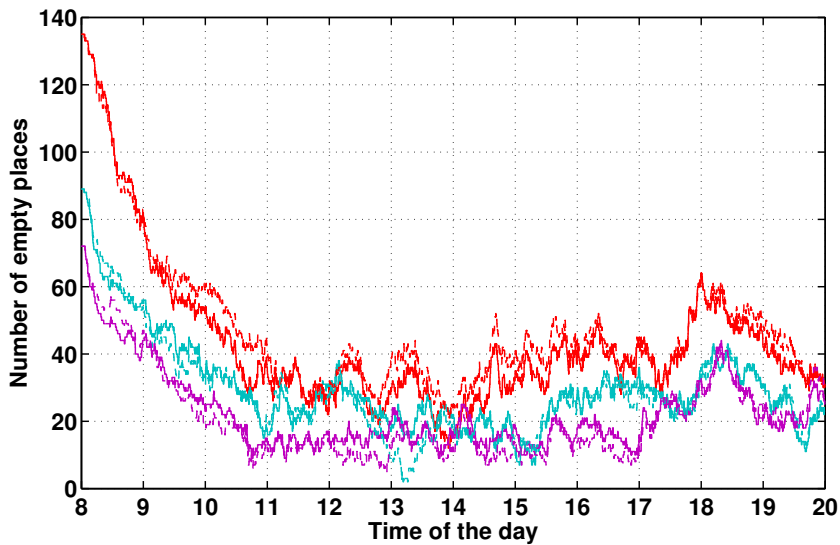


Figure 9.20.: The algorithm performs in a very similar manner in case some of the vehicles follow a slightly different static algorithm to choose their car park.
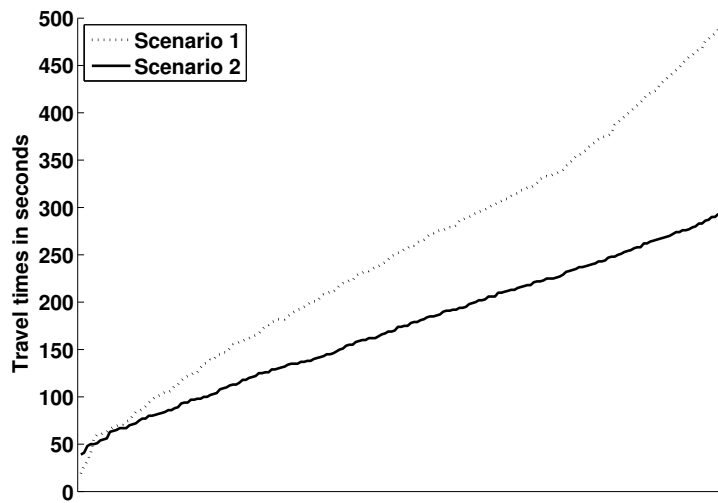
Figure 9.21.: All recorded travel times sorted by length to the first car park measured from both scenarios.

will have to use public transport or walk to arrive at car park A. (ii) All vehicles use our stochastic assignment rule which in this scenario simplifies to tossing a fair coin to decide which car park to go to. Once one of the car parks is full, all later arriving vehicles have to drive to the other car park. All drivers that end up in car park B again have to use public transport to reach car park A. In the first scenario 100 cars have to relocate from car park A to car park B, while in the second scenario this number is significantly smaller. In our simulation, only two cars had to relocate, a significant reduction in unnecessary travel, saving time and reducing congestion and pollution. Further, even without the additional relocation journeys, the second scenario is already superior. Fig. 9.21 reports the travel times from the original position to the first car park that is reached in both scenarios. It can clearly be seen that almost all travel times are higher in the first scenario. In fact the average travel time is 258 seconds in the first scenario and 174 seconds in the second scenario. This is due to local congestion around car park A due to the mass of vehicles driving there. A short video showing the junction at which car park A is located and the surrounding streets can be found at http://www.hamilton.ie/aschlote/sumo_movie.mov.

## 9.4. Shared Bicycle Systems

In this section we propose the last application area for our decentralised stochastic load balancing approach in the ITS context, although there are many more potential applications. This last application area is in shared bicycle systems. Shared bicycle

systems are expected to grow significantly in the coming years and take part in several important aspects of modern transport. They differ from previously discussed systems in a number of ways. Most importantly if a customer picks up a bicycle at a certain station it is very likely that he will return it to a different station after completion of his requirements. He has thus changed the location of available servers in the system. Another important difference is that customers may take two choices: where to pick up a bicycle and where to return the bicycle. In the presence of delay between choice and arrival, given that bike stations have a limited number of slots at which bicycles can be docked, both choices can lead to unsatisfactory situations. We propose that for both decisions customers can use our approach. To keep the system in balance, the bicycle provider company has to relocate bicycles using service vehicles. We do not choose to consider this aspect here, but it should be noted that, it is possible for them to change the number of bicycles in the system so that the number of rented bicycles plus the number of available bicycles at all stations is larger than the total number of bike slots in times of high demand.

### 9.4.1. Model

We now present a modelling approach for shared bicycle systems and describe our intended system in detail. For this we make the following assumptions.

1. Bike sharing stations are instrumented and the number of free bikes and free bike stands at each station is accessible by the users.

2. Customers can choose between several bike stations to pick up a bike and do this taking into account real time and historical occupancy data.

3. Customers that are currently using a rented bicycle can choose between several bike stations to return the bike and do this taking into account real time and historical occupancy data.

4. After having decided upon a bike station the customer has to walk or ride to that station. This is associated with a delay.

Let us consider the process according to which bicycles are being picked up. We have assumed that customers arrive to the system according to a non-homogeneous Poison process. Upon his arrival, the $n'$th customer immediately chooses a bike station based on the available information and then walks to the chosen station, $i$, which takes a time of $\tau_{n,i}$. This means that if we assume that $\tau_{n,i} = \tau_i \in \mathbb{R}$ is a constant number independent of the customer, then the rate at which bicycles are picked up at station $i$ at time $t$ depends on the number of available bicycles at all stations at time $t - \tau_i$. If we relax this assumption and let $\tau_{n,i}$ be a positive random variable, then the rate at which bicycles are being picked up at station $i$ at time $t$ potentially depends on the whole previous history of the system. Another important aspect his how the customers react to the event of arriving to a bike station and experiencing a lack of available bikes. There

are three simple options: the customer leaves the system and cancels his trip or reverts to other modes of transport, the customer waits until such times that a bike becomes available, or the customer walks to a different bike station in the hope of finding a more favourable situation there. For our simulations we make the additional assumption that a fixed fraction of these customers will wait and the rest will walk to the other station. The difficulty of analysing this model is that although the arrival process of customers to the system is Poisson, the process according to which bicycles are being picked up at individual stations is not. The exact same questions arise when defining the process according to which bicycles are being returned to the stations.

An important aspect of a bike sharing system is how decision are made. According to our assumptions each customer has much data available on which he can base his decision. An intuitive approach to making decisions is summarised in Algorithms 9.5 and 9.6. Again, this is similar to a join the shortest queue approach,

---
**Algorithm 9.5** Deterministic Station Choice - Bike Pick-Up
___

$N_i \leftarrow$ available bikes at station $i$
$j \leftarrow \text{argmax}_i N_i$
Attempt to pick up a bike at station $j$.

---

---
**Algorithm 9.6** Deterministic Station Choice - Bike Return
___

$N_i \leftarrow$ available bikes at station $i$
$j \leftarrow \text{argmax}_i C_i - N_i$
Attempt to return bike at station $j$

---

We believe that these intuitive algorithms reflect well how humans think. As we shall see in the following section, the time delay in the system between the choice of a station and arrival at that station induces undesired effects to the system. We will show that this intuitive algorithm exhibits sub-optimal behaviour.

---
**Algorithm 9.7** Randomised Station Choice - Bike Pick-Up
___

$N_i \leftarrow$ available bikes at station $i$
$p_i \leftarrow \frac{\sum_{j=1}^{i} N_j}{\sum_{j=1}^{n} N_j}$
$q \leftarrow \text{RAND}$
**if** $p_{i-1} \leq q \leq p_i$ **then**
    Attempt to pick up a bike at station $i$
**end if**

---

We propose again our simple stochastic algorithms that are superior at balancing load over bicycle sharing stations. Put into words, Algorithm 9.7 lets newly arriving customer gather information about the number of available bicycles at all accessible stations and

---
**Algorithm 9.8** Randomised Station Choice - Bike Return

---

$N_i \leftarrow$ available bikes at station $i$

$p_i \leftarrow \frac{\sum_{j=1}^{i}(C_j - N_j)}{\sum_{j=1}^{n}(C_j - N_j)}$

$q \leftarrow \text{RAND}$

**if** $p_{i-1} \leq q \leq p_i$ **then**

    Attempt to return bike at station $i$

**end if**

---

then randomly choose one among them with a probability proportional to this number. In the case when all stations are empty, one is picked randomly with uniform probability. This case has been omitted in Algorithm 9.7 to ease exposition. It should be noted that the efficacy of our proposed approach depends on the relationship between how long it takes on average to arrive at the chosen station after making the decision and the rate at which customers arrive. Clearly, if arrival is instantaneous or a customer usually arrives to the station before the next customer even makes a decision, then it is optimal to simply pick the station with the largest number of bicycles, or any station with an available bike, if balancing load between the stations is not an objective. If it takes a long time to arrive at the station after making the decision and a significant number of other customers arrive to make a decision in the mean time, then a state-independent assignment like the one proposed in Algorithm 9.7 is superior. Algorithm 9.8 is the equivalent to Algorithm 9.7 for returning bikes to the stations.

### 9.4.2. Stability of Proposed Approach

We now present a model for analysis of a shared bicycle system. We believe that this system is best described in a queueing theoretical framework. Consider a system of $n$ stations. Let $C_i$ be the capacity of station $i$, i.e. the total number of slots for bikes. Let $X_i(t)$ be the number of available bikes at station $i$ at time $t > 0$. We make the following additional assumptions

  (i) Customers arrive to the system according to a non-homogeneous Poisson process.

  (ii) Bikes are returned to the system according to a non-homogeneous Poisson process.

 (iii) Customers choices are described by Algorithms 9.7 and 9.8.

 (iv) The delays between decisions and arrival at the chosen station can be described by positive i.i.d random variables.

The system arising from the proposed stochastic assignment rule is hard to analyse. As a first step in this direction, in this section we prove stability of the fluid limit of this system. In this we follow the approach presented in Section 9.3. In this system the occupancies of all stations are continuous and customers can be seen as a fluid that flows into and out of the bicycle station system. The incoming stream of customers has rate

$\lambda$ and is split and allocated to the stations according to Algorithm 9.7 after finishing the bike ride, the customers flow out of the individual stations and combine to leave the system as a steady stream of rate $\lambda$. This model should be a good approximation for systems in which the capacity and occupancy of all station is rather large and customers arrive with a high frequency.

The evolution of the fluid limit is described by the following delay differential equation.

$$\dot{X}_i(t) = \lambda \frac{C_i - X_i(t - \tau_{j,2})}{\sum_j (C_j - X_j(t - \tau_{j,2}))} - \lambda \frac{X_i(t - \tau_{j,1})}{\sum_j X_j(t - \tau_{j,1})} \tag{9.16}$$

Summing Equation (9.16) over all $i$ yields $\sum_{i=1}^n \dot{X}_i(t) = 0$ and hence the total number of available bicycles in the system is constant. Let $N$ be this constant and let $C = \sum_{i=1}^n C_i$. We can thus rewrite Equation (9.16) as

$$\dot{X}_i(t) = \lambda \frac{C_i - X_i(t - \tau_{j,2})}{C - N} - \lambda \frac{X_i(t - \tau_{j,1})}{N} \tag{9.17}$$

This equation is of the form (8.42) and we can apply Corollary 8.7. There are two conditions in this corollary that ensure exponential stability of the constant solution $X_j(t) = N\frac{C_j}{C}$ for all $t \geq 0$. These conditions are $\tau_{j,1}\lambda < \frac{N}{2}$ and $\tau_{j,2}\lambda < \frac{C-N}{2}$. $\tau\lambda$ is the expected number of arrivals to the system in one delay period of length $\tau > 0$ and thus these conditions say that these expected values have to be smaller than the total number of available bikes and available bike spots respectively. This is in close agreement with intuition. According to Corollary 8.7, as in the previous section, the constant solution to which the system converges is the fair situation where customers frequent each station with probability proportional to the stations capacity.

### 9.4.3. Simulations

We now show the efficacy of our algorithm using simulations based on real data. To this end we present two simulations of a scenario with two stations in a shared bicycle system. In the first we use the deterministic assignment rule, in the second we use our proposed stochastic rule.

We present a simple simulation based on real data that was obtained from the Dublin Bikes website (`http://www.dublinbikes.ie`). It provides real-time information about the availability of bikes. The data in our study is collected by repeatedly crawling the website, we record the number of currently available bikes and number of free stands at 44 bike stations within the bike sharing system. Following the approach presented in [Chen et al., 2013] we can use this data to compute average arrival rates of customers and bikes to each of the stations. For the sake of this paper we have picked two bike stations near Trinity College in Dublin, located in Leinster Street and Pearse Street. The stations have capacity for 20 and 23 bikes respectively. For our simulations, we consider only the customers using these two stations. As the arrival and departure rates for our

non-homogeneous Poisson processes for bike rental and bike return we use the aggregate rates to the two stations. We consider the following scenario: Customers arrive to the general area (for example by public transport). Upon arrival they query their smartphone app for a bike station with available bikes. They will make a decision based on the smartphone advice and then walk to the designated bike station. We assume that this walking time is log-normally distributed with parameters $\mu = \log(300) - 1$ and $\sigma^2 = 2$ such that the mean walking time is 5 minutes. Upon arrival at the station they take a bike if one is available. In the absence of bikes, half of the customers will wait until a new bike arrives and the other half will commence a walk to the other bike station. Again this walking time is assumed to be log-normally distributed with mean 5 minutes.

**Simulation 1 :** For the first simulation we assume that customers use Algorithms 9.5 and 9.6 for decision making. We are interested in how balanced the system is. One possible measure of balance is the occupancy of one station in situations, when a customer arrives to the other station and finds no bike. We repeated the simulation 1000 times to reduce the effect of random events and Fig. 9.22 depicts a histogram of these values accumulated over all simulation runs.



Figure 9.22.: Histogram of positive number of bikes at the other station whenever a customer arrives to find one station empty with deterministic assignment.

**Simulation 2 :** We now repeat the above simulation with the stochastic customer assignment rules given in Algorithms 9.7 and 9.8. The results are summarised in Figure 9.23. It can clearly be seen that the stochastic approach improves the performance of the system. Most notably, where in the original simulation a total of 2828 passengers

arrive to an empty station, when there is at least one bicycle available at the other station, with the new proposed approach this number is reduced to 2002, a reduction of about 30%.

free spaces at the other station when a customer arrives to see no bike – random
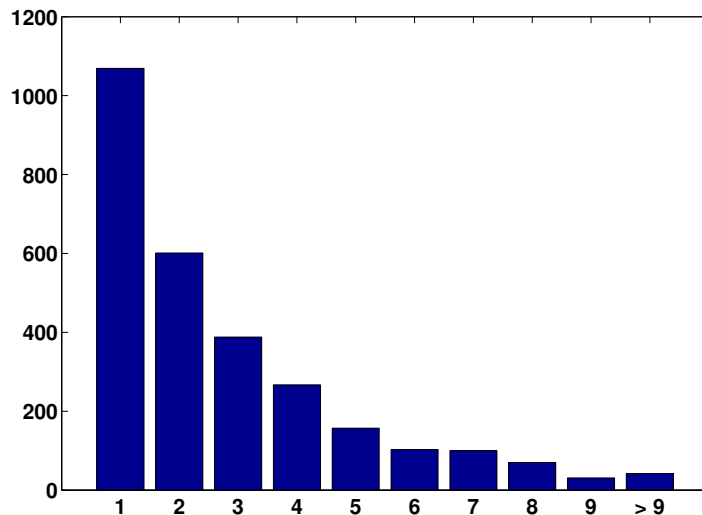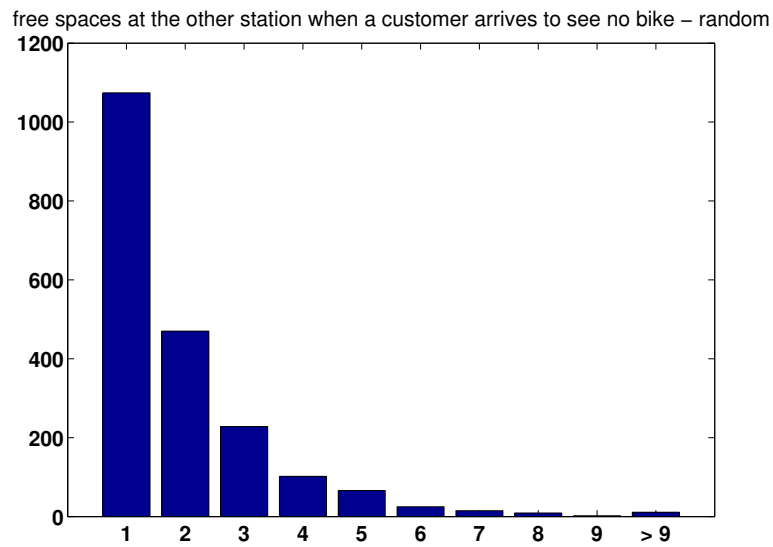


Figure 9.23.: Histogram of positive number of bikes at the other station whenever a customer arrives to find one station empty with randomised user assignment.

# 10. Conclusions

***Abstract :*** *In this chapter we conclude the thesis. We summarise the contributions contained in each chapter. Finally, we outline some possible future research questions.*

## 10.1. Summary

In Chapter 1 we introduced the reader to some typical problems arising from modern transport systems, especially in an urban context. The reader was alerted to the five main issues with high traffic densities: congestion, accidents, green house gas emissions, noise, and pollution. Also, some initial insight was given into the three main application areas investigated in this thesis. We also briefly outlined classes of historic traffic models and their draw-backs. It was explained how a recent convergence of technologies, leading to the so-called connected vehicle, has enabled a new class of approaches to deal with and improve traffic. We also highlighted the research challenges accompanying the new developments. A large number of ongoing research initiatives and current publications that make use of the new technologies was discussed. It was concluded that a certain kind of feedback has been ignored in all of these approaches; namely the effect of making data available to travellers, which influences their decisions and hence the system dynamics.

### 10.1.1. Markov Chain Traffic Model

In Part I we presented a Markov chain model of urban traffic network dynamics to the reader. The main objective of this work is to develop models of large-scale urban networks. These models are data driven and are capable of dealing with huge amounts of real time data typically made available by modern ICT. Modern computers, such as the ones built into cars and smartphones, can easily handle these models. The model's ability to predict the effect of changes to the system on macroscopic network properties allows us to design interesting new control applications.

Chapter 2 explained and motivated the choice of using Markov chains for this purpose. They are able to deal with the huge amounts of data in a way that allows fast computa-

tions that yield meaningful results. We reported a recently proposed model and discussed some of the applications in the context of congestion modelling. We also raised some of the model's shortcomings. For example, travellers in this model could not enter or leave the network, condemning them to keep travelling forever. To alleviate this, in Chapter 3 we showed how a number of restrictive assumptions made in the original model can be relaxed. Also, we further developed the analytical tools with which insights into the network can be gained through the Markov chain. For example, we wished to apply the chain to the consumption of battery power in electric vehicles. In this context it is possible to regain battery power on some road segments with negative slope. This leads to costs with different signs, which the original model could not incorporate. We also used ideas from closed queueing networks to allow travel patterns and origins and destinations for drivers. Finally in Chapter 4 we proposed two completely new application areas for the Markov chain model - the pollution chain and the aforementioned electric energy consumption chain. In both contexts we investigated how meaningful information can be extracted from the Markov chain using characteristic parameters of the chain. We proposed a number of novel applications of the Markov chain that enable highly relevant control approaches, such as avoiding pollution peaks, keeping emissions away from sensitive areas (e.g. schools, hospitals), as well as investigating several new routing strategies for whole communities of drivers that lead to social optima and routing strategies for individuals that take into account the specific constraints and characteristics of EVs.

### 10.1.2. Collaborative Emission Control

In Part II we presented a collaborative approach to controlling emissions in fleets of hybrid vehicles. We discussed the typical constraints that have to be taken into account when dealing with the modelling and control of ITS applications in Chapter 5. These constraints regard communication technology, computational ability and infrastructure involvement. In Chapter 6 we proposed a pollution control strategy for fleets of HEVs. Our approach involved measuring pollution levels and sending simple broadcast signals from a central infrastructure to a fleet of hybrid vehicles. This signal is used to influence the switching between the two engines in each vehicle. Via the construction of a proof-of-concept vehicle, it was shown that our approach is easily implemented . Finally, this work was extended in Chapter 7 to include V2V communication, which enables the fleet to allocate the available pollution budget in a manner that is both fair and decentralised.

For our work in Part II we have considered fleets of hybrid vehicles. It is however important to note that the basic ideas in the chapter can be extended in many ways to more conventional vehicle types. For example, speed limits can be adjusted based on aggregate emission levels, as can access to specific areas and routes for particular vehicle types. The later idea basically proposes regulating the mix of vehicles in a particular area (EV, HEV, buses, ICE, etc.) based on emission levels. However, these approaches have the disadvantage of affecting individual driver behaviour. The hybrid vehicle on the other hand allows implementations in which the control actions do not affect individual drivers' choices.

### 10.1.3. Feedback Modelling in Assignment Systems with Delay

In Part III we discussed a stochastic assignment strategy of customers to service stations, which works well in systems that cannot perform or enforce reservations. This strategy involves obtaining information about the occupancy of each available service station and assigning newly arriving customers in a stochastic way that favours less occupied stations. This assignment in general is difficult in the presence of delays between the choice of a service station and the arrival at the service station. In this context our strategy balances the stations' occupancies, which increases efficiency of the overall system and improves the experience for the customers. In Chapter 8 we motivated the need for new models for assignment systems in the ITS context. We showed through examples that in these systems a kind of feedback arises that has to be taken into account in models, as it can cause oscillations and decrease the overall system performance. We formulated the assignment problem in a general framework and developed a number of tools to analyse systems with the described kind of delay. This allows to obtain performance measures in a wide range of scenarios. In Chapter 9 we then discussed three application areas for this assignment strategy. These areas are: the assignment of electric vehicles to charging stations, the assignment of vehicles to car parks, and the assignment of customers to bike sharing stations.

To the best of our knowledge, in the context of EVs and charging stations, the effect of queueing has been ignored by previous literature. It is one contribution of our work to raise awareness to the importance of this effect for large-scale deployments of electric vehicles. Compared to deterministic methods, our approach is fully distributed, scalable and has a plug-and-play character, meaning that charging stations could enter and leave the system without any adjustment of the system.

## 10.2. Open Questions and Proposed Research Directions

### 10.2.1. Markov Chain Traffic Model

The work in this thesis concentrated on applying the Markov chain model to urban traffic networks and cars. Work is already carried out to apply the model to different kinds of traffic networks. We apply the model to public transport, where the model describes the movement of passengers between bus stops and train stations along fixed routes. Another application is bike sharing stations and the description of the dynamics of rented bicycles between the stations. In this context we also investigate the combination of different modes of transportation under this framework, such as a combined bus and train model, or a train and bike model. This then enables us to apply our model to multi-modal traffic networks.

In parallel we are also working on a number of extensions to the set of theoretic tools. For example, we wish to find ways to optimise the mean first passage times. We are also interested in further studying the clustering properties of Markov chains using spectral

analysis. In particular, we wish to investigate what information about sub-communities in the underlying directed graph can be gleaned from the eigenvectors of the Markov chain transition matrix.

The Markov chain model has been well received by the research community. [Morimura et al., 2013; Yang et al., 2013; Bekkerman et al., 2013] investigate how properties of the Markov chain can be recovered using only partial observations. [Moosavi and Hovestadt, 2013] further validated the Markov chain model using GPS data from taxis in Beijing.

### 10.2.2. Collaborative Emission Control

The algorithms proposed in Chapters 6 and 7 are innovative and elegant yet very simple. Therefore, the work has reached a very mature state. Testing and implementation are the next steps to be endeavoured. Work is going on to interface our demonstration vehicle with the Sumo simulator. The goal is to create large-scale hardware-in-the-loop simulations [Griggs and Shorten, 2013].

An extension worth investigating is the following. Our approach allocates an emission budget to each vehicle. It may be advantageous for the vehicles to use the available budget along certain legs of its route. To this end, the vehicle can gather information about traffic densities, travel speeds and other relevant factors. This information can be used to optimise where the available emission is being used. If the route, for example, leads through residential and non-residential roads, it is advantageous to favour the usage of the ICE along the non-residential roads.

### 10.2.3. Feedback Modelling in Assignment Systems with Delay

Among the many tools we have presented to analyse our assignment systems was a fluid limit argument. Future work will have to investigate under which circumstances and to what degree the stability results hold also for the original stochastic system. Further, the transfer of the quality of service measures developed in the single service station scenario is an important next step.

The assignment of vehicles to car parks is currently being considered for implementation and testing in TEAM [TEAM Consortium, 2014].

# Bibliography

Aizhu, C. (2013). Beijing to slap tougher emission standards on vehicles: Xinhua. Available online `http://www.reuters.com/article/2013/01/23/us-beijing-emission-idUSBRE90M10R20130123`, last accessed 20/01/2014.

Alvarez, I., Poznyak, A., and Malo, A. (2008). Urban traffic control problem a game theory approach. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 2957–2961.

Anderson, S. P. and de Palma, A. (2004). The economics of pricing parking. *Journal of Urban Economics*, 55:1–20.

Arnott, R. (2006). Spatial competition between parking garages and downtown parking policy. *Transport Policy*, 13(6):458–469.

Artmeier, A., Haselmayr, J., Leucker, M., and Sachenbacher, M. (2010). The shortest path problem revisited: optimal routing for electric vehicles. In *Proceedings of the 33rd annual German conference on Advances in artificial intelligence*, KI'10, pages 309–316, Berlin, Heidelberg. Springer-Verlag.

Avineri, E. (2009). Social value orientation and the efficiency of traffic networks. In Kitamura, R., Yoshii, T., and Yamamoto, T., editors, *The Expanding Sphere of Travel Behaviour Research*, pages 725–744. Emerald, UK.

Barlow, T. J., Hickman, A. J., and Boulter, P. G. (2001). Exhaust emission factors 2000: Database and emission factors. Project report PR/SE/230/00. Technical report, TRL Limited, Crowthorne.

Barth, M., An, F., Younglove, T., Scora, G., Levine, C., Ross, M., and Wenzel, T. (2000). Development of a comprehensive modal emissions model. Final report, NCHRP Project 25-11.

Baskar, L., Schutter, B. D., Hellendoorn, J., and Papp, Z. (2011). Traffic control and intelligent vehicle highway systems: a review. *IET Intelligent Transport System*, 5(1):38–52.

BBC (28/3/2011). UK rejects EU call for city centre ban on petrol cars. Available online `http://www.bbc.co.uk/news/uk-politics-12879566`, last accessed 20/01/2014.

Bekkerman, E. N., Kataev, S. G., and Kataeva, S. S. (2013). Heuristic approximation method for a random flow of events by an MC-flow with arbitrary number of states. *Automation and Remote Control*, 74(9):1449–1459.

Berman, A. and Plemmons, R. (1994). *Nonnegative Matrices in the Mathematical Sciences*. SIAM.

Bhandarkar, S., Reddy, A. L. N., Zhang, Y., and Loguinov, D. (2007). Emulating AQM from end hosts. *ACM SIGCOMM Computer Communication Review*, 37(4):349–360.

Biem, A., Bouillet, E., Feng, H., Riabov, A., Verscheure, O., Koutsopoulos, H., Rahmani, M., and Güç, B. (2010). Real-time traffic information management using stream computing. *IEEE Database Engineering Bulletin*, 33(2):64–68.

BIKE INTERMODAL Consortium (2013). BIKE INTERMODAL - Multi-modal integration of cycling mobility through product and process innovations in bicycle design. Available online `http://www.bike-intermodal.eu`, last accessed 20/01/2014.

BMWi (2014). Mobility Services. Available online `http://www.bmw.com/com/en/insights/corporation/bmwi/mobility_services.html`, last accessed 20/01/2014.

Botea, A., Berlingerio, M., and Nikolova, E. (2013). Multi-modal journey planning in the presence of uncertainty. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS-13)*, pages 20–28.

Boulter, P., Barlow, T., and McCrae, I. (2009). *Emission factors 2009: Report 3 - exhaust emission factors for road vehicles in the United Kingdom, Published project report PPR546*. TRL limited.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Braess, D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12:258 – 268.

Braess, D., Nagurney, A., and Wakolbinger, T. (2005). On a Paradox of Traffic Planning (translation from the German original). *Transportation Science*, 39:446 – 450.

Budzisz, Ł., Stanojević, R., Schlote, A., Baker, F., and Shorten, R. (2011). On the Fair Coexistence of Loss- and Delay-Based TCP. *IEEE/ACM Transactions on Networking*, 19(6):1811–1824.

Budzisz, Ł., Stanojević, R., Shorten, R., and Baker, F. (2009). A strategy for fair coexistence of loss and delay-based congestion control algorithms. *IEEE Communications Letters*, 13(7):555–557.

Caliskan, M., Graupner, D., and Mauve, M. (2006). Decentralised discovery of free parking places. In *Proceedings of the 3rd international workshop on Vehicular ad hoc networks*, pages 30–39.

Cao, M. and Wu, C. W. (2007). Topology design for fast convergence of network consensus algorithms. In *IEEE International Symposium on Circuits and Systems*, pages 1029–1032.

car2go Deutschland GmbH (2014). Car2go. Available online `http://www.car2go.com`, last accessed 20/01/2014.

CARBOTRAF Consortium (2014). CARBOTRAF - A Decision Support System for Reducing CO2 and Black Carbon Emissions by Adaptive Traffic Management. Available online `http://carbotraf.com`, last accessed 20/01/2014.

Chatterjee, K. (1999). Modelling the impacts of transport telematics: Current limitations and future developments. *Transport Reviews*, 19(1):57–80.

Chen, B., Pinelli, F., Sinn, M., Botea, A., and Calabrese, F. (2013). Uncertainty in urban mobility: Predicting waiting times for shared bicycles and parking lots. In *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems, The Hague, Netherlands*, pages 53 – 58.

Chen, H. and Yao, D. D. (2001). *Fundamentals of queueing networks: performance, asymptotics, and optimization*, volume 46 of *Applications of mathematics: stochastic modelling and applied probability*. Springer.

Chiguma, M. L. (2007). *Analysis of Side Friction Impacts on Urban Road Links.* PhD thesis, KTH Royal Institute of Technology.

Choa, G. E. and Meyer, C. D. (2001). Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications*, 335:137–150.

City of Brussels (2012). Pollution peaks. Available online `http://www.brussels.be/artdet.cfm/6357`, last accessed 20/01/2014.

Cone, T. (2008). California air pollution kills more people than car crashes, study shows. *Huffington Post.*

Crisostomi, E., Kirkland, S., Schlote, A., and Shorten, R. (2011a). Markov chain based emissions models: A precursor for green control. In Kim, J. H. and Lee, M. J., editors, *Green IT: Technologies and Applications*, pages 381–400. Springer Berlin Heidelberg.

Crisostomi, E., Kirkland, S., and Shorten, R. (2011b). A Google-like model of road network dynamics and its application to regulation and control. *International Journal of Control*, 84(3):633–651.

Crisostomi, E., Kirkland, S., and Shorten, R. (2011c). Robust and risk-averse routing algorithms in road networks. In *Proceedings of the 18th IFAC World Congress, Milano, Italy*, pages 9512–9517.

de Abreu, N. M. M. (2007). Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423(1):53–73.

de Weerdt, M. M., Gerding, E., Stein, S., Robu, V., and Jennings, N. R. (2013). Intention-aware routing to minimise delays at electric vehicle charging stations. In

*Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pages 83–89.

Dellnitz, M. and Junge, O. (1999). On the approximation of complicated dynamical behaviour. *SIAM Journal of Numerical Analysis*, 36(2):491–515.

Dellnitz, M., Junge, O., Koon, W., Lekien, F., Lo, M., and Marsden, J. (2005). Transport in dynamical astronomy and multibody problems. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 3:699–727.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *The Journal of Public Transportation*, 12(4):41—56.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numererische Mathematik*, 1(1):269 – 271.

Dirks, S., Gurdgiev, C., and Keeling, M. (2010). *Smarter cities for smarter growth, How cities can optimize their systems for the talent-based economy.* IBM Global Business Services Executive Report.

DriveNow GmbH & Co. KG (2014). DriveNow. Available online `www.drive-now.com`, last accessed 20/01/2014.

Driver, R. D. (1977). *Ordinary and delay differential equations.* Springer-Verlag.

Dublin City Council (2011). Movement. Available online `http://www.dublincity.ie/WaterWasteEnvironment/StrategyandProjects/Documents/Movement.pdf`, last accessed 20/01/2014.

ECODRIVER Consortium (2014). ECODRIVER - Supporting the driver in conserving energy and reducing emissions. Available online `http://www.ecodriver-project.eu`, last accessed 20/01/2014.

EcoGem Consortium (2014). EcoGem - Cooperative Advanced Driver Assisted System for Green Cars. Available online `http://www.softeco.it/ecogem/home.aspx`, last accessed 20/01/2014.

ECONAV Consortium (2014). ECONAV - Ecological Aware Navigation: Usable Persuasive Trip Advisor for Reducing CO2-consumption. Available online `http://cordis.europa.eu/projects/rcn/100231_en.html`, last accessed 20/01/2014.

Ehsani, M., Gao, Y., Gay, S. E., and Emadi, A. (2010). *Modern electric, hybrid electric and fuel cell vehicles : fundamentals, theory and design.* CRC Press.

Elfrink, W. (2012). The smart city solution. *McKinsey Quarterly.*

ELVIRE Consortium (2014). ELVIRE - ELectric Vehicle communication to Infrastructure, Road services and Electricity supply. Available online `http://www.elvire.eu`, last accessed 20/01/2014.

Enhanced WISETRIP Consortium (2014). Enhanced WISETRIP - Enhancing Intermodality of Content, Personalised Information and Functionality of WISETRIP Network of Journey Planning Engines. Available online `http://www.wisetrip-eu.org`, last accessed 20/01/2014.

Erol-Kantarci, M. and Mouftah, H. (2010). Prediction-Based Charging of PHEVs from the Smart Grid with Dynamic Pricing. In *IEEE Workshop on Smart Grid Network Infrastructure*, pages 1032–1039.

European Commission (2011). Roadmap to a single european transport area – towards a competitive and resource efficient transport system. White paper.

European Commission (2014). Horizon2020. Available online `http://ec.europa.eu/programmes/horizon2020`, last accessed 20/01/2014.

European Environment Agency (2008). Climate for a transport change. Available online `http://www.eea.europa.eu/publications/eea_report_2008_1`, last accessed 20/01/2014.

Ezell, S. (2010). Explaining International IT Application Leaderhip: Intelligent Transportation. Technical report, The Information Technology & Innovation Foundation.

Farrington, R. B. and Rugh, J. P. (2000). Impact of Vehicle Air-Conditioning on Fuel Economy, Tailpipe Emissions, and Electric Vehicle Range; Preprint available online `www.nrel.gov/docs/fy00osti/28960.pdf`. In *Earth Technologies Forum, Washington, D.C.*

Federal Republic of Germany (2014). Umweltzonen in Deutschland. Available online `http://gis.uba.de/website/umweltzonen/umweltzonen.php`, last accessed 20/01/2014.

Fiore, M. and Härri, J. (2008). The networking shape of vehicular mobility. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 261–272.

Fishman, E., Washington, S., and Haworth, N. (2013). Bike share: A synthesis of the literature. *Transport Reviews: A Transnational Transdisciplinary Journal*, 33(2):148–156.

Floyd, S. and Jacobson, V. (1993). Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413.

Fricker, C. and Gast, N. (2013). Incentives and redistribution in bike-sharing systems with stations of finite capacity. Preprint, available online `http://arxiv.org/abs/1201.1178`, last accessed 20/01/2014.

Friends of the Earth (1999). Road transport, air pollution and health. Briefing paper, available online `http://www.foe.co.uk/sites/default/files/downloads/road_air_pollution_health.pdf`, last accessed 20/01/2014.

Froyland, G. (2001). *Extracting dynamical behavior via Markov models.* Nonlinear dynamics and statistics. Birkhäuser Boston.

Garavello, M. and Piccoli, B. (2009). On fluido-dynamic models for urban traffic. *Networks and Heterogeneous Media*, 4(1):107–126.

Gavron, C. (1998). An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C*, 9(3):393–407.

Geng, Y. and Cassandras, C. G. (2011). Dynamic resource allocation in urban settings: A "smart parking" approach. In *IEEE International Symposium on Computer Aided Control System Design*, pages 1–6.

Gkatzoflias, D., Kouridis, C., Ntziachristos, L., and Samaras, Z. (2007). COPERT 4. Computer program to calculate emissions from road transport. User Manual (version 5.0). Technical report, European Environment Agency.

Göhring, D., Latotzky, D., Wang, M., and Rojas, R. (2013). Semi-autonomous car control using brain computer interfaces. In Lee, S., Cho, H., Yoon, K.-J., and Lee, J., editors, *Intelligent Autonomous Systems 12*, volume 194 of *Advances in Intelligent Systems and Computing*, pages 393–408. Springer Berlin Heidelberg.

Griggs, W. and Shorten, R. (2013). Embedding real vehicles in SUMO for large-scale ITS scenario emulation. In *2nd International Conference on Connected Vehicles and Expo*.

Grinstead, C. and Snell, J. L. (2003). *Introduction to Probability.* American Mathematical Society, 2nd edition.

Hager, W., Hearn, D., and Pardalos, P. (1994). *Large Scale Optimization: State of the Art.* Springer.

Hartenstein, H. and Laberteaux, K., editors (2010). *VANET – Vehicular applications and internetworking technologies.* Wiley.

Häusler, F., Crisostomi, E., Schlote, A., Radusch, I., and Shorten, R. (2012). Stochastically balanced parking and charging for fully electric and plug-in hybrid vehicles. In *International Conference on Connected Vehicles and Expo (ICCVE)*, pages 341–342, Los Alamitos, CA, USA. IEEE Computer Society.

Häusler, F., Crisostomi, E., Schlote, A., Radusch, I., and Shorten, R. (2013a). Stochastic park-and-charge balancing for fully electric and plug-in hybrid vehicles. *accepted for publication in IEEE Transactions on Intelligent Transportation Systems*, pages 1–6.

Häusler, F., Faizrahnemoon, M., Crisostomi, E., Schlote, A., and Shorten, R. (2013b). A framework for optimal real-time emissions trading in large-scale vehicle fleets. In *9th ITS European Congress*.

HAVEit Consortium (2014). HAVEit - Highly automated vehicles for intelligent transport. Available online `http://www.haveit-eu.org`, last accessed 20/01/2014.

Hinrichsen, D. and J.Pritchard, A. (2004). *Mathematical Systems Theory I. Modelling, state space analysis, stability and robustness.* Springer Berlin.

Hodgkinson, S. (2011). Is your city smart enough. Technical report, CISCO White Paper.

Hoogendoorn, S. P. and Bovy, P. H. (2001). State-of-the-art of vehicular traffic flow modelling. *Journal of Systems and Control Engineering*, 215(4):283–303.

Horn, R. and Johnson, C. (1990). *Matrix Analysis.* Cambridge University Press.

Huisinga, W. (2001). *Metastability of Markovian Systems: A Transfer Operator Based Approach in Application to Molecular Dynamics.* PhD thesis, Freie Universität Berlin.

Hunter, J. (2006). Mixing times with applications to perturbed Markov chains. *Linear Algebra and its Applications*, 417(1):108–123.

IBM (2007). Delivering smarter transportation systems. Technical report, IBM White paper.

IBM (2013). Dublin City Council - Leveraging the leading edge of IBM Smarter Cities research to reduce traffic congestion. Available online `http://www.ibm.com/midmarket/ie/en/att/pdf/IBM_DCC_130715.pdf`, last accessed 20/01/2014.

ICT4EVEU Consortium (2014). Ict4eveu. Available online `http://www.ict4eveu.eu`, last accessed 20/01/2014.

Jacobson, V. (1988). Congestion avoidance and control. *ACM SIGCOMM Computer Communication Review*, 18(4):314–329.

JCDecaux (2013). Latest Expansion News. Available online `http://www.dublinbikes.ie/Expansion-News`, last accessed 20/01/2014.

Jeong, M.-H., Duckham, M., Kealy, A., Miller, H. J., and Peisker, A. (2014). Decentralized and coordinate-free computation of critical points and surface networks in a discretized scalar field. *International Journal of Geographical Information Science*, 28(1).

Kellet, C., Shorten, R., and Leith, D. (2006). Sizing Internet Buffers, Active Queue Management, and the Lur'e Problem. In *Proceedings of IEEE Conference on Decision and Control*, pages 650–654.

Kelly, F. P. (1999). Mathematical Modelling of the Internet. In *Fourth International Congress on Industrial and Applied Mathematics*, pages 105–116, Edinburgh, UK.

Kemeny, J. G. and Snell, J. L. (Princeton, 1960). *Finite Markov Chains.* Van Nostrand.

175

Khalil, H. (2001). *Nonlinear Systems*. Prentice Hall.

King, C., Griggs, W., Wirth, F., Quinn, K., and Shorten, R. (2013). Alleviating a form of electric vehicle range anxiety through on-demand vehicle access. Preprint, available online `http://arxiv.org/abs/1312.5939`, last accessed 20/01/2014.

Klappenecker, A., Lee, H., and Welch, J. L. (2012). Finding available parking spaces made easy. *Ad Hoc Networks*, 12:243–249.

Knorn, F., Corless, M., and Shorten, R. (2011a). A result on implicit consensus with application to emissions control. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 1299 – 1304.

Knorn, F., Corless, M., and Shorten, R. (2011b). Results in cooperative control and implicit consensus. *International Journal of Control*, 84(3):476–495.

Kokolaki, E., Karaliopoulos, M., and Stavrakakis, I. (2011). Value of information exposed: Wireless networking solutions to the parking search problem. In *Proceedings of Eighth International Conference on Wireless On-Demand Network Systems and Services*, pages 187–194.

Koon, W., Lo, M., Marsden, J., and Ross, S. (2000). Heteroclinic connections between periodic orbits and resonance transitions in celestial mechanics. *Chaos*, 10(2):427–469.

Koon, W., Marsden, J., Ross, S., and Lo, M. (2002). Constructing a low energy transfer between Jovian moons. In *Celestial Mechanics*, volume 292, pages 129–145.

Krajzewicz, D., Bonert, M., and Wagner, P. (2006). The open source traffic simulation package SUMO. In *RoboCup 2006 Infrastructure Simulation Competition, RoboCup 2006, Bremen, Germany*, pages 1–5.

Krauß, S. (1998). *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics*. PhD thesis, Universität zu Köln.

Langville, A. N. and Meyer, C. D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ.

Levene, M. and Loizou, G. (2002). Kemeny's constant and the random surfer. *American Mathematical Monthly*, 109:741–745.

Lighthil, J. and Whitham, G. (1955). On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):317–345.

MacKay, D. (2008). *Sustainable Energy - Without the Hot Air*. UIT Cambridge Ltd.

Mackechnie, C. (2013). The last mile problem. Available online `http://publictransport.about.com/od/Transit_Planning/a/The-Last-Mile-Problem.htm`, last accessed 20/01/2014.

Mariyasagayam, M. N. and Kobayashi, Y. (2013). Electric vehicle route assistance using forecast on charging station. In *ENERGY 2013: The Third International Conference on Smart grids, Green Communications and IT Energy-aware Technologies*, pages 134–142.

Meyer, C. D. (1975). The Role of the Group Generalized Inverse in the Theory of Finite Markov Chains. *SIAM Review*, 17:443–464.

Meyer, C. D. (1989). Uncoupling the Perron eigenvector problem. *Linear Algebra Appl.*, 114-115:69–94.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Midgley, P. (2009). The role of smart bike-sharing systems in urban mobility. *Journeys*, 2:23–31.

Midgley, P. (2011). Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. Technical report, United Nations, Department of Economic and Social Affairs.

Mobi.Europe Consortium (2014). Mobi.Europe - Integrated and Interoperable ICT Applications for Electro-Mobility. Available online `http://www.mobieurope.eu`, last accessed 20/01/2014.

Mobility2.0 Consortium (2014). Mobility2.0 - Co-operative ITS Systems for Enhanced Electric Vehicle Mobility. Available online `http://mobility2.eu`, last accessed 20/01/2014.

MOBINCITY Consortium (2014). MOBINCITY - Smart Mobility in Smart City. Available online `http://www.mobincity.eu`, last accessed 20/01/2014.

Moosavi, V. and Hovestadt, L. (2013). Modeling urban traffic dynamics in coexistence with urban data streams. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, pages 1–7.

Morimura, T., Osogami, T., and Idé, T. (2013). Solving inverse problem of Markov chain with partial observations. In *27th Annual Conference on Neural Information Processing Systems*, pages 1655–1663.

Morris, E. (2007). From horse power to horsepower. *ACCESS: The magazine of the University of California Transportation Center*, 30:2–9.

Moya, S. and Poznyak, A. S. (2009). Extraproximal Method Application for a Stackelberg-Nash Equilibrium Calculation in Static Hiearchical Games. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 39(6):1493–1504.

Murphy, E. and Usher, J. (2013). The Role of bicycle-sharing in the city: Analysis of the Irish Experience. *International Journal of Sustainable Transportation*, to appear.

Nagel, K. and Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. *Journal de Physique I*, 2(12):2221–2229.

Nedic, A. and Ozdaglar, A. (2009). Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228.

Norris, E. S. and Chormaic, S. G. N. (2002). Development of a portable carbon monoxide optical sensor based on an extended cavity diode laser at 1564 nm. In *Proceedings of SPIE Opto-Ireland 2002: Optics and Photonics Technologies and Applications*, volume 4876, pages 923–929.

Norris, J. (1998). *Markov Chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.

Olfati-Saber, R., Fax, J. A., and Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233.

OSM Contributors (2014). Open street map. Available online `http://www.openstreetmap.org`, last accessed 20/01/2014.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford University.

ParkatmyHouse (2014). ParkatmyHouse - Smarter Parking. Available online `https://www.parkatmyhouse.com`, last accessed 20/01/2014.

Pavone, M., Frazzoli, E., and Bullo, F. (2011). Adaptive and distributed algorithms for vehicle routing in a stochastic and dynamic environment. *IEEE Transactions on Automatic Control*, 56(6):1259–1274.

Peden, M., McGee, K., and Sharma, G. (2002). *The Injury Chart Book: A Graphical Overview of the Global Burden of Injuries*. WHO, Geneva.

Pfrommer, J., Warrington, J., Schildbach, G., and Morari, M. (2013). Dynamic vehicle redistribution and online price incentives in shared mobility systems. Preprint, available online `http://arxiv.org/abs/1304.3949`, last accessed 20/01/2014.

Piccoli, B. and Garavello, M. (2006). *Traffic Flow on Networks*. American Institute of Mathematical Sciences.

Pucher, J. and Buehler, R. (2008). Making cycling irresistible: Lessons from Europe. *Transport Reviews*, 28(4):495–528.

Pucher, J. and Buehler, R. (2012). *City Cycling*. MIT Press.

Pucher, J., Buehler, R., and Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of Cycling Trends and Policies. *Transportation Research Part A: Policy and Practice*, 45(6):451–475.

Rainer-Harbach, M., Papazek, P., Hu, B., and Raidl, G. (2013). Pilot, graSp, and vns approaches for the static balancing of bicycle sharing systems. Technical report, Institute of Computer Graphics and Algorithms, Vienna University of Technology.

REDUCTION Consortium (2014). REDUCTION - Reducing Environmental Footprint based on Multi-Modal Fleet management System for Eco-Routing and Driver Behaviour Adaptation. Available online `http://www.reduction-project.eu`, last accessed 20/01/2014.

Richards, P. I. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51.

Sadler Consultants Ltd. (2013). Low Emission Zones in Europe. Available online `http://www.lowemissionzones.eu`, last accessed 20/01/2014.

Schlote, A. (2009). Stability of TCP models. Diplomarbeit, Bayerische Julius-Maximilians-Universität Würzburg.

Schlote, A., Chen, B., Sinn, M., and Shorten, R. (2013a). The Effect of Feedback in the Assignment Problem in Shared Bicycle Systems. In *International Conference on Connected Vehicles and Expo (ICCVE)*.

Schlote, A., Crisostomi, E., Kirkland, S., and Shorten, R. (2011). A Google-like Model of Road Network Dynamics. In *The 15th Yale Workshop on Adaptive and Learning Systems, Yale University, New Haven*.

Schlote, A., Crisostomi, E., Kirkland, S., and Shorten, R. (2012a). Traffic modelling framework for electric vehicles. *International Journal of Control*, 85(7):880–897.

Schlote, A., Crisostomi, E., and Shorten, R. (2012b). Balanced Routing for EVs. In *International Conference on Connected Vehicles and Expo*, pages 343–344.

Schlote, A., Häusler, F., Hecker, T., Bergmann, A., Crisostomi, E., Radusch, I., and Shorten, R. (2012c). Cooperative Regulation of Emissions Using Plug-in Hybrid Vehicles. In *Proceedings of the International Conference on Connected Vehicles and Expo*, pages 201–202.

Schlote, A., Häusler, F., Hecker, T., Bergmann, A., Crisostomi, E., Radusch, I., and Shorten, R. (2013b). Cooperative Regulation of Emissions Using Plug-in Hybrid Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1572–1585.

Schlote, A., King, C., Crisostomi, E., and Shorten, R. (2013c). Stochastic Algorithms for Parking Space Assignment. In *9th ITS European Congress, Dublin, Ireland*.

Schlote, A., King, C., Crisostomi, E., and Shorten, R. (2014). Stochastic Algorithms for Parking Space Assignment. *to appear in IEEE Transactions on Intelligent Transportation Systems*.

Schuijbroekm, J., Hampshire, R., and van Hoeve, W.-J. (2013). Inventory rebalancing and vehicle routing in bike sharing systems. Available online `http://www.andrew.cmu.edu/user/vanhoeve/papers/bike_sharing.pdf`, last accessed 20/01/2014.

Schütte, C. and Huisinga, W. (2003). *Biomolecular conformations can be identified as metastable sets of molecular dynamics*, volume x of *Handbook of Numerical Analysis*. North-Holland, Amsterdam.

SFMTA Municipal Transport Authority (2014). Sfpark. Available online `http://sfpark.org`, last accessed 20/01/2014.

Shaheen, S., Guzman, S., and Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present and future. Technical report, Transportation Research Record.

Shoup, D. (2006). Cruising for parking. *Transport Policy*, 13:479–486.

Shoup, D. (2007). Cruising for parking. *ACCESS: The magazine of the University of California Transportation Center*, 30:16–22.

Siljak, D. D. (2007). *Large-Scale Dynamic Systems: Stability and Structure*. Dover.

SMART-VEI Consortium (2011). SMART-VEI - The smart-vehicle. Available online `http://cordis.europa.eu/projects/rcn/86390_en.html`, last accessed 20/01/2014.

Smith, A. (2013). Smartphone Ownership 2013. Available online `http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013/Findings.aspx`, last accessed 20/01/2014.

Softeco Sismat S.r.l (2014). EMERALD - Energy ManagEment and RechArging for efficient eLectric car Driving. Available online `http://www.research.softeco.it/emerald.aspx`, last accessed 20/01/2014.

Spence, A., Turksma, S., Schelling, A., Benz, T., Medevielle, J.-P., McCrae, I., Jaaskelainen, J., and Boethius, E. (2009). Methodologies for assessing the impact of ITS applications on $CO_2$ emissions. Technical report, EC-METI Task Force.

Srikant, R. (2003). *The mathematics of Internet Congestion Control*. Systems & Control: Foundations & Applications. Birkhäuser.

Stanojević, R. (2007). *Router-based algorithms for improving Internet Quality of Service*. PhD thesis, National University of Ireland, Maynooth.

Stanojević, R. and Shorten, R. (2009a). Generalized distributed rate limiting. In *Proceedings of 17th IEEE International Workshop on Quality of Service*, pages 1–9, Charlston, SC, USA.

Stanojević, R. and Shorten, R. (2009b). Load balancing vs. distributed rate limiting: an unifying framework for cloud control. In *Proceedings of IEEE International Conference on Communications*, pages 1–6.

Stanojević, R., Shorten, R., and Kellett, C. (2006). Adaptive Tuning of Drop-Tail Buffers for Reducing Queueing Delays. *IEEE Communications Letters*, 10(7):570–572.

Stüdli, S., Crisostomi, E., Middleton, R., and Shorten, R. (2012a). AIMD-like algorithms for charging electric and plug-in hybrid vehicles. In *IEEE International Conference on Electric Vehicles, Greenville*, pages 1–8.

Stüdli, S., Crisostomi, E., Middleton, R., and Shorten, R. (2012b). A flexible distributed framework for realising electric and plug-in hybrid vehicle charging policies. *International Journal of Control*, 85(8):1130–1145.

SUPERHUB Consortium (2014). SUPERHUB - SUstainable and PERsuasive Human Users moBility in future cities. Available online `http://superhub-project.eu`, last accessed 20/01/2014.

Tate, E., Harpster, M., and Savagian, P. (2008). The electrification of the automobile: From conventional hybrid, to plug-in hybrids, to extended-range electric vehicles. In *SAE World Congress & Exhibition, Detroit. Reprinted from: Advanced Hybrid Vehicle Powertrain.*

TEAM Consortium (2014). TEAM - Tomorrow's Elastic Adaptive Mobility. Available online `http://www.collaborative-team.eu`, last accessed 20/01/2014.

Teodorović, D. and Lucić, P. (2006). Intelligent parking systems. *European Journal of Operational Research*, 175(3):1666–1681.

The Economist (2012). Open-air computers - cities are turning into vast data factories. Available online `http://www.economist.com/news/special-report/21564998-cities-are-turning-vast-data-factories-open-air-computers`, last accessed 20/01/2014.

Theakston, F., editor (2011). *Burden of disease from environmental noise - Quantification of healthy life years lost in Europe.* World Health Organisation.

Transport for London (2013). Congestion Charging. Available online `http://www.tfl.gov.uk/roadusers/congestioncharging`, last accessed 20/01/2014.

UK Department of Transport (2007). Design manual for roads and bridges, volume 11, section 3, part 1, air quality. Technical report, Highways Agency, Scottish Executive Development Department, Welsh Assembly Government, Department for Regional Development Northern Ireland.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

Wade, J. (2012). Dublin start-up wins at European Space Agency technology awards. Available online `http://businessetc.thejournal.ie/dublin-start-up-wins-at-european-space-agency-technology-awards-652257-Oct2012`http://arxiv.org/abs/1304.3949.

Weikl, S. and Bogenberger, K. (2012). Relocation strategies and algorithms for free-floating car sharing systems. In *International IEEE Conference on Intelligent Transportation Systems*, pages 355–360.

Worley, O., Klabjan, D., and Sweda, T. M. (2012). Simultaneous vehicle routing and charging station siting for commercial electric vehicles. In *Proceedings of IEEE Electric Vehicles Conference, Greenville*, pages 1–5.

Wu, C. W. (2005). Algebraic connectivity of directed graphs. *Linear and Multilinear Algebra*, 53(3):203–223.

Yang, B., Kaul, M., and Jensen, C. S. (2013). Using incomplete information for complete weight annotation of road networks. *accepted for publication in IEEE Transactions on Knowledge and Data Engineering.*

Zegeye, S. K. (2011). *Model-Based Traffic Control for Sustainable Mobility.* PhD thesis, Technische Universiteit Delft.

# A. On the Fair Coexistence of Loss- and Delay-Based TCP

***Abstract :***

*In this chapter we report an additional piece of work that was conducted in part during my PhD. This work was conducted jointly with Łukasz Budsisz, Rade Stanojević, Robert Shorten and Fred Baker. It was published in [Budzisz et al., 2011].*

Many algorithms have been proposed that avoid packet losses in communication networks and instead rely on delay measurement signals to adapt the window size. The reasoning is that if the time it takes to deliver a packet increases, this is caused by a buffer that is filling up and hence a packet loss is imminent. The window sizes can then be reduced to avoid packet losses. While the performance of delay-based algorithms is better when they are isolated, they do not compete well with loss-based variants. A basic question that has occupied researchers over the past three decades concerns methods to make delay based algorithms coexist fairly with loss-based ones. In [Budzisz et al., 2009] one such algorithm was proposed. It is based on non-linear queue dynamics and the construction of unstable equilibria. This algorithm is capable of efficient delay-based operation, but switches to a loss-based operation in the presence of loss-based algorithms. It is based on an idea presented in [Bhandarkar et al., 2007]. The algorithm is reported in Algorithm A.1. $\delta_{min}, \delta_{th}, \delta_{max}$ are positive design parameters with $\delta_{max} > \delta_{th} > \delta_{min}$. A further design parameter is $p_{max} \in [0, 1]$.

The function $p$ used in the algorithm is plotted in Figure A.1. If only Algorithm A.1 is used in a network, than all flows operate in the left part of the plot (zone A), where delays are small. When loss-based algorithms are present, the delay-based algorithms operate on the right site (zone B) of the plot, where few random back-off events happen and the algorithm operates on loss signals. Further if the loss-based algorithms leave the network, Algorithm A.1 reverts to operating in zone A. The interested reader is referred to [Budzisz et al., 2011] for full details. We now concentrate on the mathematical analysis of this algorithm.

**Algorithm A.1** Pseudo-code for algorithm

If a packet was successfully delivered:
Estimate the delay it experienced: $\delta$
Set $p(\delta)$:

$$p(\delta) = \begin{cases} p_{max} \frac{\delta - \delta_{min}}{\delta_{th} - \delta_{min}} & \text{for} \quad \delta_{min} < \delta \leq \delta_{th} \\ p_{max} \left( \frac{\delta_{max} - \delta}{\delta_{max} - \delta_{th}} \right)^4 & \text{for} \quad \delta_{th} < \delta < \delta_{max} \end{cases}$$

Pick a random number $rand$, uniformly from 0 to 1
**if** $rand < p$ **then**
    reduce $cwnd$ by $0.5 \cdot cwnd$
**else**
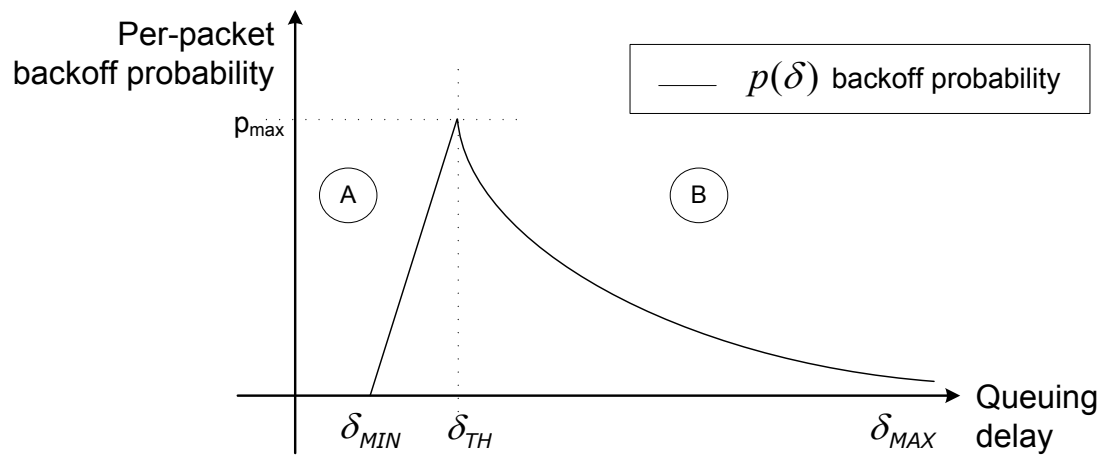    increment $cwnd$ by $1/cwnd$
**end if**



Figure A.1.: Per-packet back-off probability as a function of the observed delay.

## A.1. Mathematical Analysis

[Budzisz et al., 2011] presents stability proofs for systems using the above algorithm in two scenarios. In the first scenario a number of users compete for resources on a single link. In the second scenario a network of several links is considered and each user uses a subset of all links. We presented initial analysis for both scenarios in [Schlote, 2009]. However, the analysis relied on a very restrictive assumption, namely, this assumption was that $\delta(t)$ in the first term on the right hand side of Equation (A.2) is in fact independent of $t$. We removed this assumption in the analysis presented here for the single link scenario. What we prove here is that if the system parameters are chosen correctly, there is an attractive and stable equilibrium operating point in zone A in Figure A.1, while on the right side of that plot any equilibrium must be unstable. We refer again to [Budzisz et al., 2011] for a full discussion of these results.

We show, using a fluid (Kelly) like argument, that the we can identify stable and unstable equilibria in the system. To this end for $i = 1, \ldots, N$ let $\delta(t)$ and $W_i(t)$ be the queueing delay and congestion window size (*cwnd*) of flow $i$ at time $t$. The minimal time it takes to deliver a packet for user $i$ is $RTT_i$.

We essentially use a mean field fluid model to demonstrate the plausibility of our approach. Clearly, an accurate modelling of TCP dynamics merits a much more complicated discrete event model. Nevertheless, our approach follows accepted practice in the community, and these models do provide a reasonable approximation. Our analysis is based on the following assumption. The evolution of the average window size over one RTT is determined by: (i) the probability of a backoff; and (ii) the rate at which the window variable grows. When analysed over one small time interval $\Delta t$, the equation describing this evolution is the same for all flows and is given by: $\Delta W_i(t) = \frac{\Delta t}{RTT_i + \delta(t)} - q_0^i \cdot \frac{W_i(t)}{2}$, where $q_0^i$ is the probability that during the time interval $(t, t + \Delta t)$ a back-off occurred for flow $i$. Dividing by $\Delta t$ yields

$$\frac{\Delta W_i(t)}{\Delta t} = \frac{1}{RTT_i + \delta(t)} - \frac{q_0^i}{\Delta t} \cdot \frac{W_i(t)}{2}, \tag{A.1}$$

We denote by $M_0^i$ the number of packets the source of flow $i$ with congestion window size of $W_i(t)$ sends in the interval $(t, t + \Delta t)$. Then $M_0^i = \frac{\Delta t \cdot W_i(t)}{RTT_i + \delta(t)}$, and we can approximate $q_0^i$ as $q_0^i = 1 - (1 - p)^{M_0^i} \approx p M_0^i = \frac{p \Delta t \cdot W_i(t)}{RTT_i + \delta(t)}$. In the limit $\Delta t \to 0$ this yields

$$\dot{W}_i(t) = \frac{1}{RTT_i + \delta(t)} \left( 1 - \frac{p}{2} W_i^2(t) \right). \tag{A.2}$$

We are going to use this very general equation (A.2) to model our system also in the multiple bottleneck case.

We start our analysis by considering a scenario with $N$ flows competing at a single bottleneck. The dynamic system we are going to analyse is given by equation (A.2) for

all $i = 1, \ldots, N$, and we model queue dynamics as

$$
\dot{q} = \begin{cases} \sum_{i=1}^{N} \frac{W_i}{\tau_i + \delta} - C & \text{for} \quad q > 0 \\ \left\{ \sum_{i=1}^{N} \frac{W_i}{\tau_i + \delta} - C \right\}^+ & \text{for} \quad q = 0 \end{cases}, \tag{A.3}
$$

where $q$ is the queue length. For any function $f$ that maps into the real numbers, we denote by $\{f\}^+$ the non-negative part of $f$, that is $\{f\}^+ = \max\{f, 0\}$. $q$ and $\delta$ are coupled as $\delta = \frac{q}{C}$. If we set $\tau_i = RTT_i + \delta_{min}$ and abusing notation $\delta = \delta - \delta_{min}$ and accordingly $q = q - C\delta_{min}$, then the back-off probability function $p$ is of the form

$$
p(\delta) = \begin{cases} K\delta & \text{for} \quad 0 \le \delta \le \delta_{th}, \\ p_0 - \hat{K}\delta & \text{for} \quad \delta_{th} \le \delta \le \delta_{max}, \end{cases} \tag{A.4}
$$

for positive constants $K, p_0, \hat{K}$.

For mathematical convenience we have assumed a piecewise linear function $p(\delta)$. The following analysis can be carried out for any $p(\delta)$ with the same qualitative properties; namely it strictly increases in $(0, \delta_{th})$ and strictly decreases in $(\delta_{th}, \delta_{max})$. Some of the details change, but the approach extends in complete generality.

**Lemma A.1.** *The system* (A.2), (A.3), (A.4) *can be designed such that it has two equilibria in the positive orthant. One for $\delta \in [0, \delta_{th}]$ and one for $\delta \in [\delta_{th}, \delta_{max}]$.*

*Proof.* Setting Equation (A.2) equal to 0 yields $W_i^2 = \frac{2}{p(\delta)}$ and thus $W_i = \sqrt{\frac{2}{p(\delta)}}$ as we are only interested in physically feasible solutions. That implies that in equilibrium all window sizes are the same: $W_1^* = \cdots = W_N^* =: W^*$. Now setting equation (A.3) to 0 yields $C = \sqrt{\frac{2}{p(\delta)}} \sum_{i=1}^{N} \frac{1}{\tau_i + \delta}$. For $\delta \in [0, \delta_{th}]$ we have $p(\delta) = K\delta$ and thus

$$
C = \sqrt{\frac{2}{K\delta}} \sum_{i=1}^{N} \frac{1}{\tau_i + \delta}. \tag{A.5}
$$

It can be seen that the right hand side of this equation is strictly decreasing in $\delta$, hence there is at most one $\delta$ satisfying equation (A.5). By choosing $K$ appropriately we can assure its existence in the aspired position.

For $\delta \in [\delta_{th}, \delta_{max}]$ we have $p(\delta) = p_0 - \hat{K}\delta$ and thus

$$
C\sqrt{\frac{p_0 - \hat{K}\delta}{2}} = \sum_{i=1}^{N} \frac{1}{\tau_i + \delta}. \tag{A.6}
$$

It is easy to see, that for positive $\delta$ the left hand side of this equation is strictly concave, while the right hand side is strictly convex. Accordingly there are no more then 2 possible

values for $\delta$ to solve equation (A.6). Bearing in mind that we need $p_0 - \hat{K}\delta_{th} = K\delta_{th}$ for $p(\delta)$ to be continuous, the parameters $\hat{K}, p_0$ can now be chosen such that the largest value of $\delta$ is in $\delta \in [\delta_{th}, \delta_{max}]$ while the other is not and will be cut off as a result. $\qquad\square$

**Lemma A.2.** *The equilibrium with $\delta \in [0, \delta_{th}]$ is locally asymptotically stable.*

*Proof.* We use the following coordinate transformation. $\tilde{W}_i = W_i - W^*$ for all $i = 1, \ldots, N$ and $\tilde{q} = q - q^*$. This shifts the equilibrium the point with all coordinates equal to 0 and enables us to apply a standard Lyapunov argument. Accordingly we define $\tilde{\delta} = \delta - \delta^*$.

We will further write equation (A.3) in the form

$$
\begin{aligned}
\dot{\tilde{q}} = \dot{q} + \dot{q}^* = \dot{q} &= \sum_{i=1}^{N} \frac{W_i}{\tau_i + \delta} - C \\
&= \sum_{i=1}^{N} \frac{\tilde{W}_i + W_i^*}{\tau_i + \delta} - C = \sum_{i=1}^{N} \frac{\tilde{W}_i}{\tau_i + \delta} + \sum_{i=1}^{N} \frac{W_i^*}{\tau_i + \delta} - C \\
&= \sum_{i=1}^{N} \frac{\tilde{W}_i}{\tau_i + \delta} - \sum_{i=1}^{N} \frac{W_i^* \tilde{\delta}}{(\tau_i + \delta^* + \tilde{\delta})(\tau_i + \delta^*)} + \sum_{i=1}^{N} \frac{W_i^*}{\tau_i + \delta^*} - C \\
&= \sum_{i=1}^{N} \frac{\tilde{W}_i}{\tau_i + \delta} - \sum_{i=1}^{N} \frac{W_i^* \tilde{\delta}}{(\tau_i + \delta^* + \tilde{\delta})(\tau_i + \delta^*)}, \quad\quad\quad (A.7)
\end{aligned}
$$

where we used the identity $\frac{1}{\tau_i+\delta} = \frac{1}{\tau_i+\tilde{\delta}+\delta^*} = \frac{-\tilde{\delta}}{(\tau_i+\tilde{\delta}+\delta^*)(\tau_i+\delta^*)} + \frac{1}{\tau+\delta^*}$ and in the last step we used that at the equilibrium $\sum_{i=1}^{N} \frac{W_i^*}{\tau_i+\delta^*} = C$. And we write equation (A.2) in the form

$$
\begin{aligned}
\dot{\tilde{W}}_i = \dot{W}_i + \dot{W}_i^* = \dot{W}_i &= \frac{1}{\tau_i + \delta}\left(1 - \frac{p(\delta)}{2} W_i^2\right) \\
&= \frac{1}{\tau_i + \delta}\left(1 - \frac{K\frac{q}{C}}{2} W_i^2\right) \\
&= \frac{1}{\tau_i + \delta}\left(1 - \frac{K\frac{\tilde{q}+q^*}{C}}{2}\left(\tilde{W}_i + W_i^*\right)^2\right) \\
&= \frac{1}{\tau_i + \delta}\left(1 - \frac{K\frac{q^*}{C}}{2}(W_i^*)^2\right) \\
&\quad - \frac{1}{2}\frac{K}{\tau_i + \delta}\frac{q}{C}\left(\tilde{W}_i^2 + 2\tilde{W}_i W_i^*\right) - \frac{1}{2}\frac{K}{\tau_i + \delta}\frac{\tilde{q}}{C}(W_i^*)^2 \\
&= -\frac{1}{2}\frac{K}{\tau_i + \delta}\frac{q}{C}\left(\tilde{W}_i^2 + 2\tilde{W}_i W_i^*\right) - \frac{1}{2}\frac{K}{\tau_i + \delta}\frac{\tilde{q}}{C}(W_i^*)^2 \quad\quad (A.8)
\end{aligned}
$$

Thus for all $i = 1, \ldots, N$ the systems dynamics can be described by

$$\dot{\tilde{W}}_i = -\frac{1}{2}\frac{K}{\tau_i + \delta}\frac{q}{C}\left(\tilde{W}_i^2 + 2\tilde{W}_i W_i^*\right) - \frac{1}{2}\frac{K}{\tau_i + \delta}\frac{\tilde{q}}{C}(W_i^*)^2 \tag{A.9}$$

$$\dot{\tilde{q}} = \sum_{i=1}^N \frac{\tilde{W}_i}{\tau_i + \delta} - \sum_{i=1}^N \frac{\tilde{\delta}W_i^*}{(\tau_i + \delta)(\tau_i + \delta^*)}. \tag{A.10}$$

Now we show that the function $V : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}_+$ given by $V(\tilde{W}_1, \ldots, \tilde{W}_N, \tilde{q}) = \sum_{i=1}^N \frac{\tilde{W}_i^2}{(W_i^*)^2} + \frac{K}{2C}\tilde{q}^2$ is a Lyapunov function with which we can prove asymptotic stability of the equilibrium. To this end we look at its time derivative.

$$\dot{V}(\tilde{W}_1, \ldots, \tilde{W}_N, \tilde{q}) = \sum_{i=1}^N \frac{\tilde{W}_i}{(W_i^*)^2}\dot{\tilde{W}}_i + \frac{K}{C}\tilde{q}\dot{\tilde{q}}$$

$$= -\sum_{i=1}^N \frac{\tilde{W}_i}{(W_i^*)^2}\frac{1}{2}\frac{K}{\tau_i + \delta}\frac{q}{C}\left(\tilde{W}_i^2 + 2\tilde{W}_i W_i^*\right)$$

$$\quad - \sum_{i=1}^N \frac{\tilde{W}_i}{(W_i^*)^2}\frac{1}{2}\frac{K}{\tau_i + \delta}\frac{\tilde{q}}{C}(W_i^*)^2 \quad + \frac{K}{2C}\tilde{q}\sum_{i=1}^N \frac{\tilde{W}_i}{\tau_i + \delta}$$

$$\quad - \frac{K}{2C}\tilde{q}\sum_{i=1}^N \frac{\tilde{\delta}W_i^*}{(\tau_i + \delta)(\tau_i + \delta^*)}$$

$$= -\sum_{i=1}^N \frac{\tilde{W}_i^2}{(W_i^*)^2}\frac{1}{2}\frac{K}{\tau_i + \delta}\frac{q}{C}\left(\tilde{W}_i + 2W_i^*\right)$$

$$\quad - \frac{K}{2C^2}\tilde{q}^2\sum_{i=1}^N \frac{W_i^*}{(\tau_i + \delta)(\tau_i + \delta^*)} \leq 0 \tag{A.11}$$

Also $\dot{V}(\tilde{W}_1, \ldots, \tilde{W}_N, \tilde{q}) = 0$ only in the equilibrium. Thus by [Hinrichsen and J.Pritchard, 2004, Theorem 3.2.25] the equilibrium is locally asymptotically stable.

$\square$

**Comment:** Note that the equilibrium Equation (A.6) can be used to determine the maximum number of flows that can be supported in a stable fashion the network. To do this, one establishes the maximum $N$ for which a stable equilibrium state exists in the network; namely, the maximum $N$ such that equation (6) has a solution ($\delta$) to the left of the apex of the probability curve (for a given set of network parameters). See *model validation* for more discussion on this matter. Once such an equilibrium exists, one can also solve the equation for $\delta^*$ to find the equilibrium delay.

**Lemma A.3.** *The equilibrium with $\delta \in [\delta_{th}, \delta_{max}]$ is unstable.*

*Proof.* Here $p(\delta) = p_0 - \hat{K}\delta$ and the dynamic is described by

$$\dot{W}_i = \frac{1}{\tau_i + \delta}\left(1 - \frac{p_0 - \hat{K}\delta}{2}W_i^2\right) \tag{A.12}$$

for all $i = 1, \ldots, N$, and

$$\dot{q} = \sum_{i=1}^{N}\frac{W_i}{\tau_i + \delta} - C. \tag{A.13}$$

We will now argue that the subset $\mathcal{S}$ of the statespace where $\dot{q} > 0$ and $\dot{W}_i > 0$ for all $i = 1, \ldots, N$ is a positive invariant set, which means that any solution starting in $\mathcal{S}$ can never leave $\mathcal{S}$, and that in every neighbourhood of the equilibrium there are points in $\mathcal{S}$ that are strictly larger than the equilibrium. This then shows that the equilibrium not attractive.

We assume that solutions of (A.12), (A.13) are continuous, thus solutions starting in $\mathcal{S}$ cannot leave $\mathcal{S}$ without passing through its boundary. Regard any boundary point $(W_1, \ldots, W_N, q)$ of $\mathcal{S}$. Either $(W_1, \ldots, W_N, q)$ is the equilibrium or one of the $N+1$ equtions (A.12), (A.13) is positive. Let $\dot{q} = 0$ then we know that $\dot{W}_i > 0$ for one $i$ and as $W_i$ increases $\dot{q}$ is pushed into the positive numbers. Let now $\dot{q} > 0$ and consider any $i = 1, \ldots, N$ for which $\dot{W}_i = 0$. As $q$ increases $p_0 - \hat{K}\frac{q}{C}$ decreases, which increases $\dot{W}_i$. We conclude that $\mathcal{S} \cup \{(W_1^*, \ldots, W_N^*, q^*)\}$ is invariant. We now show that $\mathcal{S} \neq \emptyset$. To this end consider $q = \lambda q^*$ and $W_i = \sqrt{\frac{2}{p_0 - \hat{K}q}}$ for all $i = 1, \ldots, N$. This yields $\dot{W}_i = 0$ for all $i = 1, \ldots, N$ and

$$\dot{q} = \sqrt{\frac{2}{p_0 - \hat{K}\lambda q^*}}\sum_{i=1}^{N}\frac{1}{\tau_i + \frac{\lambda q^*}{\hat{K}}} - C. \tag{A.14}$$

This converges to $+\infty$ for $\lambda \to \frac{p_0}{\hat{K}q^*}, \lambda < \frac{p_0}{\hat{K}q^*}$. Hence, $\dot{q} > 0$ for all $1 < \lambda < \frac{p_0}{\hat{K}q^*}$, otherwise there would have to be another equilibrium in $[\delta_{th}, \delta_{max}]$, which we excluded.

In order to show that solutions starting in $\mathcal{S}$ are unbounded, we are going to assume the opposite. To this end let $x_0 \in \mathcal{S}$ and assume that starting in $x_0$ the system is bounded. Then the solution has two accumulation points, because it needs to have one and if there was only one, then that would be an equilibrium point. One of these accumulation points needs to be strictly larger than the other in at least one component. Using that all components of the solution are monotonically increasing it is easy to see that this scenario is impossible.

□