Title
Smart Traffic Operation: from Human-Driven Cars to Mixed Vehicle Autonomy

Author
Zahedi Mehr, Negar Zahedi

Publication Date
2019

Peer reviewed|Thesis/dissertation

Smart Traffic Operation: from Human–Driven Cars to Mixed Vehicle Autonomy

by

Negar Zahedi Mehr

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Roberto Horowitz, Chair
Professor Francesco Borrelli
Professor Murat Arcak

Summer 2019

Smart Traffic Operation: from Human–Driven Cars to Mixed Vehicle Autonomy

Abstract

Smart Traffic Operation: from Human–Driven Cars to Mixed Vehicle Autonomy

by

Negar Zahedi Mehr

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Roberto Horowitz, Chair

The goal of my research is to enhance urban mobility by developing reliable and efficient traffic control and management strategies. As cities grow everywhere, and urban roadways become overburdened, the need for the development of such strategies becomes more evident. With the prevalence of smart sensing devices, such as smart phones and smart intersections, cities are becoming smart. Moreover, with the emergence of new and inevitable technologies, such as autonomous and connected vehicles, electric vehicles, and mobility on demand systems, smart cities are rapidly evolving. As we experience the arrival of such technologies, there is an opportunity to reclaim urban mobility. However, a blind utilization of these technologies may deflect us from reaching this goal. In this dissertation, we study the efficient operation of smart cities via management strategies that can guarantee overall societal benefits both in the cities of today and future.

We focus on two natural instances of this agenda. In the first part, we tackle some of the existing challenges in the smart operation of traffic networks which are solely shared by human–driven cars. If all vehicles are human–driven, there is room for improving the efficiency of traffic networks by appropriate coordination and control of traffic signal lights. For these networks, we develop signal control algorithms that are capable of minimizing the number of stop–and–go movements, encoding fairness among vehicular arrivals, and are robust to the knowledge of system parameters. In the second part, we analyze fundamentals of traffic networks with mixed vehicle autonomy, where both human–driven and autonomous cars coexist on roadways. We study the mobility implications of selfish autonomy, i.e. autonomous cars that are not concerned about their overall impact and simply attempt to optimize their own travel benefits. Having shown the negative consequences that the increased deployment of selfish autonomy may have, we develop a pricing mechanism which can guarantee the overall societal–scale efficiency of traffic networks with mixed vehicle autonomy. Finally, we show how autonomy can act altruistically, i.e. by taking into account the decision making process of humans, autonomous cars can potentially plan for their actions in the favor of the overall good.

To Maman, Baba, Neshat, Ali, Tida,
and Payam.

Contents

Contents ii

List of Figures iv

I Introduction 1

1 Introduction 2
1.1 Thesis Approach . . . 3
1.2 Contributions . . . 4

2 Preliminaries 9
2.1 PointQ Model of Traffic Networks . . . 9
2.2 Travelers' Selfish Route Choice . . . 14

II Networks of Human–Driven Cars 23

3 Bandwidth Maximization 24
3.1 Path Relative Offsets . . . 25
3.2 Path Bandwidth . . . 27
3.3 Problem Formulation . . . 29
3.4 Problem Relaxation . . . 30
3.5 Numerical Experiments . . . 33
3.6 Case Study . . . 33
3.7 Chapter Summary . . . 35

4 Joint Perimeter and Signal Control 37
4.1 Control Algorithm . . . 38
4.2 Numerical Experiment . . . 42
4.3 Chapter Summary . . . 44

5 Signalization Control with Unknown System Parameters 46

5.1 Control Algorithm . . . 47
5.2 Numerical Example . . . 50
5.3 Chapter Summary . . . 52

III Networks with Mixed Vehicle Autonomy **54**

6 Mobility in the Presence of Selfish Autonomy **55**
6.1 Equilibrium Uniqueness . . . 57
6.2 Networks with a Single O/D Pair . . . 58
6.3 Networks with Multiple O/D Pairs . . . 62
6.4 Bounding Performance Degradation . . . 64
6.5 Chapter Summary . . . 69

7 Pricing Mixed–Autonomy Networks **70**
7.1 Undifferentiated Prices . . . 71
7.2 Differentiated prices . . . 72
7.3 Chapter Summary . . . 76

8 Altruistic Autonomy **77**
8.1 Modeling Framework . . . 77
8.2 Equilibrium Properties . . . 80
8.3 Model Validation . . . 87
8.4 Socially Optimal Lane Choice . . . 91
8.5 Optimal Lane Choice in Mixed–Autonomy Setting . . . 92
8.6 Chapter Summary . . . 98

9 Concluding Remarks **100**

Bibliography **102**

List of Figures

2.1 Schematic of an intersection. 12

3.1 Schematic of the sample network. 26
3.2 Reference trajectories for p and p'. Note that $\mathcal{M}^p = \{m_1, m_3, m_5, m_7\}$ and $\mathcal{M}^{p'} = \{m_2, m_4, m_6, m_8\}$. 28
3.3 A linear arterial. The three colored paths are samples of the generated paths. . 33
3.4 Histogram of positive bandwidths. 34
3.5 Topology of a traffic Network in San Diego. 35
3.6 Network total travel time for different number of paths and different sets of offsets. 36

4.1 Topology of the example network. 42
4.2 Arrivals and green ratios of the queues obtained from the utility maximization algorithm. 43
4.3 Sum of all queues vs. time. 44

5.1 Sum of all queues vs. time. 49
5.2 Green duration of stages in node 6. 50
5.3 Sum of all queues as a function of time when the service rates were varied. . . . 51
5.4 Green duration of the stages in node 6 when the service rates are varied. 52

6.1 A network with a single O/D pair and two paths. 57
6.2 A network with a single O/D pair (A to D) and three paths from A to D. 60
6.3 Maximum and minimum social delay at equilibrium for Example 2. 61
6.4 Maximum and minimum social delays at equilibrium for Example 4. 62
6.5 A network with three O/D pairs. 63
6.6 Social delay in Example 5 for different fractions of autonomous cars traveling along O/D pair AB when cars along all other O/D pairs are human–driven. . . 64

7.1 A network with two O/D pairs from A to B and A to C. 71

8.1 Example of a traffic diverge with two destination links P and Q. For exit link P, a steadfast car (blue car) constituting x_s^P and a bypassing car (orange car) forming x_b^P are shown. 78
8.2 Three possible configurations of $J_s^i(.)$ and $J_b^i(.)$. 81
8.3 The traffic diverge used in our simulations: the blue cars will take the upper (left) exit link; while the magenta cars will ultimately take the lower (right) exit. Using Krauss' car following model, a blue vehicle is bypassing right before the bifurcation. 87
8.4 The fraction of bypassing cars, x_b^i, predicted by our macroscopic model and the values measured from microscopic simulations as a function of demand fractions f^i's. 90
8.5 The fraction of bypassing cars, x_b^i, from simulation and the values of bypassing cars required for the social optimality as a function of f_i. 93
8.6 The fraction of human–driven bypassing cars versus the fraction of autonomous cars that are commanded to remain steadfast β for different values of autonomy fraction α. 95
8.7 The social cost per total demand versus the fraction of commanded autonomous steadfast cars β, for two different values of autonomy fractions α. 97
8.8 The fraction of mixed bypassing cars versus the fraction of commanded steadfast cars β for the autonomy fraction $\alpha = 0.5$. 99

Acknowledgments

First and foremost, I would like to thank my research advisor Professor Roberto Horowitz for his constant support. Roberto has always provided me with the freedom to explore various directions to find the ones that interest me the most, while guiding me in every project to ensure that I make the most out of it. Roberto's trust in my abilities as a researcher has always made me confident in pursuing what I once considered to be "too ambitious to achieve".

I have been extremely fortunate to learn from Professor Pravin Varaiya throughout my graduate years. I have had the privilege of receiving Pravin's feedback and criticism on my research work. Now, I know how rewarding it is to hear an "Excellent!" comment from him at the end of a presentation. I also would like to thank Professor Murat Arcak for his constant support and feedback to all members of our transportation group meeting including me. Murat's smiles and nods have always been my hope source during my presentations. Many of my work has been either in close collaboration with or inspired by Professor Ramtin Pedarsani. When starting graduate school, Ramtin had a vision for my future that I hadn't even dared dream about. Ramtin, I am very thankful for your unsparing advice and support.

Especial thanks to the faculty who helped me establish my research and Controls knowledge. Masayoshi Tomizuka and Claire Tomlin for their support, advice, and getting me introduced to Controls. Kameshwar Poolla for his advice and feedback in our fruitful conversations which have always provided me with confidence at the times when I have needed it the most. Andrew Packard for getting me excited about linear algebra and spending days and nights on mathematical proofs. Francesco Borrelli who made me determined to pursue Controls in his course Model Predictive Control.

I am also grateful for the members of our transportation group meeting. Especial thanks to Dr. Alex Kurzhanskiy and Dr. Gabriel Gomes for their continual mentorship, patience, and feedback. I would not have been able to go through my first year at Berkeley without Alex and Gabriel's support. I would like to thank Matthew Wright, Eric Kim, Ruolin Li, Stanley Smith, Mikhail Burov, and Cheng–Ju Wu for all the collaborations, brainstorms, help with preparing my talks, and being great friends of mine.

I would also like to thank all my co-authors and collaborators: Jennie Lioris, Marc Sanselme, and Nitzan Orr who helped me with some of the work that is presented in this dissertation.

I cannot thank enough my amazing and caring friends at Berkeley who have enriched my life over the last few years. I am extremely grateful for my Manville mates Zhaleh Amini and Delagah Dadbeh for our over–the–tea discussions which encouraged me to challenge myself and look at the world differently. I am especially thankful to Dorsa Sadigh and Nima Anari for always being there for me, finding time for me in their super busy schedule, and always being a source of inspiration. Thanks to all members of GAB (Girls At Berkeley) lunch to whom I owe my survival during my first year at Berkeley: Samaneh Azadi, Mohaddeseh Peyro, Maryam Farahmand, Razieh Ghorbani, Maedeh

Golshirazi, and Zeinab Heidari. Thanks to my remarkable friends who have brought joy to my life in the past few years, Yasaman Yavari, Ameneh Gholipour, Maryam Shadmehr, Maryam Moeini, Mohammad Keshavarzi, Ali Yadollahi, and Reza Abbasi–Asl. I could not have asked for better companions for all the nights we played Tichu, Avalon, and Charades together.

I devote my very special thanks to my family. My parents Shahla Koushki and Manouchehr Zahedi Mehr for their sacrifices, love, support, and teaching little 'Naanaa' how to be strong and constantly go beyond her comfort zone. My sister Neshat, for always being the 'Neshat' (joy) of my life. Thanks to Ali Nouri for his unconditional and sincere support which has always brought me with confidence, and Tida Nouri for making me the happiest auntie.

My final thanks go to Payam Delgosha who has brought delight and happiness to my life. Payam, you have raised the upper bounds of support, honesty, optimism, and patience in my mind. I started to perceive the true joy and excitement for life in 212 Cory hall, and I would not trade our 'mathematical analysis' discussions for anything.

Part I

Introduction

Chapter 1

Introduction

Mobility is the lifeblood of every city. However, as cities grow everywhere, and urban roadways become overburdened, the costs that are incurred by individuals and the infrastructure increase. Only in the U.S., congestion caused 5.5 billion hours of extra wait time and 2.9 billion gallons of extra burnt fuel in 2011 [99]. This led to an overall cost of $121 billion in the U.S. in 2011 excluding the environmental costs. It is further anticipated that congestion will likely grow 40% by 2025. The significance of congestion costs as well as the expected growth of congestion costs point to the need for development of efficient strategies that are capable of reducing or even avoiding congestion. An immediate solution for alleviating congestion is to expand the physical infrastructure to increase the capacity of road networks. However, due to the over–priced and prolonged nature of road constructions, the growth rate of traffic demand is normally higher than the expansion rate of the physical infrastructure. As a result, infrastructure expansion is not enough for better operation of traffic networks; further traffic management and control strategies are required to attain the maximum efficiency of the existing physical infrastructure.

With the prevalence of smart sensing and Internet of Things (IoT) devices, such as network of road sensors, smart phones, and sensors on autonomous vehicles, critical physical infrastructures that our society heavily relies on are being connected to the cyber world. As a result of this connectivity, cities are becoming smart. Cities are becoming cyber–physical systems that are competent of monitoring and controlling their operation in real time. Such competency of smart cities can be leveraged for better management and control of the operation of city infrastructures including traffic networks. Moreover, with the emergence of new and inevitable technologies, such as autonomous vehicles, mobility on demand systems, and electric vehicles smart cities are rapidly evolving. However, the sole act of introducing and injecting new technologies into a society will not necessarily solve its challenges; it may even exacerbate the situation. For example, mobility on demand systems which are widely used all over the world are now known to contribute to congestion rises in major cities such as San Francisco [82]. Or, for instance, the adoption of navigation apps such as waze has led to issues in public safety in cities such as Los Angeles as cars take more unsafe turns and more unpermitted directions due to the shortcuts

that such navigation apps suggest [81]. With this type of experience, as we prepare for the emergence and deployment of autonomous cars in the cities of future, how do we know that we are not going to face the same experience with autonomous cars? Are we guaranteed that we will not face exacerbated congestion of autonomous cars? Will the cities benefit or suffer from the presence of autonomy? There is an opportunity to reclaim urban mobility provided that a *proactive approach* is taken in the deployment of autonomy in the cities of future. We need to plan for autonomy deployment such that societal–scale mobility benefits are attained. This brings us to a key fundamental problem that we would like to study: how to leverage the connectivity that is inherent in smart cities as well as the opportunities that autonomous cars provide, to achieve efficient operation of smart cities? We study this problem in both current cities which include only human–driven cars and future cities where both human–driven and autonomous cars will coexist, i.e. traffic networks with "mixed vehicle autonomy".

1.1 Thesis Approach

In order to study mobility in smart cities, we divide this thesis into two parts. First, we focus on traffic networks that contain only human–driven cars. For these networks, the major control input that is effective, is the exogenous control that the infrastructure imposes through signal lights. Therefore, when studying such networks, the key challenge is to decide on the appropriate exogenous control such as the signal timing plans at intersections. However, since the timing plan of one intersection will affect the operation of its adjacent intersections, intersections cannot be considered in isolation. Due to this coupling effect, for better efficiency of traffic networks, *coordinated* signal control strategies that can optimize the overall performance of a traffic network are required. In the first part, we tackle some of the challenges in the development of efficient coordinated signal control strategies in urban traffic networks. In particular, we study and develop the following:

1. Coordination of signal lights via appropriate selection of signal offsets for maximization of the green wave that cars will face along their paths, aka. paths' bandwidth.

2. Coordination of signal and perimeter control for stabilizing traffic networks when traffic demand exceeds the network supply while guaranteeing fairness among vehicular demand.

3. Coordination of signal control strategies for stabilizing traffic networks when vehicular demands and link capacities are unknown to the controller.

In the second part, we study mobility in mixed–autonomy traffic networks. One of the key aspects for predicting and studying the performance of mixed–autonomy networks

is to understand how humans will behave in the presence of autonomy and in general the decision making process of humans. Cities involve competitive decision–making humans, where everyone seeks to optimize their own objective which may not necessarily be aligned with the overall performance of the system. For instance, when traveling from an origin to a destination, every traveler selects the shortest available route, regardless of how their route choice will affect the overall delay. Our approach to study cities with mixed autonomy is to model the humans' decision making and the dynamics of the physical world, and leverage these models to study how humans would change their behavior once autonomous cars are deployed. For example, how would a traveler change their route once autonomy is present in the system. Then, taking into account how humans would behave in the presence of autonomy, we can plan for control strategies that guarantee increased mobility in the cities of future. This approach is multilateral, and we bridge ideas from control theory, game theory, and transportation engineering. In particular, for mixed–autonomy networks, we discuss the following:

1. Societal–scale mobility in the presence of selfish autonomy.
2. Achieving better efficiency via exogenous control inputs such as pricing mechanisms under selfish autonomy.
3. Altruistic autonomy, i.e. taking into account how humans would decide, how autonomous systems can act in the favor of the overall mobility.

In the remainder of this chapter, we discuss the contributions of this thesis in more detail.

1.2 Contributions

Offset Selection for Bandwidth Maximization

If demand is within the capacity of a network, and all network parameters and traffic patterns are known, the green duration of fixed–time signals are determined such that stability is guaranteed. Then, signal offsets which determine the relative starting time of signals with respect to a global clock can be further determined to increase the green duration that cars will experience while traversing their paths, i.e. the bandwidths that cars on multiple routes receive. In Chapter 3, we consider the problem of coordinating signal offsets for fixed–time signals in an urban network of arbitrary shape so as to maximize the bandwidths that cars experience. Assuming that all signals have a common cycle time, we define and utilize the concept of relative path offsets and formulate the problem of maximizing a weighted sum of path bandwidths. This leads to a nonlinear optimization problem. We demonstrate how this optimization problem can be converted to a mixed–integer linear program; hence, providing a scalable computational framework. Our approach is in fact a generalization of a previous method in which the single arterial

problem was found to be equivalent to a linear program [35], and is distinct from the traditional formulation as a mixed–integer program. We further show the practicality of our approach in a case study of a traffic network in San Diego, California. This work is previously published in:

- [70] **N. Mehr**, M. Sanselme, N. Orr, R. Horowitz, G. Gomes, Offset Selection for Bandwidth Maximization on Multiple Routes, in Annual American Control Conference 2018.

Joint Perimeter and Signal Control

The majority of coordinated signal control strategies are developed for stabilizing traffic networks subject to feasible demands, i.e., the demand profiles that lie within the capacity of traffic networks. But, with the ongoing rise of demand in traffic networks, traffic networks normally face situations where the demand exceeds network capacity. This is a typical scenario during rush and peak hours in metropolitan area for instance. To deal with these situations, congestion control has become of major importance for urban areas. In Chapter 4, we introduce the notion of network utility maximization for boundary flow control of urban networks a.k.a. congestion control. We describe how maximizing the aggregate utility of the network leads to a fair allocation of network resources to different demands while maintaining system stability. We demonstrate how utility maximization problem can be solved using Alternating Direction Method of Multipliers (ADMM). We further show how our algorithm can be partially distributed such that each entry link finds its appropriate demand value for maximizing its own objective while maximizing the total utility of the network. We showcase the performance of our algorithm in an example illustrating fast convergence of our method and its capability to stabilize the network. The results of this chapter were previously included in the following publication:

- [69] **N. Mehr**, J. Lioris, R. Horowitz, R. Pedarsani, Joint Perimeter and Signal Control of Urban Traffic via Network Utility Maximization, in 20th International Conference on Intelligent Transportation Systems, 2017.

Signal Control with Unknown System Parameters

Among several signal control strategies that have been proposed in the literature, a key assumption is that system parameters including link capacities or service rates and the value of demand are known. However, this may not necessarily be the case as the average demand of vehicles is normally unknown. Furthermore, it is envisaged that in the next generation of transportation networks with mixed autonomy, system parameters such as service rates or capacities may vary as a function of autonomy penetration rate. In such

scenarios, a coordinated signal control strategy which can, in real time, learn the stabilizing signal timing is required. Aligned with this, in chapter 5, we propose a signal control strategy which, unlike previous approaches, can handle both unknown mean network demands and service rates. To this end, we use stochastic gradient projection to develop a cyclic iterative control, where at every cycle, the timing plan of the signals is updated. At each iteration, the update rule is based on the measured changes in the network queue lengths. If the network mean arrival and service rates are assumed to be constant, the proposed iterative signal control is guaranteed to converge to an optimal solution. We describe the intuition behind our algorithm, and further demonstrate through simulation studies that our iterative control scheme can successfully stabilize the system. The work discussed in this chapter has previously appeared in the following publication:

- [71] **N. Mehr**, J. Lioris, R. Horowitz, R. Pedarsani, Signal Control for Urban Traffic Networks with Unknown System Parameters, in 21st International Conference on Intelligent Transportation Systems, 2018.

Mobility in the Presence of Selfish Autonomy

As autonomous cars are becoming tangible technologies, it is expected that autonomy deployment will transform mobility in cities. One major feature of autonomous cars that is believed to dramatically affect mobility is that connected and autonomous cars are capable of maintaining a shorter headway and distance when they form platoons of cars. Thus, such technologies can potentially increase the road capacities of traffic networks. Consequently, it is envisioned that their deployment will also increase the overall network mobility. In Chapter 6, we examine the validity of this expected impact, taking into account that travelers select their routes selfishly, in traffic networks with mixed vehicle autonomy, i.e. traffic networks with both human–driven and autonomous cars. We consider a nonatomic routing game on a network with inelastic (fixed) demands for a set of network O/D pairs, and study how replacing a fraction of human–driven cars by autonomous cars will affect mobility at equilibrium. Using well known US bureau of public roads (BPR) traffic delay models, we show that the resulting Wardrop equilibrium is not necessarily unique for networks with mixed autonomy. Then, we state the conditions under which the total network delay at equilibrium is guaranteed to not increase as the fraction of autonomous cars increases. However, we show that when these conditions do not hold, counterintuitive behaviors may occur: the total network delay can grow as the fraction of autonomous cars increases. In particular, we prove that for networks with a single origin–destination (O/D) pair, if the road degrees of capacity asymmetry (i.e. the ratio between the road capacity when all cars are human–driven and the road capacity when all cars are autonomous) are homogeneous, the total delay is 1) unique, and 2) a nonincreasing continuous function of the fraction of autonomous cars in the network. We show that for heterogeneous degrees of capacity asymmetry, the total delay is not unique, and it can further grow as the fraction of autonomous cars increases. We demon-

strate that similar behaviors may be observed in networks with multiple O/D pairs. We further bound such performance degradations due to the introduction of autonomy in general homogeneous networks. Chapter 6 is written using the material presented in the following publications.

- [62] **N. Mehr**, R. Horowitz, Can the Presence of Autonomous Vehicles Worsen the Equilibrium State of Traffic Networks?, in IEEE Conference on Decision and Control (CDC), 2018.

- [63] **N. Mehr**, R. Horowitz, How Will the Presence of Autonomous Vehicles Affect the Equilibrium State of Traffic Networks?, to appear in IEEE Transactions on Control of Network Systems.

Socially–Aware Control of Mixed–Autonomy Networks

Since travelers select their routes selfishly, traffic networks normally operate at an equilibrium characterized by Wardrop conditions. However, it is well known that equilibria are inefficient in general. In addition to the intrinsic inefficiency of equilibria, as shown in Chapter 6, in mixed–autonomy networks in which autonomous cars maintain a shorter headway than human–driven cars, increasing the fraction of autonomous cars in the network may increase the inefficiency of equilibria. In Chapter 7, we study the possibility of obviating the inefficiency of equilibria in mixed–autonomy traffic networks via pricing mechanisms. In particular, we study assigning prices to network links such that the overall or social delay of the resulting equilibria is minimum. First, we study the possibility of inducing such optimal equilibria by imposing a set of undifferentiated prices, i.e. a set of prices that treat both human–driven and autonomous cars similarly at each link. We provide an example which demonstrates that undifferentiated pricing is not sufficient for achieving minimum social delay. Then, we study differentiated pricing where the price of traversing each link may depend on whether cars are human–driven or autonomous. Under differentiated pricing, we prove that link prices obtained from the marginal cost taxation of links will induce equilibria with minimum social delay if the degrees of road capacity asymmetry (i.e. the ratio between the road capacity when all cars are human–driven and the road capacity when all cars are autonomous) are homogeneous among network links. Our discussion in Chapter 7 is presented in the following publication:

- [64] **N. Mehr**, R. Horowitz, Pricing Traffic Networks with Mixed Vehicle Autonomy, in Annual American Control Conference, 2019.

Altruistic Autonomy

In traffic networks with mixed vehicle autonomy, if autonomous cars have a good model of how humans make their decisions, they can plan for their actions such that when

human–driven cars respond to their actions, the overall mobility of the system is increased, i.e. autonomous cars can plan for their actions altruistically. For instance, when it comes to lane choices, autonomous cars can choose their lanes such that the overall mobility of the system is increased. An influential instance of such lane choices occur at traffic diverges where bypassing behavior is observed. Vehicle bypassing is known to increase delay at traffic diverges. However, due to the complexities of this phenomenon, accurate and yet simple models of such lane change maneuvers are hard to develop. In Chapter 8, we first develop a model for human drivers' lane choices and then show how this model can be leveraged by autonomous cars for achieving better efficiency of the traffic system at a traffic diverge.

We develop a macroscopic model for predicting the number of vehicles that perform a bypass at a traffic diverge for taking an exit link. We interpret the bypassing maneuver of vehicles at a traffic diverge as drivers acting selfishly; every vehicle selects lanes such that its own cost of travel is minimized. We discuss how we model the costs that are incurred by the vehicles. Then, taking into account the selfish behavior of vehicles, we model the lane choice of vehicles at a traffic diverge as an equilibrium. We state and prove the properties of the equilibrium in our model. We show that there always exists an equilibrium for our model. Moreover, although our model is an instance of nonlinear asymmetrical routing games which in general have multiple equilibria, we prove that the equilibrium of our model is unique under certain assumptions that we observed to hold in all our case studies. We discuss how our model can be calibrated by running a simple optimization problem. Then, using our calibrated model, we validate it through simulation studies and demonstrate that our model successfully predicts the aggregate lane change maneuvers that are performed by vehicles at a traffic diverge.

Having shown the predictive power of our model, we discuss how our model can be employed to obtain the optimal lane choice behavior of vehicles, where the social or the overall cost of all vehicles is minimized. This analysis is motivated by the full–autonomy scenario where all vehicles are autonomous, and the lane choice of every car can be controlled in the favor of the overall mobility. Finally, we demonstrate how our model can be utilized in scenarios where a central authority can dictate the lane choice and trajectory of certain vehicles so as to increase the overall mobility at a traffic diverge. Examples of such scenarios include the case when both human driven and autonomous vehicles coexist in the network. We show how certain decisions of the central authority can affect the overall delay via examples. This work is previously published in:

- [67] **N. Mehr**, R. Li, and, R. Horowitz, A Game Theoretic Model for Aggregate Bypassing Behavior of Vehicles at Traffic Diverges, in 21st International Conference on Intelligent Transportation Systems (ITSC), 2018.
- [68] **N. Mehr**, R. Li, and R. Horowitz, A Game Theoretic Macroscopic Model of Bypassing at Traffic Diverges with Applications to Mixed–Autonomy Networks, arXiv preprint arXiv:1809.02762, 2018.

Chapter 2

Preliminaries

In this chapter, we discuss the preliminaries required for analysis of traffic networks. In particular, this chapter is mainly devoted to describing the modeling framework that we will utilize in subsequent chapters. We start by reviewing how we model the "dynamics" of traffic networks; this model will be further used for the development of appropriate traffic controllers in the first part of this thesis. Then, we will discuss how travelers' "route choices" are normally modeled in traffic networks. We will use this model for predicting how travelers will select their routes in the presence of autonomy in the second part of this thesis.

2.1 PointQ Model of Traffic Networks

To model the dynamics of a traffic network, we use the PointQ model proposed in [108]. PointQ models traffic networks as store and forward queuing systems. PointQ allows for easily characterizing feasible demand profiles and stabilizing signal controls. We let the traffic network be denoted by $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of network nodes, and $\mathcal{L}$ is the set of network edges. The set of nodes $\mathcal{N}$ and edges $\mathcal{L}$ represent the sets of network intersections and links respectively. Let $|\mathcal{N}| = N$ and and $|\mathcal{L}| = L$ be the number of network nodes, and network links respectively.

In pointQ, network links are divided into three types: entry links $\mathcal{L}_{\text{entry}}$, internal links $\mathcal{L}_{\text{inter}}$, and exit links $\mathcal{L}_{\text{exit}}$. Entry links are the links that carry exogenous arrivals or demands to the network. They are identified by the fact that entry links do not have any starting node in the network. Internal links connect network nodes; hence, they have both starting and end nodes. Finally, exit links are the ones through which vehicles leave the network; thus, exit links do not have any end nodes in the network.

We consider two settings of the PointQ model. Network parameters such as exogenous demands are modeled either as deterministic or random variables. In the stochastic setting, for each link $l \in \mathcal{L}$, let f_l represent the long run average of the flow of the vehicles that leave link l, and d_l be the mean value of the exogenous demand on link l respectively.

If network is modeled deterministically, for each link $l \in \mathcal{L}$, f_l and d_l represent the absolute values of link flows and exogenous demand. Note that since we have assumed that the exogenous demands enter the network only through entry links, $d_l = 0$ for all links $l \in \mathcal{L}_{\text{inter}} \cup \mathcal{L}_{\text{exit}}$.

At each node $n \in \mathcal{N}$, only certain movements or phases are allowed. For each such movement, a separate queue is considered in PointQ. We use queues and movements interchangeably in this thesis. Each movement is characterized by its origin and destination links. Moreover, for each pair of links $l, m \in \mathcal{L}$, we let $f(l, m)$ be the (average) flow of vehicles moving from l to m (in the stochastic setting). We further use $r(l, m)$ to represent the fraction of vehicles that join link m when they leave link l in the deterministic setting, or equivalently, in the case of stochastic demands and capacities, the probability of the event that a random vehicle joins link m when departing link l. It is assumed that $r(l, m)$'s are known a priori. Moreover, for each movement from link $l \in \mathcal{L}$ to link $m \in \mathcal{L}$, we use $\mu(l, m)$ to represent the saturation flow or service rate of the movement. In the stochastic setting, for each link $l \in \mathcal{L}$, $\mu(l, m)$ is interpreted as the mean saturation flow or service rate of the movement. In the remainder of this section, for simplicity, we discuss the model for the deterministic setting although the framework generalizes to the stochastic setting with the stochastic interpretation of the discussed parameters.

At each node $n \in \mathcal{N}$, let $I(n)$ and $O(n)$ be the incoming and outgoing links. Since at each network node, flow conservation holds, we have

$$\sum_{l \in I(n)} f_l = \sum_{m \in O(n)} f_m, \quad \forall n \in \mathcal{N}. \tag{2.1}$$

Furthermore, for each l, m, and $o \in \mathcal{L}$, link flows and the movement flows must satisfy the following:

$$f_l = d_l, \qquad f(l, m) = r(l, m) f_l, \qquad \text{if } l \in \mathcal{L}_{\text{entry}}, \tag{2.2}$$

$$f_l = \sum_{o \in \mathcal{L}} f(o, l), \qquad f(o, l) = r(o, l) f_o, \qquad \text{if } l \in \mathcal{L}_{\text{inter}} \cup \mathcal{L}_{\text{exit}}. \tag{2.3}$$

With the description of network flows, we can proceed to describe the signalization cyclic control in our PointQ model.

Signalization Cyclic Control

Assume that all signal controllers in the signalized intersections are cyclic, i.e. they operate on a common cycle time T. We assume that each actuator, allocates a fraction of its cycle time to a certain set of movements. We call such a set of movements a stage. For this type of controllers, at each node $n \in \mathcal{N}$, there exist multiple stages $s_j^n, 1 \leq j \leq S^n$, where S^n is the total number of stages at node n. Each stage s_j^n is a set of non–conflicting movements that can be actuated simultaneously. Assume that stage s_j^n receives green for g_j^n fraction of the cycle time. A single stage might contain several non–conflicting

movements. Additionally, a movement might be actuated during multiple stages. For a movement from link l to m belonging to stage s_j^n, we use $g_j^n(l,m)$ to denote the fraction of green that this movement receives during stage s_j^n. Clearly, if multiple movements are actuated during a stage s_j^n, the green durations that they receive during stage s_j^n must be equal. In other words, if movements from links l and u to links m and v respectively are two of such movements that belong to the same stage s_j^n, we must have

$$g_j^n(l,m) = g_j^n(u,v) = g_j^n, \quad \forall\,(l,m) \text{ and } (u,v) \in s_j^n. \tag{2.4}$$

Note that at each node $n \in \mathcal{N}$, stage green durations g_j^n must add up to 1; therefore, we must have

$$\sum_{j=1}^{S^n} g_j^n = 1, \quad \forall n \in \mathcal{N}. \tag{2.5}$$

Remark 1. In Equation (2.5), no clearance time between signal stages is considered. If clearance times are also taken into account, g_j^n's must add up to $1 - \epsilon^n$, where ϵ^n is the fraction of the cycle time during which "all red" undergoes at node n.

At a node $n \in \mathcal{N}$, since movements may receive green during multiple stages, for each movement from l to m, we use $p(l,m)$ to denote the total fraction of green duration that the movement receives during a cycle time

$$p(l,m) = \sum_{j=1}^{S^n} g_j^n(l,m). \tag{2.6}$$

Using this definition, a given demand pattern is called feasible if and only if there exists a set of stages for all network nodes such that for every $l, m \in \mathcal{L}$, we have

$$f(l,m) \leq \mu(l,m)p(l,m). \tag{2.7}$$

Example: To illustrate the notation, consider the intersection shown in Figure 2.1. There is only one node in this network. Let this node be indexed by 1. Thus, superscript 1 is considered for the single node of the network. The intersection has 8 links with links 2, 4, 6, and 8 being entry links, and 1, 3, 5 and 7 being exit links. There is no internal link in this example. Assuming that there exist only through and right movements, there are eight queues in the network, where the origin–destination links for all network queues are: (2,5), (4,7), (8,3), (6,1), (2,3), (4,5), (6,7) and (8,1). Assume that there are only 2 stages at the intersection such that each stage lasts half of the cycle time. Assume that the following movements are actuated during each stage.

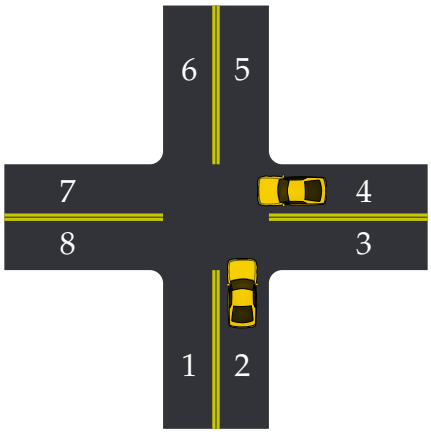


Figure 2.1: Schematic of an intersection.

First Stage: (2,5), (4,5), (2,3), (6,1), (8,1), and (6,7).
Second Stage: (4,7), (6,7), (4,5), (8,3), (8,1), and (2,3).
Then, using our notation, for the first stage, we have

$$g_1^1(2,5) = g_1^1(4,5) = g_1^1(2,3) = g_1^1(6,1) = g_1^1(8,1) = g_1^1(6,7) = g_1^1 = 0.5;$$

while, for the second stage, we have

$$g_2^1(4,7) = g_2^1(6,7) = g_2^1(4,5) = g_2^1(8,3) = g_2^1(8,1) = g_2^1(2,3) = g_2^1 = 0.5.$$

Note that following (2.5), we must have

$$g_1^1 + g_2^1 = 1.$$

Finally, the total green fraction of each movement is the following

$$p(2,5) = p(6,1) = p(4,7) = p(8,3) = 0.5,$$
$$p(4,5) = p(2,3) = p(8,1) = p(6,7) = 1.$$

Compact Notation of the Model

For readability purposes, we introduce a compact notation of the model parameters and variables. We use a notation similar to the one adopted in [69]. We use $d \in \mathbb{R}_+^L$ and $f \in \mathbb{R}_+^L$ to represent the vectors of demands and flows for all links in the network. Each element d_l of the vector d is simply equal to the exogenous arrival on link l if l is an entry link and zero otherwise. We can also collect the turning ratios $r(l,m)$'s into the matrix $R \in \mathbb{R}^{L\times L}$ such that $R_{lm} = r(l,m)$. Using this notation, the set of Equations (2.2) and (2.3) can be simply written as:

$$f = (I - R^T)^{-1} d. \tag{2.8}$$

In addition to the vector of *link* flows, we can construct the vector of *movement* flows. Assume that there exists a total of B possible movements in the network. We let $\varphi \in \mathbb{R}_+^B$ be the vector of movement flows $f(l,m)$, for all $l,m \in \mathcal{L}$ such that a movement from link l to link m is allowed. Likewise, we collect the total fraction of green that each movement receives, $p(l,m)$'s, in the *allocation vector* $p \in \mathbb{R}^B$. We further collect the service rates of all network queues, $\mu(l,m)$'s, in the diagonal matrix M, where the i_{th} diagonal element is equal to the service rate of the i_{th} movement.

Using Equations (2.2) and (2.3), it is easy to see that the linear mapping from the vector of link flows f to the vector of movement flows φ can be encoded using a matrix $\Gamma \in \mathbb{R}^{B\times L}$

$$\varphi = \Gamma f, \tag{2.9}$$

where Γ is such that at its j_{th} row, all elements are zero except for the j_{th} element which is equal to $r(l,m)$ with l and m being the origin and destination links of the j_{th} movement. Using this compact notation and (2.8), the stability condition (2.7) can be encoded as

$$\Gamma(I - R^T)^{-1}d \leq Mp. \tag{2.10}$$

Next, to represent conditions (2.4), (2.5), and (2.6), we collect the fractions of green that movements receive during all signal stages, $g_j^n(l,m)$'s in the vector $g \in \mathbb{R}^H$, where H is the total number of such durations. Then, we can rewrite Equation (2.6) for all queues as:

$$p = A_{g\to p}g, \tag{2.11}$$

where $A_{g\to p}$ is the linear mapping of appropriate dimension. Moreover, we define the linear mapping A_{eq} to encode (2.4) for all pairs of queues that belong to the same stage at an intersection,

$$A_{\text{eq}}g = \mathbf{0}, \tag{2.12}$$

where $\mathbf{0}$ is vector of zeros. Finally, we use the linear transformation A_{sum} in order to enforce (2.5) for all nodes in the network

$$A_{\text{sum}}g = \mathbf{1}^{N\times 1}, \tag{2.13}$$

where $\mathbf{1}^{N\times 1}$ is an N dimensional vector of 1's. Thus, the requirements imposed for cyclic implementation of a timing plan are represented by Equations (2.11), (2.12), and (2.13). Equations (2.12), and (2.13) encode the properties required for a timing plan to be implementable. We concatenate all these equations in the matrix A_c and vector b_c to represent implementability conditions via the following.

$$A_c g = b_c, \tag{2.14}$$

where A_c and b_c are matrices of the appropriate dimensions.

Finally, we need to define the vector of queue lengths for all network queues. We let $q \in \mathbb{R}^B$ denote the vector of queue lengths for all queues in the network.

2.2 Travelers' Selfish Route Choice

In the previous section, the assumption was that travelers' route choices were fixed and known to the controller through the matrix of turning ratios. But, to plan for the cities of future where autonomy is integrated in the city, we need to model how humans will change their behavior in the presence of autonomy, we need to model the decision making process of travelers. In this section, we discuss how travelers' route choices in a mixed–autonomy setting are modeled via nonatomic selfish routing games.

Nonatomic Selfish Routing

We model a traffic network by a directed graph $G = (\mathcal{N}, \mathcal{L}, \mathcal{W})$, where $\mathcal{N}$ and $\mathcal{L}$ are respectively the set of nodes and links in the network. Each link $l \in \mathcal{L}$ in the network is a pair of distinct nodes (u, v) and represents a directed edge from u towards v. We assume that each link joins two distinct nodes; thus, no self loops are allowed. Define $\mathcal{W} = \{(o_1, d_1), (o_2, d_2), \cdots, (o_k, d_k)\}$ to be the set of origin destination (O/D) vertex pairs of the network. A node $n \in \mathcal{N}$ can appear in multiple O/D pairs. In a nonatomic selfish routing game, if each O/D pair has a fixed given nonzero demand, then it is called a nonatomic selfish routing game with inelastic demand. Each O/D pair consists of infinitesimally small agents where every agent decides on their path such that their own delay of travel is minimized. The delay of each path depends on how network paths are shared among different O/D pairs. For each O/D pair $w = (o_i, d_i), 1 \leq i \leq k$, we let $\mathcal{P}_w$ denote the set of all possible network paths from o_i to d_i. We assume that the network topology is such that for each O/D pair $w \in \mathcal{W}$, there exists at least one path from its origin to its destination, i.e. $\mathcal{P}_w \neq \emptyset$. We further let $\mathcal{P} = \cup_{w \in W} \mathcal{P}_w$ denote the set of all network paths.

For an O/D pair $w \in \mathcal{W}$, let r_w be the given fixed demand of cars associated with w. It is important to note that in our setting, each O/D pair w has two classes of cars: autonomous and human–driven. Consequently, for each O/D pair $w \in \mathcal{W}$, we define α_w to be the fraction of cars in r_w that are autonomous. We let $r = (r_w : w \in \mathcal{W})$ and $\alpha = (\alpha_w : w \in \mathcal{W})$ be the vectors of network demand and autonomy fraction respectively. We define $r_w^h = (1 - \alpha_w) r_w$ and $r_w^a = (\alpha_w r_w)$ to respectively be the demand of human–driven and autonomous cars for each O/D pair $w \in \mathcal{W}$. Furthermore, for a path $p \in \mathcal{P}_w$, let f_p be the total flow of O/D pair w along path p. Note that each path connects exactly one origin to one and only one destination; thereby, once a path is fixed, its origin and destination are uniquely determined. Consequently, there is no need to explicitly include path O/D pairs in the notation used for f_p. Also, for each path $p \in \mathcal{P}_w$, we use f_p^h and f_p^a to respectively denote the flow of human–driven and autonomous cars along path p. Note that for each path $p \in \mathcal{P}$, we have $f_p = f_p^h + f_p^a$. Define the network flow vector f to be a nonnegative vector of human–driven and autonomous flows along network paths, i.e. $f = (f_p^h, f_p^a : p \in \mathcal{P})$. A flow vector f is called feasible for a given network G, if for

each O/D pair $w \in \mathcal{W}$,

$$\sum_{p \in \mathcal{P}_w} f_p^h = (1 - \alpha_w) r_w, \text{ and } \sum_{p \in \mathcal{P}_w} f_p^a = \alpha_w r_w, \tag{2.15a}$$

$$f_p^h \geq 0, \text{ and } f_p^a \geq 0, \ \forall p \in \mathcal{P}_w. \tag{2.15b}$$

For each link $l \in \mathcal{L}$, f_l is the total flow of cars along link l, i.e. $f_l = \sum_{p \in \mathcal{P}: l \in p} f_p$. Since we need to decompose the total link flow into human–driven and autonomous cars, we let f_l^h and f_l^a be the total flow of human–driven and autonomous cars along link l respectively. In fact, f_l^h and f_l^a are the summation of the flow of human–driven and autonomous cars on all routes containing link l,

$$f_l^h = \sum_{p \in \mathcal{P}: l \in p} f_p^h, \text{ and } f_l^a = \sum_{p \in \mathcal{P}: l \in p} f_p^a.$$

Note that if all cars are either human–driven or autonomous for an O/D pair $w \in \mathcal{W}$, i.e. either $\alpha_w = 0$ or $\alpha_w = 1$, then, we only have a single class of cars along that O/D pair, and for each path $p \in \mathcal{P}_w$, either $f_p = f_p^h$ or $f_p = f_p^a$. If for all network O/D pairs $w \in \mathcal{W}$, the autonomy fraction $\alpha_w = 0$, then $f_l = f_l^h$ for all links $l \in L$. In fact, if all cars are human–driven, our routing game reduces to a single class game.

$$(\forall w \in \mathcal{W}, \ \alpha_w = 0) \iff \left(\forall p \in \mathcal{P}, \ f_p = f_p^h\right). \tag{2.16}$$

In order to be able to model the incurred delays when cars are routed throughout the network, it is assumed that each link $l \in L$ has a delay per unit of flow function $e_l : \mathbb{R}^2 \to \mathbb{R}$. For a given class of link delay functions $e = (e_l(f_l^h, f_l^a) : l \in \mathcal{L})$ and a given demand vector r, we let the triple (G, r, e) represent a routing game on G with demand r and link delay functions e. We further assume that the delay per unit of flow for each path $p \in \mathcal{P}$ is obtained by the summation of link delays over the links that form p,

$$e_p(f) = \sum_{l \in \mathcal{L}: l \in p} e_l(f_l^h, f_l^a). \tag{2.17}$$

Equation (2.17) implies that the delay of each path $p \in \mathcal{P}$ depends not only on the flows of human–driven and autonomous cars along path p, but also on the flows along other paths. The overall network delay or social delay is given by

$$J(f) = \sum_{p \in \mathcal{P}} f_p e_p(f). \tag{2.18}$$

A flow vector $f^* = (f_p^{h^*}, f_p^{a^*} : p \in \mathcal{P})$ is socially optimal if and only if it minimizes $J(f)$ subject to relations (2.15). The optimal social delay is denoted by $J^* = J(f^*)$.

Wardrop Equilibrium

It is well known in the transportation literature that if there are many noncooperative agents, namely, flows that behave selfishly [95], a network is at an equilibrium if the Wardrop conditions hold [109]. The Wardrop conditions state that at equilibrium, no user has any incentive for unilaterally changing their path. This implies that for an equilibrium flow vector f, if there exists a path $p \in \mathcal{P}_w$ such that either $f_p^h \neq 0$ or $f_p^a \neq 0$, we must have that $e_p(f) \leq e_{p'}(f)$ for all paths $p' \in \mathcal{P}_w$.

Definition 1. *For a routing game (G, r, e), a feasible flow vector f is a Wardrop equilibrium if and only if for every O/D pair $w \in \mathcal{W}$ and every pair of paths $p, p' \in \mathcal{P}_w$,*

$$f_p^h \left(e_p(f) - e_{p'}(f) \right) \leq 0, \tag{2.19a}$$

$$f_p^a \left(e_p(f) - e_{p'}(f) \right) \leq 0. \tag{2.19b}$$

Note that an implication of the above definition is that for each O/D pair $w \in \mathcal{W}$, and any two paths $p, p' \in \mathcal{P}_w$ such that $f_p \neq 0$ and $f_{p'} \neq 0$, we must have that $e_p(f) = e_{p'}(f)$.

Definition 2. *Given an equilibrium flow vector f for a routing game (G, r, e), we define the delay of travel for each O/D pair $w \in \mathcal{W}$ to be*

$$e_w(f) := \min_{p \in \mathcal{P}_w} e_p(f). \tag{2.20}$$

Motivated by the above discussion, $e_w(f)$ is precisely the delay across all paths $p \in \mathcal{P}_w$ which have a nonzero flow. Moreover, the equilibrium conditions imply that for a path $p \in \mathcal{P}_w$ with zero flow, we have $e_p(f) \geq e_w(f)$.

It is worth mentioning that when there are no autonomous cars, i.e. for all O/D pairs $w \in \mathcal{W}, \alpha_w = 0$, since $f_p^h = f_p$ for all paths $p \in \mathcal{P}$, Conditions (2.19) reduce to

$$f_p \left(e_p(f) - e_{p'}(f) \right) \leq 0, \quad \forall w \in \mathcal{W},\ \forall p, p' \in \mathcal{P}_w. \tag{2.21}$$

Delay Characterization

We first specify the structure of our delay model. If there is only the single class of human–driven cars in the network, the US bureau of public roads (BPR) [92] suggests the following form of delay functions.

Assumption 1. *When network links are shared by only human–driven cars, the link delay functions $e_l : \mathbb{R} \to \mathbb{R}$ are of the following form*

$$e_l(f_l) = \left(a_l + \gamma_l \left(\frac{f_l}{C_l} \right)^{\beta_l} \right), \tag{2.22}$$

where C_l is the capacity of link l, and a_l, γ_l, and β_l are nonnegative link parameters.

In practice, a_l is normally the free–flow travel time on link l, γ_l is a constant link parameter, and β_l is a positive integer ranging from 1 to 4. However, we only require that a_l and γ_l be nonnegative link parameters, and β_l be a positive integer. In order to characterize the delay function in networks with mixed autonomy, where we have two classes of cars, we first need to model the impact of autonomous cars on link capacities. It was shown in [50] that in networks with mixed autonomy, C_l depends on the autonomy fraction of link l defined as

$$\alpha_l := \frac{f_l^a}{f_l^a + f_l^h}. \tag{2.23}$$

We use $C_l(\alpha_l)$ to emphasize this dependence. Let m_l and M_l be the capacity of link l when all cars are human–driven and autonomous respectively. Since autonomous cars are capable of maintaining shorter headways, it is normally the case that $\frac{m_l}{M_l} \leq 1$. When the two classes of human–driven and autonomous cars are present in the network, using the results in [50], we have

$$C_l(\alpha_l) = \frac{m_l M_l}{\alpha_l m_l + (1-\alpha_l) M_l}. \tag{2.24}$$

We adopt this capacity model throughout this thesis. Since for each link $l \in L$, $\alpha_l = \frac{f_l^a}{f_l^a + f_l^h}$ and $f_l = f_l^a + f_l^h$, using (2.24), for networks with mixed autonomy, the delay function (2.22) can be modified as

$$e_l(f_l^h, f_l^a) = \left(a_l + \gamma_l \left(\frac{f_l^h + f_l^a}{\frac{m_l M_l (f_l^h + f_l^a)}{m_l f_l^a + M_l f_l^h}} \right)^{\beta_l} \right) \tag{2.25}$$

$$= \left(a_l + \gamma_l \left(\frac{f_l^a}{M_l} + \frac{f_l^h}{m_l} \right)^{\beta_l} \right). \tag{2.26}$$

Note that when only human–driven cars are present in the network, for each link $l \in \mathcal{L}$, since $f_l = f_l^h$, the link delay function reverts to

$$e_l(f_l) = \left(a_l + \gamma_l \left(\frac{f_l^h}{m_l} \right)^{\beta_l} \right). \tag{2.27}$$

Link Prices

To exploit the potential mobility benefits of autonomous cars, regulatory mechanisms such as toll collection, may be required. In the simplest possible form, tolls are collected

for traversing network links. We review the selfish routing game setting in the presence of link prices. If one decides to set link prices in a mixed–autonomy setting, for each link $l \in \mathcal{L}$, define $\tau_l^h \geq 0$ and $\tau_l^a \geq 0$ to respectively be the price for human–driven and autonomous cars along link l. Let $\tau = (\tau_l^h, \tau_l^a : l \in \mathcal{L})$ be the vector of link prices. The price of human–driven cars along a path $p \in \mathcal{P}$ is defined as $\tau_p^h = \sum_{l \in \mathcal{L}: l \in p} \tau_l^h$. Likewise, define $\tau_p^a = \sum_{l \in \mathcal{L}: l \in p} \tau_l^a$ to be the price of autonomous cars along path p.

Note that when prices are set, there are both travel time and monetary costs for each agent. Hence, for traversing a path $p \in \mathcal{P}$, an agent experiences a delay $e_p(f)$ and pays a price equal to either τ_p^h or τ_p^a. Thus, assuming that all agents value travel delays and monetary costs identically, the *cost* of an agent along a path p is either $e_p(f) + \tau_p^h$ or $e_p(f) + \tau_p^a$ dependent on whether the agent is human–driven or autonomous. We define the link traversal *cost* functions c_l^h and c_l^a to be the following

$$c_l^h(f_l^h, f_l^a) = e_l(f_l^h, f_l^a) + \tau_l^h, \tag{2.28a}$$

$$c_l^a(f_l^h, f_l^a) = e_l(f_l^h, f_l^a) + \tau_l^a. \tag{2.28b}$$

Similarly, we define the cost of traversing a path $p \in \mathcal{P}$ for human–driven and autonomous cars to respectively be

$$c_p^h(f) = e_p(f) + \tau_p^h, \tag{2.29a}$$

$$c_p^a(f) = e_p(f) + \tau_p^a. \tag{2.29b}$$

For a given vector of link prices τ inducing a vector of link cost functions $c = (c_l^h, c_l^a, l \in \mathcal{L})$, we define a nonatomic selfish routing by a triple of the form (G, r, c).

Remark 2. For every link $l \in \mathcal{L}$, we use the term link *cost* for human–driven or autonomous cars to respectively refer to $c_l^h(f_l^r, f_l^a)$ or $c_l^a(f_l^r, f_l^a)$, where both the travel delays and link prices are included. However, we use the term link *delay* to refer solely to $e(f_l^r, f_l^a)$, which is the delay of travel along link l. Note that the cost of traversing a link $l \in \mathcal{L}$ might be different for human–driven and autonomous cars; but, the delay of traversing link l is the same for both classes of cars.

Remark 3. When a price vector τ is set, although the traversal cost of a link perceived by every agent may be different from the delay of travel along that link, the overall performance of the system is still measured via the overall delay incurred by all agents. The goal of setting link prices is to find prices such that the overall delay of the system perceived by the society is minimized.

The overall *cost* of a routing game (G, r, c) is defined by

$$C(f) = \sum_{p \in \mathcal{P}} f_p^h c_p^h(f) + f_p^a c_p^a(f). \tag{2.30}$$

For a priced network, Wardrop equilibria are defined via the following.

Definition 3. *For a priced routing game (G, r, c), a feasible flow vector $f = (f_p^h, f_p^a : p \in \mathcal{P})$ is an equilibrium if and only if for every O/D pair $w \in \mathcal{W}$ and every pair of paths $p, p' \in \mathcal{P}_w$, we have*

$$f_p^h \left(c_p^h(f) - c_{p'}^h(f) \right) \leq 0, \tag{2.31a}$$

$$f_p^a \left(c_p^a(f) - c_{p'}^a(f) \right) \leq 0. \tag{2.31b}$$

Remark 4. In general, despite the classical setting of a single vehicle class where Wardrop equilibrium is unique [101], in our mixed–autonomy setting, there may exist multiple Wardrop equilibria satisfying (2.31) .

Notice that equations (2.31) imply that if for an O/D pair $w \in \mathcal{W}$, and two paths $p, p' \in \mathcal{P}_w$, the flows f_p^h and $f_{p'}^h$ are nonzero, then, we must have $c_p^h(f) = c_{p'}^h(f)$ (we can argue similarly for autonomous cars). Moreover if at equilibrium, the flow along a path is zero, its travel cost cannot be smaller than that of the other paths with nonzero flow of the same vehicle class. Therefore, we can define the following.

Definition 4. *For a priced routing game (G, r, c), if f is an equilibrium flow vector, for each O/D pair $w \in \mathcal{W}$ define the cost of travel for human–driven and autonomous cars to be*

$$c_w^h(f) = \min_{p \in \mathcal{P}_w} c_p^h(f), \tag{2.32a}$$

$$c_w^a(f) = \min_{p \in \mathcal{P}_w} c_p^a(f). \tag{2.32b}$$

Since at equilibrium, for each O/D pair w and each class of cars, the cost of travel along the paths that have nonzero flow of that class is the same and equal to cost of travel for that class, we have

$$C(f) = \sum_{w \in \mathcal{W}} r_w^h c_w^h(f) + r_w^a c_w^a(f). \tag{2.33}$$

Relevant Prior Results

In the remainder of this chapter, we review some prior results in the literature of routing games. We will use these results in part III of this thesis to prove and analyze some properties of mixed–autonomy networks. Some of these results hold for routing games with only a single class of cars, but since our setting has two classes of cars, these results do not directly apply to our setting of mixed vehicle autonomy. However, we will discuss in part III how these results can be leveraged for the mixed–autonomy setting.

Existence of Equilibrium

We state the following proposition from the theorem in [12] which studies the conditions under which a Wardrop Equilibrium exists for a multiclass traffic network.

Proposition 1. *Given a routing game (G, r, e), if the link delay functions are continuous and monotone in the link flow of each vehicle class; then, there exists at least one Wardrop equilibrium.*

Corollary 1. *Using (2.26), in the mixed–autonomy setting, since our assumed delay functions are nonnegative, continuous, and monotone in the flow of each vehicle class, Proposition 1 implies that there always exists at least one Wardrop equilibrium for a routing game with mixed autonomy.*

Uniqueness of Equilibrium

Now, we review some known results regarding the uniqueness of Wardrop equilibria. When there is a single class of cars in the network (e.g. only human–driven or autonomous cars), equilibrium uniqueness holds in a weak sense (See Theorem 3 from [101]).

Proposition 2. *Given a routing game G with a single class of human–driven cars for each O/D pair, if the delay functions are of the form (2.27), for a given demand vector r, we have*

1. *The equilibrium is unique in a weak sense, i.e. for each link l, the total flow f_l is unique for all Wardrop equilibrium flow vectors f.*
2. *For each O/D pair $w \in \mathcal{W}$, the delay of travel $e_w(f)$ is unique for all Wardrop equilibrium flow vectors f. Thus, the delay of travel for each O/D pair at equilibrium, i.e. $e_w(f)$, only depends on network demand vector r. Hence, we may unambiguously define $e_w(r)$ to denote this unique value.*

In general, in our mixed–autonomy setting, the equilibrium may not be unique. However, we will use the following result from [4] in Chapter 7 to establish some properties of equilibria in the mixed–autonomy setting in the presence of prices.

Proposition 3. *For a priced routing game (G, r, c), if along each link $l \in \mathcal{L}$, the link traversal cost functions c_l^h and c_l^a are strictly increasing functions of the total flow along that link $f_l = f_l^h + f_l^a$, and the link cost functions c_l^h and c_l^a are identical up to additive constants, then, at equilibrium, the total flow along each link $l \in \mathcal{L}$ is unique.*

It is important to mention that in our mixed–autonomy setting, since the link cost functions (2.28) depend on the flow of each vehicle class, not the total flow along the link, Proposition 3 does not apply to our setting, but we will further use Proposition 3 in proving some of our results.

Monotonicity of the Delay of Travel

As we discussed above, in general, the equilibrium may not be unique. However, if the conditions of Proposition 2 hold for a network, the social delay and the delay of travel for each O/D pair are unique. For a single class routing game on $G = (\mathcal{N}, \mathcal{L}, \mathcal{W})$, we recall the following from Theorem 3 in [38].

Proposition 4. *Consider a routing game (G, r, e), where there exists only one class of cars. Assume that for each link $l \in L$, $e_l(.)$ is continuous, positive valued, and monotonically increasing. Then, for each O/D pair $w \in \mathcal{W}$, the delay of travel $e_w(r)$ is a continuous function of the demand vector r. Furthermore, $e_w(.)$ is nondecreasing in r_w when all other demands $r_{w'}, w' \neq w$, are fixed.*

Marginal Cost Pricing

The following proposition regarding marginal cost pricing is a generalization of the results in [22].

Proposition 5. *For a priced routing game (G, r, c), let f^* be an optimizer of the network social delay J in Equation (2.18), and J^* be the minimum social delay of the network. If for each link $l \in \mathcal{L}$, link prices τ are set to be*

$$\tau_l^h = (f_l^{h*} + f_l^{a*}) \left(\frac{\partial}{\partial f_l^h} e_l(f_l^h, f_l^a) \right) \Big|_{f^*}, \tag{2.34a}$$

$$\tau_l^a = (f_l^{h*} + f_l^{a*}) \left(\frac{\partial}{\partial f_l^a} e_l(f_l^h, f_l^a) \right) \Big|_{f^*}, \tag{2.34b}$$

then, there exists at least one equilibrium flow vector f for the routing game (G, r, c) such that the network social delay is optimal at this equilibrium, i.e. $J(f) = J^*$.

Proof. It is easy to verify that (2.34) renders f^* an equilibrium flow vector by verifying the KKT conditions at the optimal point f^*. For completeness, we have included the proof of Proposition 5 in here.

Since the social delay defined by (2.18) is a continuous function of flows along network paths, and the set of feasible flows satisfying (2.15) is compact, there exists a flow vector f^* that optimizes the social delay, i.e. f^* is the optimizer of the following optimization problem

$$\begin{aligned} \min_{f} \quad & J(f) \\ \text{subject to} \quad & \forall p \in \mathcal{P} : f_p^h \geq 0,\ f_p^a \geq 0, \\ & \forall w \in \mathcal{W} : \sum_{p \in \mathcal{P}_w} f_p^h = r_w^h,\ \sum_{p \in \mathcal{P}_w} f_p^a = r_w^a. \end{aligned} \tag{2.35}$$

Note that all constraints in (2.35) are affine functions of the decision variable f. Therefore, the regularity conditions hold for (2.35) (see, for instance, Theorem 5.1.3 and Lemma 5.1.4 in [9]). Thus, the optimizer f^* must satisfy the KKT conditions. For each path $p \in \mathcal{P}$, let $\lambda_p^h \geq 0$ and $\lambda_p^a \geq 0$ be the Lagrange multipliers associated with the nonnegativity constraint of the flows of human–driven and autonomous cars along path p respectively.

Similarly, for each O/D pair $w \in \mathcal{W}$, let ν_w^h and ν_w^a be the Lagrange multipliers associated with the flow conservation constraints for human–driven and autonomous cars along O/D pair w, respectively. Then, for a fixed path $p \in \mathcal{P}_w$ associated to an O/D pair $w \in \mathcal{W}$, for the flow of human–driven cars, the stationarity condition imposes the following

$$\frac{\partial}{\partial f_p^h} J(f)\Big|_{f^*} = \lambda_p^h - \nu_w^h. \tag{2.36}$$

From (2.17), (2.18), and (2.34), we have

$$\begin{aligned}\frac{\partial}{\partial f_p^h} J(f)\Big|_{f^*} = \\ \sum_{l \in \mathcal{L}: l \in p} \Big(e_l(f_l^h, f_l^a) + (f_l^h + f_l^a) \frac{\partial}{\partial f_l^h} e_l(f_l^h, f_l^a) \Big)\Big|_{f^*} \\ = e_p(f^*) + \tau_p^h.\end{aligned}$$

Using this together with (2.36), we can conclude that for every path $p \in \mathcal{P}$, we have

$$e_p(f^*) + \tau_p^h = \lambda_p^h - \nu_w^h. \tag{2.37}$$

On the other hand, complementary slackness requires that for every path $p \in \mathcal{P}$

$$\lambda_p^h f_p^h = 0. \tag{2.38}$$

Now, for a fixed O/D pair $w \in \mathcal{W}$, consider a pair of paths $p, p' \in \mathcal{P}_w$. If $f_p^h > 0$, from (2.38), we must have $\lambda_p^h = 0$. Then, from (2.37) and nonnegativity of $\lambda_{p'}^h$, we have

$$e_p(f^*) + \tau_p^h = -\nu_w^h \leq \lambda_{p'}^h - \nu_w^h = e_{p'}(f^*) + \tau_{p'}^h, \tag{2.39}$$

where in the last equality, we have used (2.37) for the path p'. Similarly, for autonomous cars along the two paths p and p', if $f_p^a > 0$, we must have

$$e_p(f^*) + \tau_p^a = -\nu_w^a \leq \lambda_{p'}^a - \nu_w^a = e_{p'}(f^*) + \tau_{p'}^a. \tag{2.40}$$

Note that (2.39) implies (2.31a), the reason being that if $f_p^h = 0$, (2.31a) automatically holds, and if $f_p^h > 0$, (2.31a) holds since as we showed above, $c_p^h(f^*) \leq c_{p'}^h(f^*)$. Likewise, (2.40) implies (2.31b). Hence, once prices are set according to (2.34), the optimal flow f^* is a Wardrop equilibrium for the game (G, r, c). Thus, there exists at least one induced equilibrium with minimum social delay once prices are obtained from (7.2). □

Part II

Networks of Human–Driven Cars

Chapter 3

Bandwidth Maximization

It is well known that the parameters of traffic signals such as cycle length, green times, and offsets play a key role in shaping traffic streams and network efficiency in general. The focus of this chapter is on how to select the offsets of signalized intersections in an urban network in order to reduce stop–and–go waves. Traffic signals are normally designed to guarantee that conflicting movements are not allowed at the same time at an intersection. In addition to these safety requirements, it is desired to design traffic signals such that the number of stops that cars make at intersections is minimized, which leads to increases in average traffic speed. Early traffic signal control systems focused primarily on maintaining the speed limit along a single path through the network.

The problem of selecting signal offsets for maximizing the size of uninterrupted platoons, that is, maximizing the two–way bandwidth along an arterial route, was first formulated in [79]. This work used a simple model where every signal was assumed to have only two stages. Moreover, it was assumed that no stage allowed for simultaneous actuation of through and left–turn movements. It was shown that maximization of the bandwidth could be achieved by an exhaustive search of order of 2^n, where n is the number of intersections. The computational hindrances of [79] were addressed in [83] by utilizing the symmetry of the bandwidth definition introduced in [79].

In [58], traffic signal synchronization was first formulated as a mathematical program. The approach allowed the optimization of other control parameters, in addition to the offsets, such as cycle length and speeds. This resulted in a mixed–integer linear program. An improved numerical approach was proposed in [88]. An algorithm based on the concept of interference minimization was introduced at IBM and implemented in [10]. This approach was further extended in [73] to enable the consideration of more than two stages. Cross–street flows were incorporated into the problem with work on multi–bandwidth methods beginning with [32]. The implementation of this approach can be found in MULTIBAND [102]. This approach was also extended in [31] to capture complex networks going beyond the arterial case. In [18, 19], the network queueing process was approximated by sinusoidal waves, and a semidefinite relaxation for the offset optimization problem was proposed.

As a result of the ongoing rise of interest in online adaptation of signal offsets, various algorithms for online tuning of offsets such as SCATS, SCOOT, OPAC and RHODES have been developed [61, 41, 29, 76]. The majority of these programs operate by making small adjustments to the parameters such that the timing plans evolve to match the measured traffic data. In [33], an adaptive algorithm for online tuning of offsets was proposed. The utilized framework was a data–driven approach where the collected data from the on–site sensors was utilized to find offsets such that the majority of the traffic was accommodated by the green window.

Recently, a new formulation of bandwidth maximization was proposed in [36]. This formulation was a generalization of the class of methods that derive from [58], in that it considered as given the distributions of car arrivals, rather than the green times. In this framework, the case of given green times is equivalent to an assumption of "pulse" car arrival functions, or non–dispersing platoons. It was shown that bandwidth maximization problem can be formulated as a linear program in the case of pulse arrival functions, reducing the computational complexity with respect to previous methods. However, the approach only considered a single two-way arterial, and did not generalize to networks of arbitrary topology.

In the present chapter, we generalize the framework of [36] and develop a formulation that is applicable to arbitrary networks. We first derive an explicit formula for the bandwidth as a function of the so–called "relative offsets". The formulation assumes given a set of "target routes". The target routes may be the routes with larger demands or the major routes taken by cars in the network. The goal of the problem is to select the offsets for the intersections encountered along these routes, so as to maximize a weighted sum of the route (or path) specific bandwidths. Note that these paths correspond to the directions taken by the aggregate flow of cars; they are not specific to a single car. The formulation also assumes that green times, cycle length, and phase sequences are fixed and given. We show that this optimization problem can be converted into a mixed–integer linear program. The efficacy of the approach is demonstrated with several examples and a case study. It is important to mention that although [36] can handle general car arrival functions (corresponding to different platoon dispersion characteristics and signal actuation), this chapter is restricted to the traditional case in which car arrivals are pulsed.

3.1 Path Relative Offsets

We formulate the problem of finding the offsets of network nodes in terms of *path relative offsets*. To define relative offsets, we need to define the following. Let $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ be the graph of traffic network. Each link $l \in \mathcal{L}$ is characterized by its travel time $t(l) > 0$. We assume that link travel times, $t(l)$ are given for all links $l \in \mathcal{L}$ and the current traffic conditions. It is assumed that a set of paths $\mathcal{P}$ through the network is given. Each path $p \in \mathcal{P}$ consists of an *ordered* sequence of links $\mathcal{L}^p$ and nodes $\mathcal{N}^p$ leading from the start to the end of the path. It is further assumed that these paths are linear, connected, and

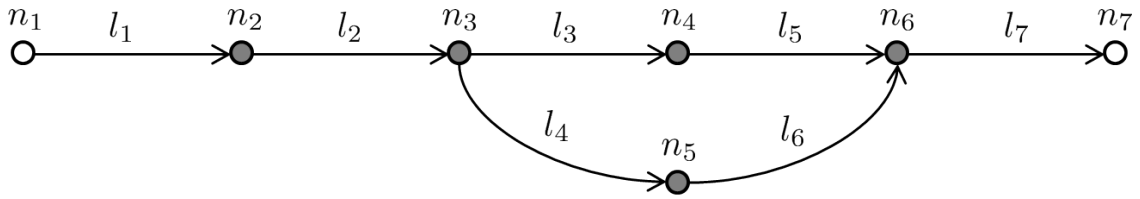


Figure 3.1: Schematic of the sample network.

contain no loops.

Let $\mathcal{N}_s \subseteq \mathcal{N}$ denote the subset of the nodes that are signalized. A *movement* is defined to be an input/output link pair on a signalized node, e.g. the north-bound left-turn movement. The set of movements for signalized node $n \in \mathcal{N}_s$ is denoted with $\mathcal{M}_n$. A signalized node with 4 incoming links and 4 outgoing links might have up to sixteen movements, but typically has only eight since U–turns and right–turns are often unsignalized. We let $\mathcal{M}^p \subseteq \mathcal{M}$ denote the set of signalized movements encountered by a vehicle following path p. Throughout the paper, superscripts refer to paths whereas subscripts refer to nodes.

Example: Consider the network shown in 3.1. There are two paths p and p' on this network, with

$$\begin{aligned}
\mathcal{L}^p &= \{l_1, l_2, l_3, l_5, l_7\} \\
\mathcal{L}^{p'} &= \{l_1, l_2, l_4, l_6, l_7\} \\
\mathcal{M}^p &= \{(l_1, l_2), (l_2, l_3), (l_3, l_5), (l_5, l_7)\} \\
\mathcal{M}^{p'} &= \{(l_1, l_2), (l_2, l_4), (l_4, l_6), (l_6, l_7)\}
\end{aligned}$$

It is assumed that all of the signals operate on a periodic fixed–time schedule with a common cycle time C, where the controller provides a certain amount of green time to each movement within a cycle. It is further assumed that each movement receives a green indication only once during a cycle. The green time for each movement $m \in \mathcal{M}_n$ is characterized by its duration $g(m) > 0$ and its offset $\theta(m)$ with respect to a global periodic clock. Each movement offset $\theta(m)$ is the time measured to the midpoint of the movement's green period. In each signalized node $n \in \mathcal{N}_s$, for each pair of movements $m, m' \in \mathcal{M}_n$, $\delta(m, m')$ is defined as:

$$\delta(m, m') = \big(\theta(m) - \theta(m')\big) \bmod^* C. \tag{3.1}$$

The $\bmod^* C$ is a modulo operator which returns the signed distance from the nearest multiple of C, as in [36]. This implies that for any real number x, $x \bmod^* C$ always lies in the interval $[-\frac{C}{2}, \frac{C}{2}]$. As an example, $6 \bmod^* 5 = 1$ while $4 \bmod^* 5 = -1$. In fact, $\delta(m, m')$ measures the time between the midpoint of the green periods for movements m and m' in the same signalized node. Note that $\delta(m, m')$ is a signed variable whose sign indicates

the ordering of the green periods of movements m and m'. Positive $\delta(m, m')$ implies that actuation of the midpoint of the green period belonging to movement m happens after the actuation of the same quantity for movement m'. The assumption of given green durations, cycle time, and phase sequences implies that the values of $\delta(m, m')$'s are fixed and given in the problem. Hence, we are only allowed to select a single offset for each signalized node $n \in \mathcal{N}_s$, which determines the displacement of the given timing plan with respect to a fixed clock.

For a movement m encountered while traversing a path p, we let $\omega^p(m)$ represent the relative offset of this movement with respect to path p. For a path p, as illustrated in Figure 3.2, path relative offsets measure the center of the green period of a movement m with respect to a coordinate frame that moves along path p at the speed of traffic (i.e. in accordance with the travel time $t(l)$ in each link l), but without stopping at red lights. Thus, for each movement m on path p ($m \in \mathcal{M}^p$), belonging to the node n ($m \in \mathcal{M}_n$), we have:

$$\omega^p(m) = \left(\theta(m) - T_n^p\right) \bmod^* C, \tag{3.2}$$

where T_n^p is the travel time from the path p's origin to node n. It is computed as the sum of the travel times $t(l)$ for all links $l \in \mathcal{L}^p$ preceding node n:

$$T_n^p = \sum_{\substack{l \in \mathcal{L}^p \\ l \text{ precedes } n}} t(l). \tag{3.3}$$

The relative offset defined in Equation (3.2) is in fact the actuation time of the midpoint of the green period of movement m in the reference frame of path p. In other words, it is the time between the middle of the green phase utilized by path p at node n, and the arrival time of a test vehicle along that path. Figure 3.2 shows reference trajectories for the paths p and p' from our previous example. The relative offsets are measured with respect to these reference trajectories, and to the center of the green periods (rectangles).

3.2 Path Bandwidth

As mentioned previously, in the case where there exist multiple paths, we define a bandwidth for each path. The bandwidth for a path $p \in \mathcal{P}$ is defined as the portion of the cycle during which a vehicle may start a journey along p and complete it without stopping. Normally, there will be only one interval of time during which an uninterrupted trip can be initiated. If there are several disconnected such intervals, then, we will define the bandwidth as the largest one. For long paths (or poorly coordinated signals) the bandwidth may reduce to zero, meaning that it is impossible to traverse the entire path without encountering a red light. Our goal for signal coordination is to maximize a weighted sum of the bandwidths for the desired paths included in $\mathcal{P}$. The weights assigned to the paths determine the priority or importance of some paths over the others. This allows us to present a generalization of the bandwidth definition provided in [36].

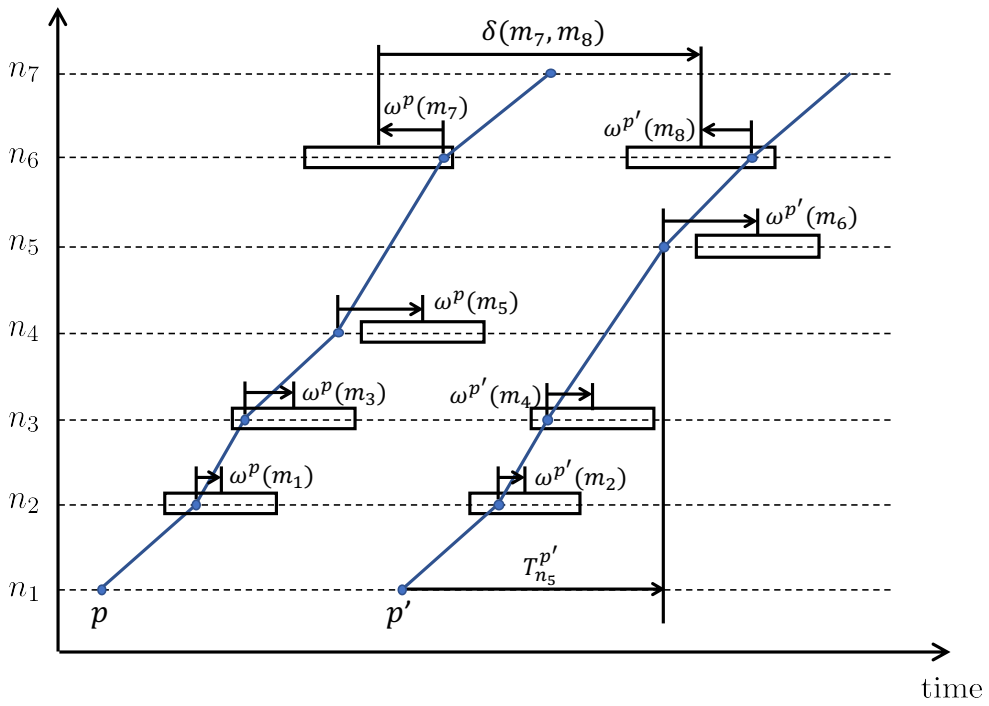


Figure 3.2: Reference trajectories for p and p'. Note that $\mathcal{M}^p = \{m_1, m_3, m_5, m_7\}$ and $\mathcal{M}^{p'} = \{m_2, m_4, m_6, m_8\}$.

In [36], maximization of bandwidth was proposed for the basic case of a single road with two paths: inbound and outbound.

The computation of bandwidth for a single path is analogous to the problem of finding the intersection of q intervals on the real line: $\{[\alpha_j, \beta_j], 1 \leq j \leq q\}$. The length of the intersection of these q intervals is:

$$\max\left(0, \min_{j,k}(\beta_j - \alpha_k)\right). \tag{3.4}$$

A path bandwidth b^p is the intersection of all green intervals for the movements belonging to path p. The green interval for a movement m utilized by path p, can be expressed in terms of the path relative offsets as $[\omega^p(m) - \frac{g(m)}{2}, \omega^p(m) + \frac{g(m)}{2}]$. Hence, b^p can be expressed in terms of relative offsets as the intersection of green intervals for all movements belonging to path p:

$$b^p(\boldsymbol{\omega}) = \bigcap_{m \in \mathcal{M}^p} \left([\omega^p(m) - \frac{g(m)}{2}, \omega^p(m) + \frac{g(m)}{2}]\right), \tag{3.5}$$

where $\boldsymbol{\omega}$ denotes the vector of relative offsets for all movements and paths in the network. Therefore, using Eq. (3.4), replacing β's and α's with $\omega^p(m) \pm g(m)/2$, we obtain the

bandwidth for path p as a function of the offsets:

$$b^p(\boldsymbol{\omega}) = \max\left(0, \min_{m,m' \in \mathcal{M}_p} \left(\omega^p(m) - \omega^p(m') + \bar{g}(m,m')\right)\right), \tag{3.6}$$

where $\bar{g}(m,m')$ is the average of $g(m)$ and $g(m')$. Note that since movement green times $g(m)$ and $g(m')$ are always positive, $\bar{g}(m,m')$ is also always positive.

3.3 Problem Formulation

Our objective is to find the node offsets such that a weighted sum of the bandwidths for a given set of network paths is maximized. This is expressed as a mathematical program as follows.

$$\begin{aligned} \underset{\boldsymbol{\omega}}{\text{maximize}} \quad & \sum_{p \in \mathcal{P}} \lambda^p \, b^p(\boldsymbol{\omega}) \\ \text{subject to} \quad & \boldsymbol{\omega} \in \Omega, \end{aligned} \tag{3.7}$$

Here λ^p is the weight associated with path p. Moreover, Ω is the set of feasible relative offsets, which following Eq. (3.2) is a hypercube in $\mathbb{R}^r$ with sides of length C and centered at the origin, where r is the number of unknown relative offsets. As stated earlier, the assigned weights reflect the importance of certain paths and directions in the network. For instance, one may assign more weight to the paths with larger demands.

Optimization problem (3.7) allows for finding relative offsets independent of each other. However, if fixed timing plans are given, as has been assumed, relative offsets of movements must be such that they conform with the given timing plans. Thus, we need additional constraints to encode this requirement.

To encode the dependence amongst relative offsets in a single intersection, we use the following. For each pair of distinct paths p and p' that have a common signalized intersection $n \in \mathcal{N}_s$, with respective movements $m, m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}$, we have

$$\begin{aligned} \omega^p(m) - \omega^{p'}(m') = \left(\theta(m) - T_n^p\right) \bmod^* C - \\ \left(\theta(m') - T_n^{p'}\right) \bmod^* C. \end{aligned} \tag{3.8}$$

Equation (3.8) can be rewritten as:

$$\begin{aligned} \left(\omega^p(m) - \omega^{p'}(m')\right) \bmod^* C = \Big(\left(\theta(m) - \theta(m')\right) - \\ \left(T_n^p - T_n^{p'}\right)\Big) \bmod^* C. \end{aligned} \tag{3.9}$$

which can be expressed in terms of $\delta(m,m')$ (which is fixed and given). Introduce the notation,

$$\Delta^{p,p'}(m,m') = \left(\delta(m,m') - \left(T_n^p - T_n^{p'}\right)\right) \bmod^* C \tag{3.10}$$

and note that the value of $\Delta^{p,p'}(m,m')$ is known and fixed by the assumption of fixed timing plans. Thus, using (3.10) and (3.9), we have

$$\Big(\omega^p(m) - \omega^{p'}(m')\Big) \bmod^* C = \Delta^{p,p'}(m,m'). \tag{3.11}$$

Adding this constraint to the optimization problem, we obtain,

$$\begin{aligned} \underset{\boldsymbol{\omega}}{\text{maximize}} \quad & \sum_{p\in\mathcal{P}} \lambda^p\, b^p(\boldsymbol{\omega}) \\ \text{subject to} \quad & \boldsymbol{\omega} \in \Omega, \\ & \Big(\omega^p(m) - \omega^{p'}(m')\Big) \bmod^* C = \Delta^{p,p'}(m,m'), \\ & \qquad \forall n,\ \forall m,m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}. \end{aligned} \tag{3.12}$$

The objective function in the above optimization problem is non-linear, non-convex, and non-differentiable. The constraints are also nonlinear. Next, we will describe how the above optimization problem can be solved.

3.4 Problem Relaxation

In order to solve (3.12), we show that it can be encoded as a mixed–integer linear program. For each path p, define the binary variable α^p such that $\alpha^p = 1$ if $b^p > 0$ and $\alpha^p = 0$ if $b^p = 0$. In fact, α^p is an indicator variable for the "max" operator in Eq. (3.6), taking the value 1 when $\min_{m,m'\in\mathcal{M}^p}(\omega^p(m) - \omega^p(m') + \bar{g}(m,m'))$ is positive, and 0 otherwise. With the introduction of these binary variables and using Equation (3.6), we can rewrite (3.12) as:

$$\begin{aligned} \underset{\boldsymbol{\omega},b^p,\alpha^p}{\text{maximize}} \ & \sum_{p\in\mathcal{P}} \alpha^p\, \lambda^p\, b^p \\ \text{subject to:} \ & \\ & \boldsymbol{\omega} \in \Omega, \\ & \alpha^p \in \{0,1\}, \\ & \Big(\omega^p(m) - \omega^{p'}(m')\Big) \bmod^* C = \Delta^{p,p'}(m,m'), \\ & \qquad \forall n,\ \forall m,m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}, \\ & b^p \le \omega^p(m') - \omega^p(m'') + \bar{g}_{m',m''}, \\ & \qquad \forall p,\ \forall m',m'' \in \mathcal{M}^p. \end{aligned} \tag{3.13}$$

Note that the decision variables in optimization problem (3.13) are relative offsets $\boldsymbol{\omega}$, path bandwidths b^p and binary variables α^p's while the solution of the above optimization problem is equivalent to that of (3.12). Now, the objective function of (3.13) is mixed

bilinear. Alternatively, we can move the binary variables, α^p's, to the constraints:

$$\begin{aligned}
&\underset{\boldsymbol{\omega}, b^p, \alpha^p}{\text{maximize}} \sum_{p \in \mathcal{P}} \lambda^p \, b^p \\
&\text{subject to:} \\
&\qquad \boldsymbol{\omega} \in \Omega, \\
&\qquad \alpha^p \in \{0, 1\}, \\
&\qquad \left(\omega^p(m) - \omega^{p'}(m')\right) \bmod^* C = \Delta^{p,p'}(m, m'), \\
&\qquad\qquad \forall n, \forall m, m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}, \\
&\qquad b^p \leq \alpha^p \left(\omega^p(m') - \omega^p(m'') + \bar{g}_{m',m''}\right), \\
&\qquad\qquad \forall p, \forall m', m'' \in \mathcal{M}^p.
\end{aligned} \tag{3.14}$$

The above conversion is true because any b^p whose α^p is set to zero will naturally rise to its upper bound at $b^p = 0$. Note that the inequality constraints in the optimization problem (3.14) are encoded for every pair of movements including the repeated movements. It is crucial to enforce these constraints for the repeated pairs of movements too so as to make sure that the path bandwidth is always smaller than or equal to the green duration of every movement belonging to that path.

The next step will be to relax the bilinear constraints to an equivalent mixed–integer form. Define the continuous variables $K^p(m', m'')$ as:

$$K^p(m', m'') := \omega^p(m') - \omega^p(m'') + \bar{g}_{m',m''}. \tag{3.15}$$

We know that $K^p(m', m'')$ is bounded above by $2C$, since $|\omega| \leq C/2$ and $g \leq C$. Now, as suggested in [34], the inequality constraints $b^p \leq \alpha^p \, K^p(m', m'')$ in (3.14) can be replaced by

$$b^p \leq \alpha^p (2C), \tag{3.16}$$

$$b^p \leq (1 - \alpha^p)(2C) + K^p(m', m''). \tag{3.17}$$

When $\alpha_p = 0$, the first inequality becomes $b_p \leq 0$ and the second is redundant; while, when $\alpha_p = 1$, the second inequality is $b_p \leq K(p, m', m'')$ and the first is redundant. Thus, we have converted our mixed bilinear problem into the following problem:

$$
\begin{aligned}
&\underset{\boldsymbol{\omega}, b^p, \alpha^p}{\text{maximize}} \sum_{p \in \mathcal{P}} \lambda^p \, b^p \\
&\text{subject to:} \\
&\boldsymbol{\omega} \in \Omega, \\
&\alpha^p \in \{0, 1\}, \\
&\Big(\omega^p(m) - \omega^{p'}(m')\Big) \bmod^* C = \Delta^{p,p'}(m, m'), \\
&\qquad \forall n, \forall m, m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}, \\
&b^p \leq \alpha^p (2C), \quad \forall p, \\
&b^p \leq (1 - \alpha^p)(2C) + K^p(m', m''), \\
&\quad \forall p, \forall m', m'' \in \mathcal{M}^p, \\
&K^p(m', m'') = \omega^p(m') - \omega^p(m'') + \bar{g}_{m',m''} \\
&\quad \forall p, \forall m', m'' \in \mathcal{M}^p.
\end{aligned}
\tag{3.18}
$$

Now, what remains for being able to solve (3.18) is to encode the equality constraints such that the resulting optimization problem becomes a mixed–integer linear program. It is important to note that $\Delta^{p,p'}(m, m')$ can be easily computed for every movement pair at an intersection given the timing plans and the path travel times. Now, note that in order for the equality constraints to be true, we must have that:

$$
\omega^p(m) - \omega^{p'}(m') = \beta^{p,p'}(m, m')C + \Delta^{p,p'}(m, m'), \tag{3.19}
$$

where $\beta^{p,p'}(m, m') \in \mathcal{Z}$ is an integer variable. In other words, for the equality constraints, we must have that $\omega^p(m) - \omega^{p'}(m')$ is equal to a multiple of cycle time C plus $\Delta^{p,p'}(m, m')$. Since relative offsets are restricted to lie in $[-\frac{C}{2}, \frac{C}{2}]$, we are guaranteed that $\Big(\omega^p(m) - \omega^{p'}(m')\Big)$ is bounded above and below by C and $-C$. Thus, the only possible values for $\beta^{p,p'}(m, m')$ are -1, 0, and 1. Hence, we can solve the optimization problem (3.18) by solving the following:

$$
\begin{aligned}
&\text{maximize} \sum_{p \in \mathcal{P}} \lambda^p \, b^p \\
&\text{subject to:} \\
&\quad \boldsymbol{\omega} \in \Omega, \\
&\quad \omega^p(m) - \omega^{p'}(m') = \Delta^{p,p'}(m, m') + \beta^{p,p'}(m, m')C, \\
&\qquad \forall n, \forall m, m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}, \\
&\quad \beta^{p,p'}(m, m') \in \{-1, 0, 1\}, \quad \forall n, \forall m, m' \in \mathcal{M}_n \cap \mathcal{M}^p \cap \mathcal{M}^{p'}, \\
&\quad b^p \leq \alpha^p (2C), \quad \forall p \\
&\quad b^p \leq (1 - \alpha^p)(2C) + K^p(m', m''), \quad \forall p, \forall m', m'' \in \mathcal{M}^p.
\end{aligned}
\tag{3.20}
$$

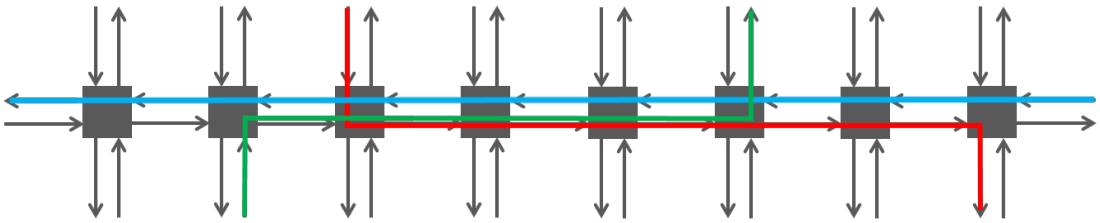


Figure 3.3: A linear arterial. The three colored paths are samples of the generated paths.

Now, optimization problem (3.20) is a mixed–integer linear program which can be easily solved via the available mixed–integer solvers. The decision variables in this optimization problem are ω^p's, b^p's, α^p's, and $\beta^{p,p'}(m,m')$'s. Once the optimal solutions are found, optimal offsets of the network intersections can be extracted from the optimal absolute movement offsets $\theta(m)$'s. It is noteworthy that when it comes to computing θ's, addition or subtraction of multiples of cycle time C to $\theta(m)$'s does not matter.

3.5 Numerical Experiments

The objective of this experiment is to demonstrate the ability of the technique to produce positive bandwidths on multiple paths simultaneously. The test network is a two–way linear arterial street with 8 intersections (Figure 3.3). One hundred instances of this network were generated, each with different segment travel times (selected uniformly between 60 and 150 seconds) and with different signal plans (selected from real–world plans found along Huntington Dr. in Arcadia, CA). Each of the sample networks was provided with 8 different paths with positive weights. Both the paths and the weights were generated randomly. The weights represent the relative magnitudes of the demands on each of the paths.

The 100 mixed integer linear programs were solved using the CPLEX software [1]. The experiment was repeated with 100 arterials with 5 intersections. Results are shown in Figure 3.4. The two histograms report the number of paths with positive bandwidth in the 5 and 8 intersection networks respectively. In the case of arterials with 5 intersections, the algorithm was able to assign positive bandwidth to an average of 6.3 paths, whereas for 8 intersections the average was 6.1. Thus, approximately 79% of paths could be traversed without stopping by some portion of the demand.

3.6 Case Study

So far, we have examined the capability of the algorithm in finding non–zero bandwidths for certain subsets of network paths. However, this does not necessarily mean that the overall network performance measures such as travel time are improved. To investigate

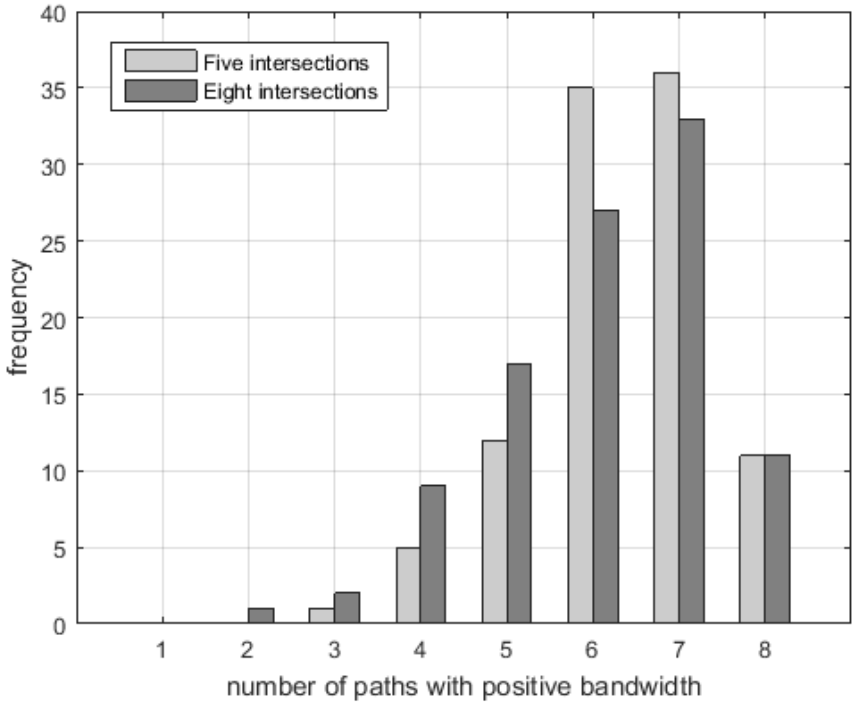


Figure 3.4: Histogram of positive bandwidths.

this, we focused on a traffic network in San Diego whose topology is shown in Figure 3.5. The grid–like geometry of the network allowed for examining the effect of bandwidth maximization in non–arterial networks. We considered 7 scenarios. In each scenario, certain numbers of paths were assumed to exist in the network. All paths were weighted equally. For each scenario, we simulated the network evolution for 1 hour using BeATS (Berkeley Advanced Traffic Simulator), in 100 experiments with randomly generated offsets for the network signals. We further used Yalmip [60] to optimize for the signal offsets such that the sum of path bandwidths in each scenario is maximized. We simulated the behavior of the network with optimized offsets for each scenario as well and computed the total travel time of the network in each experiment.

Figure 3.6, summarizes the results of our experiments. The figure shows the mean and standard deviation of the total travel time of the network for 100 random offsets versus the total travel time for the case when optimized offsets were picked in each scenario. We are excited to observe that using bandwidth maximization, we can achieve much smaller travel times in the network. In particular, when 7 paths were assumed to exist, major reduction in travel times was observed, pointing to the practicality of our framework for networks with general shapes and multiples routes. The observed reductions in total travel time further imply that although bandwidth is not a direct measure of the network delay, maximizing it can be used as a proxy for reducing the overall delays. Note that for

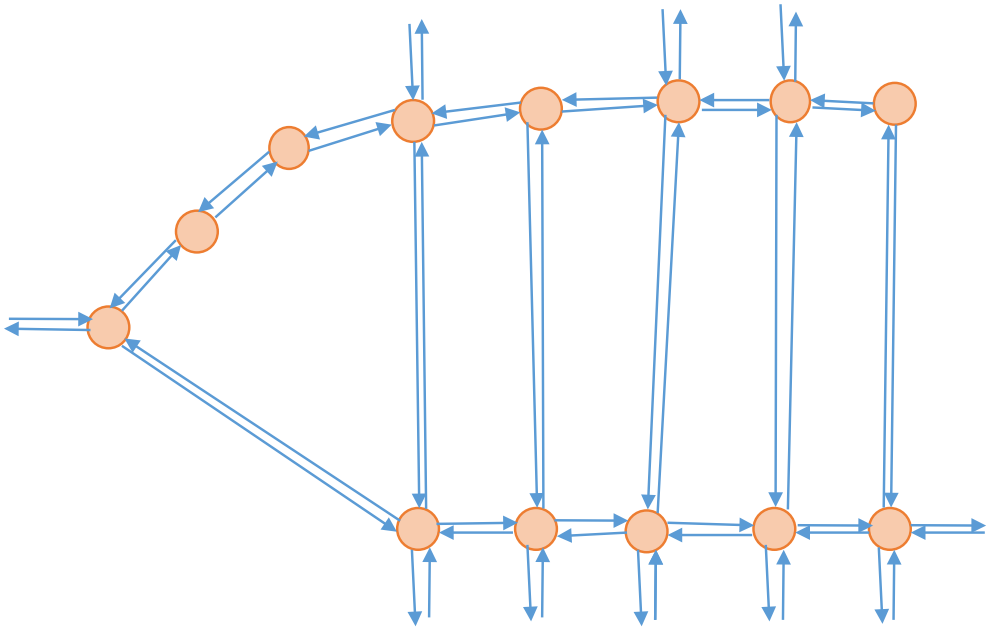


Figure 3.5: Topology of a traffic Network in San Diego.

the case of 1 and 2 paths, we recovered the offsets found by the framework of [36].

3.7 Chapter Summary

This chapter presented a first extension to the method proposed in [36]. The formulation relies on an external specification of the paths to be optimized, as well as their relative importance given by the weights. The optimization problem that resulted is non-convex and non-differentiable, but was found to be amenable to relaxations that turn the problem into a mixed–integer linear program. The experiments we have presented, although preliminary, show that the technique can be used to compute offsets that provide a significant bandwidth to a majority of the paths simultaneously. We further observed that bandwidth maximization leads to improvements in performance measure of the system such as total travel time in a case study. As a next step, it would be interesting to investigate how removing certain paths from bandwidth maximization can lead to better overall performance in the network. Moreover, it is important to analyze the effects of weights and their correlations with the total travel time of the network.

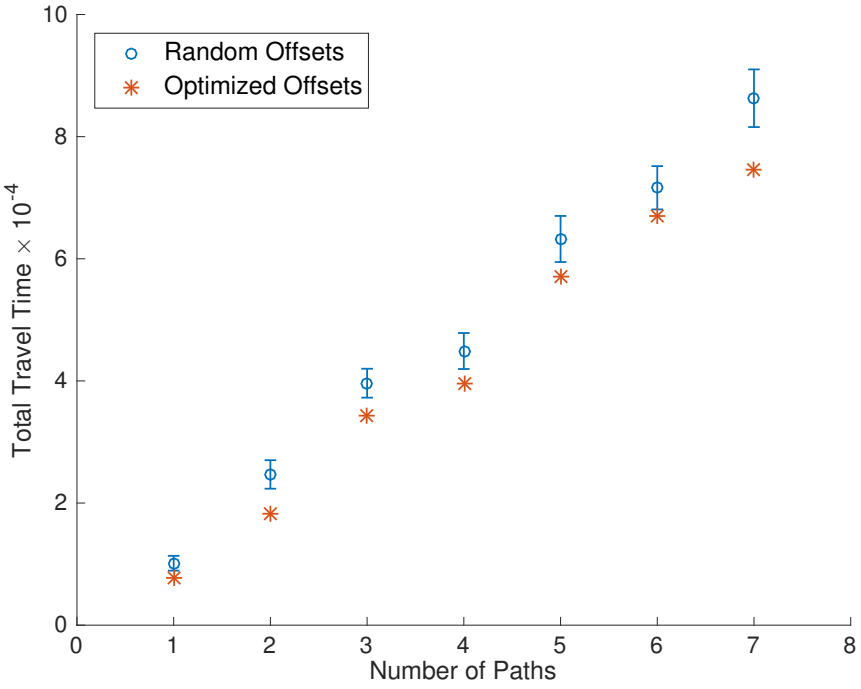


Figure 3.6: Network total travel time for different number of paths and different sets of offsets.

Chapter 4

Joint Perimeter and Signal Control

The focus of this chapter is on optimizing the performance of a network of signalized intersections in the presence of infeasible demands via scalable joint signal and perimeter control algorithms. A large body of the traffic control literature focuses on the development of signal control strategies for the regime of feasible demands, i.e. when demand lies within the capacity of the network. The simplest form of such signal controls is fixed–time control where each light operates cyclically, and each phase receives a fixed amount of green splits during the cycle. Various tools such as SYNCHRO [42], VISGAOST [105], SCOOT [93], and OPAC [30] have been proposed for determining the timing plan of fixed–time controllers either in real time or a priori. A queueing–theoretic analysis of fixed–time control policies is conducted in [80].

Other than fixed–time controllers, in [108], Max Pressure (MP) control is presented, which is a distributed control scheme that provably maximizes the network throughput and stabilizes the network in the presence of *feasible* arrivals. Using MP control, each intersection selects a stage of actuation that depends only on the length of adjacent queues. Due to nonlinearities and complexities of transportation networks, model predictive control laws have also been shown to be successful in reducing total travel time in both urban networks and freeways [54, 46, 65, 55]. Recently, synthesis from temporal logic specification has also been utilized for signal control [17, 98, 65, 72]. In such methods, the assumption is that the desired properties of the system can be encoded as formal specifications. Therefore, the control is found such that the temporal properties of interest are satisfied by the system trajectories.

The aforementioned controllers are beneficial mostly in the regime of feasible demand or arrivals. In fact, when the arrivals are not in the feasible region, regardless of the type of control that is deployed inside the network, the network is destabilized. In order to address this problem, TUC [25], which is a traffic responsive control strategy, has been proposed for handling the saturated traffic conditions. In [25], the highly nonlinear dynamics of urban roads are simplified into linear dynamics, and the feedback gains obtained from solving an infinite horizon linear quadratic regulator are implemented. In [47], traffic responsive control is developed for heterogeneous networks via perimeter control, where

the amount of boundary flow between different urban regions is determined using a PI controller. The authors in [47] model traffic evolution in each region through Macroscopic Fundamental Diagrams (MFDs).

In this chapter, we define a novel methodology for joint perimeter control and signal control of a single network for the case of infeasible arrivals. We consider a network with oversaturated arrivals and determine the timing plans and the amount of arrivals allowed to enter the network such that the network remains stable and free of congestion, the network utility is maximized, and different arrivals are treated *fairly*. Our approach is different from [47] as we consider perimeter control for a single network; thus, we do not require MFDs and partitioning the network. For a single network, we synthesize a joint congestion and signal control policy, and find the *optimal* boundary flows.

We adopt PointQ as our urban traffic model, and use the notion of utility maximization which is a well known congestion control scheme in communication networks [43, 84] for our control problem. We form an optimization problem that maximizes the aggregate utility of the network. Moreover, we demonstrate that by constructing the augmented Lagrangian and using Alternating Direction Method of Multipliers (ADMM) [11], the optimization problem can be solved iteratively such that the update step of the arrivals can be distributed while guaranteeing that network queues will remain stable. Since our iterative control utilizes ADMM for distributing its computations, it converges much faster than that of the previous work in [85] where dual decomposition is used. This fast convergence is crucial since for physical systems such as transportation networks, we do require to stabilize the system in the minimum possible number of time steps.

A unique and important feature of our work is that it allows us to introduce a notion of *fairness* among arrivals. Fairness is of paramount practical importance as vehicles in all network links must finally get the right of accessing the network regardless of where in the network they arrive. However, to the best of our knowledge, fairness has not been considered in the literature of traffic control except for our previous work, where we utilized utility maximization for fair control of freeway arrivals [66].

4.1 Control Algorithm

We use the modeling framework introduced in Section 2.1. We consider pointQ model in its deterministic setting, where arrivals and network parameters are all assumed to be deterministic. Throughout this chapter, we use demand and arrival interchangeably.

Problem Formulation

When the demand is infeasible, i.e. network cannot accommodate all vehicular arrivals, we wish to maximize the utilization of the network capacity while making sure that the network can be stabilized. At a high level, we aim to maximize the amount of flow allowed to enter the network, while network stability is preserved, and arrivals are treated

fairly. To this end, for every entry link $l \in \mathcal{L}_{\text{entry}}$, we define $U(f_l)$ to be a utility function of the total flow along link l. This utility function will be used as measure of fairness among vehicular arrivals on the network entry links. We need to decide on the amount of flow that we let enter the network d_l through every entry link $l \in \mathcal{L}_{\text{entry}}$ as well as the green duration that each movement receives during a cycle g. Using (2.10) and (2.11), the stability condition can be rewritten as

$$\Gamma(I - R^T)^{-1} d \leq M A_{g \to p} g. \tag{4.1}$$

We propose to maximize the total utility of network arrivals subject to the stability condition (4.1). In particular, we wish to solve the following optimization problem:

$$\begin{aligned} \underset{d,g}{\text{maximize}} \quad & \sum_{l \in \mathcal{L}_{\text{entry}}} U(d_l) \\ \text{subject to} \quad & \Gamma(I - R^T)^{-1} d \leq M A_{g \to p} g. \end{aligned} \tag{4.2}$$

The utility function, $U(.)$ in Equation (4.2) is a strictly concave increasing function of arrival rate d_l. Examples of such utility functions include $\log(x)$ and x^a for $a < 1$. Such functions have been extensively used for incorporating the notion of fairness among arrivals in communication networks [77]. The constraints in optimization problem (4.2) guarantee that the system stability conditions are satisfied. Optimization (4.2) can be solved by a central authority. However, one can distribute the computation of solution of (4.2). Rather than directly imposing the set of implementablility constraints (2.14) onto optimization problem (4.2) and solving it centrally, we propose to solve (4.2) iteratively to make parts of the computation distributed. Note that we have not directly included the constraints that g must satisfy to be an implementable vector of green fractions. Recall from Section 2.1 that in order for g to be implementable, g must satisfy (2.12) and (2.13). As discussed in Section 2.1, we summarize (2.12) and (2.13) by requiring g to satisfy

$$A_c g = b_c. \tag{4.3}$$

Before we proceed on how we distribute the computation of finding the optimizer of (4.2), note that we can summarize the linear inequality constraints in (4.2) by $A_d d + A_g g \leq 0$. Additionally, we convert inequality constraints to equality constraints, we utilize slack variables $0 \leq \delta \in \mathbb{R}^B_+$, to rewrite (4.2) as:

$$\begin{aligned} \underset{d,g}{\text{maximize}} \quad & \sum_{l \in \mathcal{L}_{\text{entry}}} U(d_l) \\ \text{subject to} \quad & A_d d + A_g g + \delta = 0. \end{aligned} \tag{4.4}$$

Iterative Solution of Utility Maximization Problem

The special structure of optimization problem (4.4) enables us to use ideas from Augmented Lagrangian [89] and ADMM [11] techniques to solve (4.4) iteratively such that

the update step of g is separated from d. This further leads to distributing the update step of d such that each entry link solves its own optimization problem to decide on the amount of flow it can let in. Before proceeding, we convert maximization problem (4.4) to a minimization problem

$$\begin{aligned} \underset{d,g}{\text{minimize}} \quad & -\sum_{l\in\mathcal{L}_{\text{entry}}} U(d_l) \\ \text{subject to} \quad & A_d d + A_g g + \delta = 0. \end{aligned} \tag{4.5}$$

Now, to achieve our goal, we construct the augmented Lagrangian of (4.5) as follows:

$$\begin{aligned} L_\rho = -\sum_{l\in\mathcal{L}_{\text{entry}}} U(d_l) + \alpha^T(A_d d + A_g g + \delta) + \\ \frac{1}{2}\rho\|A_d d + A_g g + \delta\|^2. \end{aligned} \tag{4.6}$$

In (4.6), $\alpha \in \mathbb{R}_+^B$ is the vector of dual variables or prices, and ρ is a finite positive number or increasing sequence penalizing for deviations from equality constraints. We can then solve (4.6) iteratively. Let α^k, d^k, and δ^k be the values of these quantities in the k_{th} iteration. Then, using ADMM, (4.5) can be solved via the following algorithm:

1. At $k = 0$, initialize α^0, d^0, δ^0, and $\rho > 0$ arbitrarily.
2. Update g^k and δ^k as follows:

$$\begin{aligned} [g^{k+1}, \delta^{k+1}] = \underset{g,\delta}{\text{argmin}}\ & {\alpha^k}^T(A_d d^k + A_g g + \delta) \\ & + \frac{1}{2}\rho\|A_d d^k + A_g g + \delta\|^2 \qquad (4.7) \\ \text{subject to}\ & A_c g = b_c. \qquad (4.8) \end{aligned}$$

3. Update d^k as follows:

$$\begin{aligned} d^{k+1} = \underset{d}{\text{argmin}} -\sum_{l\in\mathcal{L}_{\text{entry}}} U(d_l) + {\alpha^k}^T(A_d d + A_g g^{k+1} + \\ \delta^{k+1}) + \frac{1}{2}\rho\|A_d d + A_g g^{k+1} + \delta^{k+1}\|^2. \end{aligned} \tag{4.9}$$

4. Update α by:

$$\alpha^{k+1} = \alpha^k + \rho(A_d d^{k+1} + A_g g^{k+1} + \delta^{k+1}). \tag{4.10}$$

5. Apply d^{k+1} and g^{k+1} to the system and go to step 2 in the next cycle time T.

The implicit assumption in the above implementation is that the time step of the algorithm is the cycle time of the fixed–time control. In other words, at the beginning of every cycle time, we update g and d, apply them and repeat the same procedure in the next cycle time. It is important to mention that the additional control requirements are satisfied by constraining g in step 2. This further assures that the obtained green ratios satisfy the hard constraints that are essential to synthesize a valid signal plan that is implementable at each cycle.

Note that optimization problem (4.9) is an unconstrained optimization problem. Moreover, for each arrival i, its objective function consists of $U(d_i)$, quadratic terms and a linear term. In fact, optimization problem (4.9) has the following format:

$$d^{k+1} = \underset{d}{\text{argmin}} \quad \sum_{l \in \mathcal{L}_{\text{entry}}} -U(d_l) + \sum_{l,m \in \mathcal{L}_{\text{entry}}} h_{lm}(d_l, d_m), \tag{4.11}$$

with $U(d_l)$ being strictly concave and quadratic forms $h_{lm}(d_l, d_m) = \beta_{lm}(d_l d_m) + \gamma d_l$ being such that $\sum_{l,m \in \mathcal{L}_{\text{entry}}} h_{lm}(d_l, d_m)$ is in the form of convex quadratic functions. The special structure of this objective function allows us to solve (4.11) in a distributed fashion using Min Sum Message Passing Algorithm as follows [78]:

1. At $i = 0$, initialize $\tilde{d}[0] > 0$ arbitrarily.
2. Communicate $\tilde{d}[i]$s to entry links.
3. Let each entry link l update its arrival rate by:

$$d_l[i+1] = \underset{d}{\text{argmin}} \quad -U(d) + \sum_{m \in \mathcal{L}_{\text{entry}} - l} \beta_{lm}(d\tilde{d}_m) \quad + \gamma d + \beta_{ll} d^2. \tag{4.12}$$

4. Go to step 2 and repeat the procedure.

In the above algorithm, $d_l[i]$ is the value of d_l at the i_{th} communication of the message passage algorithm. This is different from d_l^k which is d_l at time step k. Note that d_l^k is implemented at time step k, whereas, $d_l[i]$ is only utilized when it converges to the optimal solution. Once the algorithm converges, we use the obtained updated arrival rates to update the dual variables via (4.10). An interesting property of (4.12) is that it has an analytical solution which eliminates the need for further computations. Let $f_{lm}(d, \tilde{d}_m) = \beta_{lm}(d\tilde{d}_m)$. It is easy to verify that the solution to (4.12) satisfies:

$$-U'(d) + \sum_{m \in \mathcal{L}_{\text{entry}} - l} \beta_{lm}\tilde{d}_m + \gamma + 2\beta_l d = 0, \tag{4.13}$$

where $\beta_{lm}\tilde{d}_m + \gamma$ is simply a constant known value and $U'(d)$ is the derivative of the utility function. Note that distributing the computation is compulsory when dealing with

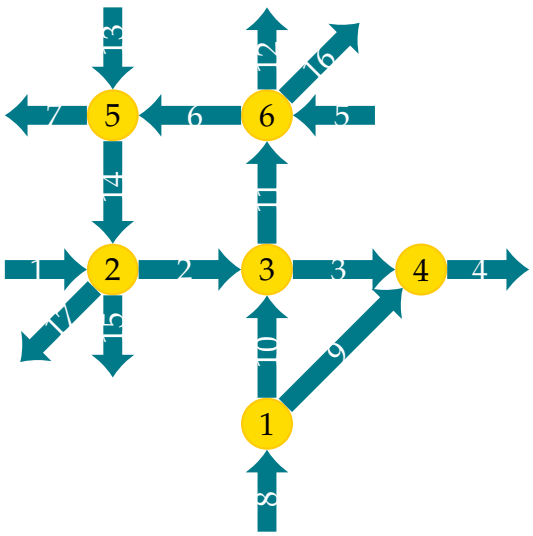


Figure 4.1: Topology of the example network.

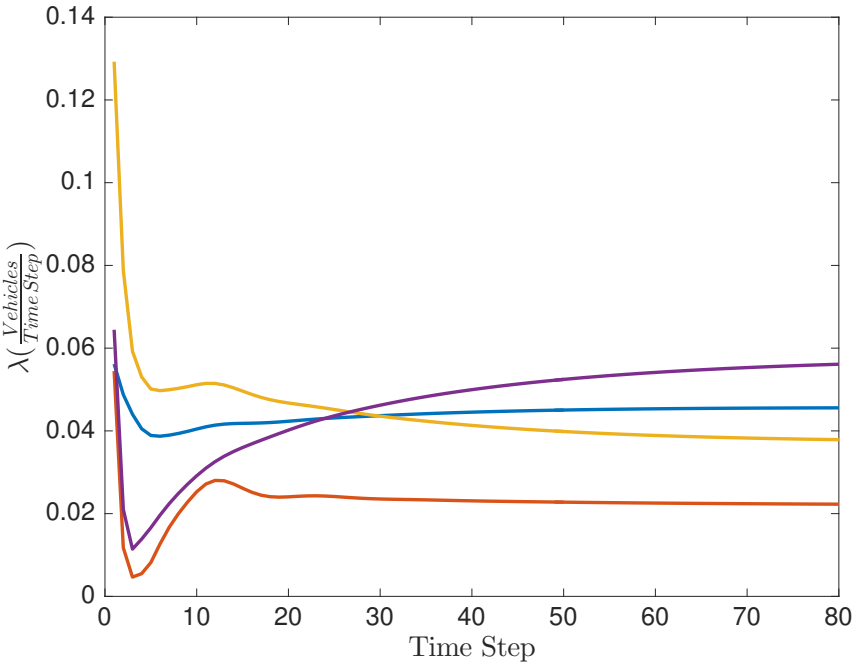
cyber–physical Systems such as transportation networks where there is normally limited computational capacity on the field; thus, we do require to formulate the problem such that it can be solved in a distributed fashion.

The solution to this problem would be fully distributed provided that (4.7) can also be distributed. Due to the hard constraints on g, it is generally hard to achieve this goal. Nonetheless, since we have introduced a quadratic program in (4.7), we can use active set methods [39] to distribute (4.7) with few number of communications and iterations as illustrated in [45]. This implies that formulating the problem such that we end up with a quadratic program in (4.7) paves the way for distributing (4.7) as well.

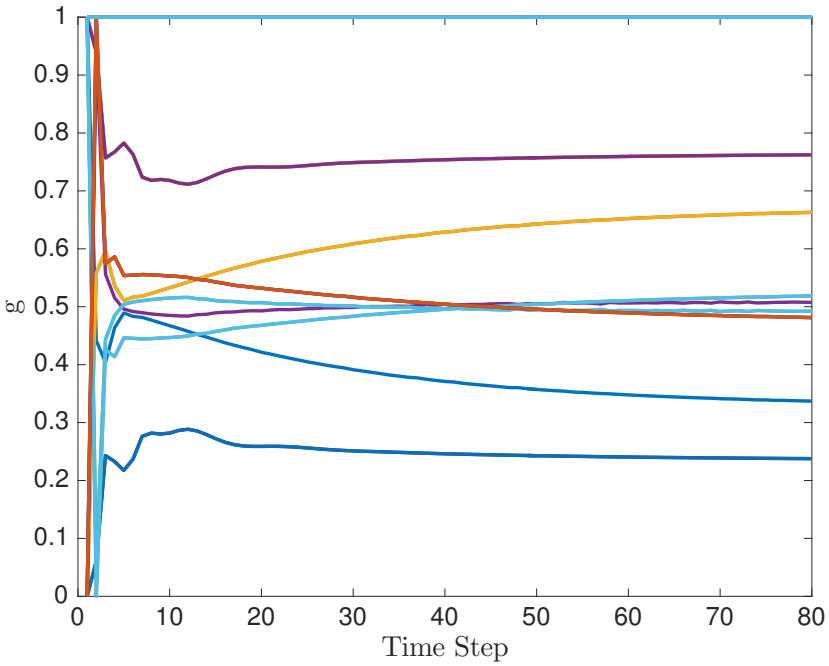
4.2 Numerical Experiment

In order to evaluate the performance of our algorithm, we utilize it for boundary flow control and signal control of the network shown in Figure 4.1. The network is subject to arrivals in links 1, 8, 5, and 13. We wish to regulate the flow that is allowed to enter through entry links by our algorithm while guaranteeing that the network is free of congestion. The network contains 17 links and 20 queues. Turning ratios at intersections are known and assumed to be constant. We used $\log(.)$ function as our utility function. The cycle time for all intersections is 90 seconds.

Figure 4.2 illustrates the arrivals and green ratios found during 80 time steps. As it can be observed, the algorithm converges to the optimal solution of maximizing the aggregate network utility in a small number of iterations. We further ran the simple dual-decomposition-based method (without the extra quadratic term in the Lagrangian) on the same network. However, it took more than 4000 time steps for the solution to converge,



(a) Arrivals versus time steps (iterations).



(b) Green ratios of queues versus time steps (iterations).

Figure 4.2: Arrivals and green ratios of the queues obtained from the utility maximization algorithm.

which makes it essentially impractical for transportation networks where the time step of the system is at the order of cycle times.

In order to verify that our control algorithm can successfully stabilize the network, we examine the queue lengths for all movements in the network to assure that they remain bounded. Figure 4.3 demonstrates the evolution of the sum of all queues in the network, which clearly remains bounded throughout the simulation demonstrating that the control successfully preserves network stability.

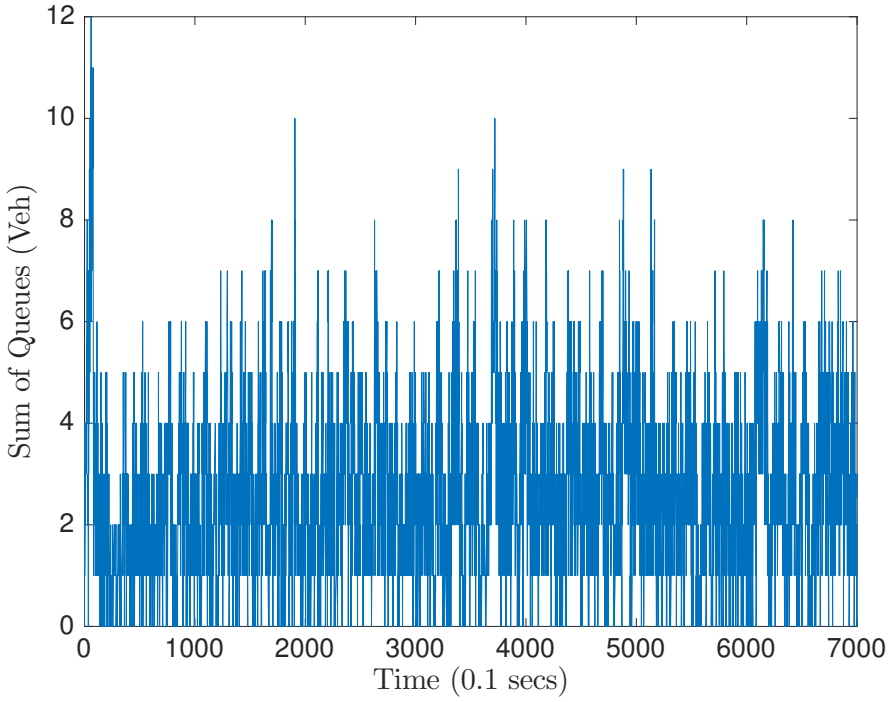


Figure 4.3: Sum of all queues vs. time.

4.3 Chapter Summary

In summary, in this chapter, we have introduced the notion of network utility maximization for fair allocation of available network resources to different arrivals that want to enter the network while stabilizing the network. To the best of our knowledge, no other control law has been proposed in the literature capable of encoding this property and synthesizing joint network congestion control and signal control. We demonstrated how our algorithm can be partially distributed to reduce the computational burden. We further showed that using ADMM, our algorithm can achieve much faster convergence rate compared to the existing dual decomposition methods for utility maximization. Finally, we illustrated that our algorithm successfully stabilizes an example network.

We note that for our solution to be fully distributed, we need to be able to distribute the computation required for updating the timings. Utilizing distributed active set or ϵ–exact penalty function [44] methods for achieving this goal can be of importance and interest. Additionally, since our iterative control algorithm can potentially adapt to the changes of system parameters, it can be employed for developing control policies that are adaptive and robust.

Chapter 5

Signalization Control with Unknown System Parameters

In this chapter, we focus on the design of signal control strategies for urban networks when some of system parameters such as arrivals and saturation flow rate of movements are unknown to the controller.

A common feature of the majority of the literature of traffic control strategies is that it is normally assumed that network parameters such as service rates of queues and network demands are known to the controller. These works range from fixed time controllers to temporal logic based controls. For instance, Max Pressure control, although robust to the knowledge of demand values, requires the knowledge of service rates [108]. However, this assumption may not necessarily hold in practice. As an example, connected vehicle technology (CVT) which has recently gained significant attention is going to be used for creating platoons of vehicles. It is shown that as the penetration rate of connected vehicles on the road varies, the service rates of network queues may vary as well [57, 50]. This implies that the higher the penetration rate is, the higher the service rates will be. Moreover, the prediction and estimation of arrivals is in general a challenging task, and it is normally the case that the demand values are not exactly known. Hence, the previous approaches which assume that these parameters are known to the controller may not be directly applicable to traffic networks. Therefore, it is important to come up with strategies that are robust to the knowledge of these parameters.

In this chapter, we propose a signal control strategy that is robust to the knowledge of both network service rates and demands similar to [85, 84, 86]. We show how the framework of [86] which was first introduced for communication networks can be applied to traffic networks when network demands and service rates are unknown. We determine the green durations of every stage such that the control converges to the timing plan which is desired for maximizing the network throughput. Our approach is different from the existing literature in that it *learns* the timing plans iteratively by measuring the changes in the queue lengths. In our approach, we still assume that turning ratios of the network are known, but this is not a restrictive assumption since network turning ratios

can be estimated by measuring link flows at intersections and running a calibration of the PointQ model [107]. We use the PointQ model introduced in Section 2.1 to formally define an urban network as a queuing system. We describe how the green duration of each stage can be updated using a gradient projection algorithm such that all the flows in the network are balanced. Our iterative control scheme is guaranteed to converge to an optimal and "balanced" signal plan when the network parameters are constant. We further demonstrate the capability and performance of our algorithm in a simulation environment.

5.1 Control Algorithm

We assume that our controller is a cyclic controller, i.e. it has a fixed cycle time, and the controller determines the green duration of each phase adaptively. We consider PointQ model in its stochastic setting where network arrivals and service rates are stochastic. We further assume that the turning ratios are known to the controller. Recall from Section 2.1 that the turning ratios define matrices R and Γ. Therefore, we assume that matrices R and Γ are known to the controller. Before we proceed to the description of our algorithm, we need to describe some notations and definitions. We define the convex set $\mathcal{C}$ to be the following

$$\begin{aligned}\mathcal{C} = \{p \in \mathbb{R}^B \mid \exists\, g \geq 0, \text{ such that } p = A_{g\to p} g,\\ A_{\text{eq}} g = 0, \text{and } A_{\text{sum}} g = 1\},\end{aligned} \tag{5.1}$$

where, as defined in Section 2.1, p is the vector of total green duration during a cycle for all movements in the network, and g is the vector of green duration for all movements during each relevant stage. Also, the matrices A_{eq} and A_{sum} encode the specifications required for implementability of a cyclic control. The set $\mathcal{C}$ is indeed the set of all p's for which there exists a corresponding vector g that satisfies the constraints required by (2.11), (2.12), and (2.13). We use k to denote the cycle or time index. Every time step of the controller is assumed to last a cycle time T. We let p_k and g_k be the aggregate green durations and stage green durations of network movements during cycle k. We let $E_k \in \mathbb{R}^{B\times B}$ be a diagonal matrix with each diagonal entry e_{ii} being 1 if the i_{th} queue is nonempty at the beginning of the k_{th} cycle and zero otherwise. We further let q_k be the vector of queue lengths for all network queues at the beginning of cycle k. We also define the sequence of step sizes or learning rates $\{\beta_k\}$ to be a decreasing sequence such that

$$\begin{aligned}\beta_k \to 0 \ \text{ as } k \to \infty,\\ \sum_{k=0}^{\infty} \beta_k = \infty,\\ \sum_{k=0}^{\infty} \beta_k^2 < \infty.\end{aligned} \tag{5.2}$$

The above conditions on β_k are standard assumptions required for the convergence of stochastic gradient projection algorithms. Notice that examples of such sequences include $\beta_k = \frac{1}{k}$ which satisfies the above conditions. We also use $\Delta q_k = q_k - q_{k-1}$ to denote the vector of difference in queue lengths at each time step k, where $k \geq 1$. Finally, we define matrix $\Lambda \in \mathbb{R}^{L \times B}$ to be such that at each link row $i \in L$, all elements of the matrix are zero except for the elements located at j_{th} columns with j being the index of the queues that originated from link i. For such elements, $\Lambda_{ij} = 1$.
Example: Consider a network for which there are two queues from link 2 to links $5, 8$. Assume that the indices of the queues for movements $(2, 5)$ and $(2, 8)$ are 3 and 6 respectively. Then, the elements Λ_{23} and Λ_{26} are equal to 1, while other elements of the second row of Λ are zero. Each row of Λ can be constructed similarly.

With the introduced notation, we are ready to state our iterative adaptive control algorithm. At the beginning of every cycle k, we update the vector of total green duration of movements p through the following algorithm:

1. Initialize p_0 with an arbitrary feasible value (i.e. $p_0 \in \mathcal{C}$).

2. At each time step $k, k \geq 1$, update vector p by

$$p_k = \left[p_{k-1} + \beta_k \left(\Gamma E_k (I - R^T)^{-1} \Lambda \Delta q_k \right) \right]_{\mathcal{C}}, \tag{5.3}$$

   where $[.]_{\mathcal{C}}$ is the convex projection on the set $\mathcal{C}$, and Γ is the matrix relating link flows to movement flows through turning ratios.

3. Apply the updated control p_k to the system, and let the system evolve to the next cycle time; then, measure Δq_k and repeat step 2.

It is important to mention that in step 3, for implementing a control law, we need g_k rather than p_k. In other words, once p_k is found, a vector g_k such that $p_k = A_{g \to p} g_k$ is required for implementing the new timing plan. Since there might be multiple vectors g_k such that $p_k = A_{g \to p} g_k$, we can obtain a vector of stage green durations g_k by simply solving a least squares problem for an updated p_k.

$$g_k = \min_x \ \left\| p_k - A_{g \to p} x \right\|^2.$$

Now, we explain the intuition behind our algorithm. If the vector of demand mean values d were known, then, the vector of movement flows φ would have been easily obtained by

$$\varphi = \Gamma (I - R^T)^{-1} d.$$

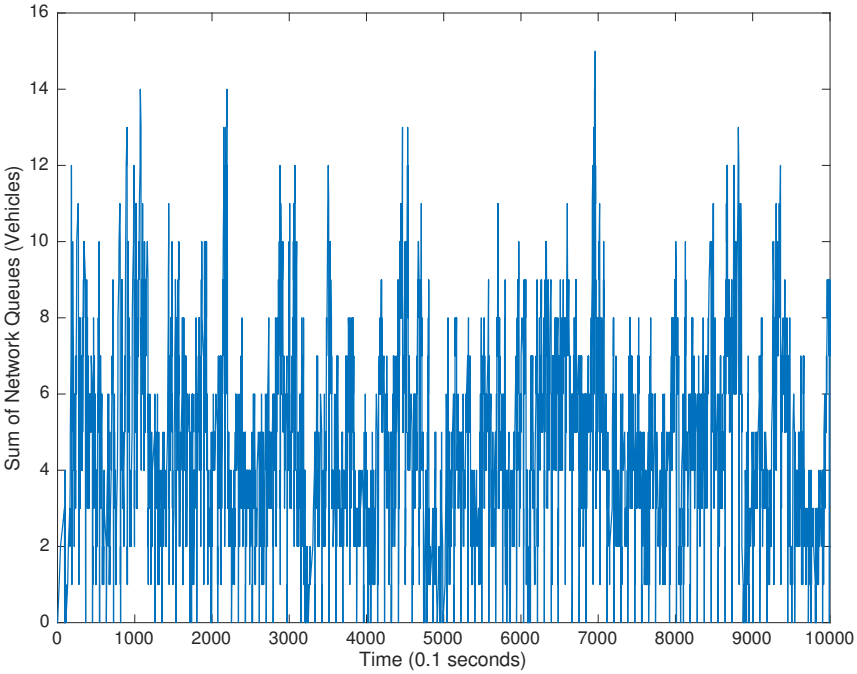


Figure 5.1: Sum of all queues vs. time.

Moreover, if the matrix of service rates M were also known, then, the gradient projection update rules of the form

$$p_k = [p_{k-1} - \beta_{k-1}(\Gamma f_{k-1} - M p_{k-1})]_{\mathcal{C}}$$

would have solved the following optimization problem:

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & \frac{1}{2}\|\varphi - Mp\|^2 \\ \text{subject to} \quad & p \in \mathcal{C}. \end{aligned} \tag{5.4}$$

Note that the solution of optimization problem (5.4) yields the optimal signal plan that would have balanced all the flows in the network while guaranteeing stability of the queues. However, when the service rates matrix M and demand vector d are unknown, the vector of movement flows φ and the matrix of service rates M are not available. Consequently, we use Δq_k to estimate $(\Gamma f - M p_k)$ as Δq_k is indeed an unbiased estimator of the gradient term with finite variance. Therefore, our update rule in (5.3) is a stochastic gradient projection algorithm for a convex optimization problem, which is guaranteed to converge for appropriate choices of the step size in (5.2). Moreover, as p_k converges to the optimal signal plan p^* that is the solution of (5.4), one can show that the network

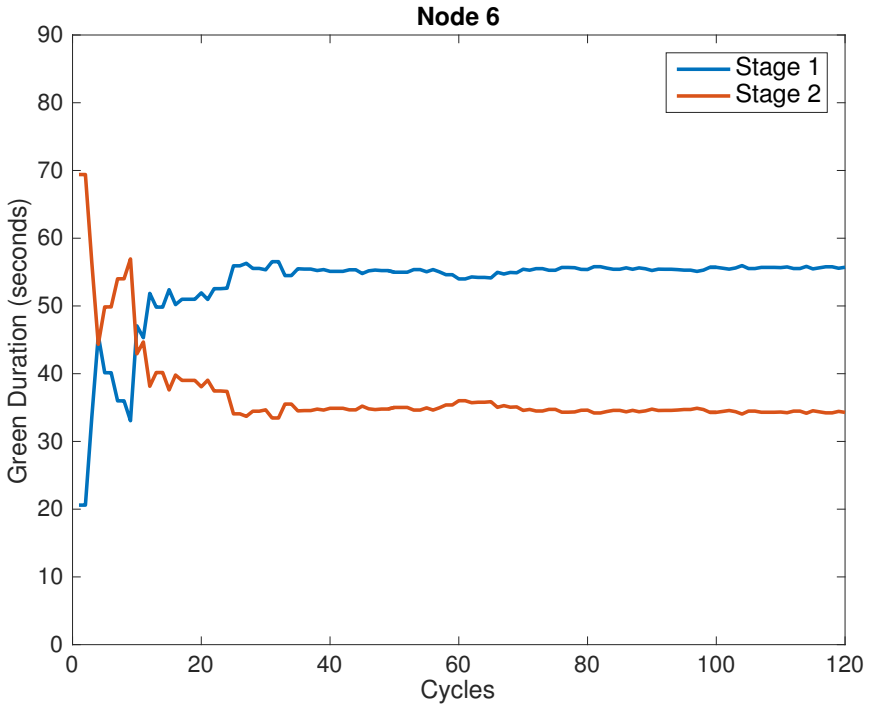


Figure 5.2: Green duration of stages in node 6.

queues remain stable using proof ideas similar to the one in [86]. The framework we have used in this work was first introduced in [86] for communication networks. In this work, we have shown how the framework can be employed for traffic networks; therefore, we do not repeat the proofs on convergence of the algorithm and its optimality. Interested reader is referred to [85, 86]. It is also important to mention that the convergence was formally guaranteed for stationary demand and service rates. However, in practice, it has been observed that as long as the changes in the network demand and service rates are slow enough, the algorithm can learn to adapt itself to the new set of system parameters.

It is noteworthy that, in general, measurement of queue length may not be available for all queues in the network. But, queue estimation methods have been proposed in the literature using loop detectors, GPS data or combination of the two for queue estimation [59, 8, 6, 110]. Hence, when queue lengths are not available, such methods can be utilized for estimating queue lengths and Δq_k as a result.

5.2 Numerical Example

To illustrate the performance of our control algorithm, consider the network shown in Figure 4.1. The network has 6 nodes, 17 links, and 20 queues. All network intersections have a common cycle time of $T = 90$ seconds. All nodes have cyclic controllers with

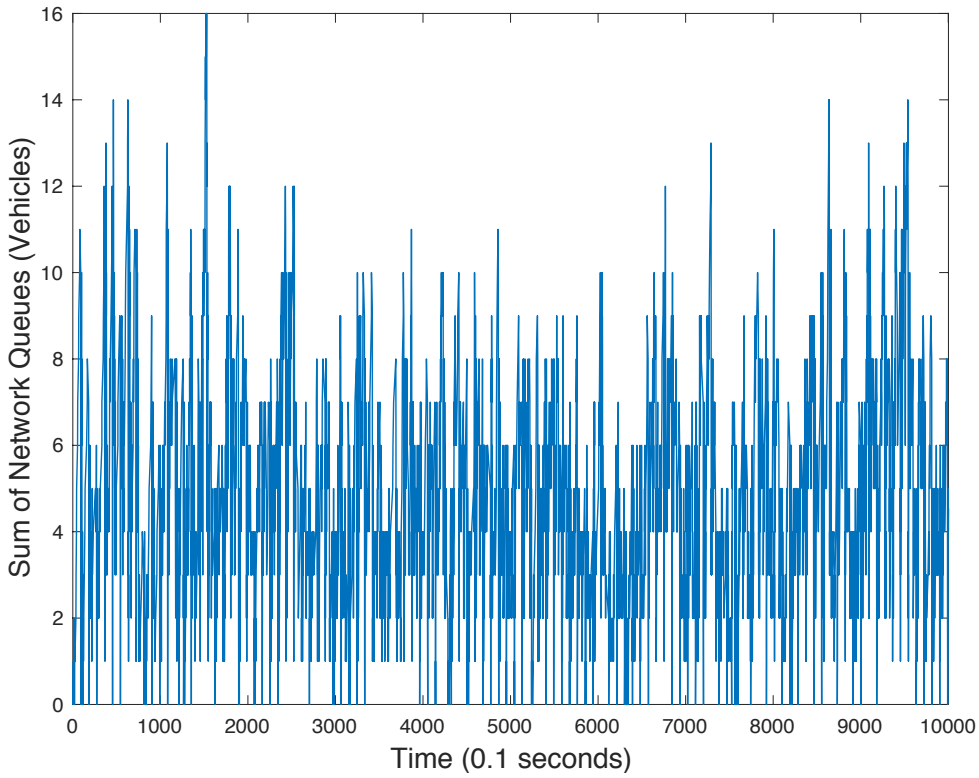


Figure 5.3: Sum of all queues as a function of time when the service rates were varied.

known stages. Nodes 2,3,5 and 6 have 2 stages, while nodes 1 and 4 are not signalized. The turning probabilities of the network queues are known to the controller. We assumed that the mean demand of the network and network service rates were unknown to the controller. For the simulation purpose, we used a typical set of demand profiles with feasible Poisson arrivals with fixed mean value. We used pointQ simulator [56] for our analysis.

We ran our control algorithm in closed loop with the simulation environment for 120 cycles. We started from a set of arbitrary yet feasible vector of stage and movement green durations. As Figure 5.1 demonstrates, the network queues remained stable (the sum of all queues in the network was not growing and remained bounded). In addition to the network stability, the convergence of network timing plans was also achieved as shown in Figure 5.2. Figure 5.2 shows the convergence of the stage green durations for node 6. Using Figure 5.2, we observe that the control algorithm learned a stabilizing timing plan over time despite the arbitrarily provided initial timing plan. Moreover, the convergence occurred quite fast, around 30_{th} cycle. The simulation was carried out for longer durations so as to showcase that convergence was achieved. Similar behavior and convergence results were achieved for other network nodes too.

In order to verify the practicality of the proposed control algorithm, we further re-

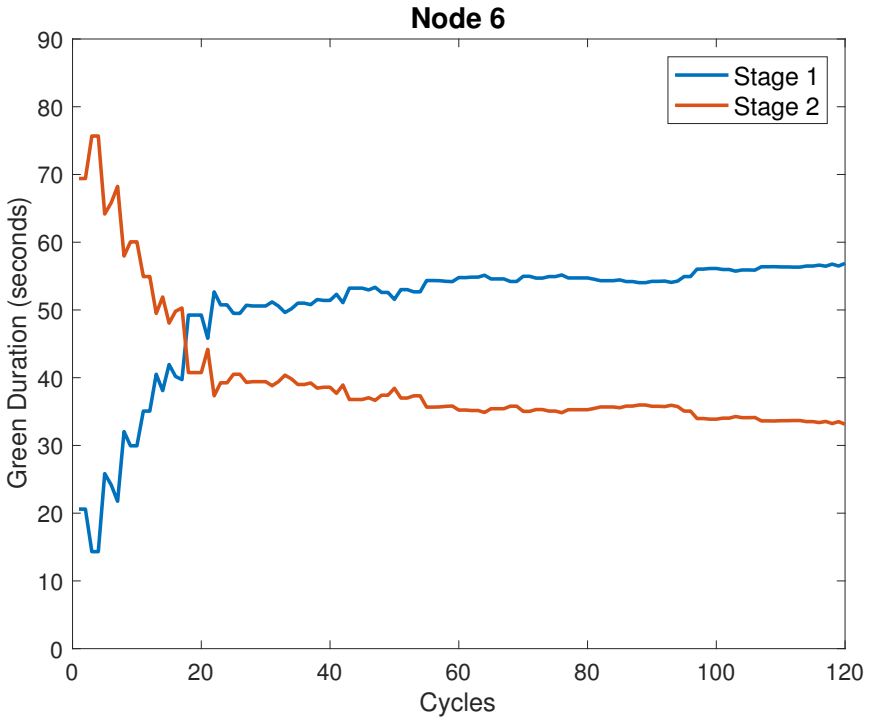


Figure 5.4: Green duration of the stages in node 6 when the service rates are varied.

peated the simulation but decreased service rates by 20% at the 20_{th} cycle to investigate how the control adapted to the new changes. As shown in Figure 5.3, since the demand was still feasible with the reduced service rates, the controller could successfully maintain the stability of the system queues. Figure 5.4 demonstrates the green durations of the two stages at node 6. The convergence of the green durations was achieved in this case too. During our simulations, we found that convergence and stability were achieved for other feasible demands too that we omit reporting due to the similarity of behavior.

5.3 Chapter Summary

In conclusion, we proposed an iterative cyclic signal control for network of signalized intersections such that the control is unaware of the network mean demands and service rates. Our control is iterative in the sense that at the beginning of every cycle time, it decides on the green durations based on the measured changes in the lengths of queues. We demonstrated through simulation studies that our controller can successfully stabilize the system. The next step of this work will be to evaluate the performance of our algorithm when estimates of the queue length rather than actual queue length are used. Also, finding the maximum tolerable rate of changes in the demand profiles will be of interest.

Moreover, since the current approach is a centralized approach, it will be interesting to study how the designed controller can be implemented in a decentralized fashion.

Part III

Networks with Mixed Vehicle Autonomy

Chapter 6

Mobility in the Presence of Selfish Autonomy

Connected and autonomous cars technology has attracted significant attention as a result of its potentials for increasing vehicular safety and drivers' comfort. Connected technologies can be used to inform drivers about the existing hazards through car to car (V2V) or car to infrastructure (V2I) communication. Aligned with these safety considerations, automobile companies have started to equip cars with autonomous capabilities. In fact, some of these capabilities, such as driver assistive technologies and adaptive cruise control (ACC) have already been deployed in cars.

The impact of these technologies is not limited to cars safety. Connected and autonomous cars technology can facilitate car *platooning*. Platoons of cars are groups of more than one car capable of maintaining shorter headways. Thus, platooning can lead to increases in the capacities of network links [57]. Such increases can be up to three–fold [57] if all cars are autonomous and connected. In addition to mobility benefits, platooning can have sustainability benefits, it can reduce energy consumption for heavy duty vehicles [2, 53, 3].

The mobility benefits of autonomous capabilities of cars are not limited to increasing network capacities. There has been a focus on how to utilize car autonomy and car connectivity to remove signal lights from intersections and coordinate conflicting movements such that the network throughput is improved [113, 106, 74, 27]. However, in order for such approaches to be implemented, all cars in the network need to have autonomous capabilities. To reach the point where all cars are autonomous, transportation networks need to face a *transient* era, when both human–driven and autonomous cars coexist in the networks. Therefore, it is crucial to study networks with mixed vehicle autonomy.

In [7], the performance of traffic networks with mixed autonomy was studied via simulations. Moreover, it was shown in multiple works that in networks with mixed autonomy, autonomous cars can be utilized to stabilize the low–level dynamics of traffic networks and damp congestion shockwaves [111, 24, 112, 90, 104]. In [67], altruistic lane choice of autonomous cars was studied in mixed–autonomy settings. In [50], the capacity

of network links was modeled in a traffic setting with mixed autonomy. This modeling framework was further used in [51] to calculate the price of anarchy of traffic networks with mixed autonomy, where the price of anarchy is an indicator of how far the equilibrium of networks with mixed autonomy is from their social optimum that could have been achieved if a social planner had routed all cars. In [52], it was shown that the local actions of the autonomous cars can lead to optimal car orderings for the global network properties such as link capacities.

It is well known that due to the selfish route choice behavior of drivers, traffic networks normally operate in an equilibrium state, where no car can decrease its trip time by unilaterally changing its route [101]. In this chapter, we wish to study how the introduction of autonomous cars in a network will affect the equilibrium state of traffic networks compared to the case when all cars are nonautonomous. We extend our initial results presented in [62]. In particular, given a fixed demand of cars, we study how increasing the fraction of autonomous and connected cars over the total number of autonomous and human–driven cars in the network, henceforth referred to as the *network autonomy fraction*, will affect the equilibrium state of traffic networks. We study the system behavior when both human–driven and autonomous cars select their routes *selfishly* to investigate the necessity of centrally enforcing autonomous cars routing by a network manager. We state the conditions under which increasing the network autonomy fraction is guaranteed to not increase the overall network delay at equilibrium. Moreover, we show that when these conditions do not hold, counterintuitive and undesirable behaviors might occur, such as the case when increasing the network autonomy fraction can *increase* the overall network delay at equilibrium. Such behaviors are similar to Braess's paradox where the construction of a new road or expanding link capacities may increase total network delay.

We model the network in a macroscopic framework where route choices are taken into account. We model the selfish route choice behavior of drivers as a nonatomic routing game [96] where drivers choose their routes selfishly until a Wardrop Equilibrium is achieved [109]. We represent a traffic network by a directed graph with a certain set of origin destination (O/D) pairs. For each O/D pair, we consider two classes of cars, human–driven and autonomous. For a given fixed demand profile along O/D pairs, we study how increasing the autonomy fraction of O/D pairs will affect the total delay of the network at equilibrium.

We first show that the equilibrium may *not* be unique. Then, we study networks with a single O/D pair and prove that if the degrees of road capacity asymmetry (i.e. the ratio between the road capacity when all cars are human–driven and the road capacity when all cars are autonomous) are homogeneous in the network, at equilibrium, the social or total delay of the network is unique, and further it is a monotone nonincreasing function of the network autonomy fraction. However, in networks with heterogeneous degrees of road capacity asymmetry, we first show that the social delay is not necessarily unique at equilibrium. Then, we demonstrate that, surprisingly, increasing the autonomy fraction of the network might lead to an *increase* in the overall network delay at equilibrium. This is a counterintuitive behavior as we may expect that having more autonomous cars in

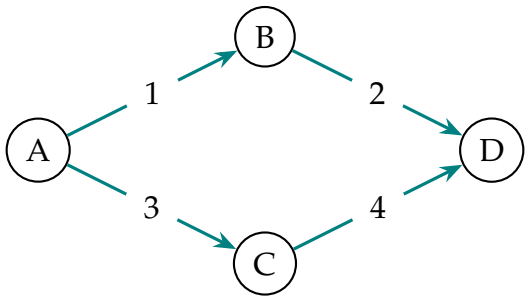


Figure 6.1: A network with a single O/D pair and two paths.

the network will always be beneficial in terms of total network delay. For networks with multiple O/D pairs, we show that similar complicated behaviors may occur, namely increasing the autonomy fraction of a single O/D pair might worsen the social delay of the network at equilibrium. Our work in fact shows that traffic paradoxes similar to the well known Braess's Paradox [13], can occur due to the capacity increases provided by autonomous cars. We further bound such performance degradations that can arise due to the presence of autonomy.

6.1 Equilibrium Uniqueness

Now, we study equilibrium uniqueness in the mixed autonomy setting. We adopt the modeling framework introduced in Section 2.2. Using Corollary 1, we know that there exists at least one equilibrium. However, since in our setting we have two classes of cars, Proposition 2 does not apply. Indeed, we demonstrate through an example that the equilibrium is *not* unique even in the weak sense introduced in Proposition 2.

Example 1. *Consider the network shown in Figure 6.1. Let p_1 and p_2 be the ABD and ACD paths respectively. For each link $l = 1, \cdots, 4$, let the link parameters be $\beta_l = 1, a_l = 1, m_l = 1$, and $M_l = 2$. Thus, for each link $l \in L$, the link delay function is $e_l(f_l^h, f_l^a) = 1 + f_l^h + \frac{f_l^a}{2}$. Assume that the demand from node A to D is $r = 2$, and $\alpha = 0.5$. The example is simple enough so that we can compute the equilibrium flows manually. Let f_1^h and f_1^a be the human–driven and autonomous flows along p_1, and f_2^h and f_2^a be the human–driven and autonomous flows along p_2. At equilibrium, using the symmetry of the network, we must have*

$$\begin{aligned} 2 + 2f_1^h + f_1^a &= 2 + 2f_2^h + f_2^a \\ f_1^h + f_2^h &= 1 \\ f_1^a + f_2^a &= 1 \\ f_1^h, f_1^a, f_2^h, f_2^a &\geq 0. \end{aligned}$$

Clearly, there is no unique solution to the above set of equations. Moreover, among the set of all possible equilibrium flow vectors, for each link, the maximum link flow at equilibrium is 1.25, whereas the minimum link flow is 0.75 at equilibrium. This implies that equilibrium uniqueness does not hold in traffic networks with mixed autonomy.

6.2 Networks with a Single O/D Pair

In this section, we study two–terminal networks which have a single O/D pair. For such networks, since there is only one O/D pair, all paths originate from a common source o and end in a common destination d. Since the set of O/D pairs $\mathcal{W}$ is singleton, we omit the subscript w from r_w, e_w and α_w throughout this section. Note that when the network has a single O/D pair, r and α are scalars.

Having observed that in the mixed–autonomy setting the equilibrium is not unique, it is important to study if the social delay at equilibrium is unique for all network equilibrium flow vectors. In the remainder of this chapter, we use the term social delay to refer to the social delay of the network *at equilibrium*. In the following, we study the properties of the social delay including its uniqueness for networks with a single O/D pair. But, before proceeding, we need to define the notion of degree of road capacity asymmetry introduced in [50] via the following.

Definition 5. *Given a network $G = (\mathcal{N}, \mathcal{L}, \mathcal{W})$, for each link $l \in \mathcal{L}$, we define $\mu_l := m_l / M_l$ to be the degree of road capacity asymmetry of link l.*

Note that since we assumed that autonomous cars' headway is less than or equal to that of human–driven cars, in the remainder of this chapter, we assume that for each link $l \in \mathcal{L}$, $\mu_l \leq 1$. Using Definition 5, in the sequel, we consider two scenarios for investigating the properties of social delay:

1. Homogeneous degrees of road capacity asymmetry, where μ_l is the same for all links, i.e. $\mu_l = \mu$ for all links $l \in \mathcal{L}$, where μ is the common value of capacity asymmetry.
2. Heterogeneous degrees of capacity asymmetry, where μ_l varies along different links.

Homogeneous Degrees of Capacity Asymmetry

In this case, we can establish the uniqueness of social delay, and characterize the relationship between social delay and network autonomy fraction.

Theorem 1. *Given a routing game (G, r, e), where G has a single O/D pair and a homogeneous degree of capacity asymmetry μ, for any value of total demand $r > 0$, we have:*

1. *For a fixed autonomy fraction $0 \leq \alpha \leq 1$, the social delay $J(f)$ is unique for all Wardrop equilibrium flow vectors f.*

2. *If for each* $0 \leq \alpha \leq 1$, *we denote the common value of social delay in the above by* $J(\alpha)$, *then* $J(.)$ *is continuous and nonincreasing.*

Proof. Fix $r > 0$ and $0 \leq \alpha \leq 1$. Recalling Corollary 1, we know that a Wardrop equilibrium exists. Let $f = (f_p^h, f_p^a : p \in \mathcal{P})$ be such an equilibrium flow vector where $f_p = f_p^a + f_p^h$ for each path p in $\mathcal{P}$. Define $e_{\min}(f) := \min_{p\in\mathcal{P}} e_p(f)$. Since the network has only one O/D pair, and the delay associated with all paths with nonzero flows are the same, denoting this uniform path delay by $e_{\min}(f)$, we realize that the social delay is given by $J(f) = re_{\min}(f)$. For each path $p \in \mathcal{P}$, define the fictitious single-class human–driven flow $\tilde{f}_p := f_p^h + \mu f_p^a$. We claim that the flow vector $\tilde{f} = (\tilde{f}_p : p \in \mathcal{P})$ is a Wardrop equilibrium for a routing game $(G, \tilde{r}, \tilde{e})$ with a single class of human–driven cars and a total demand of $\tilde{r} = r(1-\alpha) + r\alpha\mu$ with the link delay functions $(\tilde{e}_l : l \in \mathcal{L})$ defined as

$$\tilde{e}_l(\tilde{f}_l) = \left(a_l + \gamma_l \left(\frac{\tilde{f}_l}{m_l}\right)^{\beta_l}\right).$$

To see this, for each $p \in \mathcal{P}$, we show that relations (2.21) hold. Fix $p, p' \in \mathcal{P}$ and note that since f was a Wardrop equilibrium in the original setting, we know that $f_p^h(e_p(f) - e_{p'}(f)) \leq 0$, and $f_p^a(e_p(f) - e_{p'}(f)) \leq 0$. Multiplying the latter by the positive constant μ and adding the two inequalities, for every pair of paths $p, p' \in \mathcal{P}$, we have

$$\tilde{f}_p(e_p(f) - e_{p'}(f)) \leq 0. \tag{6.1}$$

Now, we claim that for all $p \in \mathcal{P}$, we have $e_p(f) = \tilde{e}_p(\tilde{f})$. Note that for each link $l \in \mathcal{L}$, we have $\tilde{f}_l = f_l^h + \mu f_l^a$. Using the fact that $\mu = m_l/M_l$ for all links $l \in \mathcal{L}$, we get

$$\begin{aligned}\tilde{e}_p(\tilde{f}) &= \sum_{l\in p}\left(a_l + \gamma_l\left(\frac{f_l^h + \frac{m_l}{M_l}f_l^a}{m_l}\right)^{\beta_l}\right)\\ &= \sum_{l\in p}\left(a_l + \gamma_l\left(\frac{f_l^h}{m_l} + \frac{f_l^a}{M_l}\right)^{\beta_l}\right) = e_p(f).\end{aligned} \tag{6.2}$$

Substituting into (6.1), we realize that for each pair of paths $p, p' \in \mathcal{P}$, we have

$$\tilde{f}_p(\tilde{e}_p(\tilde{f}) - \tilde{e}_{p'}(\tilde{f})) \leq 0, \tag{6.3}$$

which means that $\tilde{f}$ is an equilibrium flow vector. Clearly, the total demand of this new routing game is $\tilde{r} = \sum_{p\in\mathcal{P}} \tilde{f}_p = \sum_{p\in\mathcal{P}} f_p^h + \mu f_p^a = (1-\alpha)r + \mu\alpha r$. Moreover, define $\tilde{e}_{\min}(\tilde{f})$ to be the minimum of $\tilde{e}_p(\tilde{f})$ among $p \in \mathcal{P}$. Since w is the single O/D pair of the network, $\tilde{e}_{\min}(\tilde{f})$ is indeed equal to $\tilde{e}_w(\tilde{f})$, the travel delay of the single O/D pair of the network associated with $\tilde{f}$. Note that Proposition 2 implies that $\tilde{e}_{\min}(\tilde{f})$ is a function of $\tilde{r}$ only. On

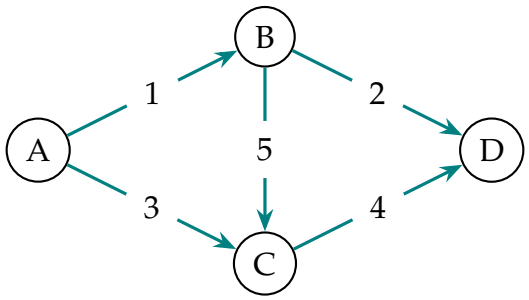


Figure 6.2: A network with a single O/D pair (A to D) and three paths from A to D.

the other hand, (6.2) implies that $\tilde{e}_{\min}(\tilde{f}) = e_{\min}(f)$. Putting these together, we realize that

$$J(f) = re_{\min}(f) = r\tilde{e}_{\min}(\tilde{f}) = r\tilde{e}_w(\tilde{r}). \tag{6.4}$$

Note that the right hand side of the above identity does not depend on f, which establishes the proof of the first part of the theorem. In fact, this shows that

$$J(\alpha) = r\tilde{e}_w\left(r(1-\alpha) + \alpha\mu r\right).$$

From Proposition 4, we know that $\tilde{e}_w(.)$ is continuous and nondecreasing. Also, since $\mu \leq 1$, the map $r \mapsto r(1-\alpha) + \alpha\mu r$ is continuous and nonincreasing. This completes the proof of the second part of the theorem. □

Heterogeneous Degrees of Capacity Asymmetry

Now, we allow μ_l to vary among network links. We show that this makes the behavior of the system more complex. First, we show via the following example that the social delay is not necessarily unique in this case.

Example 2. *Consider the network shown in Figure 6.2. Assume that $\gamma_l = 1, \beta_l = 1$, for $l = 1, 2, \cdots, 5$. Let other link parameters be the following: $\{a_1 = 1, m_1 = 1, M_1 = 1\}$, $\{a_2 = 2, m_2 = 1, M_2 = 3\}$, $\{a_3 = 1, m_3 = 1, M_3 = 2\}$, $\{a_4 = 1, m_4 = 1, M_4 = 4\}$, and $\{a_5 = 3, m_5 = 1, M_5 = 3\}$. Moreover, let the total flow from origin A to destination D be 2. Now, if we compute the social delay for this network for any $\alpha > 0$ at the different equilibria of the system, we observe that the social delay is* not *unique. In particular, Figure 6.3 shows the plots of the maximum and minimum social delay of the system at equilibrium for every value of α. To obtain Figure 6.3, we solved two optimization problems for finding maximum and minimum social delay subject to the equilibrium constraints using Mathematica. As Figure 6.3 shows, as soon as α starts to increase from 0, the uniqueness of social delay is lost. Once, $\alpha = 1$, the uniqueness of social delay is again preserved.*

Now, we study the effect of increasing the fraction of autonomous cars on the social delay. In the previous example, both the maximum and minimum social delays decreased

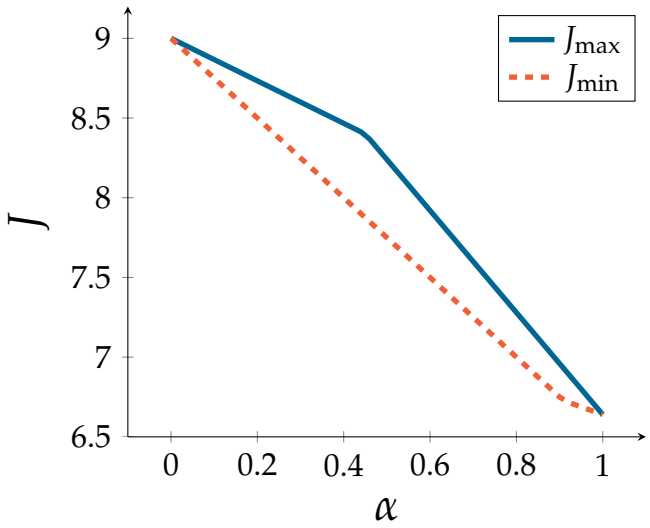


Figure 6.3: Maximum and minimum social delay at equilibrium for Example 2.

as a function of α. But, is this necessarily the case? We use the following examples to demonstrate that it may not be true in general, as increasing network autonomy may increase social delay in some networks.

Example 3. *Consider the network shown in Figure 6.2. Let $\gamma_l = 1$ and $\beta_l = 1$ for all links. Select the other network parameters to be the following, $\{a_1 = 0, m_1 = 0.1, M_1 = 0.1\}$, $\{a_2 = 50, m_2 = 1, M_2 = 1\}$, $\{a_3 = 50, m_3 = 1, M_3 = 1\}$, $\{a_4 = 0, m_4 = 0.1, M_4 = 0.1\}$, $\{a_5 = 10, m_5 = 0.5, M_5 = 1\}$. Let the total O/D demand be $r = 6$. In the absence of autonomy ($\alpha = 0$), the social delay is $J = 504.3$. However, if we increase the autonomy fraction to $\alpha = \frac{1}{10}$, $J = 518.6$. Clearly, in this case, the social delay increases when the network autonomy fraction α is increased. Note that since $\mu_l = 1$ for $l = 1, 2, 3, 4$ and $\mu_5 = 0.5 < 1$, this can be viewed as an instance of the classical Braess's Paradox [13], where an increase in the capacity of the middle link of a Wheatstone network can paradoxically lead to an increase in the social delay.*

One might argue that if we allow μ_l to be strictly less than 1 for all network links $l \in \mathcal{L}$, the network social delay will decrease by increasing the autonomy fraction. We use the following example to show that even in this case, increasing autonomy can worsen social delay.

Example 4. *Consider the previous example with the total flow $r = 6$, but change M_l's to be, $M_1 = \frac{1}{9}$, $M_2 = 1.1$, $M_3 = 1.1$, $M_4 = \frac{1}{9}$, and $M_5 = 1$. In this case, clearly, $\mu_l < 1$, for all $l \in \mathcal{L}$. We computed the maximum and minimum social delay at equilibrium for every autonomy fraction α. Figure 6.4 shows the maximum and minimum social delay in this example for different values of α. Figure 6.4 demonstrates that the maximum social delay increases as we increase α from 0, until we reach a local maximum. The minimum social delay decreases as we increase α from 0, until we reach a local minimum, and then, it increases sharply to values that are higher than the social delay*

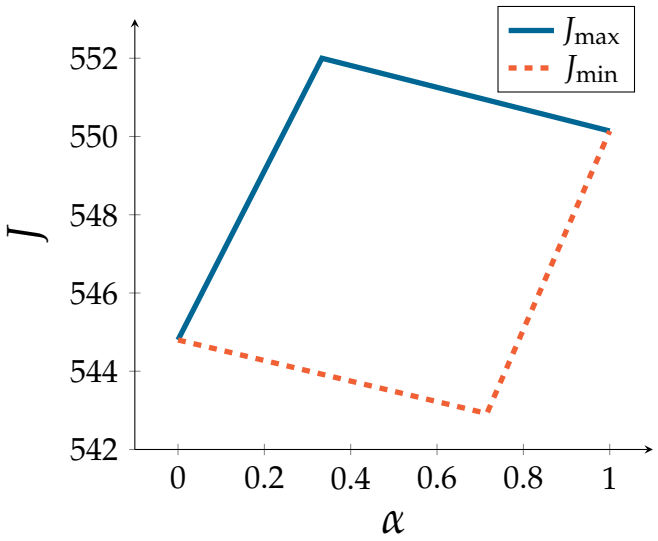


Figure 6.4: Maximum and minimum social delays at equilibrium for Example 4.

at $\alpha = 0$. Surprisingly, when all cars are autonomous ($\alpha = 1$) the social delay is greater than the social delay when $\alpha = 0$, i.e. $J(\alpha = 1) > J(\alpha = 0)$. This might be counterintuitive as we expect the network with full autonomy to have smaller social delay. However, this example shows that when capacity increases are heterogeneous across the network, the selfish behavior of cars when making their route choices might actually lead to worsening the social delay of the network.

As mentioned previously, the increase in social delay due to an increase in the fraction of autonomous cars is in fact similar to Braess's paradox. Braess's Paradox is the counter-intuitive but well known fact that removing edges from a network or increasing the delay functions on certain links can improve the delay of all cars at equilibrium [95]. It was shown in previous studies that Braess's paradox is prevalent in traffic networks [103] as the occurrence of Braess's paradox heavily depends on network topology and the parameters of link delay functions [94, 37, 75]. However, the phenomenon that we observed in this chapter for mixed–autonomy networks is different from the classical Braess's paradox in that link capacities are a function of the flows along links in a multiclass routing game. In other words, link capacity variations depend on how flows are routed throughout the network.

6.3 Networks with Multiple O/D Pairs

So far, we have seen that even in a network with only one O/D pair, the introduction of autonomous cars can result in complex behaviors. Thus, it should be expected that a general network with multiple O/D pairs will exhibit similar counterintuitive behaviors. In the previous section, we saw that in a network with a single O/D pair, the existence

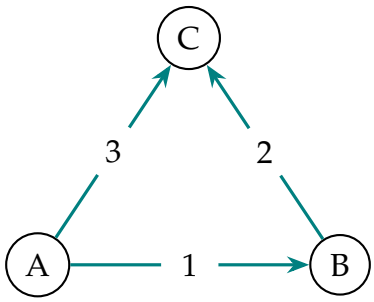


Figure 6.5: A network with three O/D pairs.

of a homogeneous degree of capacity asymmetry throughout the network is sufficient for guaranteeing improvements in the social delay by increasing the fraction of autonomous cars. We now show, via the following example, that this is not the case for networks with multiple O/D pairs.

Example 5. *Consider the network shown in Figure 6.5 which was first introduced in [28]. There are three O/D pairs,* $W = \{(A,B), (B,C), (A,C)\}$. *The total demand of the network O/D pairs are* $r_{AB} = 17, r_{AC} = 20$, *and* $r_{BC} = 90$. *Assume that* $\gamma_l = 1$, $\beta_l = 1$, *for all links* $l \in \mathcal{L}$. *Let the other link parameters be* $\{a_1 = 0, m_1 = 1, , M_1 = 4\}$, $\{a_2 = 0, m_2 = 1\}$, *and* $\{a_3 = 90, m_3 = 1\}$. *Let the cars that travel from A to C, and from B to C be all human–driven cars, i.e.* $\alpha_{AC} = \alpha_{BC} = 0$. *Figure 6.6 shows a plot of the network social delay versus the fraction of autonomous cars traveling along O/D pair AB denoted by* α_{AB}. *As the figure shows, as the autonomy fraction along O/D pair AB increases, so does the social delay. Intuitively, by increasing the autonomy fraction along O/D pair AB, the travel delay along link 1 decreases. As a result, compared to the no–autonomy case, a higher number of the cars traveling from A to C are encouraged to take advantage of the lower delay along link 1. This in turn will lead to an increase in the travel delay along O/D pair BC. This increase combined with the high demand of cars along O/D pair BC leads to an increase in the overall social delay. This example shows that allowing vehicle autonomy along certain network O/D pairs can result in worsening the overall or social delay of the network even if the road degrees of capacity asymmetry are homogeneous. This is of paramount importance in practice. For instance, if O/D pair AB belongs to a high–income neighborhood, autonomous cars may first be deployed along this O/D pair, while other neighborhood or O/D pairs may still travel via human–driven cars. Then, although the early adoption of autonomous cars along O/D pair AB may lead to a decrease in travel delay of O/D pair AB, it worsens the social delay in the network and increases the delays experienced by users along other O/D pairs. This example shows that even with homogeneous degrees of capacity asymmetry, when there exist multiple O/D pairs, different autonomy fractions along network O/D pairs can be another source of heterogeneity in the network.*

It was shown in [28, 23] that a decrease in the total demand of a single O/D pair, might lead to an increase in delay of travel along other network O/D pairs and the social delay. In the previous example, we showed that a similar behavior can also be observed due

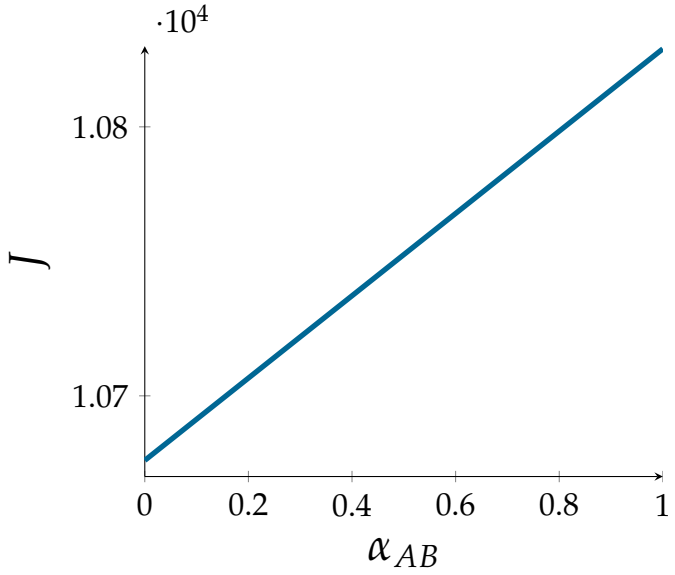


Figure 6.6: Social delay in Example 5 for different fractions of autonomous cars traveling along O/D pair AB when cars along all other O/D pairs are human–driven.

to the presence of autonomous cars. In fact, what we have shown so far is that the long known paradoxical traffic behavior resulting from constructing more roads or reducing demands can actually happen in networks with mixed autonomy due to the presence of autonomous cars. Thus, the mobility benefits of increasing car autonomy in a network are not immediate.

6.4 Bounding Performance Degradation

So far, we have shown that the social delay can increase as a consequence of the presence of autonomous cars in networks with mixed autonomy. Now, we wish to study whether we can bound this degradation in the network performance to estimate how much social delay can degrade by increasing the fraction of autonomous cars. To answer this, we derive a bound on the performance degradation that can result from all possible demand and autonomy fraction vectors in general networks that have a *homogeneous* degree of capacity asymmetry. To this end, for a given routing game (G, r, e), define the vector of fictitious reduced demand $\tilde{r} = (\tilde{r}_w : w \in \mathcal{W})$ to be $\tilde{r}_w = (1 - \alpha_w) r_w + \mu \alpha_w r_w$ for each O/D pair $w \in \mathcal{W}$. Consider an auxiliary fictitious routing game $(G, \tilde{r}, \tilde{e})$ with a total demand $\tilde{r}$ of only human–driven cars on G. For this auxiliary game, similar to Theorem 1, define $(\tilde{e}_l : l \in \mathcal{L})$ to be

$$\tilde{e}_l(\tilde{f}_l) = \left(a_l + \gamma_l \left(\frac{\tilde{f}_l}{m_l} \right)^{\beta_l} \right). \tag{6.5}$$

Let $\tilde{e}_w(\tilde{r})$ be the delay of travel for each O/D pair $w \in \mathcal{W}$ in this auxiliary game. Then, using the auxiliary fictitious game, we can state the following proposition.

Proposition 6. *Consider a routing game (G, r, e), where G is a general network with a homogeneous degree of capacity asymmetry $\mu \leq 1$ for all of its links. For any demand vector r with a fixed vector of autonomy fraction $\alpha = (\alpha_w : w \in \mathcal{W})$ such that $0 \leq \alpha_w \leq 1$ for all $w \in \mathcal{W}$, we have*

1. *The social delay $J(f)$ is unique for all Wardrop equilibrium flow vectors f.*
2. *The social delay of the original game is given by $J(f) = \sum_{w \in \mathcal{W}} r_w \tilde{e}_w(\tilde{r}_w)$ for all Wardrop equilibrium flow vectors f.*

Proof. Fix r and α, such that for each O/D pair $w \in \mathcal{W}$, $0 < r_w$ and $0 \leq \alpha_w \leq 1$. Recalling Corollary 1, we know that there exists at least one equilibrium. Let $f = (f_p^h, f_p^a : p \in \mathcal{P})$ be such an equilibrium flow vector for (G, r, e). For each path $p \in \mathcal{P}$, define $\tilde{f}_p := f_p^h + \mu f_p^a$. By generalizing the proof of Theorem 1, it is easy to see that $\tilde{f} = (\tilde{f}_p : p \in \mathcal{P})$ is an equilibrium for the defined auxiliary routing game on G with reduced demand $\tilde{r}$ of only human–driven cars. Moreover, for each path $p \in \mathcal{P}$, $e_p(f) = \tilde{e}_p(\tilde{f})$. Therefore, for each O/D pair $w \in \mathcal{W}$, $\tilde{e}_w(\tilde{f}) = \min_{p \in \mathcal{P}_w} \tilde{e}_p(\tilde{f}) = \min_{p \in \mathcal{P}_w} e_p(f) = e_w(f)$. Hence,

$$J(f) = \sum_{w \in \mathcal{W}} r_w e_w(f) = \sum_{w \in \mathcal{W}} r_w \tilde{e}_w(\tilde{f}). \tag{6.6}$$

Since $\tilde{f}$ contains only human–driven cars, recalling Proposition 2, for each $w \in \mathcal{W}$, the delay of travel $\tilde{e}_w(\tilde{f})$ is unique for a given $\tilde{r}$. Thus, following a derivation similar to (6.4), we have

$$J(f) = \sum_{w \in \mathcal{W}} r_w \tilde{e}_w(\tilde{r}). \tag{6.7}$$

As $\tilde{r}$ is uniquely determined for a given demand vector r and a vector of autonomy fraction α, the social delay $J(f)$ is unique for all Wardrop equilibrium flow vectors f and can be obtained via (6.7). □

The uniqueness of social delay established by Proposition 6 implies that for a fixed demand vector r, the social delay is a well defined function of autonomy fraction α. With a slight abuse of notation, we use $J(\alpha)$ to emphasize the dependence of the social delay on the vector of autonomy fraction α. Note that Proposition 6 establishes a connection between our original routing game, which has two classes of cars, with a fictitious auxiliary routing game, which has only human–driven cars and a reduced demand vector $\tilde{r}$. We exploit this connection in the remainder of the chapter. Since the auxiliary game has only one class of cars, the results in [20] hold for this game. Before proceeding, we need to adopt and review some of the definitions in [20] for the auxiliary game.

In the auxiliary game, for a given demand vector $\tilde{r}$, a flow vector $\tilde{f}$ is feasible if $\tilde{f}_p \geq 0$ for all paths $p \in \mathcal{P}$, and $\sum_{p\in\mathcal{P}_w} \tilde{f}_p = \tilde{r}_w$ for all O/D pairs $w \in \mathcal{W}$. Let $\phi \in \mathbb{R}^{|\mathcal{L}|}$ be a vector of *link* flows that result from a feasible flow vector $\tilde{f}$, where $|\mathcal{L}|$ is the number of links in the network. Also, let Φ represent the set of all feasible link flow vectors ϕ for a given reduced demand vector $\tilde{r}$. Then, for a vector of link delay functions $(\tilde{e}_l : l \in \mathcal{L})$ of the form (6.5) and any vector $v \in \Phi$, define

$$\lambda\big((\tilde{e}_l : l \in \mathcal{L}), v\big) := \max_{x\in\mathbb{R}^{|\mathcal{L}|}_{\geq 0}} \frac{\sum_{l\in\mathcal{L}} \big(\tilde{e}_l(v_l) - \tilde{e}_l(x_l)\big)x_l}{\sum_{l\in\mathcal{L}} \tilde{e}_l(v_l)v_l}, \tag{6.8}$$

where $0/0$ is considered to be 0. Additionally, let $\tilde{\mathcal{E}}$ be the class of delay functions represented by (6.5). Define

$$\lambda(\tilde{\mathcal{E}}) := \sup_{(\tilde{e}_l : l\in\mathcal{L})\in\tilde{\mathcal{E}}, v\in\Phi} \lambda\left((\tilde{e}_l : l \in \mathcal{L}), v\right). \tag{6.9}$$

It is important to mention that since the class of delay functions $\tilde{\mathcal{E}}$ is monotone, $\lambda(\tilde{\mathcal{E}}) \leq 1$ (See Section 4 in [20]). Note that $\lambda(\tilde{\mathcal{E}})$ can be easily computed for certain classes of delay functions such as polynomials. For instance, $\lambda(\tilde{\mathcal{E}}) = \frac{1}{4}$ for the class of linear delay functions.

Now, we can bound the network performance degradation due to the introduction of autonomy in homogeneous networks via the following theorem.

Theorem 2. *Consider a routing game (G, r, e) with G being a general network with a homogeneous degree of capacity asymmetry μ. Fix the demand vector r. Let J^o be the social delay when all cars are nonautonomous, i.e. $\alpha_w = 0$ for all O/D pairs $w \in \mathcal{W}$. Then, for any vector of autonomy fraction α such that $0 \leq \alpha_w \leq 1$ for all $w \in \mathcal{W}$, we have*

$$J(\alpha) \leq (1 - \lambda(\tilde{\mathcal{E}}))^{-1} J^o, \tag{6.10}$$

where $J(\alpha)$ is the social delay for a given vector of autonomy fraction α.

Proof. Fix the demand vector r. Let $f^o = (f^o_p : p \in \mathcal{P})$ be an equilibrium flow vector when all cars are human–driven. We further use f^o_l to denote the flow along link $l \in \mathcal{L}$ in this case. Note that using Proposition 2 , we know that at equilibrium, f^o_l is unique for every link $l \in \mathcal{L}$. Moreover, for each path $p \in \mathcal{P}$, we use e^o_p to represent the delay along path p when all cars are human–driven. Using Proposition 2, in the absence of autonomy, the delay of travel for each O/D pair $w \in \mathcal{W}$ is unique. Thus, in the no–autonomy case, the unique social delay is $J^o = \sum_{w\in\mathcal{W}} r_w e^o_w(r)$, where $e^o_w(r)$ is the delay of travel for O/D pair $w \in \mathcal{W}$ when all cars are human–driven.

On the other hand, when there are autonomous cars with a given autonomy fraction α in the network, as defined in Proposition 6, construct the auxiliary game $(G, \tilde{r}, \tilde{e})$ with

fictitious reduced demand $\tilde{r} = (\tilde{r}_w : w \in \mathcal{W})$ of only human–driven cars, where $\tilde{r}_w = (1-\alpha_w)r_w + \mu r_w \alpha_w$ for every $w \in \mathcal{W}$. Let $\tilde{f} = (\tilde{f}_p : p \in \mathcal{P})$ be an equilibrium flow vector for the auxiliary game $(G, \tilde{r}, \tilde{e})$. Using Proposition 6, the social delay of the network with autonomy is given by $J(\alpha) = \sum_{w\in\mathcal{W}} r_w \tilde{e}_w(\tilde{r})$. First, we claim that

$$J(\alpha) = \sum_{w\in\mathcal{W}} r_w \tilde{e}_w(\tilde{r}) \leq \sum_{l\in\mathcal{L}} f_l^o \tilde{e}_l(\tilde{r}). \tag{6.11}$$

To see this, note that for every link $l \in \mathcal{L}$, we have $f_l^o = \sum_{p\in\mathcal{P}: l\in p} f_p^o$. Furthermore, the origin and destination of each path $p \in \mathcal{P}$ are unique. Hence, each path p belongs to one and exactly one O/D pair $w \in \mathcal{W}$. Consequently, $f_l^o = \sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w: l\in p} f_p^o$, and we have

$$\begin{aligned}
\sum_{l\in\mathcal{L}} f_l^o \tilde{e}_l(\tilde{r}) &= \sum_{l\in\mathcal{L}} \left(\sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w: l\in p} f_p^o \right) \tilde{e}_l(\tilde{r}) \\
&= \sum_{w\in\mathcal{W}} \sum_{l\in\mathcal{L}} \left(\sum_{p\in\mathcal{P}_w: l\in p} f_p^o \right) \tilde{e}_l(\tilde{r}) \\
&= \sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w} f_p^o \sum_{l: l\in p} \tilde{e}_l(\tilde{r}) \\
&= \sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w} f_p^o \tilde{e}_p(\tilde{r}),
\end{aligned}$$

where $\tilde{e}_p(\tilde{r})$ is the delay of travel along path $p \in \mathcal{P}_w$ in the auxiliary game. Recalling Definition 2, for the auxiliary game, the travel delay of an O/D pair $w \in \mathcal{W}$ is given by $\tilde{e}_w(\tilde{r}) = \min_{p\in\mathcal{P}_w} \tilde{e}_p(\tilde{r})$; thus, we have

$$\begin{aligned}
\sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w} f_p^o \tilde{e}_p(\tilde{r}) &\geq \sum_{w\in\mathcal{W}} \sum_{p\in\mathcal{P}_w} f_p^o \tilde{e}_w(\tilde{r}) \\
&= \sum_{w\in\mathcal{W}} \tilde{e}_w(\tilde{r}) \sum_{p\in\mathcal{P}_w} f_p^o \\
&= \sum_{w\in\mathcal{W}} r_w \tilde{e}_w(\tilde{r}),
\end{aligned}$$

which proves our claim in (6.11). Now, since the auxiliary game has only one class of cars, we can use Lemma 4.1 from [20]. More precisely, since $\tilde{f}$ is an equilibrium for the auxiliary game, then Lemma 4.1 from [20] states that for every nonnegative vector of link flows $x \in \mathbb{R}_{\geq 0}^{|\mathcal{L}|}$ (x is not necessarily a feasible link flow vector), we have

$$\sum_{l\in\mathcal{L}} x_l \tilde{e}_l(\tilde{f}_l) \leq \sum_{l\in\mathcal{L}} x_l \tilde{e}_l(x_l) + \lambda(\tilde{\mathcal{E}}) \sum_{l\in\mathcal{L}} \tilde{f}_l \tilde{e}_l(\tilde{f}_l). \tag{6.12}$$

Since f_l^o is nonnegative for every link $l \in \mathcal{L}$, substituting x_l by f_l^o in (6.12), we get

$$\sum_{l\in\mathcal{L}} f_l^o \tilde{e}_l(\tilde{f}_l) \leq \sum_{l\in\mathcal{L}} f_l^o \tilde{e}_l(f_l^o) + \lambda(\tilde{\mathcal{E}}) \sum_{l\in\mathcal{L}} \tilde{f}_l \tilde{e}_l(\tilde{f}_l). \tag{6.13}$$

Now, note that since both the auxiliary game and the game with no autonomy have only human–driven cars, utilizing (6.5), we realize that

$$\begin{aligned}\tilde{e}_l(f_l^o) &= \left(a_l + \gamma_l \left(\frac{f_l^o}{m_l}\right)^{\beta_l}\right) \\ &= e_l^o(f_l^o).\end{aligned}$$

Thus,

$$\sum_{l\in\mathcal{L}} f_l^o \tilde{e}_l(f_l^o) = \sum_{l\in\mathcal{L}} f_l^o e_l^o(f_l^o) = J^o. \tag{6.14}$$

Now, since $J(\alpha) = \sum_{w\in\mathcal{W}} r_w \tilde{e}_w(\tilde{r})$, using (6.11), (6.13), and (6.14), we realize that

$$J(\alpha) \leq J^o + \lambda(\tilde{\mathcal{E}}) \sum_{l\in\mathcal{L}} \tilde{f}_l \tilde{e}_l(\tilde{r}). \tag{6.15}$$

As $\tilde{f}$ is an equilibrium for the auxiliary routing game, $\sum_{l\in\mathcal{L}} \tilde{f}_l \tilde{e}_l(\tilde{r}) = \sum_{w\in\mathcal{W}} \tilde{r}_w \tilde{e}_w(\tilde{r})$. Since for each O/D pair $w \in \mathcal{W}$, we assumed that $\alpha_w \leq 1$, we can conclude that for each O/D pair $w \in \mathcal{W}$, we have $\tilde{r}_w \leq r_w$. Therefore, using Proposition 6, we realize that

$$\sum_{w\in\mathcal{W}} \tilde{r}_w \tilde{e}_w(\tilde{r}) \leq \sum_{w\in\mathcal{W}} r_w \tilde{e}_w(\tilde{r}) = J(\alpha). \tag{6.16}$$

Using (6.16) and (6.15), we get

$$J(\alpha) \leq J^o + \lambda(\tilde{\mathcal{E}}) J(\alpha). \tag{6.17}$$

Hence, for the our monotone class of delay functions $\tilde{\mathcal{E}}$ with $\lambda(\tilde{\mathcal{E}}) < 1$, we can conclude that

$$J(\alpha) \leq (1 - \lambda(\tilde{\mathcal{E}}))^{-1} J^o,$$

which completes the proof. □

Theorem 2 provides an upper bound on the severity of increases in traffic delays when a fraction of human–driven cars is replaced by autonomous cars.

We now postulate, as an analogous concept to the price of anarchy [97], the price of vehicle autonomy in homogeneous networks under every demand vector r as follows:

$$\eta := \max_{\alpha:\ \forall w,\ 0\leq\alpha_w\leq 1} \frac{J(\alpha)}{J^o}. \tag{6.18}$$

Theorem 2 indicates that $\eta \leq (1 - \lambda(\tilde{\mathcal{E}}))^{-1}$. For polynomial delay functions of degree less than or equal to 4, $(1 - \lambda(\tilde{\mathcal{E}}))^{-1} = 2.151$ [20]. The bound that we have derived for the price of car autonomy is similar to the bounds derived for the price of anarchy of routing games with a single class of users in [97]. However, unlike the bounds for price of anarchy, the tightness of our bound for η needs further investigation.

6.5 Chapter Summary

In this chapter, we studied how the coexistence of autonomous and human–driven cars in traffic networks will affect network mobility when all cars select their routes selfishly. We compared the total network delay at a Wardrop equilibrium in networks with mixed autonomy with that of the networks with only human–driven cars. Having shown that the equilibrium is not unique in the mixed–autonomy setting, we proved that the total delay is unique when the road degree of capacity asymmetry, which is the ratio between the roadway capacity with only human–driven cars and the roadway capacity with only autonomous cars, is homogeneous among its roadway. We further proved that the total delay is a nonincreasing and continuous function of the fraction of autonomous cars on the roadways (a.k.a. the autonomy fraction) when the network has only one O/D pair. However, we showed that allowing for heterogeneous degrees of capacity asymmetry or multiple O/D pairs in the network results in counterintuitive behaviors such as the fact that increasing network autonomy fraction can worsen the network total delay. Finally, we derived an upper bound for the "price of vehicles autonomy" in networks with a homogeneous degree of capacity asymmetry, which estimates the worst possible increase in network social delay due to the presence of autonomous cars.

Chapter 7

Pricing Mixed–Autonomy Networks

Since cars select their routes selfishly, traffic networks tend to operate at equilibria characterized by Wardrop conditions, where vehicular flows are routed along the network paths such that no car can gain any savings in its travel time by unilaterally changing its route [109]. However, it is known that due to the selfish behavior of cars, network equilibria are in general inefficient. As an example, well known Braess's paradox [12] described an extreme scenario where adding a link to a network increased the overall network delay at equilibrium. Inefficiency of equilibria is commonly measured via the price of anarchy [97]. In [51], the price of anarchy of traffic networks with mixed vehicle autonomy was computed, and it was shown that the price of anarchy of networks with mixed vehicle autonomy is larger than that of networks with no autonomy. This implies that although the optimal overall delay of networks with mixed autonomy is lower (due to the capacity increase of autonomous cars), at equilibrium, the overall delay of networks with mixed autonomy is further from its optimum. Furthermore, in Chapter 6, we showed that under constant vehicular demand, if the fraction of autonomous cars increases, the overall network delay may grow at equilibrium.

In this chapter, we study how to cope with the inefficiency of equilibria in traffic networks with mixed vehicle autonomy such that the potential mobility benefits of autonomy is achieved, and minimum overall network delay also known as social delay is achieved at equilibrium. In particular, we study how to set prices on network links such that equilibria with minimum social delay are induced. Pricing has been extensively studied as a tool to create efficient equilibria in the previous literature (See [14, 16, 15, 40]). For traffic networks with only a single class of cars, a marginal cost taxation of network links was proposed in [87] which was proven to induce equilibria with minimum social delay. However, when there are multiple classes of cars, traffic networks exhibit complex behavior such as nonuniqueness of equilibria [21]; thereby, prices that are obtained from marginal costs of network links may not be sufficient for inducing optimality of social delay at all possible equilibria [22, 100].

In this chapter, we first study whether minimum social delay can be induced by setting undifferentiated prices on network links i.e. a set of prices that treat both human–

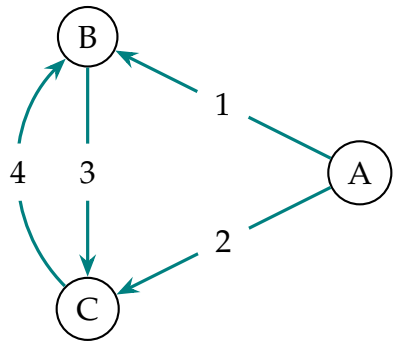


Figure 7.1: A network with two O/D pairs from A to B and A to C.

driven and autonomous cars similarly at each link. We show through an example that undifferentiated prices are not in general enough for inducing minimum social delay at equilibrium. Then, we consider the setting where differentiated prices are assigned on network links, i.e. at every network link, the price assigned to human–driven cars can be different from that of autonomous cars. We prove that despite the fact that equilibrium is not necessarily unique in traffic networks with mixed autonomy, for traffic networks with a homogeneous degree of capacity asymmetry (networks where the ratio of the link capacities when all cars are human–driven over the capacity when all cars are autonomous is uniform throughout the network), with an appropriate set of prices, the social network delay of all induced equilibria is minimum. This is of paramount importance since uniqueness of social delay guarantees that regardless of which equilibrium the network operates at, the social delay will be minimum. In the absence of such a guarantee, the social delay may be optimal only for certain subsets of the induced equilibria.

7.1 Undifferentiated Prices

When it comes to setting prices for network links, generally, an ideal set of prices is the one that induces an equilibrium flow vector that minimizes the network social delay. When there are multiple classes of cars in a network, for instance human–driven and autonomous cars in our scenario, it is important to determine whether it is possible to induce an equilibrium that optimizes social delay via prices that do not differentiate car classes. This is of practical significance because undifferentiated prices are much easier to implement. Unfortunately, in this section, we show through a counterexample that, for traffic networks with mixed autonomy, it is not always possible to induce equilibrium flows with minimum social delay by simply applying undifferentiated prices, even in networks with a homogeneous degree of capacity asymmetry.

Example 6. *Consider the network shown in Figure 7.1. Assume that the network has a homogeneous degree of capacity asymmetry* $\mu = \frac{1}{3}$. *There are two O/D pairs* $\mathcal{W} = \{AB, AC\}$. *For each O/D pair, there are two possible paths. The demand of O/D pairs are* $r^h_{AB} = 7.5$, $r^a_{AB} = 4.5$,

$r^h_{AC} = 1.2$, and $r^a_{AC} = 4.8$. For every link $l, 1 \leq l \leq 4$, assume that $\gamma_l = 1$, and $\beta_l = 1$. The other link parameters are $a_1 = 9, a_2 = 3, a_3 = 0.6$, and $a_4 = 0.6$ while $m_1 = 3, m_2 = 0.5, m_3 = 0.7$, and $m_4 = 0.5$. Note that for each link l, the parameter M_l is determined through the relation $M_l = m_l/\mu$. For this network, it is easy to compute the optimal social delay, which is $J^* = 193.54$. In order to see whether a set of undifferentiated prices can achieve this optimal social delay, we solve the following optimization problem for this network

$$\begin{aligned} \min \quad & \sum_{p \in \mathcal{P}} (f^h_p + f^a_p) e_p(f) \\ \textit{subject to} \quad & \textit{Equations (2.15)}, \\ & \textit{Equations (2.31)}, \\ & \forall l \in \mathcal{L} : \tau^h_l = \tau^a_l \geq 0, \end{aligned} \tag{7.1}$$

Note that if the minimum social delay J^* can be achieved via undifferentiated prices, the optimal value of optimization problem (7.1) must be equal to J^*. However, the minimum value of (7.1) for the network in Figure 7.1 is 195.597 which is clearly greater than J^*. This indicates that an equilibrium with socially optimal delay cannot necessarily be induced by undifferentiated link prices in general.

7.2 Differentiated prices

Having shown in Example 6 that in general, undifferentiated pricing cannot induce an equilibrium with socially optimal delay, a natural question is whether *differentiated* prices can be employed to induce an equilibrium with minimum social delay. In other words, if at every link $l \in \mathcal{L}$, we allow τ^h_l and τ^a_l to be different, does there exist a price vector $\tau = (\tau^h_l, \tau^a_l : l \in \mathcal{L})$ that induces equilibria with minimum social delay? In this section, we prove that such differentiated prices exist, and we find them.

Homogeneous Networks

The following theorem establishes the existence of optimal prices for traffic networks with a homogeneous degree of capacity asymmetry and also provides a recipe for how to find the optimal price values.

Theorem 3. *Consider a priced routing game (G, r, c) with a homogeneous degree of capacity asymmetry μ. For a given vector of autonomy fraction α, let f^* and J^* be an optimizer and the minimum value of social delay respectively. Then, for each link $l \in \mathcal{L}$, if the link prices are set to*

be

$$\tau_l^h = (f_l^{h*} + f_l^{a*}) \left(\frac{\partial}{\partial f_l^h} e_l(f_l^h, f_l^a) \right) \Big|_{f^*}, \tag{7.2a}$$

$$\tau_l^a = (f_l^{h*} + f_l^{a*}) \left(\frac{\partial}{\partial f_l^a} e_l(f_l^h, f_l^a) \right) \Big|_{f^*}, \tag{7.2b}$$

then, all induced equilibria of the game (G, r, c) have the same social delay, which is equal to J^*.

Proof. Fix the vector of autonomy fractions α. For the routing game (G, r, c) where the link prices are obtained via (7.2), Proposition 5 implies that there exists one equilibrium with minimum social delay J^*. We prove that when the above prices are set, all induced equilibria of (G, r, c) have the same social delay. This would then imply that all induced equilibria of the game (G, r, c) have the unique social delay J^* as was claimed.

Therefore, it remains to prove uniqueness of social delay at equilibria of (G, r, c) when link prices are obtained from (7.2). Recall that for a given (G, r, c) and fixed α, for each O/D pair $w \in \mathcal{W}$, the demand of human–driven and autonomous cars are obtained via

$$r_w^h = (1 - \alpha_w) r_w \tag{7.3}$$

$$r_w^a = \alpha_w r_w. \tag{7.4}$$

Now, in order to prove uniqueness of social delay, we construct an auxiliary game instance $(G, \tilde{r}, \tilde{c})$, with the same network graph G and O/D pairs $\mathcal{W}$, where the demand of O/D pairs, link traversal delays and cost functions are defined as follows. For each O/D pair $w \in \mathcal{W}$, define the demand of human–driven and autonomous cars $\tilde{r}_w^h$ and $\tilde{r}_w^a$ in the auxiliary game to be

$$\tilde{r}_w^h := r_w^h, \tag{7.5a}$$

$$\tilde{r}_w^a := \mu r_w^a. \tag{7.5b}$$

Moreover, for every link $l \in \mathcal{L}$, let the link delay functions of the auxiliary game $(G, \tilde{r}, \tilde{c})$ be defined as

$$\tilde{e}_l(\tilde{f}_l^h, \tilde{f}_l^a) := a_l + \gamma_l \left(\frac{\tilde{f}_l^h + \tilde{f}_l^a}{m_l} \right)^{\beta_l}. \tag{7.6}$$

Additionally, for every link $l \in \mathcal{L}$, with the prices as in (7.2), define the link cost functions in the auxiliary game to be

$$\tilde{c}_l^h(\tilde{f}_l^h, \tilde{f}_l^a) := \tilde{e}_l(\tilde{f}_l^h, \tilde{f}_l^a) + \tau_l^h, \tag{7.7a}$$

$$\tilde{c}_l^a(\tilde{f}_l^h, \tilde{f}_l^a) := \tilde{e}_l(\tilde{f}_l^h, \tilde{f}_l^a) + \tau_l^a. \tag{7.7b}$$

Now, let $f = (f_p^h, f_p^a : p \in \mathcal{P})$ be an equilibrium of the original game (G, r, c). For every path $p \in \mathcal{P}$, define $\tilde{f}_p^h := f_p^h$ and $\tilde{f}_p^a := \mu f_p^a$. We claim that $\tilde{f} = (\tilde{f}_p^h, \tilde{f}_p^a : p \in \mathcal{P})$ is an equilibrium flow vector for the auxiliary game $(G, \tilde{r}, \tilde{c})$. It can be easily verified that for every origin–destination pair $w \in \mathcal{W}$, we have $\sum_{p \in \mathcal{P}_w} \tilde{f}_p^h = \tilde{r}_w^h$ and $\sum_{p \in \mathcal{P}_w} \tilde{f}_p^a = \tilde{r}_w^a$. Thus, $\tilde{f}$ is a feasible flow vector for the auxiliary game. Moreover, it is easy to see that for every link $l \in \mathcal{L}$, we have $\tilde{f}_l^a = \mu f_l^a$; therefore, using the definition of $\tilde{f}$ and Equations (2.26) and (7.6), for every link $l \in \mathcal{L}$, we establish the following

$$\begin{aligned} \tilde{e}_l(\tilde{f}_l^h, \tilde{f}_l^a) &= a_l + \gamma_l \left(\frac{\tilde{f}_l^h + \tilde{f}_l^a}{m_l} \right)^{\beta_l} \\ &= a_l + \gamma_l \left(\frac{f_l^h + \mu f_l^a}{m_l} \right)^{\beta_l} \\ &= a_l + \gamma_l \left(\frac{f_l^h}{m_l} + \frac{f_l^a}{M_l} \right)^{\beta_l} \\ &= e_l(f_l^h, f_l^a). \end{aligned} \tag{7.8}$$

Thus, from (2.28), (7.7), and (7.8), for every link $l \in \mathcal{L}$, we have

$$\tilde{c}_l^h(\tilde{f}_l^h, \tilde{f}_l^a) = c_l^h(f_l^h, f_l^a) \tag{7.9a}$$

$$\tilde{c}_l^a(\tilde{f}_l^h, \tilde{f}_l^a) = c_l^a(f_l^h, f_l^a). \tag{7.9b}$$

Now since f is an equilibrium for (G, r, c), using (2.31) and (7.9), for every O/D pair $w \in \mathcal{W}$ and pair of paths $p, p' \in \mathcal{P}_w$, we have

$$f_p^h \left(\tilde{c}_p^h(\tilde{f}) - \tilde{c}_{p'}^h(\tilde{f}) \right) \leq 0, \tag{7.10a}$$

$$f_p^a \left(\tilde{c}_p^a(\tilde{f}) - \tilde{c}_{p'}^a(\tilde{f}) \right) \leq 0. \tag{7.10b}$$

Multiplying (7.10b) by the positive constant μ, we have

$$f_p^h \left(\tilde{c}_p^h(\tilde{f}) - \tilde{c}_{p'}^h(\tilde{f}) \right) \leq 0, \tag{7.11a}$$

$$\mu f_p^a \left(\tilde{c}_p^a(\tilde{f}) - \tilde{c}_{p'}^a(\tilde{f}) \right) \leq 0. \tag{7.11b}$$

Using the definition of $\tilde{f}$, from (7.11) we can conclude the following

$$\tilde{f}_p^h \left(\tilde{c}_p^h(f) - \tilde{c}_{p'}^h(f) \right) \leq 0, \tag{7.12a}$$

$$\tilde{f}_p^a \left(\tilde{c}_p^a(\tilde{f}) - \tilde{c}_{p'}^a(\tilde{f}) \right) \leq 0. \tag{7.12b}$$

Note that these are precisely the equilibrium conditions for the auxiliary game $(G, \tilde{r}, \tilde{c})$, which proves our claim that $\tilde{f}$ is an equilibrium for the auxiliary game.

Now, note that the conditions of Proposition 3 hold for the auxiliary game $(G, \tilde{r}, \tilde{c})$ since for the fixed α, at every link $l \in \mathcal{L}$, the link traversal costs of human–driven and autonomous cars $\tilde{c}_l^h$ and $\tilde{c}_l^a$ are strictly increasing functions of the total link flow $\tilde{f}_l = \tilde{f}_l^h + \tilde{f}_l^a$. Moreover, motivated by (7.7), the costs of human-driven and autonomous cars are identical up to a constant. Thus, using Proposition 3 for $(G, \tilde{r}, \tilde{c})$, for every link $l \in \mathcal{L}$, the total link flow $\tilde{f}_l = \tilde{f}_l^h + \tilde{f}_l^a$ is unique among all equilibria. Therefore, using the definition of $\tilde{f}$, the fact that $\tilde{f}$ is an equilibrium flow vector for $(G, \tilde{r}, \tilde{c})$, and the connection between f and $\tilde{f}$, we conclude that at every link $l \in \mathcal{L}$, we must have that $f_l^h + \mu f_l^a$ is unique for all equilibria of (G, r, c). Additionally, from (7.6) and (7.7), for every link $l \in \mathcal{L}$, uniqueness of the total link flow at equilibrium in the auxiliary game implies that the link traversal costs $\tilde{c}_l^h(\tilde{f}_l^h, \tilde{f}_l^a)$ and $\tilde{c}_l^a(\tilde{f}_l^h, \tilde{f}_l^a)$ are unique. Hence, from (7.9), we can conclude that in (G, r, c), for each link $l \in \mathcal{L}$, the link traversal costs $c_l^h(f_l^h, f_l^a)$ and $c_l^a(f_l^h, f_l^a)$ are also unique for all equilibrium flow vectors f. Thus, for each O/D pair $w \in \mathcal{W}$, the travel costs of both human-driven and autonomous cars c_w^h and c_w^a are unique. Consequently, from (2.33), we realize that the overall cost $C(f)$ is unique for all equilibrium flow vectors f of (G, r, c). Now, using (2.17), (2.28), and (2.30), we can rewrite the social cost of (G, r, c) as

$$C(f) = \sum_{l \in \mathcal{L}} f_l^h c_l^h(f) + f_l^a c_l^a(f) \tag{7.13}$$

$$= \sum_{l \in \mathcal{L}} (f_l^h + f_l^a) e_l^h(f) + f_l^h \tau_l^h + f_l^a \tau_l^a \tag{7.14}$$

$$= J(f) + \sum_{l \in \mathcal{L}} f_l^h \tau_l^h + f_l^a \tau_l^a. \tag{7.15}$$

Notice that under homogeneity of the network, using the special structure of (2.26), it is easy to see that Equations (7.2) imply that for every link $l \in \mathcal{L}$, we have

$$\tau_l^a = \mu \tau_l^h. \tag{7.16}$$

Substituting (7.16) in (7.15), we have

$$C(f) = J(f) + \sum_{l \in \mathcal{L}} \tau_h^l (f_h^l + \mu f_l^a). \tag{7.17}$$

Note that using our introduced auxiliary game, we proved that the overall cost $C(f)$ is unique for all equilibrium flow vectors f. Furthermore, we proved that for every link l, $f_l^h + \mu f_l^a = \tilde{f}_l^h + \tilde{f}_l^a$ is unique for all equilibria. Hence, from (7.17), we can conclude that the social delay $J(f)$ is also unique for all equilibrium flow vectors. This completes our proof. □

Note that if for each link $l \in \mathcal{L}$, the link prices are obtained from (7.2), since the link delay function (2.26) is increasing in the flow of each vehicle class, the prices that result

from (7.2) are always nonnegative which is in accordance with our initial assumption. The link prices obtained by (7.2) are in fact the extra term in the marginal cost of each vehicle class.

Heterogeneous Networks

If the road degree of capacity asymmetry is not homogeneous, but a central authority sets link prices to be obtained from (7.2), there still exists at least one induced equilibrium flow vector that achieves the minimum social delay. However, for heterogeneous networks, the social delay is not necessarily unique. Therefore, although the social delay of one induced equilibrium is optimal, the social delay of other induced equilibria might not be optimal. For such networks, optimally of the social delay may not be achieved in all induced equilibria by setting the prices to be obtained from (7.2).

7.3 Chapter Summary

We considered the problem of inducing efficient equilibria in traffic networks with mixed vehicle autonomy via pricing. We showed that minimum social delay may not be attained by imposing undifferentiated link prices, in which human–driven and autonomous vehicles are treated identically. Then, we proved that in mixed–autonomy traffic networks with a homogeneous degree of capacity asymmetry, if differentiated prices are allowed, which treat human–driven and autonomous vehicles differently, link prices can be determined such that all induced equilibria have minimum social delay.

Chapter 8

Altruistic Autonomy

In this chapter, we focus on how autonomy can act altruistically, i.e. by taking into account the decision making process of humans, autonomous cars can potentially plan for their actions in the favor of the overall system good. In particular, we consider the altruistic lane choice of autonomous cars in a mixed–autonomy setting at a traffic diverge. First, we develop a model for the lane choice of human–driven cars. Then, using this model, we discuss the altruistic lane choice of autonomy at traffic diverges.

8.1 Modeling Framework

We consider a traffic diverge where a two–lane link bifurcates into two links as depicted in Figure 8.1. This is a common scenario in freeway and arterial forks. We wish to study lane choices taken by cars in such diverges. At traffic diverges, normally, among the cars with the same target exit link, a fraction of cars choose to remain in the lane that corresponds to their intended exit, far upstream of the diverge, while the remaining fraction of the cars perform bypasses by performing a lane change behavior to the lane that corresponds to their intended exit link at a proximity close to the diverge (see Figure 8.1). Given the demand of cars for each exit link, our goal is to derive a macroscopic traffic model that can predict the fraction of cars which exhibit either of these two behaviors. Note that we wish to capture the aggregate bypassing behavior of cars in a macro scale rather than deriving the conditions under which a single car decides to perform or not perform a bypassing lane change maneuver.

Consider a traffic diverge shown in Figure 8.1. Let $I = \{P, Q\}$ be the index set of exit links. Let d^P and d^Q be the demands of cars that wish to take exit links P and Q respectively. Additionally, we use $d = d^P + d^Q$ to represent the total demand of cars upstream of the diverge. We use $f^P = \frac{d^P}{d}$ and $f^Q = \frac{d^Q}{d}$ to represent the fraction of cars whose destinations are respectively links P and Q. We describe our model in terms of these normalized demands rather than the actual demands since it simplifies our analysis. Consequently, we define $F = \{f^P, f^Q\}$ to be the normalized demand configuration for the diverge. For

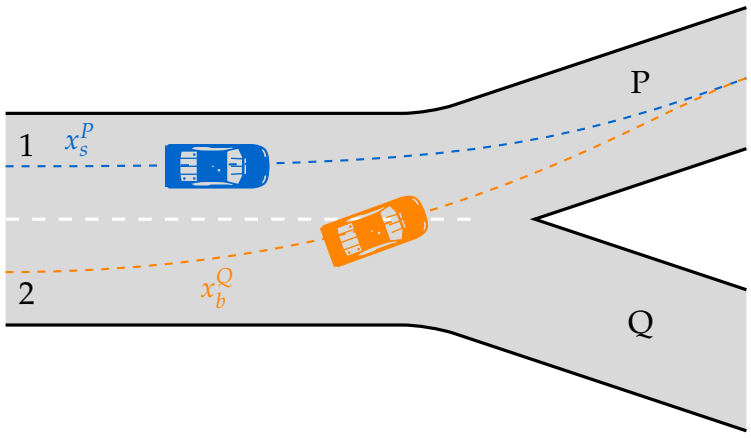


Figure 8.1: Example of a traffic diverge with two destination links P and Q. For exit link P, a steadfast car (blue car) constituting x_s^P and a bypassing car (orange car) forming x_b^P are shown.

each exit link $i \in I$, we define x_s^i to be the fraction of "steadfast" cars taking exit i among all incoming cars. These are the cars that take the lanes that correspond to their destination link i far upstream of the diverge and remain on their lanes. We also define x_b^i to be the fraction of "bypassing" cars that choose to perform a lane change maneuver to the lane that corresponds to their exit link i at a close proximity to the diverge. Figure 8.1 illustrates steadfast and bypassing cars that wish to take exit link P. We assume that cars change their lanes only once, i.e. if a car is in its target lane, it remains there. We let $\mathbf{x} = (x_s^i, x_b^i : i \in I)$ be the vector of steadfast and bypassing normalized flows for the two possible destinations of a fork. A normalized flow vector $\mathbf{x}$ is feasible for a given normalized demand configuration F if it satisfies

$$f^i = x_s^i + x_b^i, \quad \forall i \in I, \tag{8.1}$$

$$x_s^i \geq 0, \; x_b^i \geq 0, \quad \forall i \in I. \tag{8.2}$$

Example 7. *Consider the diverge shown in Figure 8.1. In this example, there are two upstream freeway lanes 1 and 2 which respectively connect to exit links P and Q. For this diverge, x_s^P is the fraction of steadfast cars that remain on lane 1 and take exit link P, whereas x_b^P is the fraction of bypassing cars that move along lane 2 and change to lane 1 at a close proximity to the diverge.*

For each exit link $i \in I$, we assume that all steadfast cars constituting x_s^i experience the same travel cost. Likewise, all bypassing cars taking an exit link i, experience the same travel cost. For each destination link $i \in I$, we let J_s^i and J_b^i be the costs incurred by the cars belonging to x_s^i and x_b^i respectively. For each pair of exit links $i \neq j \in I$, we model the cost per unit of flow of the steadfast cars by

$$J_s^i(\mathbf{x}) = C_t^i \left(x_s^i + x_b^j \right) + C_c^i x_b^i \left(x_s^i + x_b^j \right), \tag{8.3}$$

where C_i^t and C_i^c are positive constants. The constant C_i^t is associated with cost of traversing the lane that is connected to exit i. Since $(x_s^i + x_b^j)$ is the total fraction of cars that travel along the lane associated with exit i, $(x_s^i + x_b^j)$ is multiplied by C_i^t (e.g. a total of $x_s^P + x_b^Q$ traverse link 1 in Figure 8.1). Therefore, the first term in Equation (8.3) indicates that the more occupied the lanes that correspond to an exit are, the more expensive their traversal is due to the induced congestion. On the other hand, the constant C_c^i is used to reflect the negative cross effects caused by the lane change maneuvers of the bypassing fraction of cars x_b^i. This term is used to mimic the fact that as the cars in x_b^i bypass and change their lanes to take the exit i, they use the roads (resources) that join exit i; thus, they will create delays for the cars that are already in those lanes. Note that since x_s^i and x_b^j both share the target link of x_b^i up to the vicinity of the diverge, the total fraction of the cars present in the target lanes of x_b^i is $(x_s^i + x_b^j)$. Hence, C_c^i is multiplied by $(x_s^i + x_b^j)$ and x_b^i. This multiplication implies that the higher the number of cars that bypass x_b^i is, or, the more occupied the lanes that join exit i are, the larger the incurred cost by the steadfast cars will be. Note that $J_s^i(\mathbf{x})$ depends not only on x_s^i and x_b^i but also can depend on x_s^j and x_b^j.

Now, we describe how we model the costs incurred by the bypassing cars. For each pair of exit link $i \neq j \in I$, we model J_b^i via

$$J_b^i(\mathbf{x}) = C_t^j \left(x_s^j + x_b^i\right) + C_c^j x_b^j \left(x_s^j + x_b^i\right) + \gamma^i x_b^i, \tag{8.4}$$

where γ^i is a positive constant, and C_j^t and C_j^c are as previously defined. If $\gamma_i = 0$, the cost function (8.4) will be similar to (8.3) for exit j, and J_b^i would simply be equal to the cost of traversing the lanes that connect to exit j. However, $\gamma^i > 0$ models the additional discomfort cost that the bypassing cars must pay due to traversing a longer path for joining their appropriate exit and cutting in front of the cars that are already in their target lane. For a traffic diverge with two exit links, we let $\mathbf{C} = (C_t^i, C_c^i, \gamma^i : i \in I)$ be the vector of cost coefficients in our model.

Before proceeding, we need to introduce the following definition.

Definition 6. *A function $h(.) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called elementwise monotonically increasing if and only if for every $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ such that $\mathbf{x} \leq \mathbf{x}'$, where inequalities are interpreted elementwise, we have*

$$h(\mathbf{x}) \leq h(\mathbf{x}').$$

Using Equations (8.3) and (8.4), the following remark is evident.

Remark 5. For each exit link $i \in I$, the cost functions J_s^i and J_b^i, are elementwise monotonically increasing in the sense of Definition 6.

We will later use Remark 5 to guarantee certain properties of our model.

A reasonable and realistic assumption is that cars act *selfishly*, i.e. each car acts in a manner that minimizes its own cost. We now assume that each car has two options: either to choose its appropriate lane upstream of the diverge or to perform a bypass and take its target exit at a close proximity to the diverge executing a "tight" lane change. Intuitively, if for an exit link $i \in I$, $x_i^s > 0$ while $x_i^b = 0$, then cars did not have any incentive to perform a bypass, i.e. bypassing could not lead to savings in travel time i.e. $J_b^i(\mathbf{x}) > J_s^i(\mathbf{x})$. Likewise, if $x_i^s = 0$ while $x_i^b > 0$, cars did not have any incentive for acting steadfastly, i.e. $J_s^i(\mathbf{x}) > J_b^i(\mathbf{x})$. If both x_s^i and x_b^i are nonzero, we must have $J_s^i(\mathbf{x}) = J_b^i(\mathbf{x})$ since otherwise, cars will change their choice to the one with lower cost. These conditions are in fact the equilibrium conditions. Therefore, we propose to model the lane choice of cars at a traffic diverge as an equilibrium. These conditions are in fact similar to Wardrop conditions [109] in the transportation literature. In order to describe the formal definition of Wardrop conditions, let $G = (I, F, \mathbf{C})$ be a tuple enclosing the index set of exit links I, the demand configuration F, and the vector of cost coefficients $\mathbf{C}$. Then, we model the lane choice of cars as an equilibrium defined via the following.

Definition 7. *For a given $G = (I, F, \mathbf{C})$, a flow vector $\mathbf{x}$ is a Wardrop equilibrium if and only if for every exit link $i \in I$, we have:*

$$x_s^i(J_s^i(\mathbf{x}) - J_b^i(\mathbf{x})) \leq 0, \tag{8.5a}$$

$$x_b^i(J_b^i(\mathbf{x}) - J_s^i(\mathbf{x})) \leq 0. \tag{8.5b}$$

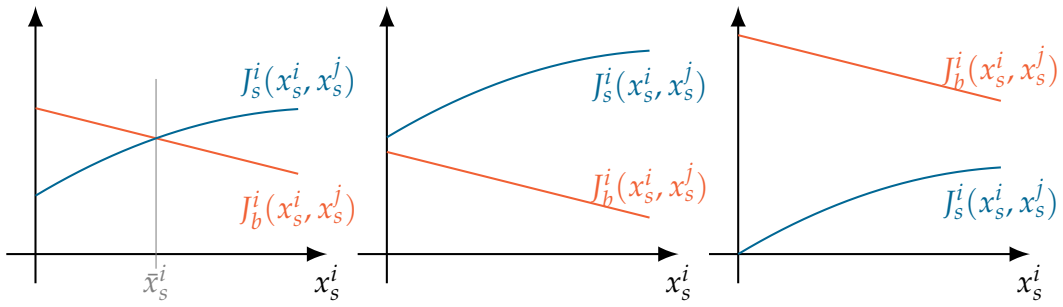
Note that Equations (8.5) imply that for an exit link $i \in I$, if $x_s^i \neq 0$ and $x_b^i \neq 0$, then at equilibrium, we must have $J_s^i(\mathbf{x}) = J_b^i(\mathbf{x})$. Alternatively, if at equilibrium $x_s^i = 0$ ($x_b^i = 0$) , we have $J_s^i(\mathbf{x}) \geq J_b^i(\mathbf{x})$ $\left(J_b^i(\mathbf{x}) \geq J_i^s(\mathbf{x})\right)$. These conditions are indeed the intuitive conditions we may expect from cars' lane choice. Note that the adoption of a Wardrop assumption implies that cars can be treated infinitesimally, i.e. the change caused by the unilateral lane change of a single car is negligible. This is in accordance with our goal of modeling the macroscopic behavior of cars at diverges

8.2 Equilibrium Properties

In this section, we state the properties of the equilibrium in our model including its existence and uniqueness.

Equilibrium Existence

Using the existence theorem in [12] for the setting of our model, we can conclude the following



(a) Case A (b) Case B (c) Case C

Figure 8.2: Three possible configurations of $J_s^i(.)$ and $J_b^i(.)$.

Proposition 7. *For a given $G = (I, F, \mathbf{C})$, if for every exit link $i \in I$, the cost functions $J_s^i(\mathbf{x})$, $J_b^i(\mathbf{x})$ are continuous and elementwise monotone in $\mathbf{x}$, then, there exists at least one Wardrop equilibrium for G.*

Remark 6. For a diverge with two exit links, using Remark 5 and continuity of $J_s^i(.)$ and $J_b^i(.)$, we can conclude that there always exists at least one equilibrium for every $G = (I, F, \mathbf{C})$.

Equilibrium Uniqueness

Once the existence of equilibrium is established, it is important to study its uniqueness. Equilibrium uniqueness is a desired and favorable property for a model to have predictive power. When the equilibrium is unique, the properties of the system at equilibrium are well defined. For instance, the overall cost at equilibrium is well defined. In this subsection, we derive the conditions under which our model has this favorable property. Equations (8.3) and (8.4) indicate that $J_s^i(.)$ depends not only on x_s^i but also on x_b^i and x_b^j. In the routing games literature, this dependence is referred to as the cost functions being *nonseparable* [20]. This nonseparability is further asymmetric, meaning that the incurred costs that are resulted from the interaction of different flow types are not the same across vehicle costs. In addition to the cost functions' asymmetric nonseparability, they are nonlinear as indicated by Equations (8.3) and (8.4). For such class of cost functions, it is known in the literature of routing games that, in general, equilibria are not unique. Equilibrium uniqueness is generally only achieved under very strong assumptions (such as strict monotonicity of the cost functions which is much stricter than elementwise montonocity), which do not hold in our model's setting [5]. Despite these complications, and the fact that none of the existing results in the literature on sufficient conditions for the uniqueness of an equilibrium can be applied to our model, we are able

to obtain the conditions under which a given $G = (I, F, \mathbf{C})$ is guaranteed to have a unique equilibrium in our model.

To prove uniqueness, we first define an auxiliary game such that there exists a connection between the Wardrop equilibrium in our model and the Nash equilibrium of the auxiliary game. For any given $G = (I, F, \mathbf{C})$, we define a two player game $\tilde{G} = \langle V, A, (\tilde{J}_v : v \in V)\rangle$, where $V = \{1, 2\}$ is the index set of players. Since both I and V are the set $\{1, 2\}$, we use a bijective correspondence between every $v \in V$ and $i \in I$. In fact, $v = 1$ ($v = 2$) implies that $i = 1$ ($i = 2$) and vice versa. Therefore, every exit link $i \in I$ is associated with a player $v \in V$. In the defined auxiliary game, $A = A_1 \times A_2$ is the action space of players, where for each $v \in V$, $A_v = [0, f^v] = [0, f^i]$ is the action set of player v. Moreover, for each player $v \in V$, $\tilde{J}_v$ is the cost associated with player v. We let $\mathbf{y} = (y_v, v \in V)$ be the vector of actions taken by the two players of the game $\tilde{G}$. To make a connection between our traffic diverge setting and the defined auxiliary game, for a given flow vector $\mathbf{x}$, we define $\mathbf{y}$ to be

$$\mathbf{y} := (x_s^i, i \in I). \tag{8.6}$$

Note that given the normalized demand configuration $F = \{f_1, f_2\}$, flow conservation requires that for every exit link $i \in I$, we have $x_b^i = f^i - x_s^i$. Thus, with a little abuse of notation, $J_s^i(\mathbf{x})$ and $J_b^i(\mathbf{x})$ can be written as $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ for every pair of exit links $i \neq j \in I$. Then, using (8.6), for every $v \neq v' \in V$, we define $\tilde{J}_v(\mathbf{y})$ to be

$$\tilde{J}_v(\mathbf{y}) = \Big(J_s^i(y_v, y_{v'}) - J_b^i((y_v, y_{v'})\Big)^2. \tag{8.7}$$

In the auxiliary game $\tilde{G}$, a vector $\mathbf{y}$ is a pure Nash equilibrium if and only if for every two players $v \neq v' \in V$, we have

$$y_v = B_v(y_{v'}) \tag{8.8}$$
$$= \text{argmin}_{y_v \in [0, f^v]} \tilde{J}_v(y_v, y_{v'}) \tag{8.9}$$

where B_v is the best response function of player v. Note that since for every player $v \in V$, $\tilde{J}(\mathbf{y})$ is a continuous function on a closed interval, a minimum is achieved. Equation (8.8) implies that if $y_{v'}$ is fixed, player v takes the best possible action that minimizes its own cost $\tilde{J}_v(\mathbf{y})$. The following proposition establishes the connection between the Wardrop equilibrium of G and the Nash equilibrium of $\tilde{G}$.

Proposition 8. *A flow vector $\mathbf{x} = (x_s^i, x_b^i : i \in I)$ is a Wardrop equilibrium for $G = (I, F, \mathbf{C})$ if and only if $\mathbf{y} = (x_s^i, i \in I)$ is a pure Nash equilibrium for $\tilde{G}$ provided that*

$$C_t^i \geq C_c^i, \ \forall i \in I. \tag{8.10}$$

Proof. First we show that for a given x_s^j, (8.10) is a sufficient condition for $J_s^i(x_s^i, x_s^j)$ to be increasing in x_s^i, and $J_b^i(x_s^i, x_s^j)$ to be decreasing in x_s^i . To see this, note that for every $i \neq j \in I$, we have:

$$\frac{\partial J_s^i(x_s^i, x_s^j)}{\partial x_s^i} = -2C_c^i x_s^i + C_t^i + C_c^i x_s^i - C_c^i(f^j - x_s^j). \tag{8.11}$$

Equation (8.11) is linear in x_s^i. Moreover, for each $i \in I$, x_s^i is allowed to only take values in interval $[0, f^i]$. Therefore, in order to obtain sufficient conditions for the positivity of (8.11), it is sufficient to guarantee that $\frac{\partial J_s^i(x_s^i, x_s^j)}{\partial x_s^i}$ is positive at all possible extreme points (x_s^i, x_s^j) which are $\{(0,0), (f^1, 0), (0, f^2), (f^1, f^2)\}$. Using the fact that the demand fractions must satisfy $f^1 + f^2 = 1$, it is easy to verify that the smallest possible value of (8.11) is attained in $(f^1, 0)$ when $f^1 = 1$. At the point $(1, 0)$, we have $\frac{\partial J_s^i}{\partial x_s^i}(1, 0) = C_t^i - C_c^i$. Therefore, (8.10) is a sufficient condition for $J_s^i(x_s^i, x_s^j)$ to be increasing in x_s^i. Similarly, we can compute $\frac{\partial J_b^i(x_s^i, x_s^j)}{\partial x_s^i}$ which is

$$\frac{\partial J_b^i(x_s^i, x_s^j)}{\partial x_s^i} = -C_t^j - \gamma^j - C_c^j(f^j - x_s^j). \tag{8.12}$$

Since $(f^j - x_s^j)$ is always greater than or equal to zero, and C_t^j and γ^j are always positive, clearly, for every $i \neq j \in I$, $J_b^i(x_s^i, x_s^j)$ is always decreasing in x_s^i for any given x_s^j.

Now, we can proceed to proving that under (8.10), every Wardrop equilibrium of G is equivalent to a Nash equilibrium of the auxiliary game $\tilde{G}$. For a player $v \in V$, consider the best response function $B_v(y_{v'})$ in (8.8). For a given $y_{v'} = x_s^j$, in order to minimize $\tilde{J}_v(y_v, y_{v'})$ over $y_v = x_s^i$, since J_s^i is increasing in x_s^i, and J_b^i is decreasing in x_s^i under (8.10), the following three scenarios may occur for a given x_s^j (see Figure 8.2):

- Case A: $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ have an intersection on the interval $(0, f^v)$. In this case, there exists a point $\bar{x}_s^i(x_s^j) \in (0, f^v)$ such that $J_s^i(\bar{x}_s^i, x_s^j) = J_b^i(\bar{x}_s^i, x_s^j)$ (See Figure 8.2, case A). Using (8.7), it can also be verified that in this case, the intersection point $y_v = \bar{x}_s^i$ is the best response for a given $y_{v'} = x_s^j$. If this is the case, Equations (8.5) are also satisfied by $\bar{x}_s^i$ for a given x_s^j. It is easy to see that the reverse is also true. Indeed, if $\bar{x}_s^i \in (0, f^i)$ satisfies (8.5) for a given x_s^j, then, $\bar{x}_s^i$ must be the intersection of $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ on the interval $(0, f^v)$. Therefore, $y_v = x_s^i$ is the best response of $y_{v'} = x_s^j$.

- Case B: $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ do not intersect on the interval $(0, f^v)$, and $J_s^i(0, x_s^j) \geq J_b^i(0, x_s^j)$ for a given x_s^j. In this case, if $y_{v'} = x_s^j$, for minimizing $\tilde{J}_v$, we must have $y_v = B_v(y_{v'}) = 0$ (See Figure 8.2, case B). Using the connection between the two games and (8.6), it is easy to see that, $x_s^i = 0$ satisfies (8.5) for a given x_s^j since if $x_s^i = 0$, then, $x_b^i = f^i$ while $J_s^i \geq J_b^i$. The reverse is also true, if $x_s^i = 0$ satisfies (8.5) for a given x_s^j, then $y_v = x_s^i = 0$ is the best response of $y_{v'} = x_s^j$.

- Case C: $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ do not intersect on the interval $(0, f^v)$, and $J_s^i(0, x_s^j) \leq J_b^i(0, x_s^j)$. In this case, if $y_{v'} = x_s^j$, then $y_v = B_v(y_{v'}) = 1$. Similar to case B, one can conclude that $y_v = x_s^i = 1$ is equal to $B_v(y_{v'})$ if and only if $x_s^i = 1$ satisfies (8.5) for a given x_s^j.

So far, we have shown that for every $v \neq v' \in V$, for a given $y_{v'}$, y_v is the best response of $y_{v'}$ if and only if $\mathbf{x} = (y_v, f^v - y_v : i \in I)$ satisfies (8.5). Therefore, $\mathbf{y} = (x_s^i : i \in I)$ is a Nash equilibrium of $\tilde{G}$ if and only if $\mathbf{x} = (x_s^i, f^i - x_s^i : i \in I)$ is a Wardrop equilibrium of G. □

Remark 7. Notice that using the three cases described in the proof of Proposition 8, for a given $y_{v'} = x_s^j$, the best response $B_v(y_{v'})$ can be found by first intersecting $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ and then projecting the intersection point $\bar{x}_s^i(x_s^j)$ onto the interval $[0, f^i]$. We will use this fact in the remainder to prove equilibrium uniqueness.

Having Proposition 8 in mind, we are ready to state and prove the following.

Theorem 4. *For a given diverge $G = (I, F, \mathbf{C})$, a Wardrop equilibrium flow vector $\mathbf{x}$ is unique if for every exit link $i \in I$,*

$$C_t^i \geq C_c^i, \tag{8.13}$$
$$\gamma^i \geq C_c^i. \tag{8.14}$$

Proof. Construct the auxiliary game $\tilde{G} = \langle V, A, (\tilde{J}_v, v \in V) \rangle$ described above. From Proposition 8, we know that if (8.13) holds, $\mathbf{x}$ is a Wardrop equilibrium for G if and only if $(y_v : v \in V) = (x_s^i : i \in I)$ is a Nash equilibrium for $\tilde{G}$. We now prove that under (8.14), $\tilde{G}$ has a unique equilibrium; thus, G must also have a unique equilibrium. To see this, note that $\mathbf{y} = (x_s^i : i \in I)$ is a Nash equilibrium for $\tilde{G}$ if and only if for every $v \neq v'$, $y_v = B_v(y_{v'})$, and $y_{v'} = B_{v'}(y_v)$. These conditions can be rewritten as

$$y_v = B_v(B_{v'}(y_v)), \tag{8.15a}$$
$$y_{v'} = B_{v'}(B_v(y_{v'})). \tag{8.15b}$$

Equations (8.15) indicate that $\mathbf{y}$ is an equilibrium if and only if for every $v \neq v' \in V$, y_v is a fixed point for $B_v\left(B_{v'}(.)\right)$. Thereby, $(y_v, y_{v'})$ is an equilibrium for $\tilde{G}$ if and only if $B_v(B_{v'}(.))$ intersects the line going through the origin with slope 1 at y_v, and $B_{v'}(B_v(.))$ intersects the line going through the origin with slope 1 at $y_{v'}$. In the remainder, we prove that under (8.14), the slope of $B_v(B_{v'}(.))$ is always positive and smaller than 1 for every $v \neq v' \in V$. Therefore, $B_v(B_{v'}(.))$ can intersect the identity line at most once. Thus, knowing that there exits at least one equilibrium, we can conclude that $\tilde{G}$ and therefore G always has a unique equilibrium if (8.13) and (8.14) hold. To prove this, it suffices to show that $0 \leq \frac{dB_v(y_{v'})}{dy_{v'}} \leq 1$, for every $v \neq v' \in V$. To see this, let x_s^j be such that $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ intersect each other at $\bar{x}_s^i(x_s^j) \in [0, f^i]$. Using (8.3) and the fact that for every $i \in I$, $x_b^i = f^i - x_s^i$, we have

$$\frac{\partial J_s^i(x_s^i, x_s^j)}{\partial x_s^j} = -C_t^i - C_c^i(f^i - x_s^i). \tag{8.16}$$

Since $(f^i - x_s^i) \geq 0$, we can conclude that $\frac{\partial J_s^i(x_s^i, x_s^j)}{\partial x_s^j} \leq 0$. Similarly, we can compute

$$\frac{\partial J_b^i(x_s^i, x_s^j)}{\partial x_s^j} = C_t^j + C_c^j(f^j - x_s^j) - C_c^j(x_s^j + f^i - x_s^i). \tag{8.17}$$

Since $(f^j - x_s^j) \geq 0$, it is easy to see that if (8.13) holds, $\frac{\partial J_b^i(x_s^i, x_s^j)}{\partial x_s^j}$ is always positive. Thus, (8.16) and (8.17) imply that as x_s^j increases, J_s^i decreases while J_b^i increases. Therefore, as x_s^j increases, $\bar{x}_s^i(x_s^j)$ can only increase. However, Remark 7 implies that if $\bar{x}_s^i(x_s^j)$ lies outside the interval $[0, f^i]$, it is projected on this interval. Since x_s^j varies on interval $[0, f^j]$, interval $[0, f^j]$ can be divided into $[0, f^j] = [0, m^j] \cup [m^j, n^j] \cup [n^j, f^j]$, such that $\bar{x}_s^i(x_s^j)$ is always 0 for $x_s^j \in [0, m^j]$, and always 1 for $x_s^j \in [n^j, f^j]$. Note that either of the intervals $[0, m^j]$, $[m^j, n^j]$ and $[n^j, f^j]$ can possibly be empty. Hence, in order to show that the slope of the best response function $B_v(x_s^j)$ is always smaller than 1, it suffices to prove that it is indeed less than 1 for x_s^j being in interval $[m^j, n^j]$ where $J_s^i(x_s^i, x_s^j)$ and $J_b^i(x_s^i, x_s^j)$ do intersect at $\bar{x}_s^i \in [0, f^i]$.

For a given $x_s^j \in [m^j, n^j]$, $\bar{x}_s^i(x_s^j)$ must satisfy

$$J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j) = 0.$$

Therefore, using implicit differentiation, $\frac{d\bar{x}_s^i(x_s^j)}{dx_s^j}$ can be computed via

$$\frac{\partial}{\partial x_s^i}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right)\frac{d\bar{x}_s^i(x_s^j)}{dx_s^j} + \frac{\partial}{\partial x_s^j}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right) = 0. \tag{8.18}$$

Using (8.3) and (8.4) , and the fact that $x_b^i = f^i - x_s^i$ for all exit links $i \in I$, we have

$$\frac{\partial}{\partial x_s^j}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right) = -C_t^i - C_c^i(f^i - x_s^i) - C_c^i + C_c^j(x_s^j + f^i - x_s^i) - C_c^j(f^j - x_s^j). \tag{8.19}$$

Since (8.19) is linear in x_s^i and x_s^j, its maximum and minimum are attained at its extreme points. It is easy to check that the maximum possible value for (8.19) is $-C_t^i - C_t^j + C_c^j$. If (8.13) holds, $-C_t^i - C_t^j + C_c^j \leq 0$. Therefore $\frac{\partial}{\partial x_s^j}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right) \leq 0$ under (8.13). Using the same approach, one can verify that under (8.13), it always the case that for every exit link $i \in I$,

$$\frac{\partial}{\partial x_s^i}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right) \geq 0.$$

Hence, using (8.18), under (8.13),

$$\frac{d\bar{x}_s^i(x_s^j)}{dx_s^j} \geq 0, \quad \forall i \neq j \in I.$$

Now that we have shown that the slope of the best response function is always positive, it only remains to prove that $\frac{d\bar{x}_s^i(x_s^j)}{dx_s^j} \leq 1$. To prove this, it suffices to show that

$$\frac{\partial}{\partial x_s^i}(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)) \geq -\left(\frac{\partial}{\partial x_s^j}\left(J_s^i(\bar{x}_s^i, x_s^j) - J_b^i(\bar{x}_s^i, x_s^j)\right)\right). \tag{8.20}$$

Substituting (8.3), (8.4), and (8.19) in (8.20) and computing the linear function at its extreme points, we observe that (8.14) is a sufficient condition for (8.20), which completes our proof. □

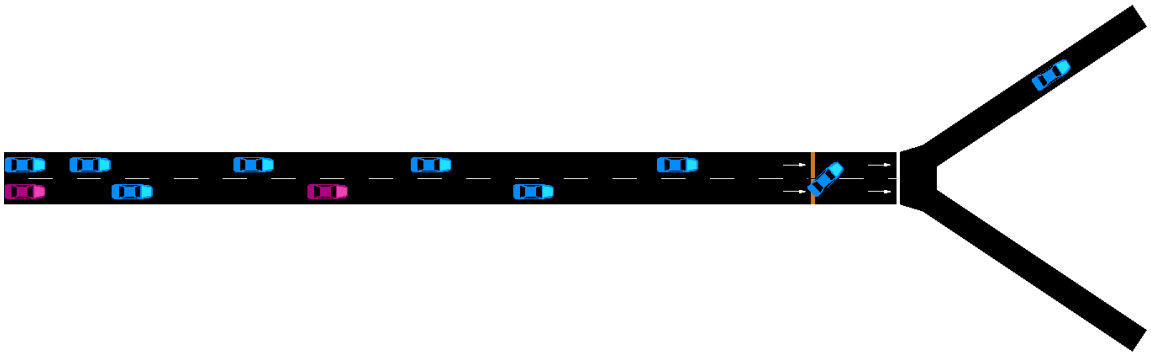


Figure 8.3: The traffic diverge used in our simulations: the blue cars will take the upper (left) exit link; while the magenta cars will ultimately take the lower (right) exit. Using Krauss' car following model, a blue vehicle is bypassing right before the bifurcation.

8.3 Model Validation

Up to now, we have described our model and its properties. In this section, we describe how our model performs in predicting the bypassing behavior of cars. We use a micro-scale traffic simulation software, SUMO [48] to generate the data required for calibrating and validating our proposed macroscopic model. By default, SUMO employs the Krauss model [49] as the car–following model and the LC2013 model [26] as the lane–changing model. Krauss's model is widely used for simulating human drivers' behavior, and it enables cars to run at their highest speed as long as safety is guaranteed, i.e., Krauss's model is collision–free. A key parameter in Krauss's model is *sigma* which characterizes the randomness of cars' behavior. Parameter sigma ranges from 0 to 1, with 0 indicating a uniform behavior. In our simulation, sigma is set by default to be 0.5 so that randomness in drivers' behavior is captured. In the LC2013 lane–changing model, there is a parameter known as lcSpeedGain, which ranges from 0 to the infinity. This parameter is an indicator of the cars' eagerness to perform lane–changing behavior for speed gains. A higher lcSpeedGain implies that we should expect more lane–changing maneuvers by more aggressive drivers. In our simulation, we set lcSpeedGain to be 10. Moreover, in the simulation, we build a sufficiently long upstream link for the diverge and we only record steadfast and bypassing cars immediately upstream of the road bifurcation and provide sufficient simulation time to SUMO to achieve steady state. An overview of the diverge in SUMO is shown in Figure 8.3.

A key element of our model, which affects its functionality, is the coefficient vector $\mathbf{C}$. Therefore, in order to study the performance of the model, it needs to be calibrated first, i.e. it is necessary to determine, using microscopic simulation data, the coefficient vector $\mathbf{C}$ that best fits the resulting microscopic traffic simulation data at the diverge.

Model Calibration

Consider a diverge with two exit links $I = \{P, Q\}$. Fix the total flow of cars $d = d^P + d^Q$ that enter the diverge. For a given fixed d, assume that there exist measurements from K possible allocations of demand d among the two exit links. Let $(\tilde{F})_k = \{(\tilde{f}^P)_k, (\tilde{f}^Q)_k\}, 1 \leq k \leq K$, be the demand configuration that corresponds to the k_{th} observed measurement, where $(\tilde{f}^P)_k$ and $(\tilde{f}^Q)_k$ denote the fractions of demand that took exits P and Q in the k_{th} observation respectively. For each measurement k, let $(\tilde{x}_s^i)_k$ and $(\tilde{x}_b^i)_k$ be the measured fraction of steadfast and bypassing flows taking exit link $i \in I$ when k^{th} demand configuration was used. We let $(\tilde{\mathbf{x}})_k$ represent the vector of measured flows for the k^{th} measurement. When using our model, the vector of cost coefficients $\mathbf{C}$ must be determined such that Equations (8.5) are satisfied by $(\tilde{x}_s^i)_k$ and $(\tilde{x}_b^i)_k$ for every $k \leq K$. But, since (8.5) contains nonlinear inequalities, finding such a $\mathbf{C}$ is nontrivial. We propose the following calibration process for the vector of cost parameters $\mathbf{C}$.

For every $k \leq K$ and $i \in I$, define the integer variables $(z_s^i)^k \in \{0, 1\}$, and $(z_b^i)^k \in \{0, 1\}$ such that

$$(\tilde{x}_s^i)_k (J_s^i((\tilde{\mathbf{x}})_k) - J_b^i((\tilde{\mathbf{x}})_k)) \leq 0 \iff (z_s^i)_k = 0 \tag{8.21a}$$
$$(\tilde{x}_s^i)_k (J_s^i((\tilde{\mathbf{x}})_k) - J_b^i((\tilde{\mathbf{x}})_k)) > 0 \iff (z_s^i)_k = 1 \tag{8.21b}$$
$$(\tilde{x}_b^i)_k (J_b^i((\tilde{\mathbf{x}})_k) - J_s^i((\tilde{\mathbf{x}})_k)) \leq 0 \iff (z_b^i)_k = 0 \tag{8.21c}$$
$$(\tilde{x}_b^i)_k (J_b^i((\tilde{\mathbf{x}})_k) - J_s^i((\tilde{\mathbf{x}})_k)) > 0 \iff (z_b^i)_k = 1 \tag{8.21d}$$

Then, letting $\mathbf{z}$ be the vector of $(z_s^i)^k$ and $(z_b^i)^k$ for all k's and all exit links $i \in I$, we propose to solve the following optimization problem for calibrating $\mathbf{C}$.

$$\begin{aligned} \underset{\mathbf{C},\mathbf{z}}{\text{minimize}} \quad & \sum_{k \in K} \sum_{i \in I} \left((z_s^i)_k + (z_b^i)_k \right) \\ \text{subject to} \quad & \text{Equations (8.21)} \\ & \mathbf{C}_r \geq 1, \end{aligned} \tag{8.22}$$

where $\mathbf{C}_r$ is the r_{th} element of $\mathbf{C}$. We use the constraint $\mathbf{C}_r \geq 1$ to prevent the optimizer from setting all the elements of $\mathbf{C}$ to be zero. It is important to note that since every term in (8.3) and (8.4), is multiplied by one and only one element of $\mathbf{C}$, and multiplying all cost functions by the same constant does not change the Wardrop conditions, scaling $\mathbf{C}$ by a single number will not affect the model. Therefore, this constraint does not affect the model. Note that for every inequality constraint that is violated in (8.22), the cost is increased by 1. Thus, (8.22) provides a penalty for not satisfying (8.5) which are the equilibrium conditions. But, how can the optimization problem (8.22) be solved where the constraints are of the form (8.21)? To answer this, we use the procedure introduced

in [91]. Let M be a large positive number, and ϵ be a small positive number close to zero. For every k, the following is equivalent to (8.21).

$$(\tilde{x}_s^i)_k(J_s^i((\tilde{\mathbf{x}})_k) - J_b^i((\tilde{\mathbf{x}})_k)) \leq M(z_s^i)_k - \epsilon, \tag{8.23a}$$
$$-(\tilde{x}_s^i)_k(J_s^i((\tilde{\mathbf{x}})_k) - J_b^i((\tilde{\mathbf{x}})_k)) \leq M(1-(z_s^i)_k) - \epsilon, \tag{8.23b}$$
$$(\tilde{x}_b^i)_k(J_b^i((\tilde{\mathbf{x}})_k) - J_s^i((\tilde{\mathbf{x}})_k)) \leq M(z_b^i)_k - \epsilon, \tag{8.23c}$$
$$-(\tilde{x}_b^i)_k(J_b^i((\tilde{\mathbf{x}})_k) - J_s^i((\tilde{\mathbf{x}})_k)) \leq M(1-(z_b^i)_k) - \epsilon. \tag{8.23d}$$

Therefore, our model can be calibrated by solving the following optimization problem

$$\begin{aligned} \underset{\mathbf{C}}{\text{minimize}} \quad & \sum_k \sum_{i\in I} \left((z_s^i)^k + (z_b^i)^k \right) \\ \text{subject to} \quad & \text{Equations (8.23)} \\ & \mathbf{C}_r \geq 1. \end{aligned} \tag{8.24}$$

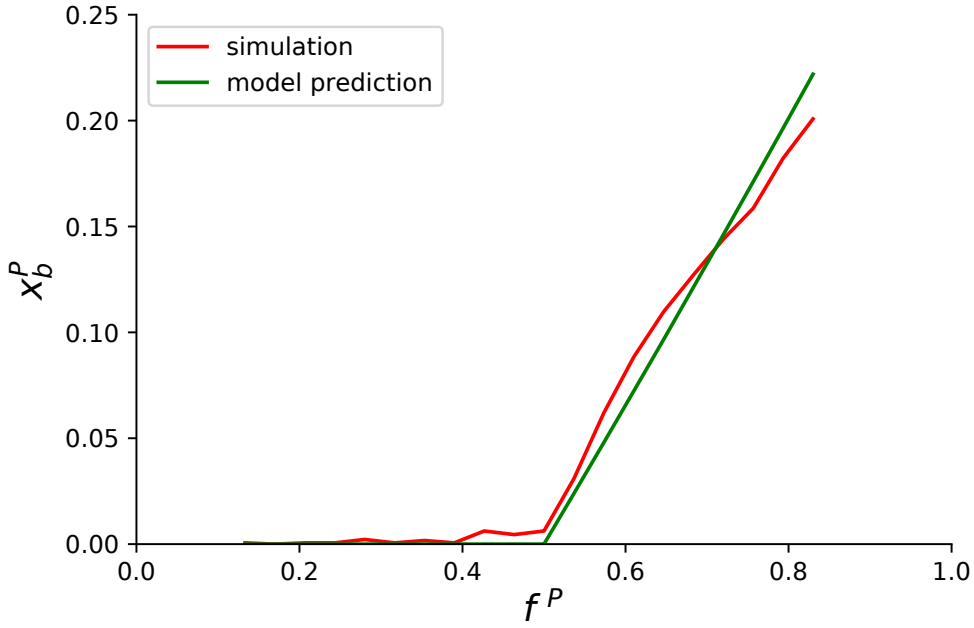
Note that (8.24) is now a mixed–integer linear program that can be easily solved using optimization packages. Since (8.24) is solved offline, and, further, the number of required integer variables is small, the computational cost for calibrating our model by solving (8.24) is not overtaxing.

Model Predictions

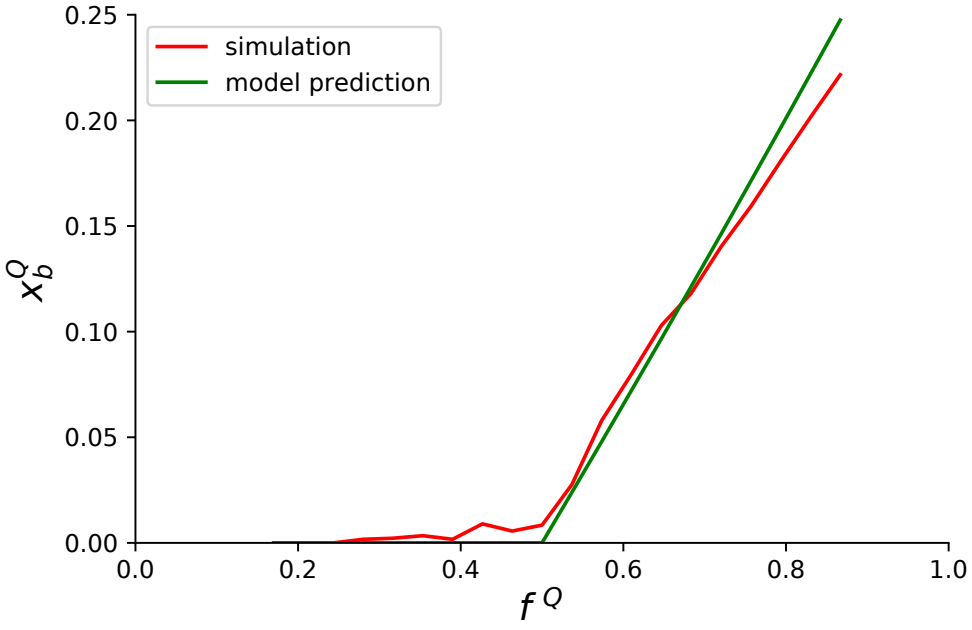
Consider the diverge shown in Figure 8.1. We used the microscopic traffic simulator SUMO [48] to simulate the traffic behavior at the diverge of Figure 8.1 for different demand configurations. A total flow of $d = 3000\frac{\text{veh}}{\text{hour}}$ enters the diverge. The capacity of each lane is $930\frac{\text{veh}}{\text{hour}}$. At every simulation, a fraction of cars f^P is assumed to take the exit link P while the remaining fraction of cars $f^Q = 1 - f^P$ is assumed to take the exit link 2. For different values of f^P, x_s^P, x_b^P, x_s^Q, and x_b^Q are measured. Then, this data set is used to calibrate the model, i.e. finding the cost parameter vector $\mathbf{C}$ that best fits the data by solving optimization problem (8.24). Since our road geometry is symmetric, we introduced the additional constraints that $C_t^P = C_t^Q$, $C_c^P = C_c^Q$, and $\gamma^1 = \gamma^2$ in (8.24), and obtained the following values for $\mathbf{C}$.

$$C_t^P = C_t^Q = 1,\ C_c^P = C_c^Q = 1,\ \gamma^P = \gamma^Q = 1.7.$$

Notice that the obtained values of $\mathbf{C}$ satisfy (8.13) and (8.14). Thus, in every scenario, Theorem 4 implies that there exits only one equilibrium. We used a total of $K = 20$ fractional demand configurations, which results in a total of 40 integer variables. The objective function of (8.24) was 4 when fitting $\mathbf{C}$, meaning that only 4 inequalities were unsatisfied among the 40 inequality constraints of our data set.



(a) x_b^P



(b) x_b^Q

Figure 8.4: The fraction of bypassing cars, x_b^i, predicted by our macroscopic model and the values measured from microscopic simulations as a function of demand fractions f^i's.

With the calibrated $\mathbf{C}$, we used our model to predict x_s^P, x_b^P, x_s^Q, and x_b^Q for the scenarios where the total flow entering the diverge is different from the value that we used to calibrate our macroscopic diverge. Figure 8.4 demonstrates such a study when the total demand to the diverge was $d = 2500\frac{\text{veh}}{\text{hour}}$. Figure 8.4 shows both the SUMO microsimulation results and our model predictions values as a function of demand configuration fractions f^i's. Note that the aggregate number of bypassing cars predicted by SUMO is obtained by running the micro-scale Krauss car–following model taking into account the randomness in drivers' behavior, while our model predictions are solely obtained by solving the equilibrium conditions using the calibrated parameter vector $\mathbf{C}$. As illustrated by Figure 8.4, our macroscopic model predicts the fraction of bypassing and steadfast cars for each destination. Notice that when the demand for exit P is low i.e. $f^P \leq 0.5$, none of the cars who aim to take exit P would take the more crowded lane 2; therefore, $x_b^P \simeq 0$. But, with the increase of f^P, some cars will take lane 2 since it will reduce their cost. This is in accordance with what we may intuitively expect, when the demand for exit P is lower than the demand for exit Q, cars taking exit P do not have any incentive for performing a bypass since lane 2 is more crowded than lane 1, but when the demand for P exceeds the demand for exit Q, bypassing behavior emerges. Our intial simulation results indicate that our model is capable of predicting the behavior of the cars. We obtained similar predictive results when the total demand that enters the diverge was varied.

8.4 Socially Optimal Lane Choice

Having shown that our model can potentially be used to predict macroscopic vehicular bypassing behavior, we can deploy it for further mathematical analysis. Intuitively, one might argue that if most cars were less selfish, and would have chosen their destination lane far upstream of the diverge, the total travel cost experienced by cars at the diverge would reduce. We now show that our model provides a powerful framework for analytically studying this conjecture. Assume that there is a central authority which can dictate the lanes that a car must travel on, such that the total cost of cars at the diverge is minimum; or equivalently, that the social optimum is achieved. For example, if all cars are autonomous, then, their lane choices could be encoded so that the sum of the costs of all cars is minimized. The total or social cost experienced by all cars can be computed using our macroscopic model as

$$J_{\text{soc}} = d\left(\sum_{i\in I}\left(x_s^i J_s^i + x_b^i J_b^i\right)\right), \tag{8.25}$$

where d is the total demand entering the diverge. Then, the lane changing profile required for attaining the minimum social cost can be determined by solving the following optimization problem

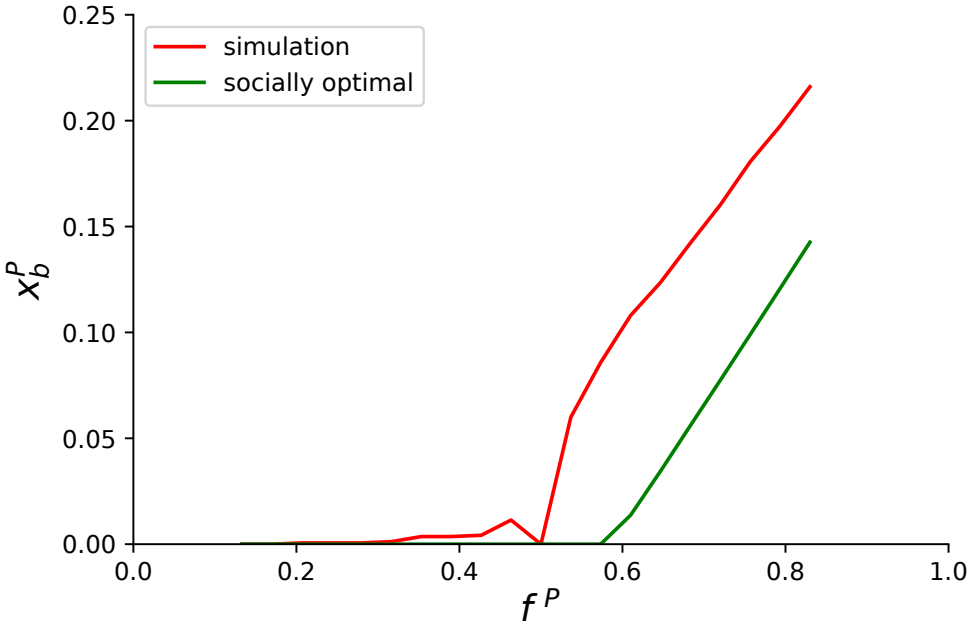
$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & J_{\text{soc}} \\ \text{subject to} \quad & x_s^i + x_b^i = f^i, \quad \forall i \in I, \\ & x_s^i \geq 0, \; x_b^i \geq 0, \quad \forall i \in I. \end{aligned} \tag{8.26}$$

Optimization (8.26) can be solved to find the optimal lane choice and bypassing behavior. Note that in (8.26), the decision variables are x_s^i, x_b^i, for every exit link $i \in I$. Thus, the objective function of (8.26) is a 3rd order polynomial in the decision variables. Optimization problem (8.26) can be easily solved using commercial solvers. This simplicity should be contrasted to the existing methods, where strategies for finding better lane choices are heuristically determined through simulation studies.

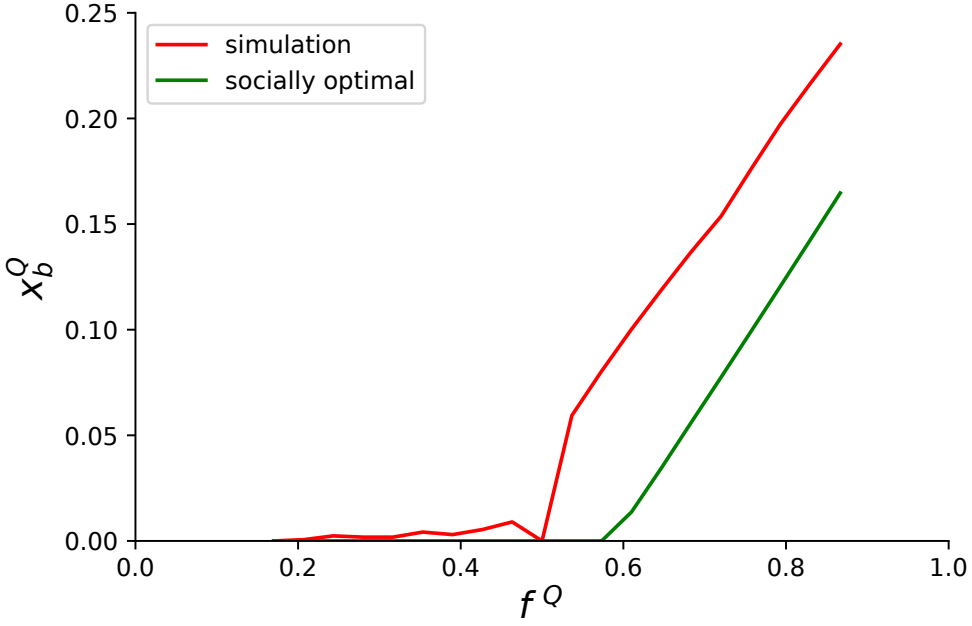
Using the calibrated vector of cost parameters vector $\mathbf{C}$ that was obtained from our model calibration, we solved (8.26) for the case when the total flow d entering the diverge was $3000\frac{\text{veh}}{\text{hour}}$. Figure 8.5 demonstrates the amount of bypassing that is required for achieving the optimal overall cost of cars versus the amount of bypassing observed in microsimulation using Krauss' model. As Figure 8.5 shows, for every fractional demand configuration, in the microsimulations, since cars select their lanes selfishly, the number of bypassing cars is larger than the number of bypassing cars required for optimality. Moreover, a key outcome of the determination of the vehicular lane change behavior that minimizes the overall or social cost in (8.28), is that bypassing maneuvers are necessary, as flow demands increase in order to achieve the minimum socially–optimal cost, as illustrated in Figure 8.5. This result is non–intuitive as it is not necessarily conform to conventional wisdom. Note that the bypassing cars gain a speed increase by their maneuver, while their maneuver incurs a cost for the steadfast cars. For optimality the trade–off between these two effects must be found. As it could be observed from Figure 8.5, when the demand for one of the exit links exceeds certain threshold, it makes sense for a proportion of the cars to perform a bypass to gain speed increases. However, due to the extra cost that these cars would incur on the steadfast cars, the amount of bypassing required for optimality is smaller than the bypassing behavior under selfish lane choices. Therefore, as illustrated in Figure 8.5, the socially optimal bypassing proportions and the resulting lane choices are often smaller than those determined by the Wardrop equilibrium, but are nonzero. Our model allows for quantitatively analyzing this trade–off, which we believe has not been previously captured in the literature.

8.5 Optimal Lane Choice in Mixed–Autonomy Setting

The mathematical macroscopic model that was derived in previous section allows traffic engineers to perform further analysis, involving efficient traffic management policies under a mixed autonomy setting. In particular, a novel and important potential use of our model becomes apparent if we assume that a central authority has control over a fraction of the cars, and is able to to dictate control actions and lane choices to these compliant



(a) x_b^P

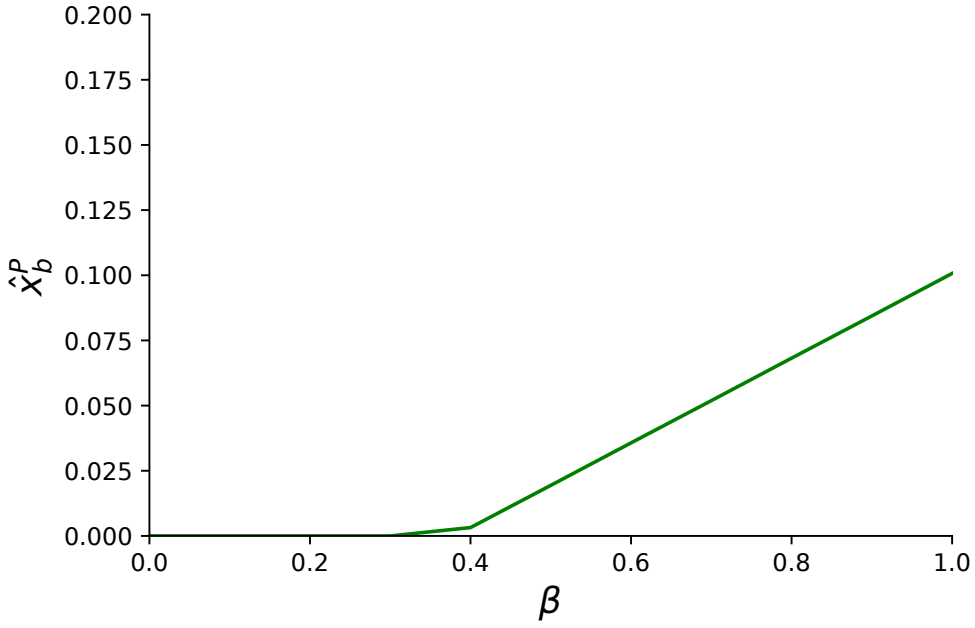


(b) x_b^Q

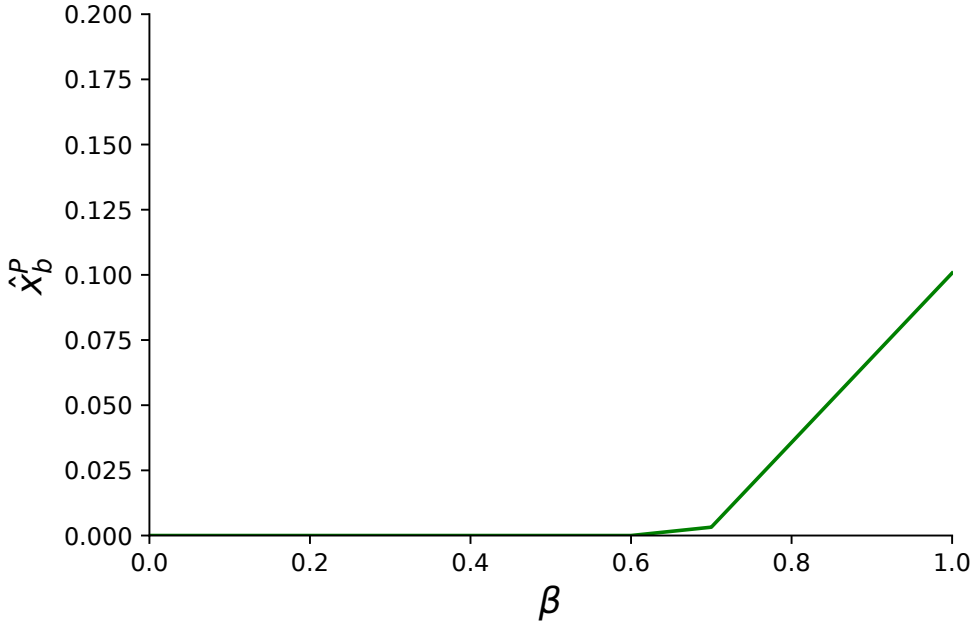
Figure 8.5: The fraction of bypassing cars, x_b^i, from simulation and the values of bypassing cars required for the social optimality as a function of f_i.

cars such that, when the noncompliant cars react to the actions of the compliant cars, the overall network performance is improved. In the context of routing games, such control mechanisms are known as Stackelberg control of cars. As autonomous cars are becoming tangible technologies, it is expected that in the near future both human–driven and autonomous cars will coexist in traffic networks. It was shown in [62] that only replacing human–driven cars by autonomous cars might not be enough for improving network mobility, and further control mechanisms are required to exploit the mobility benefits of autonomous cars. In the context of lane choice of autonomous cars in a mixed–autonomy setting, it is crucial to note that human–driven cars will respond selfishly to the actions of autonomous cars. Thus, autonomous cars must take the potential response of human drivers into account and plan their actions accordingly. Since our model predicts the lane choice of cars, it can be used for such planning purposes. Aligned with this scenario, in order to increase the mobility of traffic networks with mixed autonomy, lane choice and bypassing decisions for autonomous cars must be developed such that when human drivers act selfishly, a decrease in the overall cost of cars is achieved. In this section, we use our model to study how the lane choice behavior of autonomous cars in the mixed–autonomy setting can change the social cost at a traffic diverge with two exit links, through an example.

Consider the traffic diverge shown in Figure 8.1. Assume that our proposed model has been calibrated with the cost parameter vector $\mathbf{C}$. Fix the total demand d and the normalized demand configuration $F = \{f^P, f^Q\}$. Let α be the fraction of cars that are autonomous among all cars that wish to take exit P. We also assume in this example that all cars that take exit Q are human–driven. This implies that an αf^P percent of the total cars are autonomous. We also assume that a central authority is able to command a β fraction of autonomous cars to be steadfast cars, and the remaining $(1-\beta)$ fraction of autonomous cars to be bypassing cars i.e. $(1-\beta)$ fraction of autonomous cars are commanded to bypass at the diverge to take exit link P, and the remaining β fraction of the autonomous cars are commanded to choose lane 1 far upstream of the diverge and remain on this lane. Let $w = (1-\beta)\alpha f^P$ and $z = \beta\alpha f^P$ denote the fractions of bypassing and steadfast autonomous cars respectively with respect to the total demand of cars. For a fixed w and therefore z, the remaining human–driven cars will react selfishly to the lane choices of the autonomous cars, and a new Wardrop equilibrium will be achieved. Thus, every choice of w and z induces a new Wardrop equilibrium of human–driven cars. For each exit link $i \in I$, we use $\hat{x}_s^i$ and $\hat{x}_b^i$ to represent the fraction of steadfast and bypassing human–driven cars in the induced equilibrium. Note that in this case, for a given w and z, flow conservation requires that $\hat{x}_s^P + \hat{x}_b^P = f^P - w - z$ and $\hat{x}_s^Q + \hat{x}_b^Q = f^Q$. Let $\hat{\mathbf{x}} = (\hat{x}_s^i, \hat{x}_b^i : i \in I)$ represent the vector of human–driven flows. In this case, the cost of steadfast and bypassing cars are



(a) $\alpha = 0.25$



(b) $\alpha = 0.5$

Figure 8.6: The fraction of human–driven bypassing cars versus the fraction of autonomous cars that are commanded to remain steadfast β for different values of autonomy fraction α.

$$\hat{J}_s^P(\hat{\mathbf{x}}, w, z) = C_t^P\left(\hat{x}_s^P + \hat{x}_b^Q + z\right) + C_c^P(\hat{x}_b^P + w)\left(\hat{x}_s^P + \hat{x}_b^Q + z\right),$$
$$\hat{J}_b^P(\hat{\mathbf{x}}, w, z) = C_t^Q\left(\hat{x}_s^Q + \hat{x}_b^P + w\right) + C_c^Q\hat{x}_b^Q\left(\hat{x}_s^Q + \hat{x}_b^P + w\right) + \gamma^P(\hat{x}_b^P + w),$$
$$\hat{J}_s^Q(\hat{\mathbf{x}}, w, z) = C_t^Q\left(\hat{x}_s^Q + \hat{x}_b^P + w\right) + C_c^Q\hat{x}_b^Q\left(\hat{x}_s^Q + \hat{x}_b^P + w\right),$$
$$\hat{J}_b^Q(\hat{\mathbf{x}}, w, z) = C_t^P\left(\hat{x}_s^P + z + \hat{x}_b^Q\right) + C_c^P(\hat{x}_b^P + w)\left(\hat{x}_s^P + z + \hat{x}_b^Q\right) + \gamma^Q\hat{x}_b^Q.$$

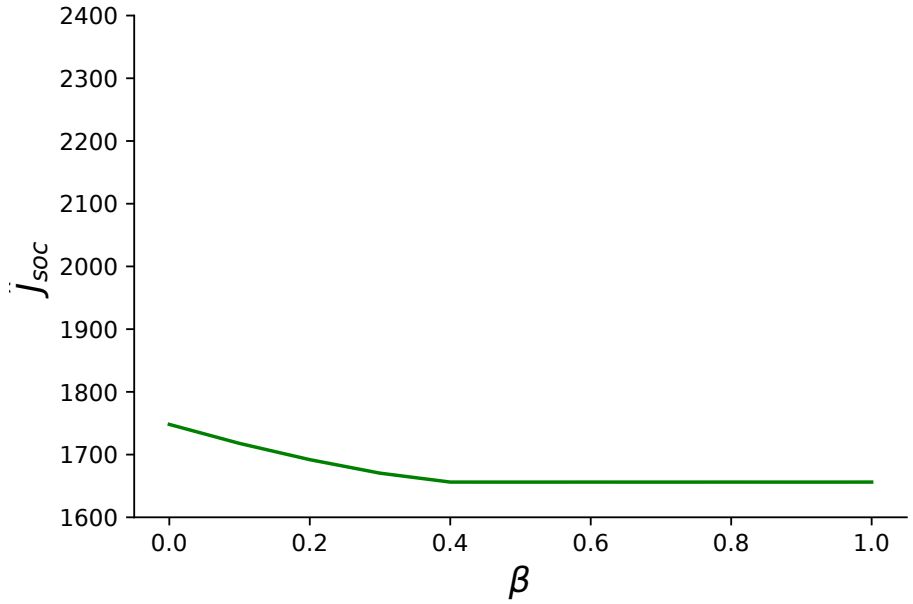
Using these modified cost functions, a vector of human–driven flows $\hat{\mathbf{x}}$ is an induced equilibrium if for every exit link $i \in I$, it satisfies

$$\hat{x}_s^i(\hat{J}_s^i(\hat{\mathbf{x}}) - \hat{J}_b^i(\hat{\mathbf{x}})) \leq 0, \tag{8.27a}$$
$$\hat{x}_b^i(\hat{J}_b^i(\hat{\mathbf{x}}) - \hat{J}_s^i(\hat{\mathbf{x}})) \leq 0. \tag{8.27b}$$

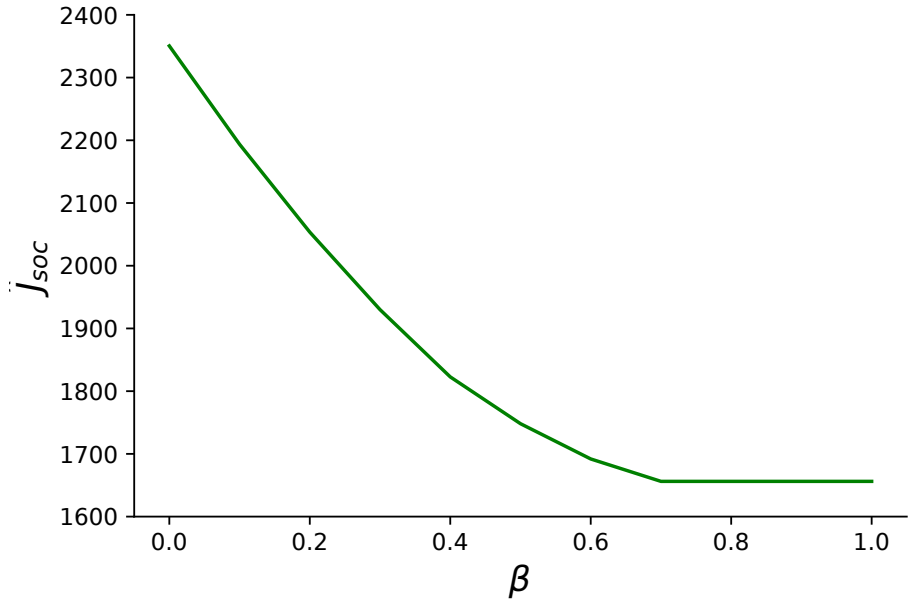
Let $\hat{J}_{\text{soc}}$ be the social cost of the users in the induced equilibrium. Then, the social cost $\hat{J}_{\text{soc}}(\hat{\mathbf{x}})$ is obtained via

$$\hat{J}_{\text{soc}}(\hat{\mathbf{x}}, w, z) = d\big((\hat{x}_s^P + z)\hat{J}_s^P(\hat{\mathbf{x}}, w, z) + (\hat{x}_b^P + w)\hat{J}_b^P(\hat{\mathbf{x}}, w, z) + \hat{x}_s^Q\hat{J}_s^Q(\hat{\mathbf{x}}, w, z) + \hat{x}_b^Q\hat{J}_2^b(\mathbf{x}, \hat{w}, z)\big). \tag{8.28}$$

To make our exposition more concrete, let the total demand be $d = 3000\,\frac{\text{veh}}{\text{hour}}$. We use the cost parameters $\mathbf{C}$ obtained from calibration. Now, fix the fractional demand configuration to be $F = \{f^P = 0.65, f^Q = 0.35\}$. We fix α and vary β from 0 to 1. For each value of β, we computed the fractions of human–driven steadfast and bypassing cars in the induced equilibrium using Equations (8.27). Figure 8.6 plots the predicted fractions of human–driven bypassing cars for different values of β. Then, using Equation (8.28), we computed the resulting social cost as a function β. As shown in Figure 8.6, in both cases, when $\beta = 0$, i.e. none of the autonomous cars were steadfast, no bypassing was observed at the equilibrium. As β increased to 0.4 for $\alpha = 0.25$ and β increases to 0.7 for $\alpha = 0.5$, the cars started to bypass at the equilibrium. Intuitively, this means that when the fraction of commanded steadfast autonomous cars went beyond a threshold, the bypassing behavior of human–driven cars emerged. Figure 8.7 plots the social cost as a function of the fraction of autonomous cars that are commanded to remain steadfast β for different values of the autonomy fraction α. As this figure shows, when $\alpha = 0.25$, the minimum social cost was achieved around $\beta = 0.4$, and then the social cost remains unchanged for $\beta > 0.4$. When the fraction of autonomous cars was increased to $\alpha = 0.5$, the minimum social delay was achieved around $\beta = 0.7$, and then the social cost remains unchanged for $\beta > 0.7$. The social cost remains the same after β achieves a given value because the fraction of mixed bypassing cars (including both human–driven and autonomous cars) remains the same



(a) $\alpha = 0.25$



(b) $\alpha = 0.5$

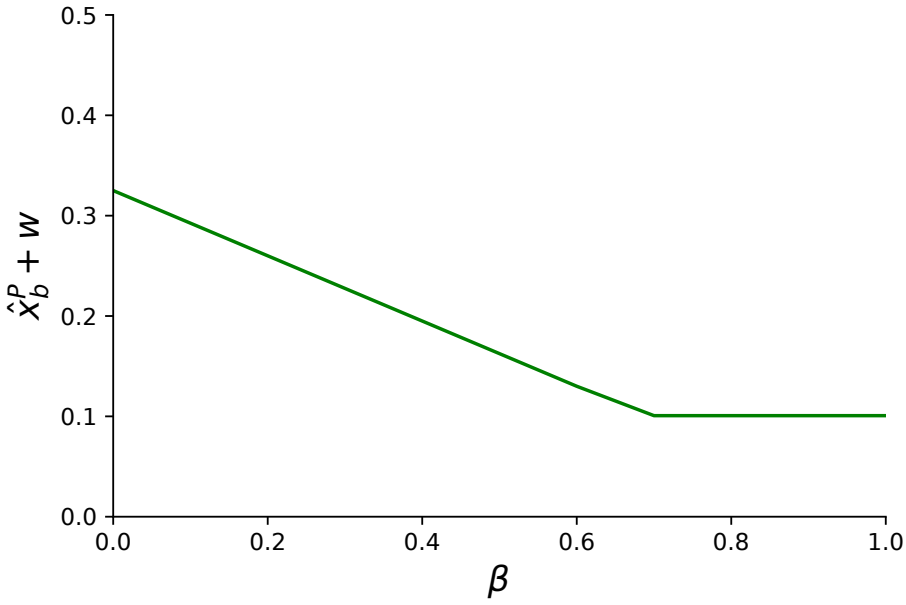
Figure 8.7: The social cost per total demand versus the fraction of commanded autonomous steadfast cars β, for two different values of autonomy fractions α.

after that value. To illustrate this, we take $\alpha = 0.5$ as an example. Figure 8.8 shows the curves for the total fraction of bypassing cars taking exit P versus β. We can see from the figure that the fractions of bypassing cars become constant for $\beta > 0.7$, therefore, the social cost remains unchanged for $\beta > 0.7$.

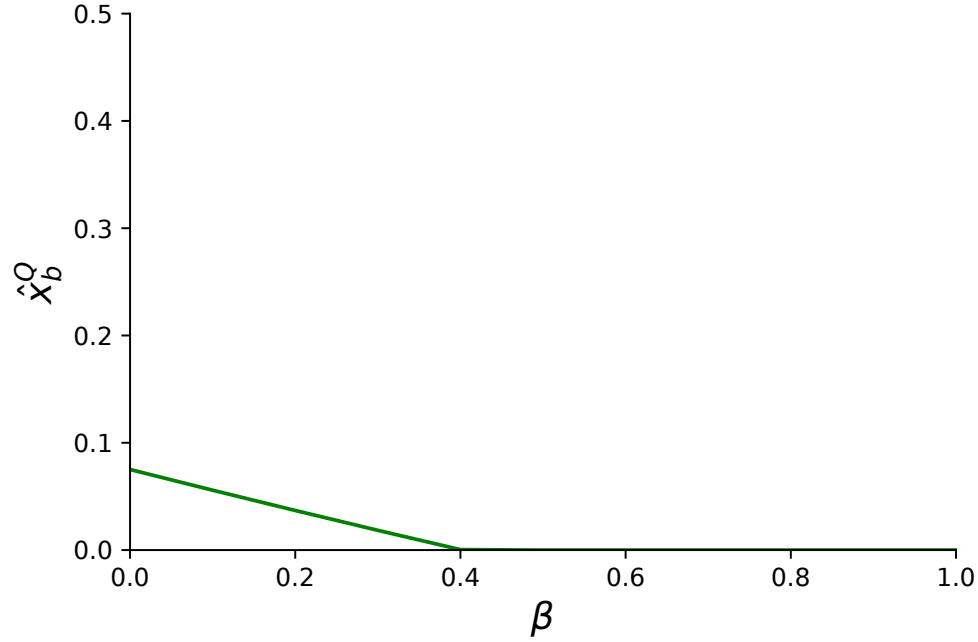
Notice how the parameter α affects the pattern of the social cost plots. For $\alpha = 0.25$, all values of β lead to approximately similar values of the social cost, whereas for $\alpha = 0.5$, β is more determinative. This is intuitive since when the penetration rate of autonomous cars is low, controlling the lance choices and bypassing does not have a sufficiently large impact on the overall system performance. It is interesting to observe that, as the number of commanded steadfast autonomous cars increases from zero, the overall social cost of the system decreases, which is what we expect when autonomous cars act steadfastly. However, when the fraction of commanded steadfast autonomous cars β increases beyond a critical value, further increases do not decrease the social cost further. Moreover, as the penetration rate of autonomous cars increases, this critical value of β increases. As this example illustrates, our proposed model provides a powerful framework for performing such traffic analysis for planning the actions of autonomous cars. In this section, we studied the impact of autonomy presence through an example. Our model can be further used to mathematically find the optimal lane choice of commanded cars for any given autonomous cars penetration rate.

8.6 Chapter Summary

We provided a game theoretic framework for macroscopically modeling the aggregate lane choice and bypassing behavior of cars at a traffic diverge, where cars were assumed to be selfish. We modeled the resulting equilibrium as a Wardrop equilibrium and proved the existence and uniqueness of this equilibrium. We described how our model can be easily calibrated and demonstrated via an initial simulation study using data provided by SUMO, a popular traffic simulation software, that our model could yield promising results in predicting at the macroscopic level, the lane change and bypassing behavior of cars. We also showed how our model can potentially be used to determine traffic management policies that reduce the social cost in traffic networks with mixed vehicle autonomy.



(a) $\hat{x}_b^P + w$



(b) $\hat{x}_b^Q$

Figure 8.8: The fraction of mixed bypassing cars versus the fraction of commanded steadfast cars β for the autonomy fraction $\alpha = 0.5$.

Chapter 9

Concluding Remarks

We envision an evolution in how future smart cities will serve the society. To mention a few of such evolutions, it is anticipated that 1) Mobility will no longer be only about owning a car, and more people will be moved with fewer vehicles. 2) There will be a strong reduction in the demand for street parking. 3) Streets will be managed in real time. 4) Streets will be safe for both vehicles and pedestrians. 5) Vehicles will be electrified, and controlled such that emissions and energy consumption are minimized. To reach this state, we need to deploy new technologies *proactively*, and enable the cooperation between different infrastructures such as transportation and power systems, as well as the collaboration among public and private sectors such as cities and ride hailing companies.

Among many emerging technologies that can affect transportation systems, it is expected that autonomous cars will transform the transportation system in the cities of future. This thesis has been a key step towards understanding the potential system–level impact of autonomy on mobility. In this chapter, we would like to conclude this dissertation by discussing some of the limitations of our framework and future directions.

Our work on the mobility implications of selfish autonomy in Chapter 6 reveals that if autonomous cars are not cognizant of their system–level impact, and like human–driven cars act selfishly, their presence may even exacerbate traffic congestion. In particular, Chapter 6 demonstrates that such exacerbation may occur due to the fact that *humans will change their behavior in the presence of autonomy*; human–driven cars may change their routes as soon as autonomy is adopted. This implies that anticipated benefits of autonomy adoption are in accordance with the assumption that humans will act as they do. However, once this assumption is violated, the consequences of autonomy adoption may not be intuitive. We believe that the results presented in Chapter 6 indicate that the expected mobility benefits resulting from the wide spread utilization of autonomous vehicles in traffic networks may not be immediate. Thus, in order to take advantage of the potential mobility benefits of autonomy, it will be necessary to make infrastructure enhancements that will homogenize the degree of capacity asymmetry in the network, and study traffic management and control strategies for mixed–autonomy networks. Further control and management must be developed for these networks such that the system is

steered to the equilibria that have lower total delay. Therefore, revisiting routing and pricing strategies for networks with mixed vehicle autonomy is essential.

Aligned with these results, we showed in Chapter 8 that if we have a good model of how humans behave, taking into account how humans would respond to the actions of autonomy, autonomous systems can act altruistically; they can plan for their actions in the favor of society rather than their own objective function. We showed in Chapter 8 the benefits of such altruistic actions for the simple example of a traffic diverge. However, it is crucial to study how such altruistic actions can be extended to different scenarios at different scales. How efficiently can the system operate if autonomous cars exhibit altruistic route, lane, speed and acceleration choices? To achieve societal–scale benefits from the presence of autonomy, we believe that autonomous cars must be socially–aware; they need to be aware of the externalities their presence will incur and compensate for it by their altruistic actions.

Even if all autonomous systems act altruistically, still, humans are an inseparable part of cities. Indeed, there may still be a human taking an autonomous ride to reach her destination. However, humans are not actuators to follow the commands that a central authority encodes, they need incentive/disincentives for changing their behavior. For instance, a traveler served by an autonomous car may not desire the altruistic route choice of her autonomous car unless appropriate incentives are provided. We believe that a key component required for the efficient operation of transportation system is to provide the right incentives/disincentives for humans. In Chapter 7, we studied pricing as a disincentive mechanism for mixed–autonomy systems in the simple form of pricing network links. Nevertheless, it may not be practical to price all links in a traffic network. We believe that autonomous cars will allow for more subtle forms of incentive mechanisms which can be leveraged for better efficiency of transportation systems. For example, autonomous cars allow for path–based pricing rather than link prices. It is essential to study the impact of such subtle forms of incentive mechanisms to achieve further objectives such as collecting the minimum possible monetary value or obtaining a fair incentive mechanism.

Bibliography

[1] URL: https://www.ibm.com/support/knowledgecenter/SSSA5P_12.5.0/ilog.odms.cplex.help/CPLEX/MATLAB/topics/cplex_matlab_overview.html.

[2] Assad Al Alam, Ather Gattami, and Karl Henrik Johansson. "An experimental study on the fuel reduction potential of heavy duty vehicle platooning". In: *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE. 2010, pp. 306–311.

[3] Assad Alam et al. "Heavy-duty vehicle platooning for sustainable freight transportation: A cooperative method to enhance safety and efficiency". In: *IEEE Control Systems* 35.6 (2015), pp. 34–56.

[4] Eitan Altman and Hisao Kameda. "Equilibria for multiclass routing in multi-agent networks". In: *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*. Vol. 1. IEEE. 2001, pp. 604–609.

[5] Eitan Altman et al. "A survey on networking games in telecommunications". In: *Computers & Operations Research* 33.2 (2006), pp. 286–311.

[6] Zahra Amini et al. "Queue-length estimation using real-time traffic data". In: *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE. 2016, pp. 1476–1481.

[7] Armin Askari et al. "Measuring impact of adaptive and cooperative adaptive cruise control on throughput of signalized intersections". In: *arXiv preprint arXiv:1611.08973* (2016).

[8] Xuegang Jeff Ban, Peng Hao, and Zhanbo Sun. "Real time queue length estimation for signalized intersections using travel times from mobile sensors". In: *Transportation Research Part C: Emerging Technologies* 19.6 (2011), pp. 1133–1156.

[9] Mokhtar S Bazaraa, Hanif D Sherali, and Chitharanjan M Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.

[10] Robert L Bleyl. "A practical computer program for designing traffic-signal-system timing plans". In: *Highway Research Record* 211 (1967).

[11] S. Boyd et al. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Foundations and Trends in Machine Learning* 3.1 (Apr. 2011), pp. 1–122. DOI: 10.1561/2200000016.

[12] D Braess and G Koch. "On the existence of equilibria in asymmetrical multiclass-user transportation networks". In: *Transportation Science* 13.1 (1979), pp. 56–63.

[13] Dietrich Braess. "Über ein Paradoxon aus der Verkehrsplanung". In: *Unternehmensforschung* 12.1 (1968), pp. 258–268.

[14] Ron Cocchi et al. "Pricing in computer networks: Motivation, formulation, and example". In: *IEEE/ACM Transactions on networking* 1.6 (1993), pp. 614–627.

[15] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. "How much can taxes help selfish routing?" In: *Journal of Computer and System Sciences* 72.3 (2006), pp. 444–467.

[16] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. "Pricing network edges for heterogeneous selfish users". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM. 2003, pp. 521–530.

[17] Samuel Coogan et al. "Controlling a network of signalized intersections from temporal logical specifications". In: *American Control Conference (ACC), 2015*. IEEE. 2015, pp. 3919–3924.

[18] Samuel Coogan et al. "Offset optimization for a network of signalized intersections via semidefinite relaxation". In: *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE. 2015, pp. 2187–2192.

[19] Samuel Coogan et al. "Offset optimization in signalized traffic networks via semidefinite relaxation". In: *Transportation Research Part B: Methodological* 100 (2017), pp. 82–92.

[20] José R Correa, Andreas S Schulz, and Nicolás E Stier-Moses. "A geometric approach to the price of anarchy in nonatomic congestion games". In: *Games and Economic Behavior* 64.2 (2008), pp. 457–469.

[21] Stella C Dafermos. "The traffic assignment problem for multiclass-user transportation networks". In: *Transportation science* 6.1 (1972), pp. 73–87.

[22] Stella C Dafermos. "Toll patterns for multiclass-user transportation networks". In: *Transportation science* 7.3 (1973), pp. 211–223.

[23] Stella Dafermos and Anna Nagurney. "On some traffic equilibrium theory paradoxes". In: *Transportation Research Part B: Methodological* 18.2 (1984), pp. 101–110.

[24] Swaroop Darbha and KR Rajagopal. "Intelligent cruise control systems and traffic flow stability". In: *Transportation Research Part C: Emerging Technologies* 7.6 (1999), pp. 329–352.

[25] Christina Diakaki, Markos Papageorgiou, and Kostas Aboudolas. "A multivariable regulator approach to traffic-responsive network-wide signal control". In: *Control Engineering Practice* 10.2 (2002), pp. 183–195.

[26] Jakob Erdmann. "Lane-Changing Model in SUMO". In: *SUMO2014*. Vol. 24. Reports of the DLR-Institute of Transportation SystemsProceedings. Deutsches Zentrum für Luft- und Raumfahrt e.V., May 2014, pp. 77–88. URL: https://elib.dlr.de/89233/.

[27] S Alireza Fayazi and Ardalan Vahidi. "Mixed Integer Linear Programming for Optimal Scheduling of Autonomous Vehicle Intersection Crossing". In: *IEEE Transactions on Intelligent Vehicles* (2018).

[28] Caroline Fisk. "More paradoxes in the equilibrium assignment problem". In: *Transportation Research Part B: Methodological* 13.4 (1979), pp. 305–309.

[29] Nathan H Gartner. *OPAC: A demand-responsive strategy for traffic signal control*. 906. 1983.

[30] Nathan H Gartner, Farhad J Pooran, and Christina M Andrews. "Implementation of the OPAC adaptive control strategy in a traffic signal network". In: *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE. 2001, pp. 195–200.

[31] Nathan H Gartner and Chronis Stamatiadis. "Arterial-based control of traffic flow in urban grid networks". In: *Mathematical and computer modelling* 35.5-6 (2002), pp. 657–671.

[32] Nathan H Gartner et al. "A multi-band approach to arterial traffic signal optimization". In: *Transportation Research Part B: Methodological* 25.1 (1991), pp. 55–74.

[33] Douglas Gettman et al. "Data-driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems". In: *Transportation Research Record: Journal of the Transportation Research Board* 2035 (2007), pp. 1–9.

[34] Fred Glover. "Improved linear integer programming formulations of nonlinear integer problems". In: *Management Science* 22 (1975).

[35] Gabriel Gomes. "Bandwidth maximization using vehicle arrival functions". In: *IEEE Transactions on Intelligent Transportation Systems* 16.4 (2015), pp. 1977–1988.

[36] Gabriel Gomes. "Bandwidth maximization using vehicle arrival functions". In: *IEEE Transactions on Intelligent Transportation Systems* 16 (2015).

[37] Jane N Hagstrom and Robert A Abrams. "Characterizing Braess's paradox for traffic networks". In: *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE. 2001, pp. 836–841.

[38] Michael A Hall. "Properties of the equilibrium state in transportation networks". In: *Transportation Science* 12.3 (1978), pp. 208–216.

[39] Michael Hintermüller, Kazufumi Ito, and Karl Kunisch. "The primal-dual active set strategy as a semismooth Newton method". In: *SIAM Journal on Optimization* 13.3 (2002), pp. 865–888.

[40] Xin Huang, Asuman E Ozdaglar, and Daron Acemoglu. "Efficiency and Braess' Paradox under pricing in general networks". In: *IEEE Journal on Selected Areas in Communications* 24.5 (2006), pp. 977–991.

[41] PB Hunt et al. "The SCOOT on-line traffic signal optimisation technique". In: *Traffic Engineering & Control* 23.4 (1982).

[42] David Husch and John Albeck. "Synchro 6: Traffic signal software, user guide". In: *Albany, California* (2003).

[43] Frank P Kelly, Aman K Maulloo, and David KH Tan. "Rate control for communication networks: shadow prices, proportional fairness and stability". In: *Journal of the Operational Research society* 49.3 (1998), pp. 237–252.

[44] Solmaz S Kia. "Distributed optimal resource allocation over networked systems and use of an e-exact penalty function". In: *IFAC-PapersOnLine* 49.4 (2016), pp. 13–18.

[45] Sarah Koehler, Claus Danielson, and Francesco Borrelli. "A primal-dual active-set method for distributed model predictive control". In: *Optimal Control Applications and Methods* (2016).

[46] Sarah Koehler et al. "Stable hybrid Model Predictive Control for ramp metering". In: *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE. 2016, pp. 1083–1088.

[47] Anastasios Kouvelas, Mohammadreza Saeedmanesh, and Nikolas Geroliminis. "Feedback perimeter control for heterogeneous urban networks using adaptive optimization". In: *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. IEEE. 2015, pp. 882–887.

[48] Daniel Krajzewicz et al. "Recent Development and Applications of SUMO - Simulation of Urban MObility". In: *International Journal On Advances in Systems and Measurements*. International Journal On Advances in Systems and Measurements 5.3&4 (Dec. 2012), pp. 128–138. URL: http://elib.dlr.de/80483/.

[49] Stefan Krauß. "Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics". PhD thesis. Dt. Zentrum für Luft-und Raumfahrt, 1998.

[50] Daniel A Lazar, Samuel Coogan, and Ramtin Pedarsani. "Capacity modeling and routing for traffic networks with mixed autonomy". In: *Decision and Control (CDC), 2017 IEEE 56th Annual Conference on*. IEEE. 2017, pp. 5678–5683.

[51] Daniel A Lazar, Samuel Coogan, and Ramtin Pedarsani. "The Price of Anarchy for Transportation Networks with Mixed Autonomy". In: *arXiv preprint arXiv:1710.07867* (2017).

[52] Daniel A Lazar et al. "Maximizing Road Capacity Using Cars that Influence People". In: *arXiv preprint arXiv:1807.04414* (2018).

[53] Kuo-Yun Liang, Jonas Mårtensson, and Karl Henrik Johansson. "When is it fuel efficient for a heavy duty vehicle to catch up with a platoon?" In: *7th IFAC Symposium on Advances in Automotive Control, Tokyo, Japan, September 4-7, 2013*. 2013.

[54] Shu Lin et al. "Efficient network-wide model-based predictive control for urban traffic networks". In: *Transportation Research Part C: Emerging Technologies* 24 (2012), pp. 122–140.

[55] Shu Lin et al. "Fast model predictive control for urban road networks via MILP". In: *IEEE Transactions on Intelligent Transportation Systems* 12.3 (2011), pp. 846–856.

[56] Jennie Lioris, Alex A Kurzhanskiy, and Pravin Varaiya. "Control Experiments for a Network of Signalized Intersections Using the'. Q'Simulator." In: *WODES*. 2014, pp. 332–337.

[57] Jennie Lioris et al. "Platoons of connected vehicles can double throughput in urban roads". In: *Transportation Research Part C: Emerging Technologies* 77 (2017), pp. 292–305.

[58] John DC Little. "The synchronization of traffic signals by mixed-integer linear programming". In: *Operations Research* 14.4 (1966), pp. 568–594.

[59] Henry X Liu et al. "Real-time queue length estimation for congested signalized intersections". In: *Transportation research part C: emerging technologies* 17.4 (2009), pp. 412–427.

[60] Johan Lofberg. "YALMIP: A toolbox for modeling and optimization in MATLAB". In: *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*. IEEE. 2004, pp. 284–289.

[61] PR Lowrie. "Scats, sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic". In: (1990).

[62] Negar Mehr and Roberto Horowitz. "Can the Presence of Autonomous Vehicles Worsen the Equilibrium State of Traffic Networks?" In: *2018 IEEE Conference on Decision and Control (CDC)*. IEEE. 2018, pp. 1788–1793.

[63] Negar Mehr and Roberto Horowitz. "How Will the Presence of Autonomous Vehicles Affect the Equilibrium State of Traffic Networks?" In: *to appear in IEEE Transactions on Control of Network Systems* (2019).

[64] Negar Mehr and Roberto Horowitz. "Pricing Traffic Networks with Mixed Vehicle Autonomy". In: *2019 Annual American Control Conference (ACC)*. IEEE. 2019.

[65] Negar Mehr and Roberto Horowitz. "Probabilistic Freeway Ramp Metering". In: *rN* 1.f1 (2016), p. 1.

[66] Negar Mehr, Roberto Horowitz, and Ramtin Pedarsani. "Low-complexity ramp metering for freeway congestion control via network utility maximization". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. 2017, pp. 5672–5677.

[67] Negar Mehr, Ruolin Li, and Roberto Horowitz. "A game theoretic macroscopic model of bypassing at traffic diverges with applications to mixed autonomy networks". In: *arXiv preprint arXiv:1809.02762* (2018).

[68] Negar Mehr, Ruolin Li, and Roberto Horowitz. "A game theoretic macroscopic model of bypassing at traffic diverges with applications to mixed autonomy networks". In: *arXiv preprint arXiv:1809.02762* (2018).

[69] Negar Mehr et al. "Joint Perimeter and Signal Control of Urban Traffic via Network Utility Maximization". In: *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*. IEEE. 2017.

[70] Negar Mehr et al. "Offset Selection for Bandwidth Maximization on Multiple Routes". In: *2018 Annual American Control Conference (ACC)*. IEEE. 2018, pp. 6366–6371.

[71] Negar Mehr et al. "Signal Control for Urban Traffic Networks with Unknown System Parameters". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 2171–2176.

[72] Negar Mehr et al. "Stochastic predictive freeway ramp metering from Signal Temporal Logic specifications". In: *American Control Conference (ACC), 2017*. IEEE. 2017, pp. 4884–4889.

[73] Carroll J Messer et al. "A variable-sequence multiphase progression optimization program". In: *Highway Research Record* 445.1973 (1973), pp. 24–33.

[74] David Miculescu and Sertac Karaman. "Polling-systems-based control of high-performance provably-safe autonomous intersections". In: *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE. 2014, pp. 1417–1423.

[75] Igal Milchtaich. "Network topology and the efficiency of equilibrium". In: *ICM Millennium Lectures on Games*. Springer, 2003, pp. 233–266.

[76] Pitu Mirchandani and Fei-Yue Wang. "RHODES to intelligent transportation systems". In: *IEEE Intelligent Systems* 20.1 (2005), pp. 10–15.

[77] Jeonghoon Mo and Jean Walrand. "Fair end-to-end window-based congestion control". In: *IEEE/ACM Transactions on Networking (ToN)* 8.5 (2000), pp. 556–567.

[78] Ciamac C Moallemi and Benjamin Van Roy. "Convergence of min-sum message-passing for convex optimization". In: *IEEE Transactions on Information Theory* 56.4 (2010), pp. 2041–2050.

[79] John T Morgan and John DC Little. "Synchronizing traffic signals for maximal bandwidth". In: *Operations Research* 12.6 (1964), pp. 896–912.

[80] Ajith Muralidharan, Ramtin Pedarsani, and Pravin Varaiya. "Analysis of fixed-time control". In: *Transportation Research Part B: Methodological* 73 (2015), pp. 81–90.

[81] U.S. News. *Why Some Cities Have Had Enough of Waze*. 2018. URL: https://www.usnews.com/news/national-news/articles/2018-05-07/why-some-cities-have-had-enough-of-waze (visited on 05/07/2018).

[82] NPR. *Uber And Lyft Caused Major Traffic Uptick In San Francisco, Study Says*. 2019. URL: https://www.npr.org/2019/05/08/721139488/uber-and-lyft-caused-major-traffic-uptick-in-san-francisco-study-says (visited on 05/08/2019).

[83] Natale Papola and Gaetano Fusco. "Maximal bandwidth problems: a new algorithm based on the properties of periodicity of the system". In: *Transportation Research Part B: Methodological* 32.4 (1998), pp. 277–288.

[84] Ramtin Pedarsani. "Robust Scheduling for Queueing Networks". PhD thesis. EECS Department, University of California, Berkeley, May 2015. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-58.html.

[85] Ramtin Pedarsani, Jean Walrand, and Yuan Zhong. "Robust scheduling and congestion control for flexible queueing networks". In: *Computing, Networking and Communications (ICNC), 2014 International Conference on*. IEEE. 2014, pp. 467–471.

[86] Ramtin Pedarsani, Jean Walrand, and Yuan Zhong. "Robust scheduling for flexible processing networks". In: *Advances in Applied Probability* 49.2 (2017), pp. 603–628.

[87] Arthur Pigou. *The economics of welfare*. Routledge, 2017.

[88] Rekha S Pillai, Ajay K Rathi, and Stephen L Cohen. "A restricted branch-and-bound approach for generating maximum bandwidth signal timing plans for traffic networks". In: *Transportation Research Part B: Methodological* 32.8 (1998), pp. 517–529.

[89] Michael JD Powell. "A method for nonlinear constraints in minimization problems". In: *Optimization* (1969), pp. 283–298.

[90] Rattaphol Pueboobpaphan and Bart van Arem. "Driver and vehicle characteristics and platoon and traffic flow stability: Understanding the relationship for design and assessment of cooperative adaptive cruise control". In: *Transportation Research Record: Journal of the Transportation Research Board* 2189 (2010), pp. 89–97.

[91] Vasumathi Raman et al. "Model predictive control with signal temporal logic specifications". In: *2014 IEEE 53rd Annual Conference on Decision and Control (CDC)*. IEEE. 2014, pp. 81–87.

[92] BUREAU OF PUBLIC ROADS. "Traffic assignment manual". In: *US Department of Commerce* (1964).

[93] Dennis I Robertson and R David Bretherton. "Optimizing networks of traffic signals in real time-the SCOOT method". In: *IEEE Transactions on vehicular technology* 40.1 (1991), pp. 11–15.

[94] Tim Roughgarden. "Designing networks for selfish users is hard". In: *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*. IEEE. 2001, pp. 472–481.

[95] Tim Roughgarden. "On the severity of Braess's paradox: designing networks for selfish users is hard". In: *Journal of Computer and System Sciences* 72.5 (2006), pp. 922–953.

[96] Tim Roughgarden. *Selfish routing*. Tech. rep. CORNELL UNIV ITHACA NY DEPT OF COMPUTER SCIENCE, 2002.

[97] Tim Roughgarden and Éva Tardos. "How bad is selfish routing?" In: *Journal of the ACM (JACM)* 49.2 (2002), pp. 236–259.

[98] Sadra Sadraddini and Calin Belta. "A Provably Correct MPC Approach to Safety Control of Urban Traffic Networks". In: *arXiv preprint arXiv:1602.01028* (2016).

[99] David Schrank, Bill Eisele, and Tim Lomax. "TTI's 2012 urban mobility report". In: *Texas A&M Transportation Institute. The Texas A&M University System* 4 (2012).

[100] MJ Smith. "The marginal cost taxation of a transportation network". In: *Transportation Research Part B: Methodological* 13.3 (1979), pp. 237–242.

[101] Mt J Smith. "The existence, uniqueness and stability of traffic equilibria". In: *Transportation Research Part B: Methodological* 13.4 (1979), pp. 295–304.

[102] Chronis Stamatiadis and Nathan Gartner. "MULTIBAND-96: a program for variable-bandwidth progression optimization of multiarterial traffic networks". In: *Transportation Research Record: Journal of the Transportation Research Board* 1554 (1996), pp. 9–17.

[103] Richard Steinberg and Willard I Zangwill. "The prevalence of Braess' paradox". In: *Transportation Science* 17.3 (1983), pp. 301–318.

[104] Raphael E Stern et al. "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments". In: *Transportation Research Part C: Emerging Technologies* 89 (2018), pp. 205–221.

[105] Jelka Stevanovic et al. "Stochastic optimization of traffic control and transit priority settings in VISSIM". In: *Transportation Research Part C: Emerging Technologies* 16.3 (2008), pp. 332–349.

[106] Pavankumar Tallapragada and Jorge Cortés. "Coordinated intersection traffic management". In: *IFAC-PapersOnLine* 48.22 (2015), pp. 233–239.

[107] Fatma Yildiz Tascikaraoglu et al. "PointQ model of an arterial network: calibration and experiments". In: *arXiv preprint arXiv:1507.08082* (2015).

[108] Pravin Varaiya. "Max pressure control of a network of signalized intersections". In: *Transportation Research Part C: Emerging Technologies* 36 (2013), pp. 177–195.

[109] John Glen Wardrop. "Some theoretical aspects of road traffic research". In: *Inst Civil Engineers Proc London/UK/*. 1952.

[110] Matthew A Wright, Simon FG Ehlers, and Roberto Horowitz. "Neural-Attention-Based Deep Learning Architectures for Modeling Traffic Dynamics on Lane Graphs". In: *arXiv preprint arXiv:1904.08831* (2019).

[111] Cathy Wu et al. "Emergent Behaviors in Mixed-Autonomy Traffic". In: *Conference on Robot Learning*. 2017, pp. 398–407.

[112] Jingang Yi and Roberto Horowitz. "Macroscopic traffic flow propagation stability for adaptive cruise controlled vehicles". In: *Transportation Research Part C: Emerging Technologies* 14.2 (2006), pp. 81–95.

[113] Yue J Zhang, Andreas A Malikopoulos, and Christos G Cassandras. "Optimal control and coordination of connected and automated vehicles at urban traffic intersections". In: *American Control Conference (ACC), 2016*. IEEE. 2016, pp. 6227–6232.