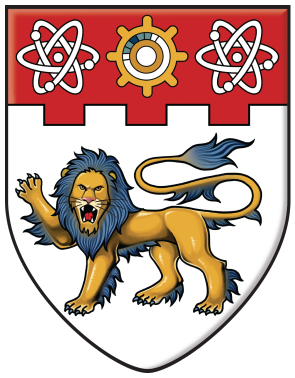
REGION-BASED URBAN TRAFFIC MANAGEMENT



NANYANG TECHNOLOGICAL UNIVERSITY

ANTONIOS LENTZAKIS

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2017

Acknowledgements

I wish to thank my supervisor, Professor Rong Su for his contributions, as well as his unfailing support, constantly providing motivation and believing in my potential to be a good researcher. I would also like to thank my co-supervisor Professor Changyun Wen for his helpful comments and suggestions, always providing structure and clarity of meaning to the concepts being discussed, allowing me to convey my thoughts and move forward.

I wish to thank Nanyang Technological University and in particular the SMART Future Urban Mobility and the NTU-NXP Smart Mobility Test Bed research groups for providing me with opportunities to conduct research in a highly conducive environment with accomplished and bright researchers. I also wish to thank Professor Justin Dauwels, for his encouragement to apply for the SINGA scholarship and essentially kickstarting my PhD journey, from faraway Greece to the city of Lions, Singapore.

I wish to express my most sincere gratitude from the bottom of my heart to the following people:

- Dr. Simon Ware for our very fruitful collaboration, as well as our long weekly discussions which went well beyond our collaborative work and which I thoroughly enjoyed.

- Dr. Liyong Lin for helping with my mathematical formulations, as well as keeping my mind keen with our weekly Go/Weiqi matches.

- Dr. Nikitha Radhakrishnan, without whose advice, inspiration, and most of all, motivation, I would not have made it this far. Literally the most helpful and supportive friend anyone could ever ask for!

- Alireza Barzergar and his wife Leila, for helping me relieve stress through long conversations, as well as introducing me to the most delicious homemade Iranian food!

- Dr. Khu Vu and Zsolt Szabo, my old flatmates, who took time to make me feel welcome every time I came home from the lab, as well as sharing their authentic Vietnamese and Hungarian food with me, even though I could only reciprocate their generosity with Greek salad!

Last but never least, I want to thank for their material and emotional support, as well as their unconditional love, my wife Kimmy, my parents Fanourios and Athanasia, my siblings Emilia and George and all of my friends back home.

Contents

List of Figures

List of Tables

List of Abbreviations

ATIS Advanced Traveler Information System

ATT Average Travel Time

CBD Central Business District

CV Coefficient of Variation

DSO Dynamic System Optimum

DSRC Dedicated Short Range Communications

DTA Dynamic Traffic Assignment

FD Fundamental Diagram

FR Fixed Routing

GPS Global Positioning System

IRP Incremental Route Planning

ITR Incomplete Trips Rate

LTA Land Transport Authority

MFD Macroscopic Fundamental Diagram

MPR Market Penetration Rate

MPR1 Market Penetration Rate for 1^{st} traveler class

MPR2 Market Penetration Rate for 2^{nd} traveler class

MPR3 Market Penetration Rate for 3^{rd} traveler class

MSA Method of Successive Averages

NC Non-compliance Rate

NSC Normalized Spectral Clustering

NTM Network Transmission Model

RSU Road Side Units

S1 Scenario 1

S2 Scenario 2

SC Silhouette Coefficient

TP Total Performance

TVT Total Vehicle Travel time

PAR Periodically Adjusted Routing

PRM Proxy Regret Matching

PRMNC Proxy Regret Matching with Non-compliance

PTD Public Transit Diversion

RGIS Route Guidance and Information System

VMS Variable Message Sign

WKHM Weighted k-Harmonic Means

WKM Weighted k-Means

Abstract

Quality of life in cities, especially for vehicle owners, is closely correlated with the delay experienced by drivers, a direct consequence of congestion. Costs associated with delays incurred to travelers during peak periods include the loss of man-hours, increased vehicle emissions, increased fuel consumption and more. The magnitude of the socio-economic impact of congestion is usually very great. Unfortunately, due to limited space availability in modern urban regions, infrastructure upgrade and extension costs are prohibitive, with optimal utilization of available infrastructure being the only viable alternative. This work describes different aspects of a proposed Region-based Urban Traffic Simulation and Management Framework, which takes advantage of the Urban-scale Macroscopic Fundamental Diagram concept. Concretely, weighted feature extensions of center-based clustering methods are introduced and compared with an implementation of normalized spectral clustering. A new, feature weighted extension of k-harmonic means clustering outperforms all other methods when applied to an urban traffic network encapsulating Singapore's Central Business District. The partitioning of the urban traffic network into homogeneously congested regions is an essential requirement for the existence and validity of Urban-scale Macroscopic Fundamental Diagram. Subsequently, a multi-class extension of a regional dynamic traffic model, the Network Transmission Model, is proposed and subsequently used to simulate regional traffic dynamics. Two innovative region-based prescriptive route guidance approaches are proposed, i.e., a

non-predictive strategic learning-based routing method and a predictive simulation-based routing method. They are applied on a 16-region, diamond-shaped grid network which makes use of the Network Transmission Model to simulate the regional traffic dynamics. Their performance is compared to multinomial logit routing, a realistic reference case which models self-interested traveler route choice behavior. It should be noted that, all routing methods come integrated with a Public Transit Diversion mechanism. As expected, the predictive routing method outperforms all other routing methods.

A market penetration scheme is chosen to be applied to the new multi-class extension of the Network Transmission Model, where three traveler classes are defined. The 1^{st} traveler class makes use of autonomous vehicles, the 2^{nd} traveler class makes use of conventional vehicles equipped with Route Guidance and Information System and the 3^{rd} traveler class makes use of unequipped conventional vehicles. The 1^{st} class is represented by the predictive routing method with full compliance and under two information provision scenarios S1 (imperfect information),S2 (perfect information), that affect travel time prediction accuracy. The 2^{nd} class is represented by the non-predictive routing method with the possibility of non-compliance. The 3^{rd} class is represented by multinomial logit routing, modeling route choice under imperfect travel time perception. Regional traffic dynamics for simultaneous application of all aforementioned routing methods are simulated, for various combinations of traveler class market penetration rates and non-compliance rates. An analysis of the overall network performance gains, as well as individual performance metric results, was conducted for each set of market penetration rate combinations with promising results about the potential impact that autonomous vehicles will have on the overall network performance, as well as the individual benefits provided to all other traveler classes.

Chapter 1

Introduction

Quality of life in cities, especially for vehicle passengers, is profoundly correlated to the delay they experience daily, as a consequence of congestion. The European Commission [4] estimated that the costs incurred from congestion in the European Community will exceed $90 billion annually. Traffic congestion in the United States in 2009 cost the economy $114.8 billion [5]. Congestion can be caused by temporary deterioration of network performance, in case of an incident. There can also be recurrent congestion, whereby the travel demand bound for certain regions within the urban traffic network exceeds the capacity of said regions, for specific time periods. Due to limited space availability in modern urban areas, costs associated with infrastructure upgrade and extension are prohibitive. Optimizing the supply of existing infrastructure resources, through use of traffic light control [6–10], route guidance [11–14], as well as restricting network travel demand through toll pricing [15–19], at different degrees of granularity (individual network links, subnetworks), can help alleviate congestion phenomena. Integration of intelligent sensor technologies, such as RSU (Road Side Units) with DSRC (Dedicated Short Range Communications) capability, to the network infrastructure can help increase the performance of the methods used to increase network capacity or restrict network

travel demand, especially during peak hour.

1.1 Motivation

In the not-so-distant future, self-driving cars are expected to comprise a significant portion of the vehicle population in metropolitan areas, more concretely, experts are forecasting a 70% market penetration rate for autonomous vehicles by 2040 [20]. Additionally, the US Secretary of Transportation, during a visit at the Frankfurt Auto Show, predicted the ubiquitous presence of self-driving cars by as early as 2025 [21]. One of the main driving forces behind autonomous vehicle technology adoption is the improvement of passenger safety through the elimination of human error. Human error accounts for 94% of road accidents [22], which in turn account for 10%-30% of delays due to congestion [23]. One of the main obstacles facing the adoption of autonomous vehicles is the uncertainty regarding their successful integration in human driver-populated traffic networks. Conventional vehicles equipped with navigation devices are able to receive several types of guidance. Descriptive guidance involves prevailing traffic conditions, predicted travel times and incident information. Prescriptive guidance involves specific shortest route information provision. Route guidance can provide user-oriented benefits. Given that vehicles remain compliant to their prescribed routes [24], route guidance can also lead to system-wide near-optimal network performance. Autonomous vehicles are expected to always maintain compliance, and through coordination with traffic control centers, can lead to system-wide benefits. Region-based routing can be considered as routing at an aggregated level. Any large urban traffic network can be appropriately partitioned into regions [25–29]. Region-based routing then presents a set of region sequences that drivers can follow from their Origin region to their desired Destination region.

This has been shown to lead to significant decrease of total vehicle travel time and better utilization of available resources, i.e., the regions of the urban traffic network, in [30]. In transportation economics literature, travelers are classified into certain broad categories, according to their value of time, type of activity (commuting to work, leisure) or other criteria and are given the option of changing their desired departure times, which becomes an additional control parameter [31]. However, this classification introduces additional complexity and stochasticity to the route selection process. In addition, many studies fail to consider the bounded rationality of travelers as decision makers [32]. A public transit diversion mechanism, as integrated to each routing method presented in this dissertation, can be considered as an alternative route choice which does not compromise travelers' departure time preferences. This feature would not be as readily justifiable in a link-by-link routing approach, where the Destination location might be a considerable distance away from a public transit station.

The main objective of this work is to introduce a Region-based Urban Traffic Management framework, employing the concept of the Urban-scale Macroscopic Fundamental Diagram. Some of the components of this framework presented in this dissertation include urban traffic network partitioning, region-based multi-class dynamic traffic modeling and region-based route guidance with a market penetration scheme. These framework components can be used as support tools by traffic operations centers to facilitate strategic (high-level) rather than tactical (low-level) decision-making about the distribution of class-specific traveler demand so as to alleviate congestion and mitigate the effects of random incidents such as car accidents. A holistic approach is taken, incorporating travelers employing autonomous vehicles, conventional vehicles with Route Guidance and Information Systems (RGIS), unequipped conventional vehicles, as well as public transit. In this manner, a more accurate estimation of prevailing and future traffic conditions can be achieved, by

analyzing the potential impact of autonomous vehicle technology adoption and the various traveler classes interaction.

A brief overview of past literature associated with the Urban-scale Macroscopic Fundamental Diagram follows. The dynamics of traffic flow and congestion formation have been modeled at different scales according to formulations based on fluid mechanics, cellular automata, particle systems, queuing systems, among others. The majority of these models constitute an approximation of the traffic flow behavior because the unpredictability of human behavior poses considerable limitations to the exact reproducibility of the actual system behavior. Various macroscopic models have been proposed to describe traffic flow in cities from the viewpoint of collective vehicular flow. Lighthill and Whitham [33] presented a model that represented vehicles in traffic flow as particles within a fluid. Smeed's theory [34], where the maximum traffic volume allowed within an urban center is a function of the area covered by the roads, as well as their capacity, does not describe the urban road network throughput effects when capacity is exceeded during peak traffic volume periods. Thomson [35] employed road data collection of central London for many years and arrived at a monotonically decreasing linear average speed-flow relationship. Wardrop [36] came up with a monotonically decreasing average speed-flow relationship as a function of mean street width and mean distance between junctions. Zahavi [37] employed similarity analysis for British and American cities. He aggregated data for various urban areas at identical time intervals including rush hour, and suggested the existence of a monotone inverse speed-flow relationship. While these models are capable of describing lightly congested traffic states, due to the monotonicity that characterizes them, they cannot be used to describe the peak periods in a congested city.

Godfrey [38] first proposed the concept of a Macroscopic Fundamental Diagram representing the relationship between vehicle density (veh/km) and space-mean flow

(veh/h) for urban regions. Herman and Prigogine [39] defined a Two Fluid Kinetic model later improved upon by Herman and Ardekani [40] that stated that the average speed in urban network traffic is a function of the ratio of stopped vehicles, which can be represented as a power function of network link density. Daganzo [41] defined a relationship between urban network output, representing the number of trips ending inside or outside the network, and the number of vehicles present at the network, or accumulation. This relationship should be valid under the condition that congestion is distributed in a homogeneous fashion throughout the network and as long as external conditions are varying at a slow rate. Geroliminis and Daganzo [42] further extended this by postulating that a Macroscopic Fundamental Diagram exists for urban regions characterized by homogeneously distributed congestion, the urban network output is proportional to the product of average flow and network length, or production, for brevity. See figure 1.1 for an illustration.

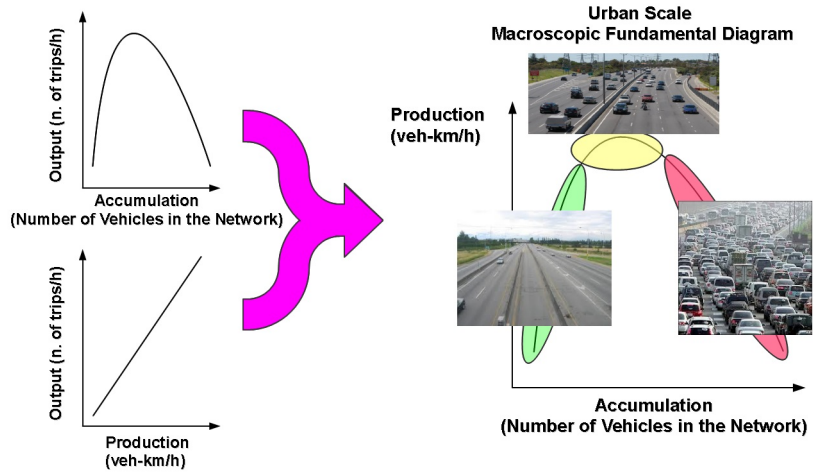


Figure 1.1: Illustration of the postulates set by Geroliminis and Daganzo [1] based on the experimental results from Yokohama and the microscopic simulation results run on a network representing Downtown San Francisco

Geroliminis and Daganzo [1] employed microscopic simulation of the San Francisco downtown area during peak period for different demand distribution scenarios, as well as analyzing experimental data derived from a 10 km^2 area in Yokohama, Japan

and were able to posit the existence of an Urban-scale Macroscopic Fundamental Diagram independent of traffic demand. Geroliminis and Sun [43], Mazloumian et al. [44], Daganzo et al. [45] have shown that variability of network link density plays a major role in the potential shape, scatter or existence of the Macroscopic Fundamental Diagram. Knoop and Hoogendoorn [46], chose to incorporate the effect of link density variability on the network performance by developing the concept of a Generalized Macroscopic Fundamental Diagram, which relates the production in an urban traffic network to the network density variability and the accumulation, defined as the weighted network average density. In order to compensate for the demand heterogeneity appearing in most real-life urban traffic networks, Ji and Geroliminis [25] implemented a partitioning of the urban traffic network into homogeneous regions with low link density variability within each region. The degree of homogeneity of density for a particular region can be expressed through link density variability. Link density variability is calculated by the standard deviation of vehicle density (veh/km) for all links within a region at a particular time instant as follows

$$\alpha_C = \sqrt{\sum_{l \in C} \frac{(\rho_l - \rho_{mean})^2}{|C|}} \tag{1.1}$$

where α represents the link density variability for cluster C, $l \in C$ represent the links that belong to cluster C and ρ_l, ρ_{mean} represent the vehicle density per link $l \in C$ and the average vehicle density for cluster C, respectively.

Centralized and distributed approaches on traffic light control of road junctions are very prominent in research literature, but the same cannot be stated for traffic management and control along the boundaries of partitions of urban traffic networks. Therein lies the motivation of partitioning urban traffic networks into homogeneous regions. Derivation of the Urban-scale Macroscopic Fundamental Diagrams for each region is the subsequent step. It should be noted that the homogeneous regions set

may not be unique, since it has been shown that different parts of the network may experience congestive regimes at distinct peak periods throughout the day, as well as different types of day, i.e. workday and weekend. From the Urban-scale Macroscopic Fundamental Diagrams, the critical accumulation value within each region will be able to be determined. This value has been used in the literature to affect feedback control under various control frameworks. In Table 1.1, an overview of papers related to the Urban-scale Macroscopic Fundamental Diagram concept is presented. In addition to the conclusions and insights presented in Table 1.1, work from [47–50] indicates that network topology, traffic signal settings, route distribution and demand can actually affect the total output of a region. Leclercq et al. [51] posited that MFD shape will change when for the same network, different route distribution is applied and proposed a framework whereby all routes within a region would be grouped into macroscopic route clusters. Keyvan-Ekbatani et al. [52], expanding Gayah et al.'s work [50], found that the integration of adaptive traffic control and network gating leads to higher network throughput as well as shorter queues on the boundary. There are two approaches in mitigating these effects. One approach is to incorporate link density variability, a measure of congestion heterogeneity, in the relationship defined by the Urban-scale MFD, deriving an MFD dependent not only on the accumulation, but also link density variability, as in [14,46,53,54]. Another approach is to partition the urban traffic network into homogeneously congested regions with well-defined Urban-scale MFDs. Examples of different partitioning schemes resulting in homogeneously congested regions can be found in [25,26,29, 55]. The concept of Urban-scale MFD has been used extensively in research into perimeter control schemes, [56–63], pricing [15–17,19,53], but to a lesser extent in routing applications, [12–14,30,64,65]. Recently, efforts have been made to combine link-based and aggregate traffic modeling for MFD-based control approaches [16, 52]. Investigations into combined types of traffic management have also gained

traction in the researcher community [66–70]. Sirmatel and Geroliminis [71] are the first to combine route guidance and perimeter control as part of an MFD-based traffic management scheme. A significant part of transportation is classification of the travel demand into multiple modes, however most of the MFD-based traffic literature have been focusing on the public transport aspect of multi-modal travel demand modeling and management (buses, trams) [72–79], however, there has been, to the author's knowledge, no research into a framework that integrates autonomous vehicles into a private ownership or ridesharing scheme.

Table 1.1: Summary of papers related to the Urban-scale MFD concept (after Knoop et al. [3])

Paper	data type	network type	remarks
Daganzo [41]	theoretical analysis	not applicable	Overcrowded networks lead to performance degradation
Geroliminis and Daganzo [1]	historical	Yokohama	Urban-scale MFDs appear in real network and there is a proportional relationship between completed trips and average network output
Daganzo and Geroliminis [80]	historical and simulated	Yokohama (real) San Fransisco (simulated)	The shape of MFDs can be theoretically explained
Buisson and Lavier [81]	historical	Toulouse urban + urban motorway	There is scatter on the MFD if detectors are not ideally placed or at the presence of inhomogeneous congestion
Ji et al. [82]	simulated	urban + motorway	Hybrid networks → scattered MFDs, inhomogeneous congestion reduces flow, should be included in network control scheme
Cassidy et al. [83]	historical	3 km motorway	MFDs on motorways hold for supercritical or subcritical traffic conditions, otherwise points appear within MFD
Mazloumian et al. [44]	simulated	urban grid - periodic boundary flows	Spatial variability of density is important in deriving performance
Geroliminis and Ji [84]	historical	Yokohama	Spatial variability of density is important in deriving performance
Wu et al. [85]	historical	900 m arterial	Arterial FD exists, influenced by timing plans
Daganzo et al. [45]	simulated	grid	Equilibrium states in a network are either free flow or highly supercritical. Rerouting increases critical density for supercritical states
Gayah and Daganzo [86]	simulated	grid/bin	Hysteresis loops appear in Urban-scale MFDs due to faster recovery of uncongested parts. Can be reduced with increased driver adaptability

1.1.1 Contributions

In this thesis, the following contributions to Urban-scale MFD literature are made:

1. The first contribution is the feature weighted extension of k-harmonic means clustering and subsequent implementation of time-dependent urban traffic network partitioning into homogeneous regions. While k-means clustering and a variant of spectral clustering have already been implemented in [25], no past traffic-related literature, to the author's knowledge, has extended and implemented the feature weighted k-harmonic means clustering, which, for the specific network test case, outperforms both aforementioned clustering methods.

2. The second contribution is the introduction and implementation of a multi-class extension of a region-based dynamic traffic model called Network Transmission Model for the purpose of mixed application of different types of route guidance, to showcase the interaction of different traveler classes, on a regional traffic network. This region-based dynamic traffic model extension integrates the capability to efficiently deal with different types of route guidance and different classes of travelers.

3. The third contribution is the development and implementation of practicable multi-type/multi-traveler-class route guidance methodologies, including a non-predictive and a predictive region-based route guidance method with pre-trip information dissemination, as well as the introduction and implementation of a region-based variant of the Multinomial Logit Routing (MLR) method [87]. The non-predictive routing method is a strategic learning approach to region-based route guidance, employing Proxy Regret Matching (PRM). The predictive simulation-based routing approach called Incremental Route Planning (IRP) comes integrated with a Public Transit Diversion (PTD) mechanism.

Both the non-predictive and the multinomial logit routing methods (PRM and MLR for short) have also been extended with Public Transit Diversion integration.

4. The last contribution is the introduction of a market penetration scheme where autonomous vehicles, RGIS-equipped and unequipped vehicles comprise the vehicle population. In past traffic-related literature, market penetration schemes are presented, however, the vehicle population is never composed of a combination of more than two of the aforementioned vehicle types. In the proposed market penetration scheme, each traveler class features distinct characteristics and is, thus, being represented by different types of route guidance. This particular combination of traveler classes has never been implemented simultaneously in a simulation framework before, which can be considered as a big gap in the literature, given the fact that autonomous vehicles will soon comprise a non-negligible part of the vehicle population in metropolitan cities. This particular market penetration scheme is used to evaluate regional urban traffic network performance for varied combinations of traveler class penetration rates and are able to demonstrate the improved performance robustness of non-predictive and predictive routing approaches as they interact with self-interested travelers employing multinomial logit routing (MLR).

In short, this dissertation contains several innovative methodologies, including predictive and non-predictive methods for region-based route guidance, which have not been used before in a region-based traffic context and most importantly, comprises of elements which have been extended and combined to construct a Region-based Urban Traffic Simulation and Management Framework. This Framework integrates such innovative concepts as a combined simulation of autonomous vehicles, RGIS-equipped vehicles and unequipped vehicles, as well as Public Transit, thus rendering said framework as a very appealing solution to transportation authorities, that wish

to plan for the inexorable advent of vehicle autonomy and its potential impact to metropolitan city networks. In view of the smart city initiatives undertaken by government organizations around the world, this work appears as most timely.

1.1.2 Thesis Overview

This thesis comprises a total of seven chapters. Brief descriptions for each individual chapter can be found below:

Chapter 2: This chapter provides concise descriptions of the modules proposed for a Region-based Urban Traffic Simulation and Management Framework. More attention is given to the modules which have already been implemented in this thesis, including the partitioning module, the macroscopic dynamics module and the optimization module. More in-depth descriptions are given regarding the traveler classes used in the multi-class extension of the Network Transmission Model, which is part of the macroscopic dynamics module and the predictive and non-predictive routing methods implemented as part of the optimization module.

Chapter 3: This chapter describes the first important step in applying the Urban-scale MFD concept on a region-based urban traffic management framework. An urban traffic network partitioned into homogeneously congested regions is a fundamental requirement for the validity or even the existence of Urban-scale MFDs. After a brief description of the theory behind center-based and spectral clustering approaches, weighted feature extensions of k-means and k-harmonic means clustering, as well as normalized spectral clustering are implemented and compared on a network consisting of about 90 junctions and 180 links, encapsulating the CBD area of Singapore. Historical data from the morning peak period are provided by the Land Transport Authority (LTA) of Singapore are used to implement time-

dependent partitioning into homogeneously congested regions.

Chapter 4: This chapter introduces the multi-class extension of a region-based dynamic traffic model, the Network Transmission Model, which is used in later chapters to simulate region-based traffic dynamics and provide traffic state information to the non-predictive and predictive routing approaches introduced in chapter 5.

Chapter 5: This chapter introduces and compares non-predictive and predictive routing methods. A strategic learning-based routing method, which employs Proxy Regret Matching (PRM) that provides a distribution of regional paths to be assigned based on their estimated average regret, and a simulation-based approach, Incremental Route Planning (IRP), which integrates a Public Transit Diversion (PTD) mechanism and comprises 2 steps, forecasting and path set computation/assignment. A Multinomial Logit Routing (MLR) method is also implemented to provide a more realistic route choice behavior. Both PRM and MLR are extended to integrate a PTD mechanism and are subsequently compared with IRP.

Chapter 6: This chapter introduces a market penetration scheme, wherein three distinct traveler classes, a class of travelers equipped with autonomous vehicles, a traveler class comprising of human-driven vehicles equipped with a Route Guidance and Information System (RGIS) and a traveler class comprising of human-driven unequipped vehicles are defined. The 1^{st} traveler class is assumed to be fully compliant and is represented by IRP, while the 2^{nd} traveler class is represented by PRM and may include non-compliant human drivers, which would be subsumed in the 3^{rd} class of travelers, which is represented by MLR. For various combinations of market penetration and non-compliance rates, overall network performance, as well as individual traveler class performance is evaluated.

Chapter 7: This chapter contains the conclusions of this thesis, as well as proposed future work.

Chapter 2

Region-based Urban Traffic Simulation and Management Framework

In this chapter, a concise description of the Region-based Urban Traffic Simulation and Management Framework is provided. The framework in question can be characterized as modular, in the sense that, most elements of the framework can be added or removed, depending on engineering requirements and limitations. Certain core modules of this framework are implemented in this dissertation. The full Region-based Urban Traffic Simulation and Management Framework is expected to be part of a project proposal in the future.

2.1 Framework Description

The proposed framework presented on Figure 2.1, comprises of several modules which will be described in the following subsections.

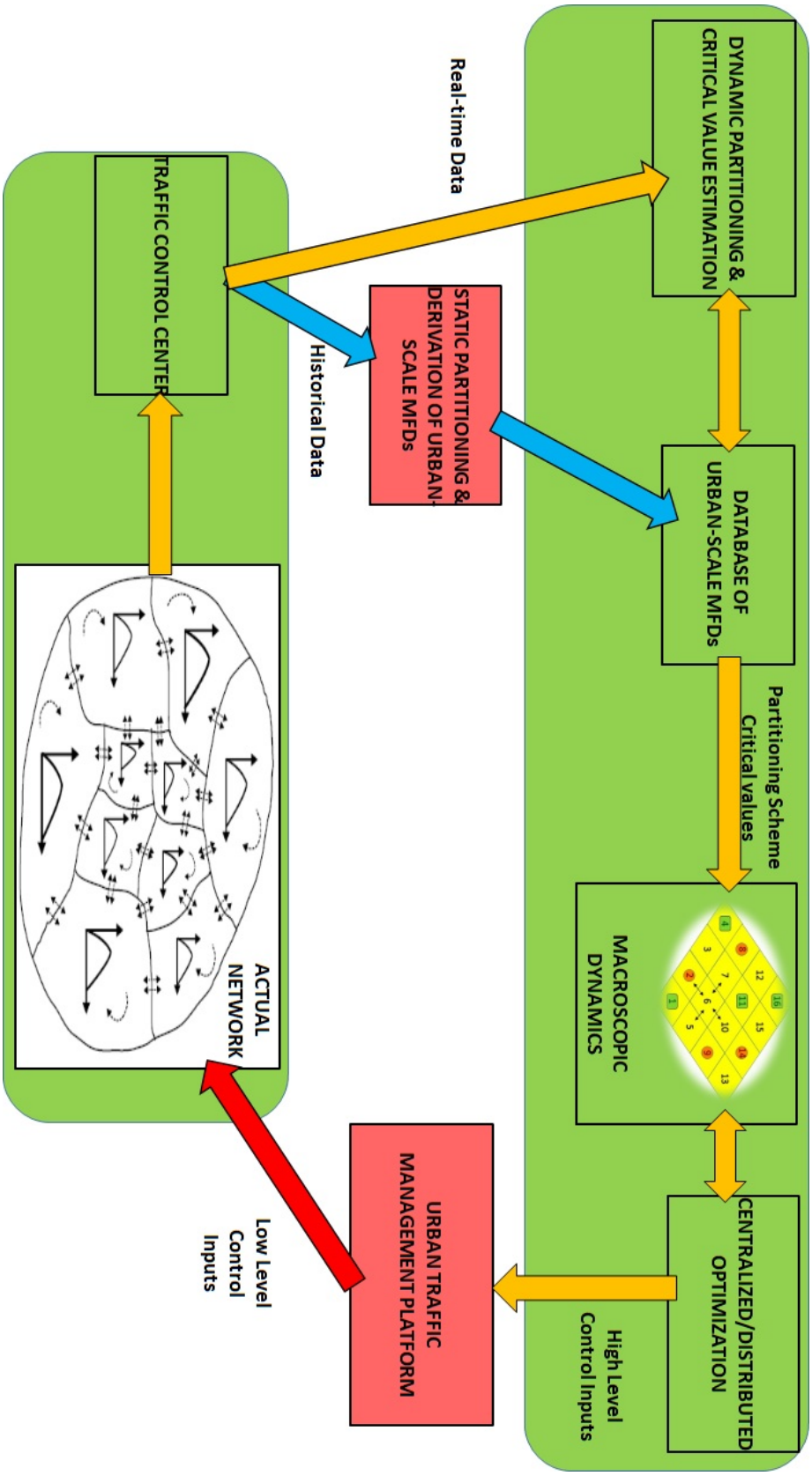


Figure 2.1: Proposed Framework for Region-based Urban Traffic Simulation and Management

2.1.1 Traffic Operations Control Center

Traffic Operations Control Centers can be found in most metropolitan areas. They are responsible for day-to-day operations including traffic monitoring, managing data collected from an array of sensors (loop detectors, traffic cameras, GPS, Road Side Units), deployment of maintenance crews in road work-zones, relaying of advisory information to travelers in real-time, through VMS (Variable Message Signs) or ATIS (Advanced Traveler Information Systems).

2.1.2 Partitioning System

The partitioning of an urban traffic network into several homogeneously congested regions can be static, based on historical data describing recurring traffic patterns, dynamic, based on real-time data describing prevailing traffic conditions. A time-dependent implementation could initially use historical data to partition the urban traffic network into homogeneously congested regions and in subsequent time steps improve the regional boundaries definition. The first steps towards a time-dependent implementation, as defined here, have already been taken in this work, using historical data provided by the Land Transport Authority (LTA) of Singapore, with very promising results pertaining to partitioning methods performance, regarding region compactness, cohesion and execution time.

2.1.3 Urban-scale MFD Database

Using the regions resulting from static partitioning, an Urban-scale MFD database can be constructed off-line. This database can be used to quickly identify critical values corresponding to specific region shapes, thus decreasing the computational effort required during the regional dynamic traffic simulation that follows. Urban-scale MFDs can also be derived for the regions resulting from time-dependent partitioning

and the critical accumulation values can similarly be updated on-line.

2.1.4 Macroscopic Dynamics

Any number of regional dynamic traffic models may be used to simulate regional traffic dynamics. The majority of these models fall under the macroscopic traffic modeling category. Microscopic traffic model implementation is not advisable for real-time applications, due to issues of scale and complexity. In this work, a multi-class extension to the Network Transmission Model is introduced and implemented. Three distinct traveler classes are subsequently defined, a class of travelers equipped with autonomous vehicles, a traveler class comprising of human-driven vehicles equipped with a route guidance and information system and a traveler class comprising of human-driven vehicles without any sort of navigation equipment. The 1^{st} traveler class is assumed to be fully compliant, while the 2^{nd} traveler class may include non-compliant human drivers, which would be subsumed in the 3^{rd} class of travelers. I consider 2 distinct market penetration rates, one for the adoption of private autonomous vehicles as the primary mode of transport, assigned to the 1^{st} traveler class (MPR1), and one for the equipment of route guidance information systems utilizing strategic learning-based routing assigned to the 2^{nd} traveler class (MPR2). Subsequently, after the respective market penetration rates for the different traveler classes have been set, a rate of non-compliance (NC) for the 2^{nd} class of human-driven vehicles is defined and after grouping the non-compliant segment together with the remainder of the vehicle population, classified as the 3^{rd} traveler class, rate (MPR3) is determined. The diagram shown in Figure 2.2 depicts the aforementioned classification process.

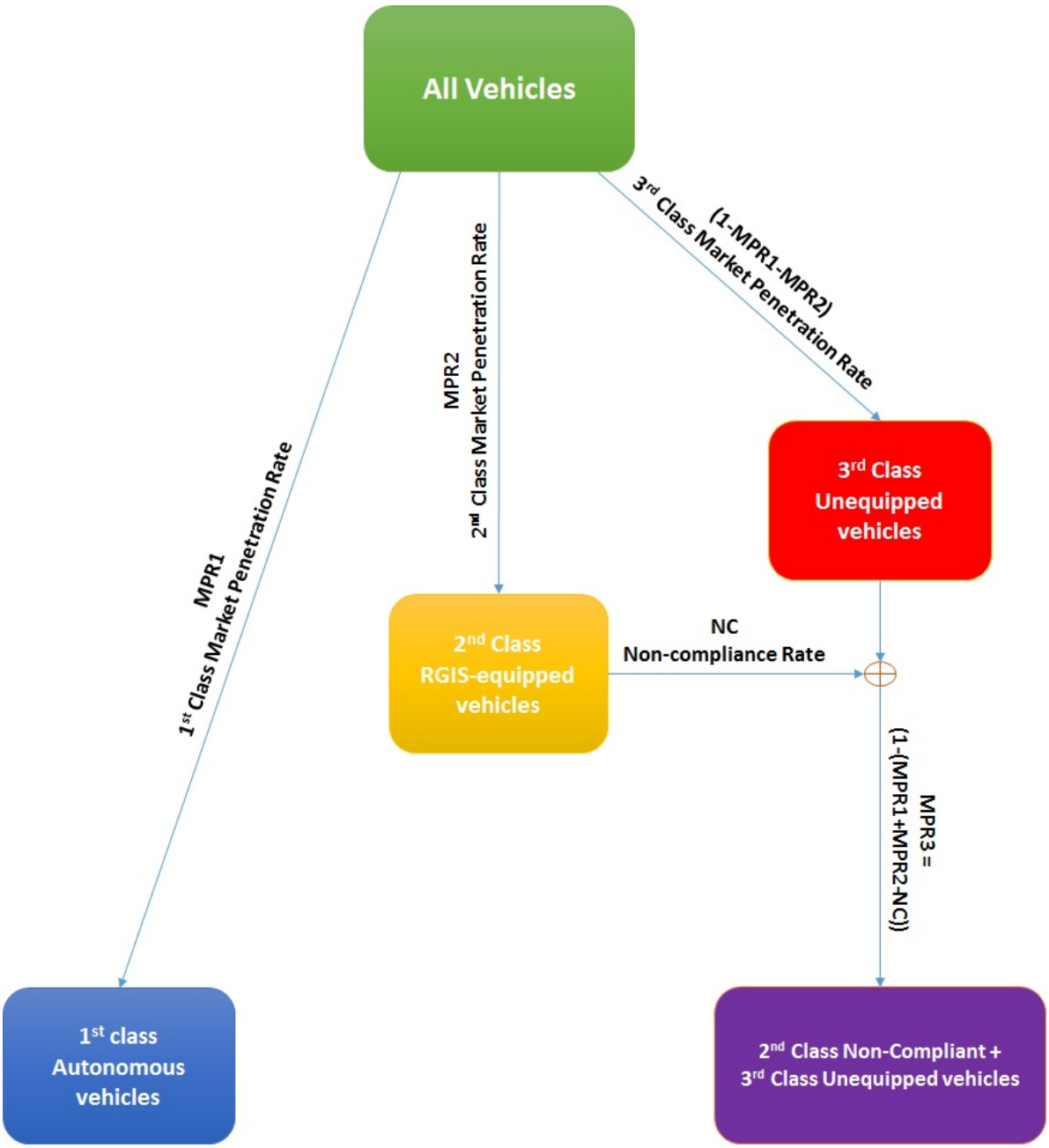


Figure 2.2: Process of assigning vehicle population to multiple traveler classes.

2.1.5 Centralized/Distributed Optimization

After having received information regarding the traffic state from the regional dynamic traffic model that has been selected for simulation, the regional path selection from Origin region to corresponding Destination region is considered as the control input and, depending on the optimization approach (centralized or distributed) and objective function (user-oriented or system-oriented) selected, appropriate route advisory information is provided to all travelers. In this work, prescriptive pretrip guidance approaches are employed, including a predictive simulation-based approach, Incremental Route Planning (IRP), which comprises 2 steps, forecasting and path set computation/assignment, for the 1^{st} traveler class, and Proxy Regret Matching (PRM), a non-predictive strategic learning-based approach, which provides a distribution of regional paths to be assigned based on their estimated average regret, for the compliant segment of the 2^{nd} traveler class. Finally, an extended version of Multinomial Logit-based Routing (MLR) was implemented, which gives the multinomial logit-based Dynamic Stochastic User Equilibrium, DSUE for brevity, representing a more realistic route choice behavior for the non-compliant segment of the 2^{nd} traveler class and the 3^{rd} traveler class of unequipped human-driven vehicles, both characterized by imperfect travel time perception.

2.1.6 Urban Traffic Management Platform

Resulting regional paths can be considered as a high level control input which can be translated into low level control input, through an urban traffic management platform that calculates time-dependent shortest link-based paths in the regions belonging to the respective regional path. It should be noted that additional constraints are required to ensure that the intraregional paths in the sequence of regions belonging to each respective regional path are connected along the region bound-

aries.

2.1.7 Chapter Overview

This chapter provides brief information about the proposed modules of a Region-based Urban Traffic Simulation and Management Framework. Additional focus falls on modules which have been partially or fully implemented in the course of this thesis, including the Partitioning System, the Macroscopic Dynamics and the Centralized/Distributed Optimization. The author's work on time-dependent urban traffic network partitioning is described in detail in the following chapter.

Chapter 3

Urban Traffic Network Partitioning

In this chapter, my implementation of urban traffic network partitioning is described, as well as some technical details about the center-based and spectral clustering approaches used, feature weighted k-means clustering, feature weighted k-harmonic means clustering and normalized spectral clustering.

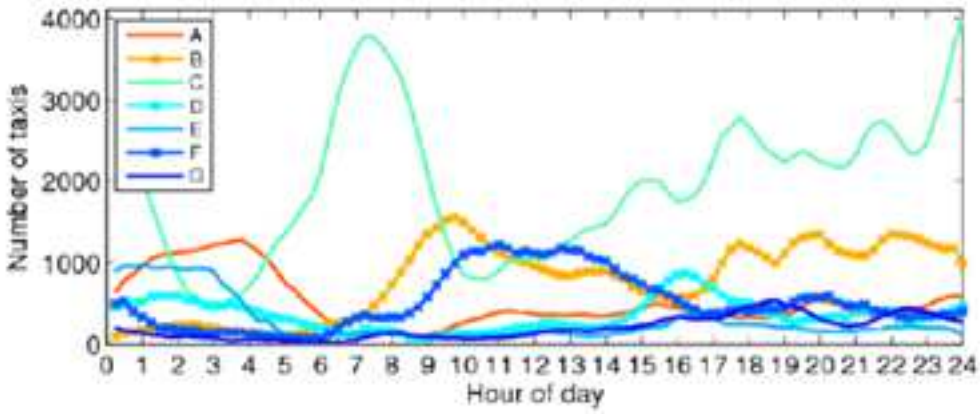
3.1 Prior work and process description

It is well documented that different traffic patterns appear throughout the day. It follows that different partitions of the same urban traffic network need to be attributed to different time periods. Implementation of partitioning can be based on historical data, real-time data, different demand scenario-based micro-simulations or any combination thereof. Partitioning into homogeneously congested regions has been investigated extensively in the last 5 years, starting with Ji and Geroliminis' seminal paper [25] and followed by several others [26,29,55]. Both simulation-based and real-life datasets have been used to derive homogeneously congested regions, either statically or dynamically. As mentioned previously, Geroliminis et al. [88]

developed a regional dynamic traffic model, which was extended in [14,54,89] to model and control for heterogeneity in MFD at the regional level. However, [14,54] perform a secondary partition of each heterogeneous region into homogeneous subregions, where MFD functions are assumed to be dependent only on the subregional accumulation. In this case, an appropriate partitioning implementation would be selected, dependent on the type of data available, as well as other parameters of the given network, so as to achieve the best possible results for two features, intraregional link density variability and regional compactness and cohesion. Coming back to Ji and Geroliminis [25] seminal work, the scenario-based micro-simulation approach was chosen for a 6.5 km^2 area of San Francisco consisting of about 100 junctions and around 400 network links. Their choice enabled them to pinpoint appropriate times when network congestion has not set yet and the traffic volumes are still high, providing them with a variety of link densities for their algorithms. My network, representing the Marina Bay area of the city of Singapore, also incorporates parts of Chinatown, Clarke Quay, Dhoby Ghaut and the Central Business District. In this chapter, I elected to use historical data from loop detectors situated upstream of the majority of junctions for all directions. While historical data can provide more accuracy than scenario-based simulations, as regards the link density values, the question arises as to how to pinpoint the appropriate times with conditions that guarantee adequate link density variability to warrant partitioning of the selected network. Aslam et al. [2] used loop detector data obtained from the Land Transport Authority in Singapore and collected travel data from taxis for the month of August 2010. Machine learning was employed for vehicle distribution and traffic volume inference for the entire island of Singapore. Thus, they were able to identify hotspot areas throughout Singapore for specific time periods. The traffic hotspot areas defined through the inferred traffic volumes as can be shown in Figure 3.1.

Certain areas of Singapore exhibit high volume throughout the entire day. Other

(a) Hotspots with different daily volume patterns



(b) Volume variation of hotspots over time

Figure 3.1: The labeled hotspot areas on the top map correspond to the volume – time-of-day plots. Aslam et al. [2]

areas are characterized by high volumes only during respective peak hours. Figure 3.1(a) pinpoints the busiest areas with respect to traffic volume. In figure 3.1(b), the traffic volumes throughout the day at the respective high volume areas can be observed. The area labeled as C exhibits the highest volumes. This particular area is located around Changi airport. It should be discounted since airports constitute a source of attraction for taxis in particular and not so much for general traffic. Their results enabled the selection of times for the test network which can guarantee the link density variability required for successful partitioning of the test network into homogeneous regions. In this chapter, I derive a partitioning for the network using weighted k-means clustering, k-harmonic means clustering and spectral clustering algorithms. More specifically, I make use of the location coordinates and the vehicle density for each link in the test network as distinct features. Subsequently, I apply weights to each feature. In this manner, the clustering algorithms can produce clusters where link density variability is low within each cluster, without sacrificing spatial cohesion. Concretely, producing clusters whose links are connected to all other links within each cluster without ever having to traverse a path that crosses a different cluster. A performance metric developed by Ji and Geroliminis [25] is utilized to evaluate the intra- and inter-cluster link density variability and I also test for the spatial cohesion of each resulting cluster. The results show that the optimal number of clusters for this particular network is 2, while k-harmonic means clustering and normalized spectral clustering provide the most stable results.

3.2 Center-based and spectral clustering algorithms

The k-means algorithm belongs to the category of center-based clustering algorithms [90]. Center-based clustering algorithms utilize a number of centers to represent and subsequently partition input data. Each center defines a cluster with a

central point and perhaps a covariance matrix. This family of algorithms usually begins with a guess about the solution and then refine the positions of centers until reaching a local optimum. These methods can work well, but they can also converge to a local minimum that is far from the global minimum, i.e. the clustering that has the highest quality according to the criterion in use. Converging to bad local optima is related to sensitivity to initialization, and is a primary problem of data clustering. A general model for the family of center-based algorithms that use iterative optimization follows.

Define a f-dimensional set of n data points $D = \{d_1, ..., d_n\}$ as the data to be clustered. Define a f-dimensional set of k centers $M = \{m_1, ..., m_k\}$ as the clustering solution that an iterative algorithm refines.

A membership function $e(m_j|d_i)$ defines the proportion of data point d_i that belongs to center m_j with constraints $e(m_j|d_i) \geq 0$ and $\sum_{j=1}^{k} e(m_j|d_i) = 1$. The k-means algorithm uses a *hard* membership function defined below

$$e_{KM}(m_l|d_i) = \begin{cases} 1 & \text{if } m_l = arg\min_j \|d_i - m_j\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

One of the reasons that k-means algorithm can converge to poor solutions is due to its *hard* membership function. The k-harmonic means algorithm described, later on, uses a *soft* membership function defined below

$$e_{KHM}(m_l|d_i) = \frac{\|d_i - m_j\|^{-q-2}}{\sum_{j=1}^{k} \|d_i - m_j\|^{-q-2}} \tag{3.2}$$

where exponent q takes even values.

A weight function $v(d_i)$ defines how much influence data point d_i has in recomputing the center parameters in the next iteration with constraint $v(d_i) > 0$. Zhang et al. [91] first introduced a dynamic weight function that gives more weight to data

points not "well-covered" by the current solution

$$v_{KHM}(d_i) = \frac{\sum_{j=1}^{k} \|d_i - m_j\|^{-q-2}}{(\sum_{j=1}^{k} \|d_i - m_j\|^{-q})^2} \tag{3.3}$$

Now a general model of iterative, center-based clustering can be defined. The steps are:

1. Initialize the algorithm with guessed centers M.

2. For each data point d_i, compute its membership $e(m_j|d_i)$ in each center m_j and its weight $v(d_i)$.

3. For each center m_j, recompute its location from all data points d_i according to their memberships and weights:

$$m_j = \frac{\sum_{i=1}^{n} e(m_j|d_i)v(d_i)d_i}{\sum_{i=1}^{n} e(m_j|d_i)v(d_i)} \tag{3.4}$$

4. Repeat steps 2 and 3 until convergence.

Center-based clustering algorithms are usually characterized by their objective function. The k-means algorithm clusters data points in k-sets, where k represents the number of cluster centers. Voronoi partitions are formed according to data point membership to each cluster. The k-means algorithm uses the Euclidean norm to compute the distance of each data point from the selected cluster centers and then updates the centers according to the means of the data points associated with that cluster. Weighted k-means clustering uses feature weights to signify the importance of one particular feature over others. It should be noted that the weights $v(d_i)$ mentioned above are applied to each individual data point, while the feature weights apply to one particular feature of the entire data set. Weighted k-means clustering uses feature weights to signify the importance of one particular feature over others.

It should be noted that the feature weights apply to one particular feature of the entire data set. In this implementation, w_c is defined as the weight for the link coordinates feature and w_ρ as the weight for the density feature, leading to the following equation:

$$WKM(D, M) = \sum_{i=1}^{n} \sum_{f \in \{c,\rho\}} \min_{j=\{1,\ldots,k\}} w_f \|d_{if} - m_{jf}\|^2 \tag{3.5}$$

where $d \in D \in \mathbb{R}^{n\times 3}$ represent the data points, $m \in M \in \mathbb{R}^{k\times 3}$ represent the cluster centers and k is the number of clusters. The reason I enforce weights is to ensure the spatial cohesion of the resulting partitions, otherwise, the resulting clusters will consist of spatially disparate links with admittedly very low density variability. One other main caveat in employing k-means clustering lies, as mentioned before, in the fact that the resulting data point clusters may not represent global minima, as regards to the feature variability meant to be minimized and are highly dependent on the initialization of the cluster centers. This weakness can be partially mitigated if one selects, using prior knowledge, particular data points as the initial cluster centers, but information at this level of detail is usually hard to obtain.

I chose to apply a weighted variation of k-harmonic means algorithm as the resulting clusters are insensitive to cluster center initialization. The k-harmonic means algorithm [91] objective function is characterized by the use of the harmonic mean instead of the Euclidean norm of the distance of each data point from the cluster centers. Although k-harmonic means includes a weight function that rewards data points with greater distance from their respective cluster centers, due to the particular nature of partitioning I aim to implement, I include feature weights as in the case of the k-means algorithm, resulting to the following equation:

$$WKHM(D,M) = \sum_{i=1}^{n} \frac{k}{\sum_{j=1}^{k} \sum_{f\in\{c,\rho\}} w_f \|d_{if} - m_{jf}\|^{-q}} \tag{3.6}$$

Even exponent q values are suggested in the literature, in this implementation, $q = 4$ was chosen.

As an alternative to the k-means clustering algorithm family, I select to implement a version of normalized spectral clustering [92], which is similar to the Normalized Cut algorithm implemented by Ji and Geroliminis [25]. The advantage of this approach is the fact that feature weights are unnecessary, because spectral clustering treats clusters as connected subgraphs, while the k-means algorithm family treats them as distinct convex sets [93]. Additionally, this approach ignores sparse interconnections among subgraphs, providing a better defined boundary.

The selected urban traffic network has two representations. First, as an undirected graph $G = (V, E)$ where each node $i \in V$ represents a link. Each node is associated with a density value ρ_i. It has to be denoted that two way road link connections are considered separate undirected edges. Link density values were derived from flow values in the following manner. The raw loop detector data was aggregated to $\tau_s = 300s$ (Sampling rate $= 1/300Hz$). The data was selected in such a way, so that the initial values for link density can be considered close to zero, i.e. at a time before the morning peak period, where flow values were very low and almost zero. Then the link density values were iteratively updated for each link as follows. First, for every link $i \in V$, I made a list of all immediately upstream links J_i. Subsequently, for each $j \in J_i$, I summed up the number of lanes λ_j. Then, the incoming flow to link i from all immediately upstream links $j \in J_i$ is $qin_i = \sum_{\forall j\in J_i} \frac{\lambda_{ji}}{\lambda_j} qout_j$, where λ_{ji} is the number of lanes from link j directed to link i, $qout_j$ is the outflow for link j and λ_j is the total number of lanes of link i. Then, link density will be $\rho_i^t = \rho_i^{t-\tau_s} + \tau_s \left(qin_i - qout_i\right) / \Lambda_i \lambda_i$, where Λ_i and λ_i represent the length and number of lanes of link i, respectively. The superscript t, indicating specific time, has been omitted

from following instances of link density ρ_i for convenience. This first representation is appropriate for the implementation of the k-means and k-harmonic means clustering algorithms. The second representation as a weighted undirected graph is better suited to the implementation of the normalized spectral clustering algorithm. The normalized spectral clustering algorithm can be described by the following steps:

1. The weights $\beta_e \in B \in \mathbb{R}^{n\times n}$ on the edges belong to the weighted adjacency matrix B and are described by the similarity function between link density values which takes the form of a Gaussian probability distribution function

$$\exp\left(-\left(\rho_i - \rho_j\right)/2\sigma^2\right), \forall i, j \in V$$

   which punishes density value differences, decreasing in value faster the larger the differences

2. After obtaining the weighted adjacency matrix B, the degree matrix needs to be calculated, a diagonal matrix $R \in \mathbb{R}^{n\times n}$ whose diagonal elements can be calculated as follows:

$$r_i = \sum_{j=1}^{n} \beta_{ij}, \forall i \in V$$

3. The normalized Laplacian matrix $L_s \in \mathbb{R}^{n\times n}$ is then calculated by

$$L_s = I - R^{-1/2} B R^{-1/2}$$

4. Subsequently, the first k generalized eigenvectors $y_1, ...y_k$ are calculated by solving the equation

$$L_s y = \lambda R y$$

5. Normalization of the rows of matrix $Y \in \mathbb{R}^{n\times k}$ to L^1-norm is followed by implementation of k-harmonic means clustering to matrix $Y_s \in \mathbb{R}^{n\times k}$.

6. Finally, a set of clusters $C = \{C_1, ..., C_k\}$ which correspond to the initial data points cluster membership is obtained.

3.3 Performance metrics

Ji and Geroliminis [25], based on metrics developed by Rousseeuw [94], derived metrics to measure the performance of the Normalized Cut algorithm in comparison with a weighted k-means clustering implementation and showed that for a set of clusters $C = \{C_1, ..., C_k\}$ one can measure the intra-cluster density variability, as well as the inter-cluster density variability conditioned on the adjacency of the clusters being compared, as follows:

$$CV(C_v) = \frac{2var(C_v)}{var(C_v) + var(C_u)\left(\mu_{C_v} - \mu_{C_u}\right)^2} \tag{3.7}$$

where C_v is the cluster whose performance I aim to measure and C_u is the adjacent cluster with the least inter-cluster variability as regards C_v. Small values for the variances denoted by $var(C)$ and large difference between the means, denoted by μ_C, gives small values for CV which means the cluster partitioning is successful. It has to be noted that the average CV over all clusters is taken to account for the difference in link density among adjacent clusters. In order to measure spatial cohesion for each cluster, I decided to use the silhouette coefficient developed by Kaufman and Rousseeuw [95], which, for each data point measures two quantities, cohesion $\alpha(x)$, which measures average distance between data points within the cluster and separation $b(x)$, which measures the minimum average distance of data points to other clusters. Then silhouette $s(x)$ is defined as

$$s(x) = \frac{b(x) - \alpha(x)}{max\left\{\alpha(x), b(x)\right\}} \tag{3.8}$$

where $s(x) \in [-1, 1]$ with values close to -1 considered as spatially poorly connected and values close to 1 considered as spatially well connected. The silhouette coefficient is the average for all data points and can be calculated using

$$SC = \sum_{i=1}^{N} s(x_i)/N \tag{3.9}$$

In addition to calculating the isolated performance for spatial cohesion, this metric can be adapted to measure the total performance of the partitioning, by simply selecting both spatial and density features for each data point x_i. TP will be used to represent this metric for the remainder of this chapter.

$$TP = \sum_{i=1}^{N} s(x'_i)/N \tag{3.10}$$

where each corresponding datapoint x'_i includes information about both spatial (geographic coordinates) and density features. It should be noted that for extreme values on either end of the spectrum, TP should have a limited impact during the clustering algorithm comparison, since its value is skewed by the extremely low or high values.

3.4 Network Setup

As mentioned before, using prior information about hotspot areas throughout Singapore for specific time periods, I selected an urban region with an area of about 2.2 km^2, located to the south of the island of Singapore, encompassing the Marina Bay area, as well as parts of the Central Business District, Chinatown, Dhoby Ghaut, and Clarke Quay. My network consisted of about 90 junctions and 180 links. It should be noted that bidirectional road segments are considered as separate links. Link lengths range from 100m to 1000m and the speed limit is set to 50km/h. In-

stead of Cartesian coordinates, Geographic coordinates (decimal degrees) are used in the plots, with longitude on the x-axis and latitude on the y-axis. Loop detector data obtained from the Land Transport Authority in Singapore for the month of August 2010 were used, based on Aslam et al. [2] results regarding peak traffic volumes in that particular area, it was decided that the tests should be implemented for a time period of 3 hours, starting from 8am and ending at 11am, based on historical data for the first Monday of August 2010. The reason I choose time 9am as the initial partitioning is that the network is partitionable at this time (optimal number of clusters is larger than 1). With a well partitioned network at a certain time, the growing and shrinking phenomena of congested regions in the network can easily be captured. On the contrary, if I use 7am or 12pm as initial partitioning when there are no obvious clusters in the network (i.e., the network is homogeneously congested or uncongested), the proposed time variant partitioning would not produce desired clusters with compact shapes when it reaches partitionable time periods. The algorithm would only work on a network that has already been roughly well partitioned after some preprocessing, such as initial segmenting and merging. The network can be shown in Figure 3.2

Figure 3.2: Illustration of the urban traffic network representing the Marina Bay Area of Singapore, which includes parts of Chinatown, Clarke Quay, Dhoby Ghaut and the Central Business District

3.5 Experimental results

The main objective was to produce clusters with low intra-cluster density variability and high inter-cluster density variability. This objective, however, must satisfy the constraint of spatial cohesion. Having a cluster with very low link density variability, which includes links with absolutely no connectivity or spatial cohesion is useless for what I am trying to achieve. Another goal was to find the minimum number of clusters required for a sufficiently homogeneous partitioning. Several test runs were implemented using all three algorithms. More specifically, I tested for different numbers of clusters, for different feature weights and for different times within the selected period of 8am to 11 am. It was determined that the optimal number of clusters is 2, which is understandable, since the encompassing area is not large enough to warrant further partitioning. After several runs, I was able to determine feature weights suitable for keeping a balance between the primary objective

and the satisfaction of the spatial cohesion constraint. Finally, for 2 clusters and a weight ratio w_c/w_ρ of 100/1, I compare all algorithms throughout the 8am to 11am time period.

The number of clusters for which the partitioning performance was investigated ranges from 2-4. There were runs for $k = 5$, but the resulting 5th cluster was empty, which provided information regarding the upper bound for this investigation. Obviously, if the performance indices for 2 clusters were also too high as regards link density variability, the only conclusion is that the selected network was homogeneously congested in its entirety. Fortunately, this was not the case, which means the region and time period selection was not made poorly. Due to the inherent tendency of k-means to reach local minima, dependent on the initial cluster center selection, a range of values for each cluster is obtained. Hence, Table 3.1 contains the average performance metric values derived from the k-harmonic means implementation, which exhibits insensitivity to initial cluster center selection. It should be noted, for this particular investigation, that the average of the range of values mentioned above and the performance metric values taken from the k-harmonic means algorithm implementation differ very slightly.

Table 3.1: Cluster number comparison

Performance Metrics	k**-number of clusters**		
	$k = \mathbf{2}$	$k = \mathbf{3}$	$k = \mathbf{4}$
Average CV	**0.8705**	1.0487	0.8183
Average SC	**0.4713**	0.6475	0.3966
Average TP	**0.0957**	-0.3217	-0.721

It is evident that 2 clusters partitioning performs better in total. It should be noted that upper thresholds for CV and lower thresholds for SC and TP are set in order to facilitate the comparison. If the average CV value exceeds 1.0000, the partitioning is considered unsuccessful due to high inter-cluster link density similarity. If the average SC or TP value is below 0.0000, the partitioning is considered unsuccessful

with respect to spatial cohesion, or a combination of low spatial cohesion and high inter-cluster similarity.

For the aforementioned reason, I elect to utilize the k-harmonic means algorithm to perform the comparison for different feature weights. Table 3.2 contains the average performance metric values for different feature weight ratios w_c/w_ρ, ranging from 10/1 to 150/1.

Table 3.2: Feature weight comparison

Performance Metrics	w_c/w_ρ				
	10	**25**	**50**	**100**	**150**
Average CV	0.0626	0.0814	0.1530	**0.8705**	0.9720
Average SC	-0.0749	-0.0578	-0.0566	**0.4713**	0.4782
Average TP	0.8740	0.8728	0.8647	**0.0957**	0.0285

Admittedly, the weighting for spatial cohesion is quite aggressive, starting from merely 10/1 and eventually reaching 10 times the initial ratio, but based on a combination of performance metric values and graphical evidence, it was deemed necessary so as to enforce the critical constraint of spatial cohesion. From the performance metric values it can be easily observed that, had the isolated spatial cohesion metric not been included, the weight ratio selection would have been off by a magnitude of 10.

I investigated the performance of each algorithm for several time slots within the selected time period of 8am to 11am. It is interesting to observe that the partitioning produced from each algorithm implementation displays the progression of traffic congestion throughout the time period. In Table 3.3 I compare all three algorithms, but, it should be pointed out that the k-means produces several sets of results for each time slot, dependent on the initial cluster center selection. Therefore, I have decided to present values to showcase that good performance can be achieved, but the instability of the results should lead to their dismissal from the algorithm comparison in the general case. For the majority of the time slots tested, k-harmonic

means clustering outperformed normalized spectral clustering. Figures 3.3 and 3.4 graphically present the clusters produced k-harmonic means and normalized spectral clustering respectively, for time slots 8am, 9am, 10am and 11am. The progression of traffic congestion throughout the selected time period can easily be observed. Observation of the clusters shows, that an ebb and flow effect is taking place for that specific time period, which is consistent with peak traffic volume observations of Aslam et al. [2].

Table 3.3: Clustering algorithm comparison

Performance Metrics	**Type of clustering algorithm**		
	WKM	**WKHM**	**NSC**
Average CV	0.9237	**0.9125**	0.9451
Average SC	0.4753	**0.4743**	0.4703
Average TP	0.1555	**0.0706**	0.0305

From the experimental results, several conclusions can be reached.First, that the minimum number of clusters is dependent on the selected network topology and size. Second, that center-based clustering algorithms can be very effective in partitioning urban traffic networks into homogeneous regions, but stability remains an issue. Third, that spectral clustering approaches are suited to urban traffic network partitioning, omitting the tedious step of calculating optimal feature weights. Spatial cohesion is satisfied by the construction of the weighted adjacency matrix B. However, there is some computational effort involved in the adjacency matrix construction, which may lead to spectral clustering approaches being slower than center-based clustering approaches such as k-means and k-harmonic means. The results also show that the weighted version of k-harmonic means outperforms normalized spectral clustering, as regards link density variability and spatial cohesion. One must consider however that in order to achieve the same level of spatial cohesion, very aggressive weighting of the coordinates feature had to be employed. Finally, after examining the different clusters produced from the clustering algorithms, one

can deduce how congestion is formed and propagated over time. This pertains to historical data, which might not be reflecting the current traffic situation for that particular area, however this approach can be extended to apply to real-time data, or a combination of historical and real-time data. These approaches can be used as a middle layer of traffic surveillance, enabling traffic engineers to observe in real-time the traffic state, without overloading them with information. They can also serve as a first step in the derivation of the Urban-Scale Macroscopic Fundamental Diagram for each separate cluster, thus enabling the application of different perimeter control strategies along the boundary of the clusters.

3.6 Discussion

Urban systems experience highly dynamic behavior and different traffic patterns may arise for different times of day (think of morning-evening commuting patterns or stochastic variations of traffic flow). In these cases, very likely one needs to identify different optimal sets of clusters depending on these patterns.The developed algorithms of this report, given their short computational time (a few seconds), can be directly applied real-time in these cases. Nevertheless, further research is needed to identify how often a partition should be adjusted. My understanding is that dynamic clustering can be performed at a time resolution that it is smaller than the control decisions, e.g. if control decisions for traffic signals through perimeter control are made every 5 min, clustering might need to be performed every 15-30 min. This problem has a strong link with the spatio-temporal propagation of congestion in transportation networks. Regarding the online part of the algorithm, one can check if the variance of the predefined offline clusters exceeds some threshold values, which would indicate that the network would require repartitioning. Note that these thresholds do not need to be extremely small. Geroliminis and Sun [43]

have noticed from the real network of Yokohama that a low scatter MFD exists even if there is some variance in the distribution of congestion (coefficient of variation in link density, i.e. dimensionless standard deviation divided by the mean, was around 0.25). Computationally speaking an online partitioning is not a problem, as the method developed in this report is fast and can be applied in real time if data are available.

A difficulty might arise if the new partition does not exist in the historical database and the shape of the MFD cannot be directly estimated from the data. Nevertheless, Mazloumian et al. [44] have shown that there is a low scatter relationship between the network flow and variance of link density (which expresses the spatial heterogeneity of congestion) for a given mean network density. Thus, the shape of the MFD can be estimated by integrating an analytical formulation (e.g. the one of Geroliminis and Daganzo [80] or Geroliminis and Boyaci [96] with the empirical spatial distribution of link density). Ampountolas and Kouvelas [97] designed a Kalman filter-based estimation algorithm that utilizes real-time measurements of circulating flow and accumulation of vehicles to produce estimates of the currently prevailing critical accumulation. The developed strategy may be valuable whenever the Urban-scale Macroscopic Fundamental Diagram is not well defined and the critical accumulation cannot accurately be specified or is subject change due to traffic-responsive signal control, traffic composition (e.g. cars versus buses), or nonrecurrent day-to-day traffic patterns. Another interesting research direction is to identify the monitoring needs to provide an accurate and dynamic partitioning that will lead to the development of efficient control strategies. Information in every link in a network is not necessary, as shown by Ji and Geroliminis [25] regarding partitioning and by Keyvan-Ekbatani et al. [60] with respect to the derivation of Urban-scale MFDs. In case of highly variable link density (e.g. high directional flows), data needs might be larger for some part of the network (especially the ones

with higher spatial heterogeneity).

If dynamic partitioning could describe the spatiotemporal characteristics of congestion, then region-based traffic management and control strategies can be extended to capture dynamic boundaries. While dynamic boundary adjustment of initial partitioning can be considered as a straightforward way to implement dynamic partitioning, several caveats can be associated with this approach. Choosing an appropriate time period for initial partitioning is difficult, especially at the beginning of congestion formation where the network is usually not partitionable. A boundary adjustment algorithm will not be able to observe the development of new pockets of congestion. Finally, the computational complexity of boundary adjustment algorithms is high and thus can be hardly developed in a real system. However, appropriate dynamic partitioning is instrumental in implementing traffic management strategies on homogeneously congested regions, since, static partitioning algorithms aim at clustering the most similar links instead of congested links. In addition, there is no time correlation regarding the formation of congestion, since static partitioning is implemented for data from distinct time instants. Ji and Geroliminis [55], using GPS data from 20000 taxis for one month in the city of Shenzhen, were able to address these issues associated with dynamic and static partitioning.

3.7 Chapter Overview

This chapter describes the implementation of time-dependent partitioning of a real-life urban traffic network in south central Singapore into homogeneous regions using several clustering approaches. Spectral clustering, k-means and k-harmonic means clustering were applied to a part of the Singapore urban traffic network encompassing the Marina Bay area, as well as parts of the Central Business District, Chinatown, Dhoby Ghaut, and Clarke Quay. The main objectives were to derive

regions with low intraregional and high interregional link density variability, as well as maintaining spatial cohesion for the resulting regions, utilizing data collected from loop detectors located on the traffic network links before, during and after the morning peak period. For this particular traffic network, k-harmonic means clustering was found to outperform both k-means clustering and spectral clustering approaches. As discussed in the previous section, k-harmonic means clustering lends itself to an online partitioning implementation, due to the low computational effort required. A multi-class extension to a traffic model, called Network Transmission Model [98], for describing regional dynamics, which will be employed in the proposed region-based urban traffic simulation and management framework, in the next chapter.

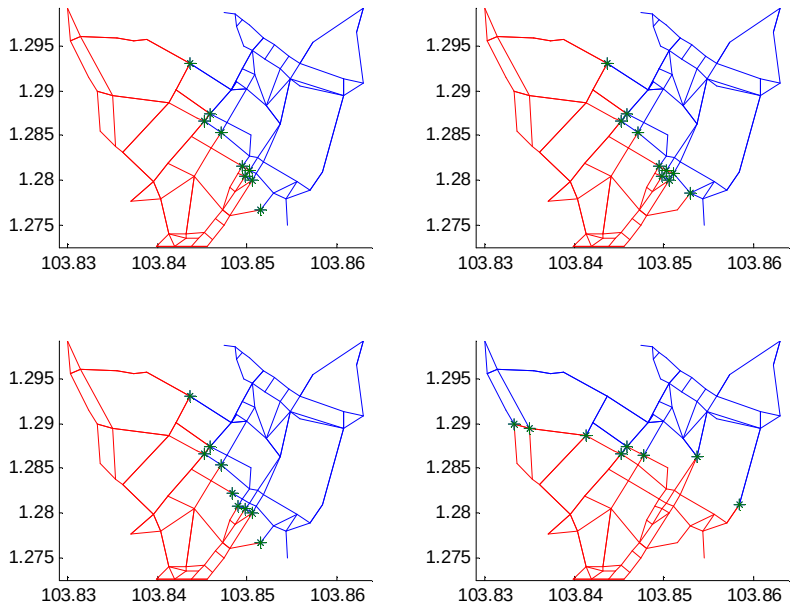


Figure 3.3: Weighted k-harmonic means clustering for time slots 8am (top left), 9am (top right), 10am (bottom left) and 11am (bottom right).

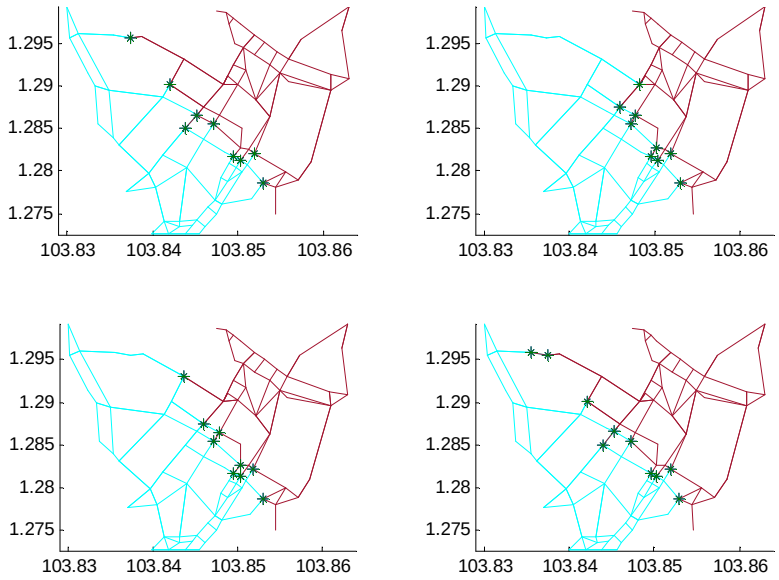


Figure 3.4: Normalized spectral clustering for time slots 8am (top left), 9am (top right), 10am (bottom left) and 11am (bottom right).

Chapter 4

Region-based Traffic Dynamics

4.1 Multi-class Network Transmission Model

In this section, the proposed multi-class extension of the Network Transmission Model (NTM) is introduced, originally developed by Knoop and Hoogendoorn [98]. In section 4.1.1, the notation and terminologies used in the proposed model are provided. In section 4.1.2, a description of the multi-class regional traffic model is provided.

4.1.1 Notations and Terminologies

Let a directed graph $G = (K, A)$ denote the regional traffic network, which comprises a set of regions $\mathcal{K}$ and a set of interregional boundaries $\mathcal{A}$. The following notations will be used to describe the proposed multi-class extension of the NTM:

Sets:

$\mathcal{K}$: set of nodes,$|\mathcal{K}| = m$, representing homogeneous regions;

$\mathcal{A} \subseteq \mathcal{K} \times \mathcal{K}$: set of arcs, $|\mathcal{A}| = l$, representing interregional boundaries;

$\mathcal{K}_o \subset \mathcal{K}$: set of origins,$|\mathcal{K}_o| = o$;

$\mathcal{K}_d \subset \mathcal{K}$: set of destinations,$|\mathcal{K}_d| = d$;

$C = [c_{i,j}]^{m \times m} \geq 0,\ c_{i,j} > 0 \iff (i,j) \in \mathcal{A}$, the boundary capacity matrix;

$\mathcal{P}_{b,e}$: set of all paths from origin b to destination e, $\forall b \in \mathcal{K}_o, \forall e \in \mathcal{K}_d$;

$\mathcal{P}_e := \bigcup_{b \in \mathcal{K}_o} \mathcal{P}_{b,e}$, set of all paths with destination e, $\forall b \in \mathcal{K}_o$;

$\mathcal{P}_b := \bigcup_{e \in \mathcal{K}_d} \mathcal{P}_{b,e}$, set of all paths with origin b,$\forall e \in \mathcal{K}_d$;

$\mathcal{P} := \mathcal{P}_b \bigcup \mathcal{P}_e$, set of all possible paths;

Ω: set of classes, representing different types of travelers;

$\mathcal{Z}_i$: the set of links in region $i \in \mathcal{K}$;

$\mathcal{H}$: set of discrete time steps.

Index variables:

$i, j \in \mathcal{K}$: region indices, where $(i,j) \in \mathcal{A}$, i.e., region i is adjacent to j;

$b \in \mathcal{K}_o$: origin region index;

$e \in \mathcal{K}_d$: destination region index;

$p \in \mathcal{P}_{b,e}$: path from origin b to destination e;

$\omega \in \Omega$: traveler class index;

$z \in \mathcal{Z}_i$: link that belongs to region i;

$h \in \mathcal{H}$: discrete time index.

Traffic network parameters:

$v_{i,f} \in \mathbb{R}^+$: average speed (km/h) for region i under free-flow conditions;

$n_{i,crit} \in \mathbb{R}^+$: critical accumulation (veh) for region i;

$Q_{i,crit} \in \mathbb{R}^+$: average network flow (veh·m/h) for region i at capacity flow;

λ_z: number of lanes on link $z \in \mathcal{Z}_i$;

Λ_z: length of link $z \in \mathcal{Z}_i$;

T_H: simulation time (s);

T_s: sample time (s);

$H = \lfloor T_H/T_s \rfloor$: simulation horizon.

Traffic state variables:

$n_z(h)$: number of vehicles on link $z \in \mathcal{Z}_i$ at step h;

$n_i(h) = \sum_{z \in \mathcal{Z}_i} n_z(h)$, accumulation in region i i at step h;

$n^{\omega}_{i,b,e}(h)$: ω-class, b-origin, e-destination specific accumulation in region i at step h;

$\zeta^{\omega}_{b,e}(h)$: ω-class, b-origin, e-destination specific travel demand at departure time step h;

$\zeta_{b,e}(h) = \sum_{\omega \in \Omega} \zeta^{\omega}_{b,e}(h)$: b-origin, e-destination specific travel demand at departure time step h;

$Q_i(n_i(h))$: approximation function to the Urban-scale MFD, according to [99];

$v_i(n_i(h))$: average network speed approximation function, according to [99];

$P_j(h)$: representing the supply, similar to [101];

$\beta^{p,\omega}_{i,j,b,e}(h)$: p-path, ω-class, b-origin, e-destination specific branching rates from region i to j, at step h, where $p \in \mathcal{P}_{b,e}$

$Z_{i,j}(h)$: unrestricted demand from region i to j at step h;

$\tilde{Z}_{i,j}(h)$: effective demand from region i to j at step h;

$\tilde{Z}^{\omega}_{i,j,b,e}(h)$: ω-class, b-origin, e-destination specific demand from region i to j at step h;

$Z_j(h)$: total demand to region j at step h;

$\epsilon_j(h)$: the flow proportion that can travel into region j;

$\chi_i(h)$: the minimum of the exit flow proportions;

$q_{i,j}(h)$: outgoing flow from region i to j at step h;

$q^{\omega}_{i,j,b,e}(h) = \chi_i(h)\tilde{Z}^{\omega}_{i,j,b,e}(h)$: ω-class, b-origin, e-destination specific outgoing flow from region i to j at step h.

4.1.2 Model Description

In order to describe region-based traffic dynamics, a multi-class extension the Network Transmission Model (NTM) is introduced. The NTM describes a traffic network at the urban traffic level as well as at the aggregated regional traffic level. NTM takes into account the dynamics of a network as expressed through the Urban-scale MFD, as well as the limited capacity of the boundaries between regions [100]. The NTM combines the concepts of Daganzo's Cell Transmission Model [101], CTM for brevity, and a node model by Jin et al. [102] and applies it to a model describing urban traffic network dynamics at a regional level. Unlike the regional dynamic traffic model developed by Geroliminis et al. [88], further extended in [14,54,89], to model and control for heterogeneity in MFD at the regional level as well as explicitly model internal traffic demand, NTM accounts for internal traffic demand implicitly. Concretely, while, for a supply/demand scheme, regional supply values work similarly to Daganzo's CTM, regional demand values follow the curve of a regional performance function, an abstraction of a link performance function, corresponding to a scaled-down Urban-scale MFD [103]. This modeling approach enables a sidestep of the uncontrollable internal flows and leads to faster, if less accurate,

analysis of the traffic management effects on the network. Here, one should mention that another way to sidestep the explicit modeling of the uncontrollable internal flows, while keeping their effects in the flow dynamics, is by introducing parameters (instead of accumulation internal variables) in the framework of Geroliminis et al. [88], see e.g. Haddad and Shraiber [104] and others. Our multi-class extension introduces class/origin/destination specific accumulations, as well as pre-trip information in the form of regional path/class branching rates for each respective origin/destination pair, adjusting the traffic dynamics accordingly. As mentioned in the previous section, in this case, classes refer to travelers represented by different types of routing approaches. In short, the dynamics of each region $i \in \mathcal{K}$ are described as follows:

$$n_i(h+1) = \sum_{e \in \mathcal{K}_d} \sum_{b \in \mathcal{K}_o} \sum_{\omega \in \Omega} n^{\omega}_{i,b,e}(h+1), \tag{4.1}$$

where total accumulation $n_i(h+1)$ for region i at step $h+1$ comprises all ω-class, b-origin, e-destination specific accumulations $n^{\omega}_{i,b,e}(h+1)$, which in turn can be derived from the following dynamic equation:

$$n^{\omega}_{i,b,e}(h+1) = n^{\omega}_{i,b,e}(h) + \frac{T_s}{\sum\limits_{z \in \mathcal{Z}_i} \lambda_z \Lambda_z} \left(\sum_{j \in \mathcal{K}, (i,j) \in \mathcal{A}} q^{\omega}_{j,i,b,e}(h) - \sum_{j \in \mathcal{K}, (i,j) \in \mathcal{A}} q^{\omega}_{i,j,b,e}(h) \right) \tag{4.2}$$

where $q^{\omega}_{j,i,b,e}(h)$ and $q^{\omega}_{i,j,b,e}(h)$ are the incoming and outgoing flow rates, respectively, for class ω, with origin b and destination e, at step h. More concretely, the urban traffic network is first partitioned into homogeneously congested regions and the Urban-scale MFD for each region is derived. In the proposed multi-class extension of the NTM, I have selected to represent the Urban-scale MFDs with an approximation function according to [99]. Specifically, the Urban-scale MFD of region i with accumulation $n_(i)$, free speed $v_{i,f}$ and critical accumulation value $n_{i,crit}$, is

$$Q_i(n_i(h)) = n_i(h) v_{i,f} exp\left(-\xi \left(\frac{n_i(h)}{n_{i,crit}}\right)^{\alpha}\right) \tag{4.3}$$

Drake et al. [99] also provide an average network speed approximation function,

$$v_i(n_i(h)) = v_{i,f} exp\left(-\xi \left(\frac{n_i(h)}{n_{i,crit}}\right)^{\alpha}\right), \tag{4.4}$$

which is used to derive regional travel times. In literature, several simplistic approximations of MFDs in the form of triangular or trapezoid functions appear for theoretical examples [45,47,50] and even some real-life applications [105]. In recent literature [3,54,88], 3^{rd} order polynomial functions, whose coefficients are tuned based on real-life data, have been shown to be more accurate approximations for MFDs of specific regions. The approximation function by [99] provides a good compromise between accuracy and computational effort. These approximation functions have been shown to be suitable for describing the dynamics of regional traffic networks [12,13,106]. Following [12,13], I set $\xi = 0.5$ and $\alpha = 2$, however ξ, α can be amenable to tuning, to better fit the Urban-scale MFD derived from real-life data. The interregional flow at each stage is determined by the Urban-scale MFD, separated in demand (unrestricted)

$$Z_{i,j}(h) = \frac{1}{n_i(h)} \sum_{e \in \mathcal{K}_d} \left(\sum_{b \in \mathcal{K}_o} \left(\sum_{\omega \in \Omega} \left(n^{\omega}_{i,b,e}(h) \sum_{p \in \mathcal{P}_{b,e}} \beta^{p,\omega}_{i,j,b,e}(h) Q_i(n_i(h)) \right) \right) \right) \tag{4.5}$$

and supply

$$P_j(h) = \begin{cases} Q_{j,crit} & \text{if } n_j(h) \leq n_{j,crit} \\ Q_j(n_j(h)) & \text{if } n_j(h) > n_{j,crit} \end{cases} \tag{4.6}$$

where regions $i, j \in \mathcal{K}, (i,j) \in \mathcal{A}$, at step h. In this supply/demand formulation, supply $P_j(h)$ and total demand $Z_j(h)$ for region j at step h, can be described through $Q_j(n_j(h))$ corresponding to the Urban-scale MFD approximation function described

above. Then, similarly to CTM, one can consider the supply $P_j(h)$ equal to regional capacity flow $Q_{j,crit}$, when vehicle accumulation $n_i(h)$ is below the critical value $n_{i,crit}$ and equal to the scaled-down Urban-scale MFD otherwise. Unlike the CTM, where demand increases in tandem with the beyond-critical-value increase to cell density, regional demand $Z_{i,j}(h)$ is *reduced* when vehicle accumulation $n_i(h)$ increases beyond $n_{i,crit}$. In this manner, internal traffic demand for each region and the resulting restriction on the regional outgoing flow are taken into consideration indirectly. For the unrestricted demand $Z_{i,j}(h)$, the boundary capacity between two regions also comes into play, resulting in effective demand

$$\tilde{Z}_{i,j}(h) = \min\{Z_{i,j}(h), c_{i,j}\}, \tag{4.7}$$

from region i to j at step h. Effective demand can be further decomposed into ω-class, b-origin, e-destination specific demands from region i to j at step h in the following manner:

$$\tilde{Z}^{\omega}_{i,j,b,e}(h) = \frac{1}{n_i(h)} \left(n^{\omega}_{i,b,e}(h) \sum_{p \in \mathcal{P}_{b,e}} \beta^{p,\omega}_{i,j,b,e}(h) \right) Q_i(n_i(h)) \left(\frac{\tilde{Z}_{i,j}(h)}{Z_{i,j}(h)} \right) \tag{4.8}$$

After consolidating all incoming demands to region j into total demand

$$Z_j(h) = \sum_{i \in \mathcal{K}, (i,j) \in \mathcal{A}} \tilde{Z}_{i,j}(h), \tag{4.9}$$

the flow proportion that can travel into region j is determined as

$$\epsilon_j(h) = \min_{j \in \mathcal{K}, (i,j) \in \mathcal{A}} \left\{ \frac{P_j(h)}{Z_j(h)}, 1 \right\} \tag{4.10}$$

Subsequently, flow proportion

$$\chi_i(h) = \min_{j \in \mathcal{K}, (i,j) \in \mathcal{A}} \{\epsilon_j(h)\} \tag{4.11}$$

will be the same for all demands $Z_{i,j}(h)$, under the assumption that traffic cannot freely transition from region to region, due to spillback events on the boundaries as well as congestion within region i. If the flow is limited by the supply, region j receives flow proportional to the demands converging to j. Tampere et al. [107] observed that Jin et al's approach of distributing flows proportional to the demand does not satisfy the conservation of branching rates requirement, however, in a regional macroscopic flow setting, this is not necessary. Outflow from region i to j

$$q_{i,j}(h) = \chi_i(h)\tilde{Z}_{i,j}(h), \tag{4.12}$$

which is assumed constant between time steps, is effectively the minimum of supply $P_j(h)$ and demand $Z_{i,j}(h)$.

As described in the above equations, the main characteristic of the proposed multi-class extension of the NTM is that, unlike the interregional splitting rates of the original, the proposed control variables are path-class-origin-destination specific branching rates, which allow for path-based, rather than transition-based, assignment. Our problem's pre-trip information dissemination structure, as presented in the following section, is usually amenable to path-based route assignment [11,69,108]. Concretely, in the proposed multi-class dynamic traffic model, route guidance is implemented through control input variables $\beta^{p,\omega}_{i,j,b,e}(h) \in [0,1]$, which represent p-path, ω-class, b-origin, e-destination specific branching rates from region i to j at step h, where $p \in \mathcal{P}_{b,e}, \omega \in \Omega$. Let $\mathcal{B}^{p,\omega}_{i,b,e}$ be the collection of all feasible p-path, b-origin, e-destination specific branching rates in region i, associated with traveler class ω. If a particular traveler class is characterized by non-compliance, with non-compliance

rates $NC \in \mathcal{O}_{NC} = [0, 1]$, the effect of the path-based route assignment provided on the resulting traffic patterns cannot be determined analytically. Rather, the traffic patterns can be described as a stochastic process over a probability space, whose sample space is $\mathcal{B}^{p,\omega}_{i,b,e} \times \mathcal{O}_{NC}$, with an induced event set $\mathcal{F} := 2^{\mathcal{B}^{p,\omega}_{i,b,e}}$, which is a σ-algebra, and a probability map $\mathcal{S}^{p,\omega}_{i,b,e}$. Owing to this stochastic nature of the branching rates for each region, it is extremely difficult, if still possible, to analyze the traffic conditions under different routing approaches.

4.2 Chapter Overview

This chapter introduces the multi-class extension of a region-based dynamic traffic model, the Network Transmission Model (NTM). The NTM has been developed by Knoop and Hoogendoorn [98] to act as a support system for real-time traffic control operations, however, traveler class heterogeneity has not been considered. Taking into consideration the fact that paths are required as input for route guidance implementations, interregional splitting rates, the control variables of the original model, are replaced with path-class-origin-destination branching rates to facilitate path-based route assignment methods. The following chapter provides a brief introduction to learning games fundamentals, and subsequently, the Dynamic Stochastic User Equilibrium (DSUE) is defined. A Multinomial Logit Routing (MLR) method for a regional network to achieved DSUE conditions is presented, followed by the introduction of a non-predictive strategic learning-based routing method, which employs Proxy Regret Matching (PRM). Last but not least, a simulation-based approach, Incremental Route Planning (IRP), which integrates a Public Transit Diversion (PTD) mechanism, is introduced and compared with PRM and MLR, which it consistently outperforms.

Chapter 5

Prescriptive Route Guidance with Pre-trip Information Dissemination

Route Guidance and Information Systems' (RGIS) primary objective is to improve drivers' route choice by providing information about prevailing traffic conditions (descriptive guidance), or directly prescribing routes which are guaranteed to have reduced travel times (prescriptive guidance). More specifically, descriptive guidance provides traffic information to the traveler, such as estimated travel times for a set of alternate route suggestions, queue lengths, average link speeds, incident location reports. The burden of choice lies with the traveler and user-behavioral models are used to model said choice. Prescriptive route guidance, unlike descriptive, provides a single route option, with the implicit guarantee of a shorter travel time. Information about traffic conditions or prescribed routes can be passed along at the beginning of the trip (pre-trip). Drivers can also receive updated information in discrete time intervals en-route to their Destination, however, this work exclusively employs pre-trip information dissemination. For each Origin-Destination pair, a set of shortest

paths is calculated and route choice probabilities are assigned for the routes corresponding to the set of shortest paths. Based on this set of probabilities, a route is appropriately prescribed to each departing traveler, representing the control input of a RGIS. The routing method in 5.5.1 comes integrated with a Public Transit Diversion (PTD) mechanism. The routing methods presented in 5.3, 5.4 have also been extended to provide the possibility for vehicle passengers to be diverted to public transit, dependent on prevailing or predicted traffic conditions. Following [109], I assume that public transit capacity can accommodate enough travelers that all public transit routes travel times are constant. For all Origin-Destination pairs, travel times on public transit routes are assumed to be twice as long as the travel times for the corresponding regional routes, under free flow conditions. This travel time difference takes into consideration possible traveler delays, when changing modes of transport, from private vehicles to public transit.

5.1 Non-predictive Routing Method

The accuracy of route travel time estimation plays an important role in the drivers' compliance rate with the route guidance system. If the route travel time estimation is based on currently prevailing or past traffic conditions, the guidance provided is considered non-predictive. The majority of RGIS (Route Guidance and Information Systems) in use today employ non-predictive guidance. A subclass of non-predictive guidance approaches employs so-called adaptive learning algorithms. In a game theoretical context, adaptive learning algorithms teach players, over a number of iterations, to take actions that constitute equilibria for their particular game setup. Adaptive learning algorithms exchange information in a distributed fashion, reducing the computational complexity associated with the optimization step. Further details regarding Game Theory fundamentals, can be found in Appendix A. In the

following section, a more detailed overview of different types of adaptive learning algorithms is presented.

5.2 Adaptive Learning Algorithms

Best response dynamics is a rudimentary example of a family of learning algorithms, which aim to maximize each player's payoff by selecting at every step their best response strategy. While easy to implement, this algorithm exhibits sensitivity to initial conditions, with guarantee of convergence limited to specific types of payoff functions only [110]. Fictitious play is representative of a family of adaptive learning algorithms, where players are able to derive the empirical frequencies of other player's actions based on past observations. At each iteration of the learning process, players give a best response to other players' mixed strategy, informed by their respective empirical frequencies. Smooth fictitious play incorporates the randomization feature, where each player gives the best response with some high probability or chooses another action with a small probability. Fictitious play cannot always guarantee convergence to a Nash equilibrium [111]. Another family of adaptive learning algorithms, called regret matching algorithms, aim to minimize the players' respective regret for their selected actions [112]. Regret in this context refers to the difference in utility of playing a particular alternative action instead of their currently selected action, given that other players' selected actions are fixed. Players select their actions by drawing from a probability distribution which is proportional to their positive regrets. Regret matching algorithms provide guarantees for convergence of the empirical probability distribution to the set of correlated equilibria. Our Strategic Learning approach to Region-based Route Guidance employs an algorithm belonging to this adaptive learning algorithm classification, called Proxy Regret Matching. Other learning approaches include approximate dynamic pro-

gramming, or reinforcement learning as it is commonly called in literature, where no prior knowledge of the system model is required, rather, the algorithm learns the state transition probabilities online while following the optimal strategy for the current iteration. Reinforcement learning can be considered an online analogue of Dynamic Programming [113].

5.3 Multinomial Logit-based Dynamic Stochastic User Equilibrium

In traffic and transportation literature, social welfare is referred to as System Optimum (SO), while the concept of equilibrium, is referred to as User Equilibrium (UE). These terms were derived from Wardrop's principles [114]. System Optimum describes the minimization of average travel time for the entire network. User Equilibrium describes how individual traffic network users seek to minimize their respective travel times through specific route selection. One can use the Price of Anarchy [115], which in a transportation setting can be viewed as the ratio of a User Equilibrium to the System Optimum to measure the efficiency of User Equilibria with respect to the performance of the entire traffic network. Unlike the deterministic User Equilibrium as expressed in the Wardrop principle, where travel costs for each traveler are assumed to be minimal, equal and known to all travelers and no one-sided deviation from each traveler's selected route can provide additional travel cost decrease, the fractional flows resulting from a Stochastic User Equilibrium do not usually satisfy the Wardrop principle of travel time equality, and all available routes will be used with some probability, however small, and never zero. Stochastic User Equilibrium takes into consideration the imperfect nature of information available to travelers. The error in travel cost perception by travelers contributes to a more realistic route choice behavior, given the fact that most travelers decide upon

which route to take based on past experience or some other psychological factor.

In past literature, multinomial probit and multinomial logit route choice models have been used to derive Dynamic Stochastic User Equilibria (DSUE), both for link-based urban traffic networks [87,116–121] and regional traffic networks [14,89]. Yildirimoglu et al. [89] introduced a region-based Dynamic Traffic Assignment (DTA) model for a regional network and achieved DSUE conditions through implementation of the Method of Successive Averages (MSA). They were able to demonstrate that congestion was accurately described by developed regional dynamics with results similar to link-based dynamics, which otherwise might be unavailable. Yildirimoglu et al. extended their work in [14], whereby they introduce heterogeneity in the MFD-based traffic model, as well as the option to have paths containing the same region more than once and most importantly route guidance based on Dynamic System Optimum (DSO) regional flow distribution. However, their method was not intended as a traffic management strategy for too many regions. The proposed route guidance system is composed of non-predictive and predictive routing methods with a high degree of efficiency, without great loss of estimation accuracy, allowing for real-time traffic management of a large number of regions. Multinomial logit route choice models lend themselves for computationally efficient implementations, as the route choice probability solutions can be expressed in closed-form and still provide a good approximation to the DSUE. Multinomial probit route choice models, however, require heuristic techniques, such as the MSA mentioned previously, to derive the route choice probability solutions. Logit models, unlike probit models, fail to consider partial path-overlap and high path length variability, however, as discussed in [122], for simulation applications with networks prone to high congestibility, as in this case study, logit models can provide solutions of comparable quality to probit models in a computationally efficient manner.

In my implementation, a multinomial logit route choice model is used to real-

istically represent self-interested driver behavior [87]. This approach provides a multinomial logit-based Dynamic Stochastic User Equilibrium (DSUE) for the regional network at every step, which can be used as a more realistic reference case for traveler assignment with imperfect travel time information, during periods of recurrent congestion. A probability is assigned to each alternate path μ with travel time TT_μ from a set of κ shortest paths, as follows:

$$P(\mu|\kappa) = \frac{e^{-\theta TT_\mu}}{\sum_{\nu \in \kappa} e^{-\theta TT_\nu}} \tag{5.1}$$

where θ is a scale parameter associated with travel time uncertainty. It should be noted that the resulting probabilities are sensitive to the units (h, min, s) selected for the travel times.

5.4 Strategic Learning Approach to Region-based Route Guidance

The use of a strategic learning algorithm called Proxy Regret Matching for non-predictive region-based route guidance is described below. A repeated game is considered, where players, in this case, Origin regions, are competing for different endpoints, in this case, Destination regions. In my case, it is assumed that the bipartite graph of Origins and Destinations is complete, i.e. all Origins can reach all Destination regions. This entails the simultaneous play of a set of repeated games and in each game all Origin regions, representing players, compete for one specific Destination region belonging to the set of Destination regions, which has been predetermined. The actions are probability distributions over the route choices available for each Origin region. For each Destination region, at each stage of the repeated game, each Origin region (*player*) selects a set of routes (*actions*) ending at said

Destination, based on an empirical distribution of the player's average past regrets. In a traffic context, these routes are assigned from a RGIS to travelers within the region, according to their Destination preferences. In the first step of this routing method, Yen's algorithm [123] is used to find the sets of κ-shortest paths which correspond to the aforementioned route choices. These routes are updated according to the prevailing traffic conditions for each region in the network. In the final step, the probability distributions for each set of κ-shortest paths are derived, following the procedure presented below, called Proxy Regret Matching. First some definitions are provided.

5.4.1 Assumptions

Certain assumptions are made regarding dynamic region-based route guidance in the network:

- In each vehicle, driver-integrated (smartphone applications) or vehicle-integrated navigation equipment (GPS device) is assumed to be present. The driver is able to receive traffic state information and store a digital representation of the urban traffic network, as well as experienced travel times (by drivers) and expected travel times (from RGIS) for available regional routes. A similar scheme for individual network links is used in [124].

- It is assumed that each departing vehicle can receive real-time pre-trip information in the form of prescribed routes, which are dependent on prevailing traffic conditions. The prescribed route information is updated at regular intervals.

- For every Origin-Destination pair, public transit is available with infinite capacity and constant travel time. A real-life example would be the Mass Rapid

Transit system of Singapore, employing trains that traverse 170.7km over a network comprising 8 lines.

- It is assumed that each driver completely follows the route prescribed to them, selected from the alternative κ-shortest routes provided at their Origin region.

It should be noted that the requirement set in the last assumption can be relaxed, by introducing the concept of traveler non-compliance, whereby a certain fraction of the traveler population will comply to the advice provided and the remaining fraction will disregard it. This approach to modeling non-compliance was originally suggested by Papageorgiou [125] and is implemented in my comparison of non-predictive routing approaches down below. These assumptions are considered to hold for large metropolitan areas and are expected to be commonplace for most urban areas in the near future.

5.4.2 Regret Matching

Regret matching algorithms aim to minimize the players' respective regret for their selected actions [112] in the setting of a repeated game. In a repeated game, each player tracks their past payoffs and computes the empirical average payoff for each action. Let the Origins be the players competing for each of the Destinations. A repeated game in a traffic context can then be defined in the following manner, $\forall e \in \mathcal{K}_d$:

Definition 5.4.1. *A repeated game tuple* $\langle \mathcal{K}_o, \mathcal{P}_e, U, h \rangle$ *where*

- $\mathcal{K}_o$: is the set of players representing the regions designated as Origins
- $\mathcal{P}_e := \bigcup_{b \in \mathcal{K}_o} \mathcal{P}_{b,e}$, representing the set of joint actions, where $P_{b,e}$ represents the set of actions available to player $b \in \mathcal{K}_o$. These can be coded as e Destination-specific route choices.

- $U_b : \mathcal{P}_e \rightarrow \mathbb{R}$ represents the utility function for player i dependent on the joint action of the players
- h represents the current stage of the repeated game

Regret in this context refers to the difference in utility of playing a particular alternative action instead of their currently selected action, given that other players' selected actions are fixed. It should be emphasized that regret affects the decision-making of Origin regions, which I have designated as players in the repeated game, rather than each individual traveler. In this regional traffic context, assuming that a traffic operations center decides in which manner to distribute the travel demand for each Origin region, regret could be the difference in the average regional speed variability, a good indicator of efficient network resources utilization. Players select their actions by drawing from a probability distribution which is proportional to their positive regrets. Regret matching algorithms provide guarantees for convergence of the empirical probability distribution to the set of correlated equilibria, however, said algorithms can be memory intensive, since, each player must keep track of the strategies of all players at every period of play, as well as being able to compute their own utility for changing their strategies during their past plays [112]. This motivated me to consider Proxy Regret Matching as a less computationally and memory intensive alternative.

For a specific Destination $e \in \mathcal{K}_d$, the average Regret of player $b \in \mathcal{K}_o$ for $y \in P_{b,e}$ to $z \in P_{b,e}$ at stage h is defined as

$$M_b^h(y,z) = max\left\{\frac{\sum_{\eta \leq h: p_b^h = y}\left[U_b^\eta(z, p_{-b}^\eta) - U_b^\eta(y, p_{-b}^\eta)\right]}{h}, 0\right\} \tag{5.2}$$

which showcases the potential loss in utility for not having selected action z every time that action y was selected in the past. Regret Matching algorithms can be memory intensive, since, each player must keep track of the strategies of all players

at every period of play, as well as being able to compute their own utility for changing their strategies during their past plays [112].

Proxy Regret Matching

Consider now that each player, after each period of play, knows only their own set of actions and their received utility but do not know what is their utility function. They are not aware of what game is being played, i.e. the number of players, their respective actions and payoffs. These are the assumptions of a modified Regret Matching algorithm, called Proxy Regret Matching [126]. For Destination $e \in \mathcal{K}_d$, the average Proxy Regret of player b for $y \in P_{b,e}$ to $z \in P_{b,e}$ at stage h

$$\hat{M}_b^h(y,z) = max\left\{\frac{1}{h}\left[\sum_{\eta \leq h: p_b^\eta = z} \frac{\sigma_b^\eta(y)}{\sigma_b^\eta(z)} U_b^\eta(p_b^\eta) - \sum_{\eta \leq h: p_b^\eta = y} U_b^\eta(p_b^\eta)\right], 0\right\} \tag{5.3}$$

with $\sum_{y=1}^{p_b} \sigma_b^\eta(y) = 1$, where $\sigma_b^\eta(y)$ denotes the play probability for player $b \in \mathcal{K}_o$ and η denotes the history of plays up to stage h. After calculating the estimated average Proxy Regret, the player adaptively updates the probability of selecting actions to achieve higher utility. If the player selects action y at stage h, the probability of selecting action z at stage $h+1$ is approximately proportional to the average Proxy Regret from y to z. The play probabilities of player b at stage $h+1$ are assigned as follows

$$\sigma_b^{h+1}(z) = \left(1 - \frac{\delta}{h^\gamma}\right) min\left\{\frac{\hat{M}_b^h(y,z)}{\mu}, \frac{1}{|P_{b,e}| - 1}\right\} + \frac{\delta}{h^\gamma p_b},\ z \neq y, z \in P_{b,e} \tag{5.4}$$

and

$$\sigma_b^{h+1}(y) = 1 - \sum_{w \in P_{b,e}: w \neq x} \sigma_b^{h+1}(w) \tag{5.5}$$

where $\delta \in [0,1]$, $\gamma < 0.25$ and inertia parameter μ large enough to guarantee that $\sigma_b^{h+1}(y) > 0$

In this approach, information provision, in the form of past proxy regret distributions, as well as shortest path set assignment, is done in a decentralized manner, even though route selection is implicitly affected by the repeated game's players, in this case, the Origin regions. Even with partial information, Proxy Regret Matching still converges to a distribution of approximate correlated equilibria. This distribution can be found in polynomial time, ($\mathcal{O}(|\mathcal{K}_o||P_{b,e}|^2)$), and there is no need to retain all other players' past strategies in memory or be required to know what their utility functions might be. In short, the computational, memory and information gains more than compensate for the loss in distribution estimation accuracy.

Comparison of non-predictive routing approaches

A single-class implementation of the proposed multi-class extension to the Network Transmission Model (NTM) presented in chapter 4, will be used to simulate the evolution of traffic and develop routing methods on a macroscopic level. The urban traffic network is partitioned into homogeneously congested regions. Then, an allocation of traffic flows through different regional routes is implemented so as to minimize total travel time and improve the rate of arrival for the respective Destinations. For simplicity, an assumption is made that vehicle trips start from the center of each region and that any vehicle traveling from their Origin to their respective Destination will only traverse each region once. A diamond-shaped grid network is considered for implementation of the proposed extension of the NTM. The network consists of 16 regions, with an area of 25 km^2 (5x5) for each region, as shown in Figure 5.1. It should be noted that, as the focus of this work is on the performance of varied routing methods, I have opted for a less sophisticated network test case, where the Urban-scale MFDs for all regions are identical.

In this section I compare five different non-predictive routing methods, including a method employing the proposed strategic learning algorithm and its variant

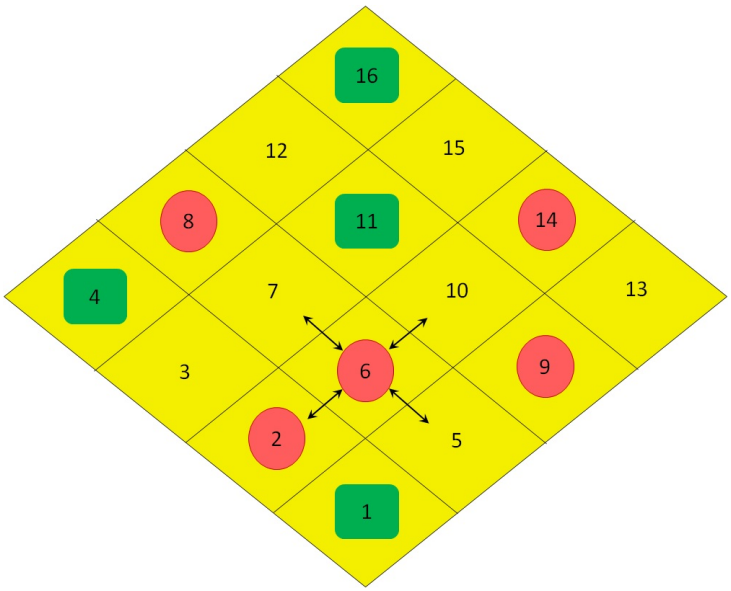


Figure 5.1: Layout of the 4x4 regional network, where green squares and red circles denote Origins and Destinations, respectively.

with non-compliance integration. All routing methods are applied in the proposed example of NTM.

- Fixed Routing (FR)
- Periodically Adjusted Routing (PAR)
- Multinomial Logit Routing (MLR)
- Proxy Regret Matching (PRM).
- Proxy Regret Matching with non-compliance (PRMNC)

Concretely, $\forall i, j \in \mathcal{K}, (i, j) \in \mathcal{A}$, the κ shortest paths are found using Yen's algorithm [123] from region i to region j. Then the shortest of the κ paths is used to implement fixed (FR) and periodically adjusted routing (PAR). In the case of fixed routing (FR), the travel times for the shortest paths are calculated and these shortest paths are retained for the duration of the simulation. Fixed routing (FR) will be the so-called uncontrolled case, against which, all subsequently presented

approaches will be compared. For all other routing methods, routing information update interval was selected to be 5 min. In the case of periodically adjusted routing (PAR), the shortest path set for every region is updated from the travel times based on the prevailing average speeds $v_i(h)$ per region $i \in \mathcal{K}$ at step h. FR and PAR are used for comparison in a similar manner to [13]. Multinomial logit routing (MLR) is also implemented for a more realistic approach that models self-interested traveler behavior. For this case study, sample time was selected to be $10s$, hence, parameter $\theta = \frac{1}{6}$. In similar fashion to periodically adjusted routing (PAR), the shortest path set for every region is updated from the travel times based on the prevailing average regional speeds $v_i(h)$. The vehicle accumulation results for respective routing approaches can be seen in Fig. 5.2.

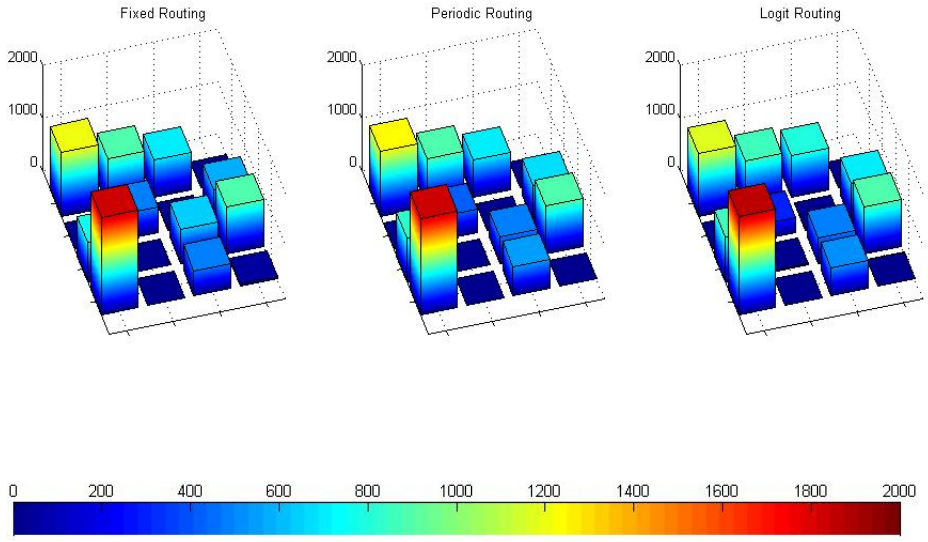


Figure 5.2: Vehicle accumulation results per region for Fixed Routing, Periodic Routing and Multinomial Logit Routing, at the end of the simulation

For comparison, 2 performance metrics are defined:

- $\mathcal{C}_{\text{TVT}} = T_s \sum_{h=0}^{H-1} \sum_{i \in \mathcal{K}} \left(\sum_{z \in \mathcal{Z}_i} \lambda_z \Lambda_z n_i(h) \right)$, the Total Vehicle Travel time for all regions

- $\mathcal{C}_{\mathrm{v}}(H-1) = \sum_{i,j \in \mathcal{K}} \left(v_i(H-1) - v_j(H-1)\right)^2$ average Speed Variability for all regions at the end of the simulation horizon $H-1$

The total vehicle time can be found using the total number of vehicles arriving at the region over a defined time interval T_s. The speed variability metric measures how evenly traffic load is distributed and, implicitly, demonstrates the phenomenon of build up of congestion to capacity in some regions, while other regions receive very little traffic. It should be noted that 50% non-compliance was also introduced as a behavioral characteristic of drivers in the proposed route guidance scheme. Non-compliance was implemented by simply selecting each path $p \in \mathcal{P}_{b,e}, \forall b \in \mathcal{K}_o, e \in \mathcal{K}_d$, multiplying the respective fractional flow by the selected non-compliance percentage and proportionately redistributing the remaining flow to all paths $p' \in \mathcal{P}_{b,e} \backslash p$.
For Fixed Routing (FR), Periodically Adjusted Routing (PAR), Multinomial Logit Routing (MLR), Proxy Regret Matching (PRM) and PRM with non-compliance (PRMNC) the results are: Fixed Routing (FR) is considered as the uncontrolled

	FR	**PAR**	**MLR**	**PRM**	**PRMNC**
C_{TVT}	1.5232	1.5196	1.5208	1.5156	1.5240
C_v	8.5190	8.4541	8.3666	6.6510	6.7700

case, and compare it to the remaining Routing Methods as follows:

$$\mathcal{C}_{\mathrm{TOT}} = w\frac{\mathcal{C}_{\mathrm{TVT,g}}}{\mathcal{C}_{\mathrm{TVT,FR}}} + (1-w)\frac{\mathcal{C}_{\mathrm{v,g}}}{\mathcal{C}_{\mathrm{v,FR}}}$$

where $g \in \{PAR, MLR, PRM, PRMNC\}$ and w an appropriate weight to denote the importance placed on the respective performance measures

w=0.5	**FR**	**PAR**	**MLR**	**PRM**	**PRMNC**
C_{TOT}	1.0000	0.9950	0.9902	0.8878	0.8976

Vehicle accumulation results for respective region-based route guidance approaches can be seen below

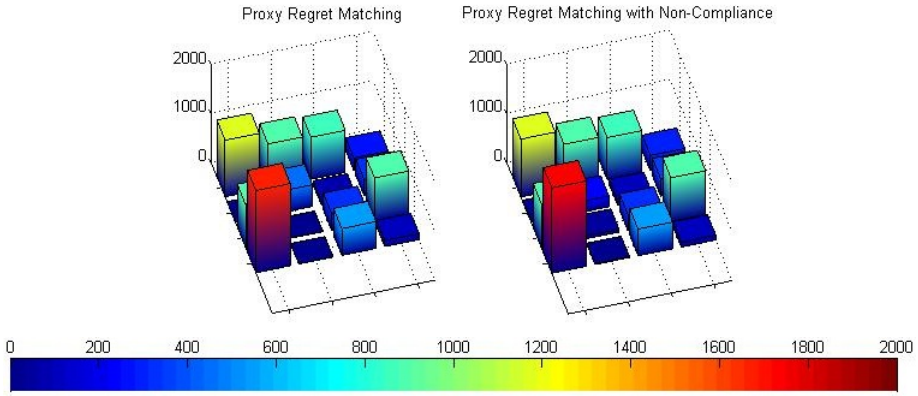


Figure 5.3: Vehicle accumulation results per region for Proxy Regret Matching (PRM) and PRM with non-compliance (PRMNC), at the end of the simulation

As is evident from the above table, at least 11% improvement can be obtained in composite index C_{TOT} performance and even with 50% non-compliance on the drivers' part, a 10% improvement can still be obtained, compared to the uncontrolled case, i.e. Fixed Routing. In the figure that follows, one can visually confirm the more uniform distribution of vehicle accumulation in the example network at the end of the simulation, both for Proxy Regret Matching and Proxy Regret Matching with non-compliance on the drivers' part.

5.5 Predictive Route Guidance

Predictive route guidance employs travel time prediction, based on simulation-based forecasting of traffic conditions. The expected route travel time predicted by the route guidance system will be closer to the experienced route travel time by the drivers than the instantaneous travel time, which is based on current traffic conditions. While the difference between expected and experienced travel time is not precisely quantified, it will be significantly lower than for any non-predictive approach applied within the same network setup. This can be ascribed to the

fact that prediction of traffic state evolution, more specifically, which regions are congested currently or in the near future, will greatly improve the quality of routes prescribed to travelers.

5.5.1 Incremental Route Planning method with Public Transit Diversion

In this section, a proposed predictive routing method is described, which is an extension of the dynamic forecast routing method presented in [127]. Certain definitions and notations necessary for the detailed description of this method are presented, including the time-expanded network representation, first introduced by [128]. A particularly appealing property of time-expanded networks is their flexibility, allowing for modeling of dynamic travel times dependent on respective regional time-varying demand and/or supply. The size of regional networks, i.e. the number of regions, can implicitly depend on many factors, as the network partitioning scheme is dependent on the Urban-scale MFDs, which in turn can be affected from network topology, traffic signal settings and Origin-Destination demand patterns [47,49,50]. Compared to conventional urban traffic networks, however, regional networks can be at least an order of magnitude smaller. Therefore, concerns regarding the exponential time-expanded network size growth related to the problem input size do not significantly affect regional networks.

For the time-expanded networks, first a planning horizon

$\bar{\mathcal{H}} = \{h_0, h_0 + g, h_0 + 2g, h_0 + 3g, ..., h_0 + \Gamma g\}$ is defined, where h_0 is the vehicle departure time, g is a small enough interval such that observed traffic conditions vary slowly and Γ a positive integer large enough such that the entire planning horizon is covered. Subsequently, a simulation-based forecasting approach to predictive region-based routing is proposed. This approach consists of 2 steps, forecasting and shortest path set computation and assignment. Unlike the proposed non-predictive routing

method, the forecasting step of the proposed predictive routing method aggregates prediction information in a centralized manner, while the subsequent shortest path set computation and assignment step is applied locally, for each respective Origin-Destination region pair.

1. The forecasting step involves:

   - Introduction of new vehicles into the regional network based on time-dependent Origin-Destination Demand Matrix.
   - Simulation of the forward movement of all currently routing vehicles using an extension of the NTM for the regional traffic dynamics, throughout the planning horizon.
   - Virtual vehicle trip generation, starting at the Origin regions and assuming route selection according to multinomial logit routing.

   More concretely, the simulation continues until all vehicles have finished their trips and virtual vehicles on Origin regions that are assigned trips based on Multinomial Logit routing are also generated. Virtual vehicle trip generation allows for more accurate travel time estimation, requiring, however, the computation and assignment of shortest paths to the demand generated at the Origin regions. Virtual vehicle trip generation also discourages the frequent use of Origin regions as part of the route assignment process.

2. For the shortest path set computation and assignment step:

   - The original regional network graph $G = (\mathcal{K}, \mathcal{A})$ is expanded to a time-expanded network representation $\bar{G} = (\bar{\mathcal{K}}, \bar{\mathcal{A}})$, where $\bar{\mathcal{K}} = \left\{ i^h | i \in \mathcal{K}, h \in \bar{\mathcal{H}} \right\}$ a set of time-expanded regions and
     $\bar{\mathcal{A}} = \left\{ (i^h, j^{h'}) | (i,j) \in \mathcal{A}, h + x_{i,j}^h = h', h_0 \leq h \leq h' \leq h_0 + \Gamma g \right\}$ a set of interregional time-dependent transitions along the set of boundaries $\mathcal{A}$.

Weight $x_{i,j}^{h} \in \mathcal{H}$ is assigned, at time h, for time-expanded regions $i^h, j^{h'} \in \bar{\mathcal{K}}, (i^h, j^{h'}) \in \bar{\mathcal{A}}$, as the number of time steps required for the departing vehicle to transition from region i to j based on the travel times derived during the forecasting step, since the number of steps required for interregional transition are time-varying.

- For each Origin region $b \in \mathcal{K}_o$, the discrete time step $h = h_0 + \nu g,\ \nu \in \mathbb{N}^+$ represents vehicle departure time at Origin time-expanded region b^h . For each Destination region $e \in \mathcal{K}_d$, the arrival time cannot be predetermined because it is dependent on traffic conditions and the route selected from the κ-shortest paths. So a dummy Destination time-expanded region $e^{\emptyset}$ is introduced, to represent trip completion with undetermined arrival time $\emptyset$. See Figure 5.4 for an example.

- The resulting graph $\bar{G}$ contains the set of all eligible paths $\bar{\mathcal{P}}_{b^h,e^{\emptyset}}$ from b^h to $e^{\emptyset}, \forall b \in \mathcal{K}_o, e \in \mathcal{K}_d$. Similarly to the non-predictive routing approach utilizing Proxy Regret Matching, aggregated regional data preprocessing is implemented, utilizing a bit vector $\bar{f}_{\mathcal{K}} = \left[\bar{f}_i\right]^{m \times 1} \in \{0,1\}, \forall i \in \mathcal{K}$ to store true (1) or false (0) values. These vector values represent each region's state of congestion and are similarly used to increase shortest path calculation efficiency, but with one specific difference. If $\tilde{H} = H - h, \Theta = \left[h, ..., h + \tilde{H}\right]$, where h is the current time step, a true (1) value is assigned for any region $i \in \mathcal{K}$, if time-expanded region $i^{\theta} | n_i(\theta) > r_{IRP} n_{i,crit}, \forall i^{\theta} \in p^{\theta} \in \bar{\mathcal{P}}_{b^{\theta},e^{\emptyset}}, \theta \in \Theta$, where r_{IRP} represents a threshold coefficient as a percentage of the critical accumulation. Subsequently, the set of eligible paths $\bar{\mathcal{P}}_{b^h,e^{\emptyset}}$ contains only regions $i | \bar{f}_i = 0, i \in \mathcal{K}$. If no such path exists, users are diverted to public transit.

- Finally, κ-shortest paths are found by applying Yen's algorithm [123] on the modified time-expanded network graph $\bar{G} = (\bar{\mathcal{K}}, \bar{\mathcal{A}})$, which contains

only eligible paths. The κ-shortest paths are used to distribute the demand $\zeta_{b,e}(h)$ departing Origin b with Destination e at time step h.

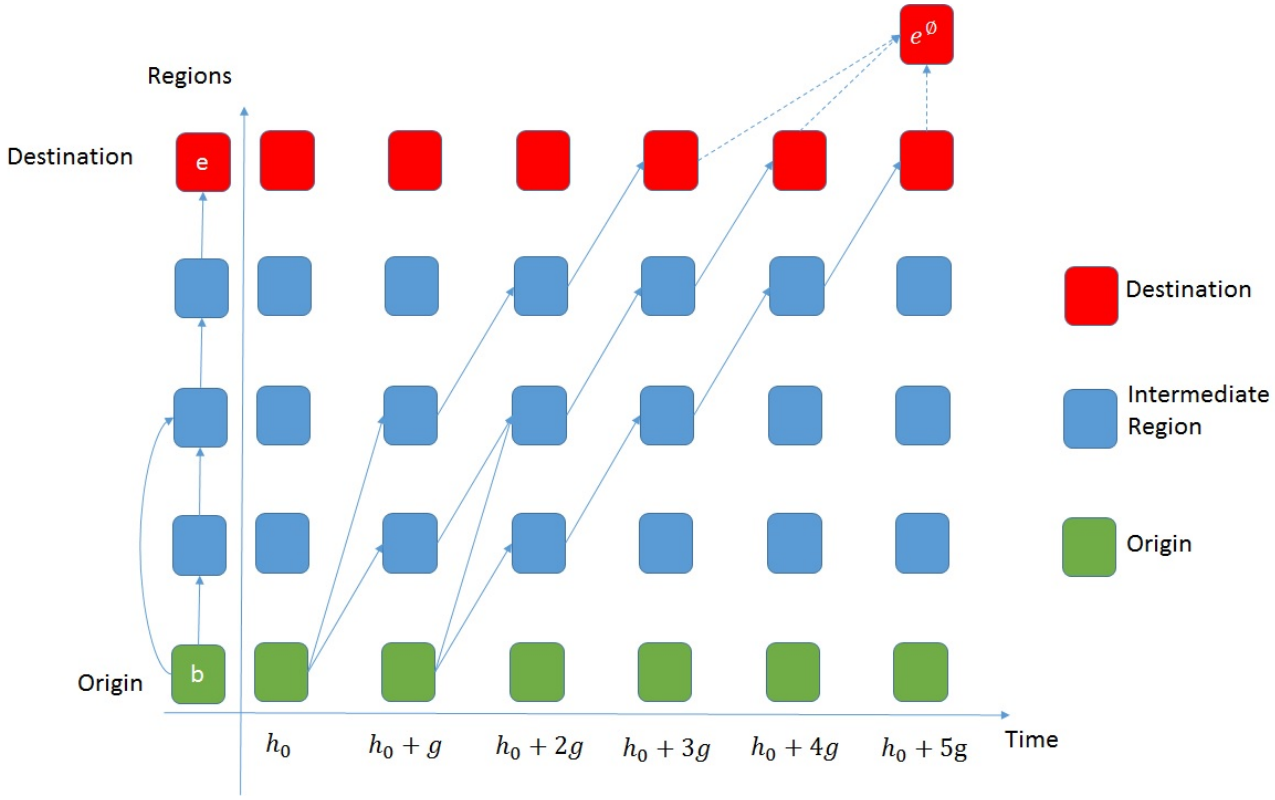


Figure 5.4: Time-expanded Network Representation.

Public Transit Diversion Mechanism integration for Proxy Regret Matching and Multinomial Logit Routing

An extension of the approach used in the previous subsection 5.4.2 is described, with the additional feature of public transit diversion of travelers at the Origin, based on prevailing traffic conditions. A prerequisite assumption for the employment of the public transit diversion mechanism, in addition to the assumptions in subsection 5.4.1, is that public transit is available for every Origin-Destination pair, with capacity large enough so that travel times can be considered constant. Similar to the approach described in 5.4, in the first step of this routing method, Yen's

algorithm [123] is utilized to find the sets of κ-shortest paths which correspond to the aforementioned route choices. These routes are updated according to the prevailing traffic conditions for each region in the network. An intermediate step is introduced for the application of the public transit diversion mechanism. Inspired by [129], in the second step of this routing method, aggregated regional data preprocessing is implemented, so as to increase shortest path calculation efficiency, but most importantly, integrate Public Transit Diversion to the proposed approach. Concretely, for any region $i \in \mathcal{K}$, true (1) or false (0) values are stored in a bit vector $f_{\mathcal{K}} = [f_i]^{m \times 1} \in \{0, 1\}, \forall i \in \mathcal{K}$. Each vector element f_i is assigned a true (1) value for any region $i | n_i(h) > r_{PRM} n_{i,crit}, i \in \mathcal{K}, h \in \mathcal{H}$, with h representing the current time step and r_{PRM} being a threshold coefficient as a percentage of the critical accumulation, allowing for a more or less conservative Public Transit Diversion mechanism. Subsequently, all regions $i | f_i = 1, i \in \mathcal{K}$ are excluded from all κ-shortest path set calculations at time step h. If no such path exists, users are diverted to public transit. In the final step, the probability distributions for each set of κ-shortest paths are derived following the procedure described in section 5.4.2. In similar fashion, the multinomial logit route choice model is extended to also account for the probability that vehicle passengers will choose to travel by public transit. While the shortest path set for every region is updated from the travel times based on the prevailing average speeds $v_i(h)$, the public transit paths from all Origins to all Destinations are assumed to have constant speed $v_{i,PT} = 0.5 v_{i,f}$ and the public transit travel times are calculated accordingly.

Network Performance Benchmarks

Several performance metrics are now defined, by which an effort will be made to compare the effects of each region-based routing method on the regional network performance:

- $\mathcal{C}_{\text{TVT}} = T_s \sum_{h=1}^{H} \sum_{i \in \mathcal{K}} \left(\sum_{z \in \mathcal{Z}_i} \lambda_z \Lambda_z n_i(h) \right)$, the Total Vehicle Travel time for all regions

- $\mathcal{C}_{\text{v}^2} = \sum_{h=1}^{H} \sum_{i,j \in \mathcal{K}} \left(v_i(n_i(h)) - v_j(n_j(h)) \right)^2$ the sum of squares of average regional speed variabilities for all regions

- $\mathcal{C}_{\text{PTD}}$, Public Transit Diversion, i.e., the fraction of total demand generated throughout the simulation, which is diverted to Public Transit

- $\mathcal{C}_{\text{ITR}}$, Incomplete Trips Rate, i.e., the fraction of total demand generated throughout the simulation, which has not completed their trips by the end of the simulation

- $\mathcal{C}_{\text{ATT}}$, the Average Travel Time for each traveler on the regional urban network

The total vehicle travel time can be found using the total number of vehicles arriving at the region over a defined time interval T_s. The speed variability metric measures how evenly traffic load is distributed and, implicitly, demonstrates the phenomenon of build up of congestion to capacity in some regions, while other regions receive very little traffic.The performance metric $\mathcal{C}_{\text{PTD}}$ represents the number of potential travelers unable to utilize the urban network, as a fraction of the number of vehicle requests from all Origins to all Destinations throughout the simulation horizon. Since the speed selected for public transit is lower than the average speed corresponding to the critical accumulation value, this metric is able to demonstrate indirectly the aggregate cost of network underutilization due to overcritical congestion levels. The Incomplete Trips Rate $\mathcal{C}_{\text{ITR}}$ implicitly showcases the network throughput performance. Finally, the Average Travel Time $\mathcal{C}_{\text{ATT}}$ is calculated as the weighted average of total travel time for all vehicles and provides an easily quantifiable measure of individual traveler benefits.

5.6 Simulation Results

A diamond-shaped grid network is considered for implementation of the proposed extension of the NTM, in similar fashion to subsection 5.4.2.

Table 5.1: Origin Destination Demand Matrix

	Demands (veh/h)			
Destination / Origin	Region 2	Region 8	Region 9	Region 14
Region 1	800	1440	1400	2400
Region 4	1520	1120	800	1120
Region 11	1360	960	1040	1040
Region 16	1600	800	800	1440

The homogeneous regions are each described by an Urban-scale MFD with critical accumulation $n_{i,crit} = 25$ veh/km, region network length $\sum_{z \in \mathcal{Z}_i} \lambda_z \Lambda_z = 10$ km and free flow speed $v_{i,f} = 100$ km/h, $\forall i \in \mathcal{K}$. Capacity along the boundaries is $c_{i,j} = 2000$ veh/h/lane, $\forall i, j \in \mathcal{K}, (i,j) \in \mathcal{A}$. For each region i, all regions j such that $(i,j) \in \mathcal{A}$ are defined as the ones that are in the northwest-southeast or northeast-southwest directions with respect to the location of the region i. The demand (veh/h) for each Origin-Destination pair is shown in Table 5.1. Disturbance to the demand through time-varying factor multiplication is introduced to increase the level of realism of the simulations. At each time step, the demand values are multiplied by a uniformly distributed random number with mean value 1 and variance 0.1. The simulation horizon H=2h30min or 9000s. The sample time T_s=10s. It should be noted, that in this case, prescriptive route guidance with pre-trip information dissemination is used. This means that, contrary to the case of en-route guidance, the drivers are not the ones updated, rather, the Origin regions are updated regarding the traffic conditions every 10 seconds and drivers at the current departure time h are assigned the corresponding updated routes. It is assumed that threshold coefficient $r_{IRP} = 1.0$ for the Public Transit Diversion mechanism in IRP (Incremental Route Planning)

routing and similarly $r_{PRM} = 1.0$ for PRM (Proxy Regret Matching) routing. For each routing approach, 10 replications were ran and the average of each respective performance metric was considered.

5.6.1 Results for application of each individual route guidance approach

Each routing method, when it is applied to the entire network individually, is compared to all others. That would mean all travelers use the same routing approach,predictive in the case of IRP and non-predictive in the case of PRM. When compared to MLR (Multinomial Logit Routing), which is designated as the realistic route choice model for self-interested travelers with imperfect information, i.e. unreliable travel time estimation. The following Table 5.2 presents performance metric results for IRP, PRM and MLR routing methods integrated with Public Transit Diversion, each individually applied to the regional network:

Table 5.2: Comparison of network performance for each routing method when it is applied to the entire network individually

MPR=100%	$C_{TVT}(veh \cdot s)$	$C_{v^2}(km^2/h^2)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	1.282e+08	6.616e+05	18.78	0.00	704.2
PRM	1.804e+08	8.010e+06	36.04	5.07	1252.2
MLR	2.476e+08	1.321e+07	36.34	12.09	1721.0

One can visually verify from figures 5.5, 5.6, 5.7, above, that IRP allows for the most homogeneous distribution of the demand generated at the Origin regions, among all approaches. This means the region vehicle numbers never exceed the critical value $n_{i,crit}, \forall \in \mathcal{K}$. PRM allows for better distribution, compared to MLR, however there are still 2 out of 16 regions which exceed critical value $n_{i,crit}$. MLR performs the worst when it comes to demand distribution with 5 out of 16 regions exceeding the critical value $n_{i,crit}$ by a large margin, while the rest of the regions

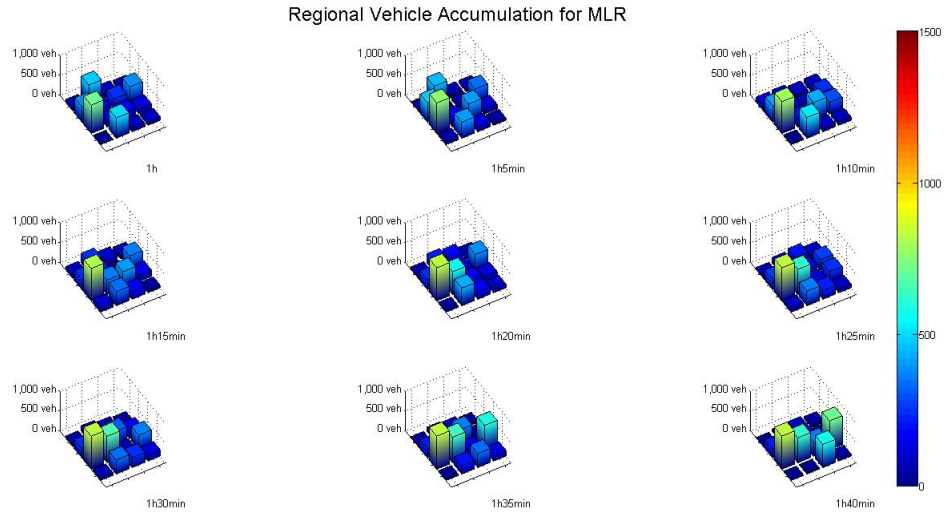


Figure 5.5: Vehicle Accumulation for MLR from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16, with numbering identical to 5.1

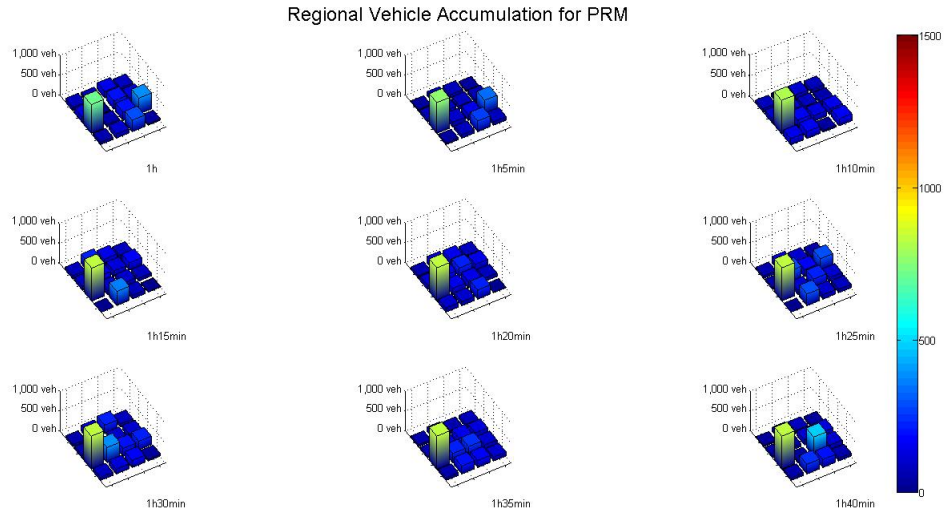


Figure 5.6: Vehicle Accumulation for PRM from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16

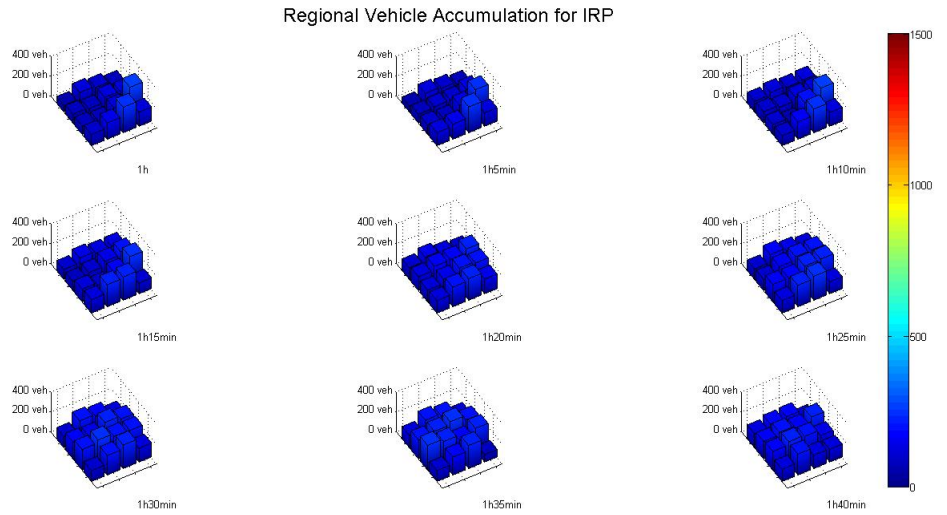


Figure 5.7: Vehicle Accumulation for IRP from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16

are hardly used. This can be attributed to the fact that MLR models the imperfect travel time perception of unguided human travelers. As is evident from the results from Table 5.2, IRP outperforms all other routing methods. This, of course, is to be expected, since IRP belongs to the predictive routing approach category, which means that travel time estimation accuracy is the highest amongst all methods compared. PRM finds itself in the middle, performing better than MLR but worse than IRP. This result is also consistent with what is expected, since PRM belongs to non-predictive routing approach category, meaning that travel time estimation accuracy is worse than it would be for any predictive routing approach. As expected, the worst performing routing method, when individually applied to the regional network, is the MLR routing method. MLR under-performs in all performance metrics.

5.7 Chapter Overview

In this chapter, the following was proposed and presented

- A region-based extension to Multinomial Logit Routing (MLR) which includes a Public Transit Diversion mechanism for travelers with limited or no traffic network information available to them.

- A new strategic learning approach to region-based route guidance (PRM), with three innovations, which, to the author's knowledge, have not appeared in traffic-related literature:

    1. The designation of the set of Origin regions, i.e. regions that generate travel demand to specific Destination regions, as a set of players in a repeated game setting. In traffic-related literature, game-theoretic applications assume each traveler is a player.

    2. The first-time implementation in a traffic context, of the modified Regret Matching algorithm [126], titled Proxy Regret Matching, as part of my strategic learning approach to region-based route guidance.

    3. The integration of a Public Transit Diversion mechanism to the Proxy Regret Matching algorithm, whereby paths are removed from the so-called set of eligible paths based on whether accumulation levels observed in regions belonging to said paths, exceed a critical value. When the eligible paths set becomes empty, travelers are diverted to Public Transit.

- A novel simulation-based approach to provide predictive region-based route guidance (IRP), with a sophisticated congestion prediction step that includes the generation and trip assignment of virtual vehicles, with varying degrees of information availability. This approach proactively diverts travel demand at the origin region to alternate regional paths, even before the intermediate regions in selected regional paths have become congested. In the least desirable case, whereupon all intermediate regions for every possible alternate regional path are predicted to become congested, travel demand is diverted

to public transit through the integrated Public Transit Diversion mechanism. While MLR is incorporated in the prediction step and Yen's algorithm is used for the derivation of eligible paths set, the majority of this route guidance methodology is completely new.

The following chapter describes a market penetration scheme, wherein three distinct classes of travelers, represented by three different routing methods, comprise the vehicle population of a region-based dynamic traffic simulation. For varied combinations of market penetration and non-compliance rates, the overall network performance is obtained, as well as the individual routing method performances corresponding to each distinct class.

Chapter 6

Network Performance under Market Penetration Scheme

In this chapter, the main focus will be on the mixed application of predictive and non-predictive types of prescriptive route guidance methods with pre-trip information dissemination. A market penetration scheme, treated as a series of alternative future scenarios, is utilized for the analysis of network performance in the simultaneous presence of autonomous, RGIS-equipped and unequipped vehicles. Individual performance for each traveler class is analyzed and the effects of each routing method, associated with the corresponding traveler class, are investigated and analyzed.

6.1 Assumptions

The following assumptions are made regarding the mixed application of varied types of route guidance:

- In each vehicle, vehicle-integrated navigation equipment (GPS device) is assumed to be present. The vehicle is able to receive traffic state information

and store a digital representation of the urban traffic network, as well as experienced travel times (by travelers) and expected travel times (from RGIS) for available regional routes.

- Each departing vehicle can receive real-time pre-trip information in the form of prescribed routes, which are dependent on prevailing or future traffic conditions. Improvements in all routing methods allow for the prescribed route information to be updated in regular intervals ($10s$ in this test case).

- Each traveler completely follows the route prescribed to them, selected from the alternative κ-shortest routes provided at their Origin region.

- For every Origin-Destination pair, public transit is available with capacity large enough so that travel times can be considered constant.

- A predictive routing method is used, in this case Incremental Route Planning (IRP), to represent the 1^{st} class of travelers.

- A non-predictive routing method is used, in this case Proxy Regret Matching (PRM), to represent the compliant segment of 2^{nd} traveler class drivers.

- Multinomial Logit Routing (MLR) is used to represent the non-compliant segment of 2^{nd} class travelers drivers, as well as the 3^{rd} class travelers.

- Regarding the forecasting step of the proposed predictive routing method, when applied simultaneously with other methods, two scenarios are considered. In the first scenario (S1), it is assumed that the 1^{st} traveler class make use of privately-owned autonomous vehicles. The information regarding the routes taken from travelers belonging to the 2^{nd} and 3^{rd} classes is unavailable to them. In the second scenario (S2), it is assumed that the 1^{st} class travelers use a one-way, ride-sharing service, which employs publicly-owned autonomous vehicles.

The information regarding the routes taken from 2^{nd} and 3^{rd} class travelers is assumed to be provided to them from the transportation authority in the form of historical or real-time data, or a fusion thereof.

Concretely, in (S1), starting from the beginning of the forecasting simulation, trips are assigned based on Multinomial Logit Routing to all departed vehicles, which are not using the proposed predictive routing method. Thus, the forecasting step takes into consideration the fact that the routing approaches used by proprietary route guidance and information systems are unknown and can only be modeled by MLR. Scenario (S2) is modeled during the forecasting step as 1^{st} traveler class possessing complete information regarding the 2^{nd} and 3^{rd} class travelers' route choices up until the current time step. Public transportation service employing autonomous vehicles in a similar manner to (S2) is already undergoing trials in the city-state of Singapore [130].

The aforementioned assumptions are considered to hold for large metropolitan areas and are expected to be commonplace for most urban areas in the near future. As justification regarding for the selection of routing method for each traveler class, the following arguments can be made.

a) A non-predictive routing method is used, based on strategic learning, to represent the compliant segment of 2^{nd} traveler class drivers, so as to account for the type of route guidance and information systems currently in use by the majority of human-driven vehicles. It is assumed that consumer inertia will ensure that this type of guidance will still be employed on metropolitan areas, at least in the foreseeable future.

b) A predictive routing method is used, which employs simulation-based forecasting, to represent the 1^{st} class of travelers, so as to account for the type of route guidance and information systems expected to be installed on autonomous ve-

hicles.

c) Finally, multinomial logit routing is used to represent the non-compliant segment of 2^{nd} class travelers, as well as the 3^{rd} class travelers, which are assumed to select their routes based on individual preference and past experience, since their route travel time perception is imperfect.

6.2 Simulation Results

The performance metrics introduced in subsection 5.5.1 are also used in this subsection. Said performance metrics provide an easily quantifiable measure of collective as well as individual traveler benefits obtained through the introduction of autonomous (1^{st} class) and RGIS-equipped (2^{nd} class) vehicles. A diamond-shaped grid network is considered for implementation of the proposed extension of the NTM, in similar fashion to subsection 5.4.2. The network consists of 16 regions, with an area of 25 km^2 (5x5) for each region, as shown in Figure 5.1. It should be noted that, as the focus of this chapter is on the interaction of varied routing methods, a less sophisticated network test case was chosen, where the Urban-scale MFDs for all regions are identical.

Table 6.1: Origin Destination Demand Matrix

	Demands (veh/h)			
Destination / Origin	Region 2	Region 8	Region 9	Region 14
Region 1	400	720	700	1200
Region 4	760	560	400	560
Region 11	680	480	520	520
Region 16	800	400	400	720

The homogeneous regions are each described by an Urban-scale MFD with critical accumulation $n_{i,crit} = 25$ veh/km, region network length $\sum_{z \in \mathcal{Z}_i} \lambda_z \Lambda_z = 10$ km and free flow speed $v_{i,f} = 45$ km/h, $\forall i \in \mathcal{K}$. Capacity along the boundaries is $c_{i,j} = 2000$

veh/h/lane, $\forall i, j \in \mathcal{K}, (i,j) \in \mathcal{A}$. For each region i, all regions j such that $(i,j) \in \mathcal{A}$ are defined as the ones that are in the northwest-southeast or northeast-southwest directions with respect to the location of the region i. The demand (veh/h) for each Origin-Destination pair is shown in Table 6.1. It is important to note that the travel demand values have been adjusted and $v_{i,f} = 45$ km/h, to increase the level of realism for the simulation, since the majority of links in urban transportation networks, including that of Singapore, have a speed limit that is less than half of the one used as $v_{i,f}$ in simulations in chapter 5. Disturbance to the demand was introduced through time-varying factor multiplication to increase the level of realism of the simulations. At each time step, the demand values are multiplied by a uniformly distributed random number with mean value 1 and variance 0.1. The simulation horizon H=2h30min or 9000s. The sample time T_s=10s. It should be noted, that in this case, prescriptive route guidance with pre-trip information dissemination is used. This means that, contrary to the case of en-route guidance, the drivers are not the ones updated, rather, the Origin regions are updated regarding the traffic conditions every 10 seconds and drivers at the current departure time h are assigned the corresponding updated routes. For the 2^{nd} class of travelers using Proxy Regret Matching routing, non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$ are considered. These values might be viewed as appropriately chosen, given that in previous studies, non-compliance rates were shown to be dependent on expected travel time reliability and at the aggregate level, travel time estimation is expected to be less accurate than at the individual link route guidance case. It is assumed that threshold coefficient $r_{IRP} = 1.0$ for the Public Transit Diversion mechanism in IRP (Incremental Route Planning) routing and similarly $r_{PRM} = 1.0$ for PRM (Proxy Regret Matching) routing. For each routing approach, 10 replications were ran, and the average of each respective performance metric was considered.

6.2.1 Results for application of each individual route guidance approach

Each routing method is compared to all others, when it is applied to the entire network individually. That would mean all travelers use the same routing approach, or in market penetration scheme terms, MPR=100%. Alternate rates $MPR \in \{10\%, 30\%, 50\%, 70\%, 90\%\}$ are considered, to showcase the performance robustness of the predictive and non-predictive routing approaches, when compared to MLR (Multinomial Logit Routing), which is designated as the realistic route choice model for self-interested travelers with imperfect information, i.e. unreliable travel time estimation. Assuming that MPR of the vehicle population employ IRP or PRM, while $(100\% - MPR)$ employ MLR, representing vehicles which are not equipped with RGIS (Route Guidance and Information Systems). The following Table 6.2 presents performance metric results for IRP, PRM and MLR routing methods integrated with Public Transit Diversion, each with a market penetration rate MPR=100%.

Table 6.2: Comparison of network performance for each routing method when it is applied to the entire network individually, i.e. MPR=100%

MPR=100%	$C_{TVT}(veh \cdot s)$	$C_{v^2}(km^2/h^2)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	1.410e+08	1.189e+06	18.44	0.01	1565.2
PRM	1.744e+08	6.603e+06	33.08	10.24	2359.9
MLR	2.428e+08	1.187e+07	26.20	27.60	2972.4

As is evident from the results from Table 6.2, IRP outperforms all other routing methods. This, of course, is to be expected, since IRP belongs to the predictive routing approach category, which means that travel time estimation accuracy is the highest amongst all methods compared. PRM finds itself in the middle, performing better than MLR in most metrics, but worse than IRP with respect to all metrics. This result is also consistent with my expectations, since PRM belongs to non-

predictive routing approach category, meaning that travel time estimation accuracy is worse than it would be for any predictive routing approach. As expected, the worst performing routing method, when individually applied to the regional network, i.e. the respective market penetration rate MPR=100%, is the MLR routing method. MLR under-performs in all performance metrics except for the Public Transit Diversion metric C_{PTD}. This has to do with the stochastic nature of the Public Transit Diversion Mechanism for MLR, which is simply represented by an alternate route choice with constant travel time for each corresponding Origin-Destination pair.

Table 6.3: IRP (Incremental Route Planning) routing for various MPR (Market Penetration Rates) and for scenarios (S1),(S2)

Scenario	**(S1)**				
$\mathbf{MPR_{IRP}\backslash MPR_{MLR}}(\%\backslash\%)$	**90\10**	**70\30**	**50\50**	**30\70**	**10\90**
$C_{TVT}(veh \cdot s)$	1.445e+08	1.447e+08	1.545e+08	1.812e+08	2.146e+08
$C_{v^2}(km^2/h^2)$	1.208e+06	1.144e+06	1.698e+06	5.258e+06	9.742e+06
$C_{PTD}(\%)$	17.35	15.32	13.28	18.46	26.01
$C_{ITR}(\%)$	0.10	0.01	0.87	6.34	17.68
C_{ATT}(s)	1584.1	1549.1	1605.6	1949.0	2566.2
Scenario	**(S2)**				
$\mathbf{MPR_{IRP}\backslash MPR_{MLR}}(\%\backslash\%)$	**90\10**	**70\30**	**50\50**	**30\70**	**10\90**
$C_{TVT}(veh \cdot s)$	1.435e+08	1.450e+08	1.529e+08	1.768e+08	2.131e+08
$C_{v^2}(km^2/h^2)$	1.191e+06	1.156e+06	1.289e+06	4.674e+06	9.654e+06
$C_{PTD}(\%)$	17.38	15.47	13.07	17.24	25.61
$C_{ITR}(\%)$	0.05	0.0	0.0	5.30	18.15
C_{ATT}(s)	1574.0	1553.1	1587.3	1870.2	2527.6

Table 6.4: PRM (Proxy Regret Matching) routing with for various MPR (Market Penetration Rates) assuming that NC=0% non-compliance rate

Non-compliance	**NC**=0%				
$\mathbf{MPR_{PRM}\backslash MPR_{MLR}}(\%\backslash\%)$	**90\10**	**70\30**	**50\50**	**30\70**	**10\90**
$C_{TVT}(veh \cdot s)$	1.729e+08	1.768e+08	1.825e+08	2.001e+08	2.239e+08
$C_{v^2}(km^2/h^2)$	6.223e+06	6.505e+06	7.555e+06	9.189e+06	1.106e+07
$C_{PTD}(\%)$	31.98	32.26	31.04	30.53	30.91
$C_{ITR}(\%)$	9.01	5.36	5.78	9.68	19.93
C_{ATT}(s)	2302.9	2369.6	2383.0	2597.7	2914.0

Even compared to the performance of IRP and PRM with varying degrees of market penetration, it can be observed that the MLR still performs the worst in the majority of performance metrics, compared to the implementations of IRP and

PRM with the lowest market penetration rates (10%). This is to be expected, since MLR travelers' behavior is characterized by self-interest, as well as imperfect travel time information provision. More specifically, from Table 6.3 one can observe that for all combinations of market penetration rates $MPR_{IRP}\backslash MPR_{MLR}$, (S2) scenario results are better than the respective results for scenario (S1). That is to be expected, since the information provided during (S2) is considered to be perfect, when it comes to the route choices of travelers not employing IRP and subsequently, the travel time estimation accuracy is higher. Finally, from Table 6.4 one can see that IRP performs consistently better than PRM, for respective market penetration rate combinations. Even for $MPR_{IRP}\backslash MPR_{MLR} = 30\%\backslash 70\%$ for both (S1), (S2), it can be observed that most performance metric results are better than for PRM with a market penetration rate $MPR_{PRM} = 100\%$.

Based on the results from Tables 6.2, 6.3, 6.4, it was decided to use MLR as the reference routing method and measure the performance gains of IRP, for scenarios (S1), (S2), as well as PRM for various non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$. The performance metric derived for said comparison is described in equation 6.1 as the weighted average of all performance metric ratios.

$$\mathcal{C}_{\text{RWA}} = \frac{\left\{ w_{TVT}\frac{\mathcal{C}_{\text{TVT,g}}}{\mathcal{C}_{\text{TVT,MLR}}} + w_{v^2}\frac{\mathcal{C}_{v^2\text{,g}}}{\mathcal{C}_{v^2\text{,MLR}}} + w_{PTD}\frac{\mathcal{C}_{\text{PTD,g}}}{\mathcal{C}_{\text{PTD,MLR}}} + w_{ITR}\frac{(1-\mathcal{C}_{\text{ITR,MLR}})}{(1-\mathcal{C}_{\text{ITR,g}})} + w_{ATT}\frac{\mathcal{C}_{\text{ATT,g}}}{\mathcal{C}_{\text{ATT,MLR}}} \right\}}{(w_{TVT} + w_v + w_{PTD} + w_{ITR} + w_{ATT})} \tag{6.1}$$

where

$g \in \{IRP(S1), IRP(S2), PRM(NC), IRP(S1)\backslash PRM(NC)\backslash MLR, IRP(S2)\backslash$ $\backslash PRM(NC)\backslash MLR\}$ and $W = \{w_{TVT}, w_{v^2}, w_{PTD}, w_{ITR}, w_{ATT}\}$ a set of appropriate weights to denote the importance placed on the respective performance measures. In this test case, all weights are set to be equal $w_{TVT} = w_{v^2} = w_{PTD} = w_{ITR} = w_{ATT}$.

After the calculation of the $\mathcal{C}_{\text{RWA}}$ values for all routing methods other than MLR, the Performance gain - Market Penetration curves for non-predictive routing method

PRM, as well as predictive routing methods IRP (S1) and IRP (S2) can be derived, see Figure 6.1. It can easily be observed that, for IRP, overall network performance exceeds that of PRM in all corresponding degrees of market penetration. Performance gains of PRM seem to increase with MPR values up to 90%. When MPR=100%, there is a 2.2% performance loss compared to MPR=90%. This can be attributed to the fact, while PRM at various degrees of market penetration will improve overall network performance, having a percentage of self-interested travelers using MLR, allows for higher network resource utilization. In the case of MLR, all available routes will be used with some probability, however small, and never zero, thus distributing demand along alternative routes that might not have been considered by PRM. One can also observe that for PRM, the weighted average values of $\mathcal{C}_{\mathrm{ITR}}$ peak at MPR=70%. From the results for individual routing method performance, as the MPR for PRM increases, the individual ITR values for PRM also increase, while the simultaneous converse effect for travelers using MLR can be observed. So MPR=70% could be described as the intersection of the increasing ITR values for PRM and decreasing ITR values for MLR. In the case of scenarios (S1), (S2) for IRP, performance gains seem to increase in tandem with the degree of market penetration, however, for MPR$\geq$50%, performance gradually decreases until, at MPR=100%, it reaches a value of $\mathcal{C}_{\mathrm{RWA}}$ $\approx$47% which lies in between values achieved for MPR=50% ($\mathcal{C}_{\mathrm{RWA}}$ $\approx$50%) and MPR=30% ($\mathcal{C}_{\mathrm{RWA}}$ $\approx$35%). These results demonstrate that performance gains are dependent on the forecasting step methodology. While higher degrees of market penetration are required for better overall network performance, the virtual vehicle generation and route assignment leads to less accurate travel time prediction, owing to the fact that MLR is employed for virtual vehicle route assignment. Therefore, having fewer unequipped vehicles as part of the vehicle population leads to diminishing performance gains. Nevertheless, performance robustness for both IRP (S1) and IRP (S2) is assured,

as is evident from the consistent gain in performance for every degree of market penetration. Figures 6.2–6.4 show vehicle accumulation levels per region starting

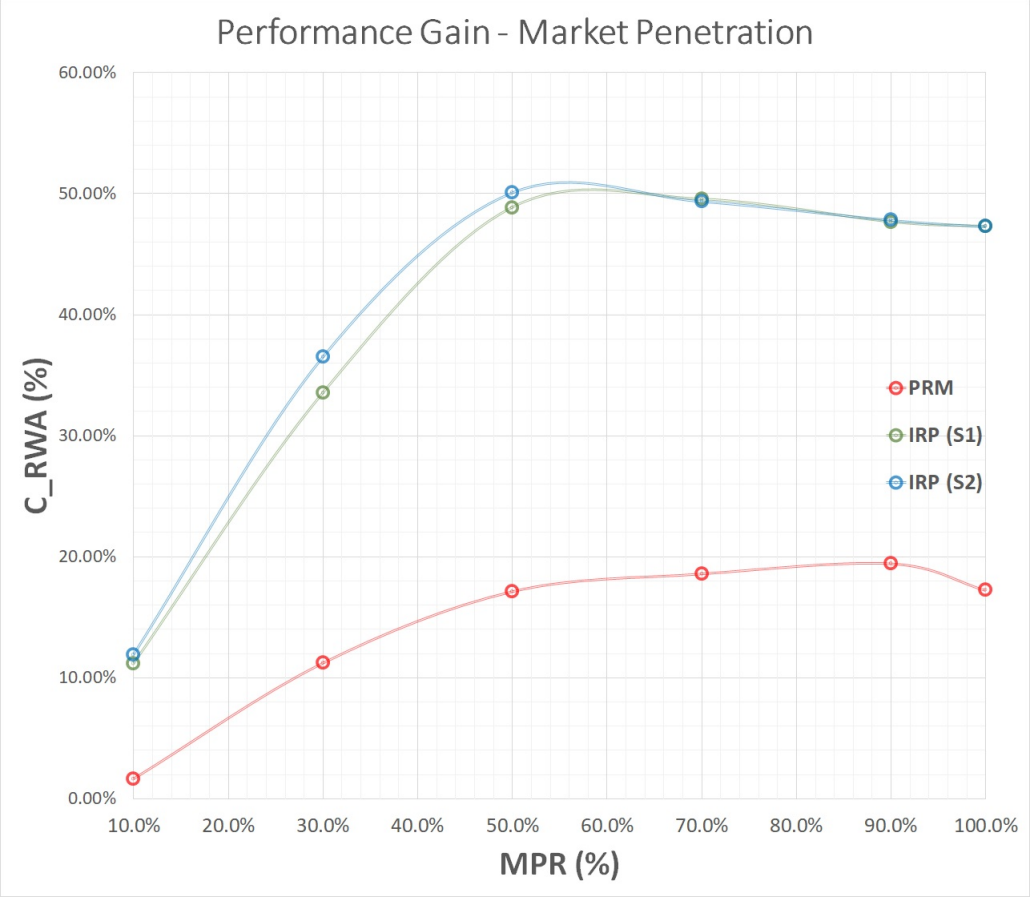


Figure 6.1: Performance gain - Market Penetration Curves of PRM, IRP (S1), IRP (S2), for penetration rates $MPR \in \{10\%, 30\%, 50\%, 70\%, 90\%\}$

from 1h (3600s) of the simulation up until 1h40min (6000s) with 5 min intervals, for 100% market penetration for each routing method (MLR, PRM, IRP) with region numbering as in Figure 5.1. This particular simulation time period was selected because it contains the simulation midpoint and the effect each routing method has, is easy to discern.

As is evident from Figures 6.2–6.4, during these 40 min of simulation, IRP manages to keep vehicle accumulation levels below 500 veh. for all regions, in addition to maintaining a more even distribution of vehicles over all regions. PRM is able to maintain vehicle accumulation levels below 500 vehicles for 14 out of 16 regions,

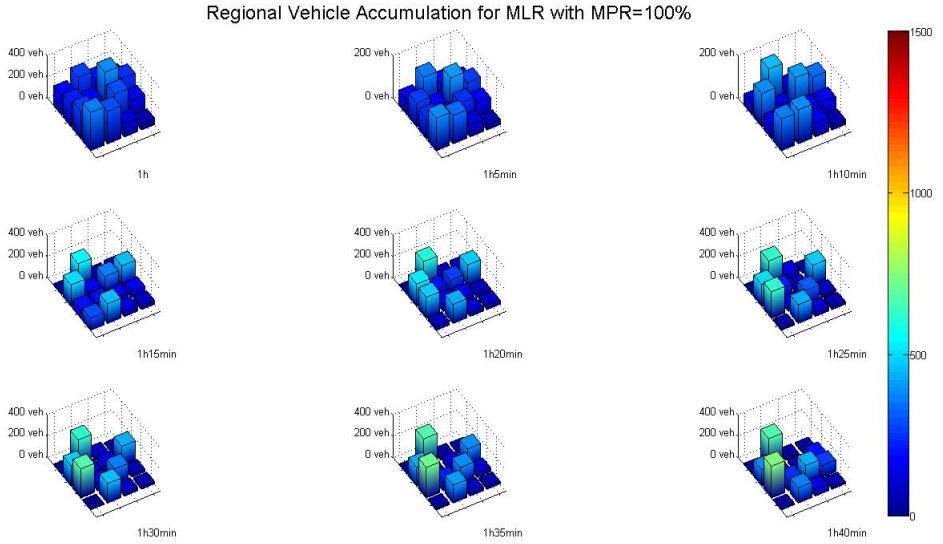


Figure 6.2: Vehicle Accumulation for MLR from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16

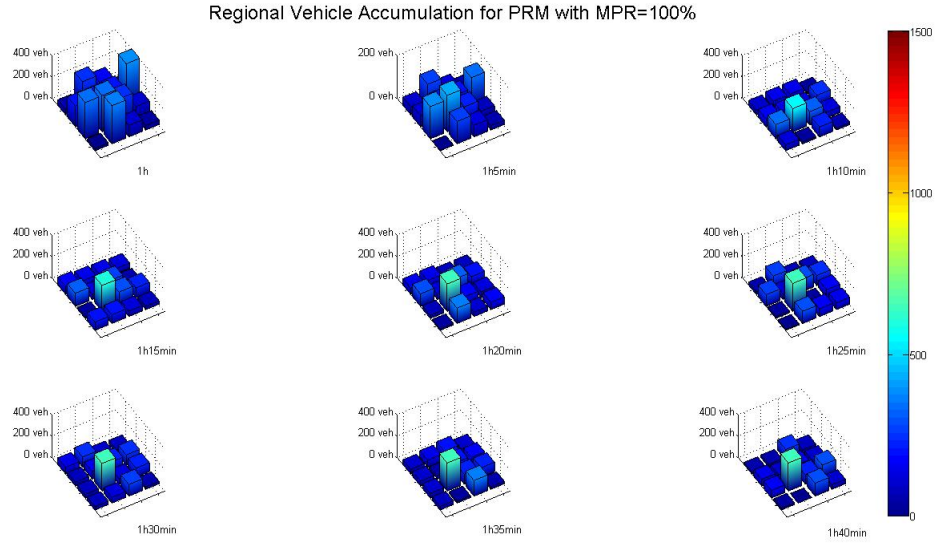


Figure 6.3: Vehicle Accumulation for PRM from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16

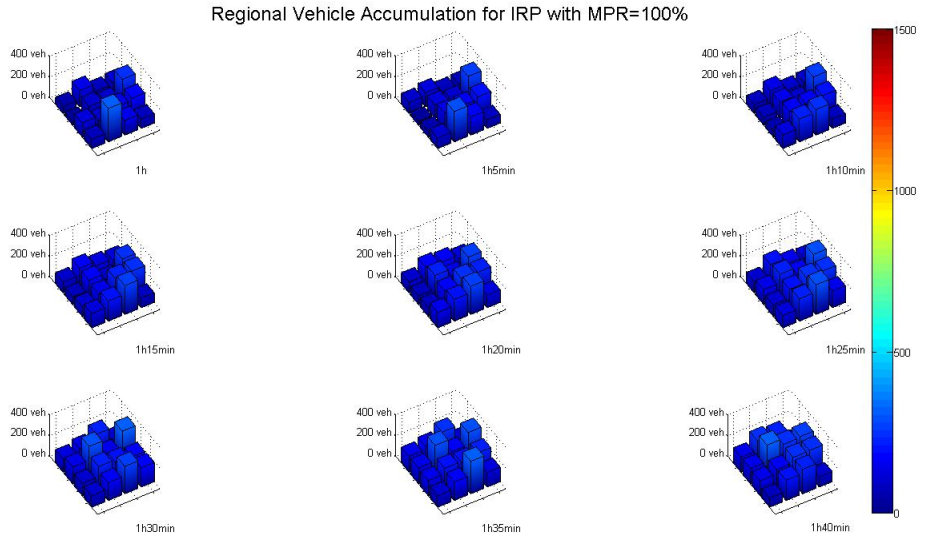


Figure 6.4: Vehicle Accumulation for IRP from 1h-1h40min of the simulation, at 5 min intervals, in regions 1-16

however the vehicle distribution is not even throughout the regional network. As is expected, MLR is not able to maintain vehicle accumulation levels below 500 veh. for 5 out of 16 regions, while the vehicle distribution over all regions is markedly uneven, compared to IRP and PRM. Based on this network setup, regional vehicle accumulation levels over 500 veh. indicate that regions are congested. Vehicle distribution is an important indicator of network resource utilization. Homogeneous vehicle distribution over all regions signifies higher network resource utilization.

6.2.2 Comparison of route guidance approaches in mixed class application with various combinations of market penetration rates

Several mixed class application cases are considered, with different combinations of market penetration rates, MPR1 and MPR2, for 1^{st} and 2^{nd} traveler classes respectively. For the prescriptive, pre-trip route guidance provided to the 2^{nd} class of travelers, the compliance model originally suggested by Papageorgiou [125] was

extended, whereby a certain fraction of 2^{nd} class travelers will comply to the advice provided and the remaining fraction will disregard it. Concretely, assuming non-compliance rate NC for the 2^{nd} class of travelers, one can say that the travel demand for each traveler class at departure time h for every Origin-Destination pair, $\forall b \in \mathcal{K}_o, e \in \mathcal{K}_d$ is:

$$\zeta_{b,e}^{IRP}(h) = MPR1 \cdot \zeta_{b,e}(h) \tag{6.2}$$

$$\zeta_{b,e}^{PRM}(h) = (MPR2 - NC) \cdot \zeta_{b,e}(h) \tag{6.3}$$

$$\zeta_{b,e}^{MLR}(h) = 1 - (MPR1 + (MPR2 - NC)) \cdot \zeta_{b,e}(h) \tag{6.4}$$

The selected 3 market penetration rates seem consistent with experts' predictions regarding the adoption of autonomous vehicle technology [20,21]. If an approximate backwards extrapolation extrapolation is made, MPR1=70% by early 2040, MPR1=40% by early 2030 and MPR1=10% by early 2020 can be safe assumptions. The market penetration values of 2^{nd} traveler class MPR2 in each mixed class routing scenario are MPR2=ϕ_{10}, $\phi_{10} \in \{10\%, 40\%, 70\%\}$, for MPR1=10%, MPR2=$\phi_{40}$, $\phi_{40} \in \{40\%, 60\%\}$, for MPR1=40% and MPR2=$\phi_{70}$=30%, for MPR1=70%. Obviously, in the last 2 scenarios, for $\phi_{40} = 60\%$ and $\phi_{70} = 30\%$, the 2^{nd} traveler class comprises the maximum possible percentage such that MPR1+MPR2=100%. Equal market penetration rates MPR1=MPR2=MPR3=33.3% are also considered, so as to more accurately evaluate the overall network performance, when the degree of penetration does not affect the result. For the 2^{nd} class of travelers, non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$ are considered. As before, threshold coefficient $r_{IRP} = 1$ for the Public Transit Diversion mechanism in IRP routing methods and $r_{PRM} = 1$ for PRM routing method. Similar to earlier, this market penetration rate combination was selected so as to rate the individual performance of each routing method, where the degree of penetration plays no role for the results, but PRM's non-compliance rate is also evaluated. Table 6.5 presents performance gain

results for all market penetration rate and non-compliance rate combinations mentioned above for mixed application of IRP (S1), IRP (S2), PRM (NC) and MLR. A subsequent analysis of the individual performance metric results for each market penetration rate combination follows, in an effort to gain better insight regarding the effect MPR1, MPR2, MPR3, explicitly and implicitly, through the implementation of non-compliance NC, have on Public Transit Diversion, Incomplete Trips Rate and Average Travel Time for each routing method individually. It should be noted that individual performance metric results are used to derive all the performance gain results as presented on Table 6.5. Similar to Figures 6.2–6.4, Figures 6.5–6.11 demonstrate regional vehicle accumulation levels for a mixed application of IRP, PRM and MLR with corresponding market penetration rates MPR1, MPR2, MPR3, during the same simulation time period of 40 min as mentioned before. Tables 6.6–6.12 present individual performance metric results for each routing method in a mixed application of IRP, PRM and MLR with corresponding market penetration rates MPR1, MPR2, MPR3 for non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$ and scenarios (S1),(S2).

Table 6.5: Various market penetration rate combinations for scenarios (S1),(S2) of IRP representing 1^{st} class travelers, PRM representing 2^{nd} class travelers with NC=0%, 50%, 70% non-compliance rate and MLR representing non-compliant segment of 2^{nd} class, as well as 3^{rd} class travelers

Ratio Weighted Average	$C_{RWA}(\%)$					
Scenarios	**(S1)**			**(S2)**		
NC(%) / **MPR1\MPR2\MPR3 for IRP\PRM\MLR** (%\%\%)	**0**	**50**	**70**	**0**	**50**	**70**
10\ 10(1-NC)\ 80(1+0.125·NC)	9.74	15.96	13.60	15.80	14.44	14.39
10\ 40(1-NC)\ 50(1+0.8·NC)	27.64	18.62	15.66	28.70	20.25	17.84
10\ 70(1-NC)\ 20(20+3.5·NC)	27.02	28.42	19.92	30.13	28.30	20.31
33.3\ 33.3(1-NC)\ 33.3(1+NC)	40.73	45.34	42.97	46.25	47.35	46.85
40\ 40(1-NC)\ 20(1+2·NC)	42.30	47.80	48.81	46.38	48.53	49.82
40\ 60(1-NC)\ 60·NC	36.12	44.19	48.85	43.73	47.36	48.87
70\ 30(1-NC)\ 30·NC	44.51	48.31	48.14	46.81	48.37	48.51

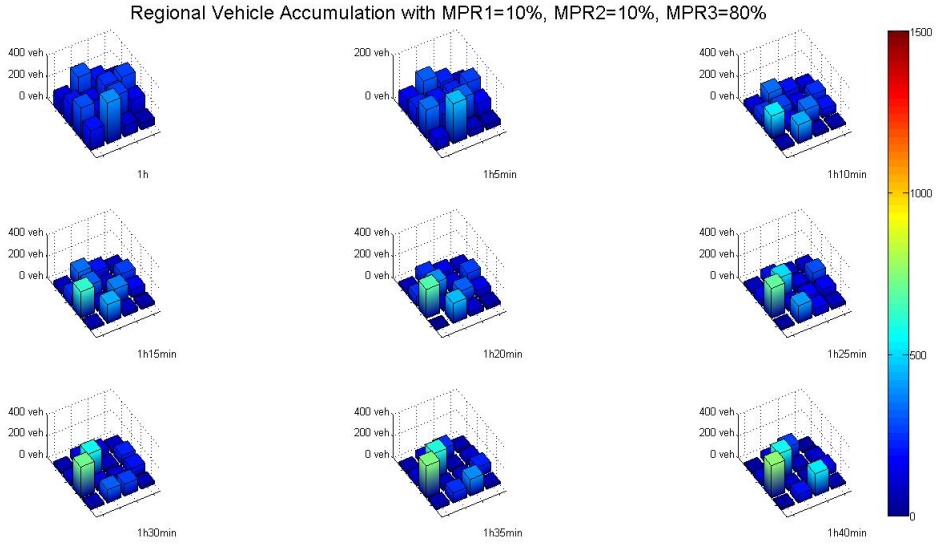


Figure 6.5: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

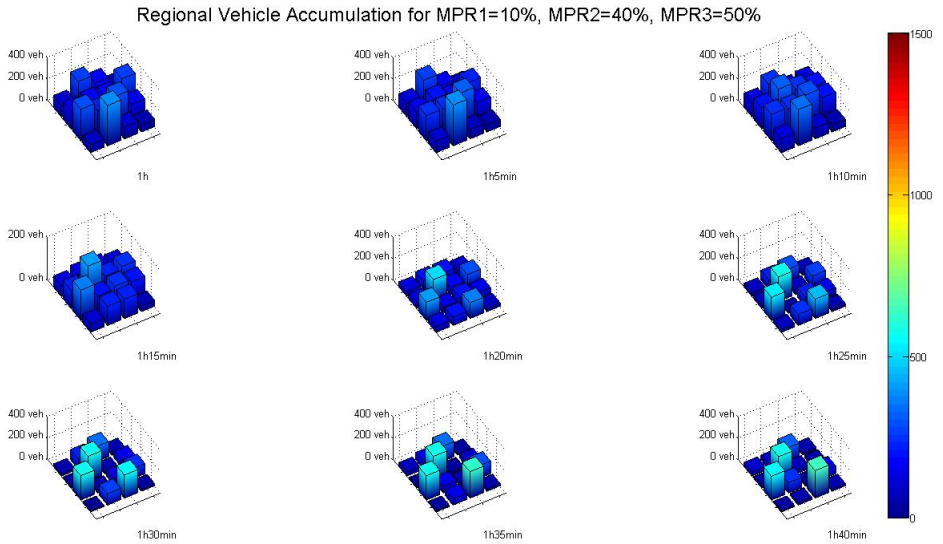


Figure 6.6: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

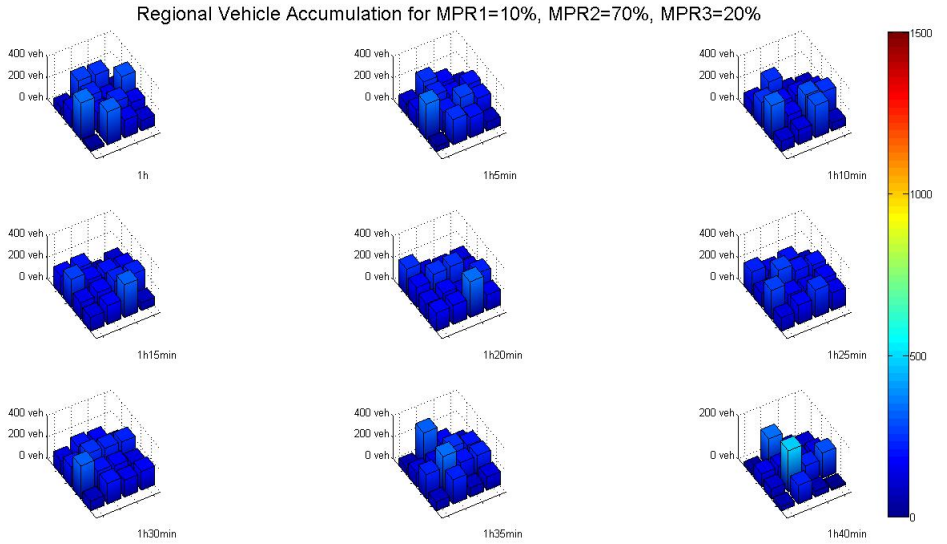


Figure 6.7: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

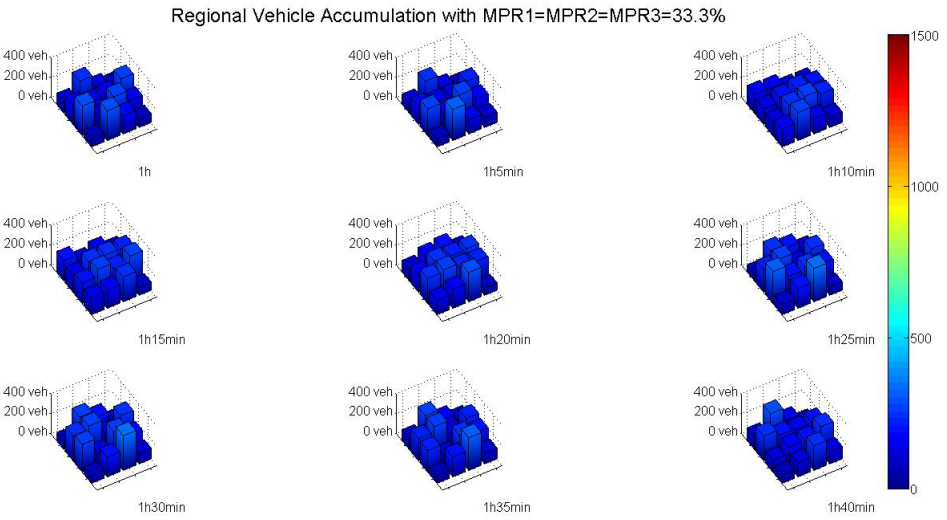


Figure 6.8: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

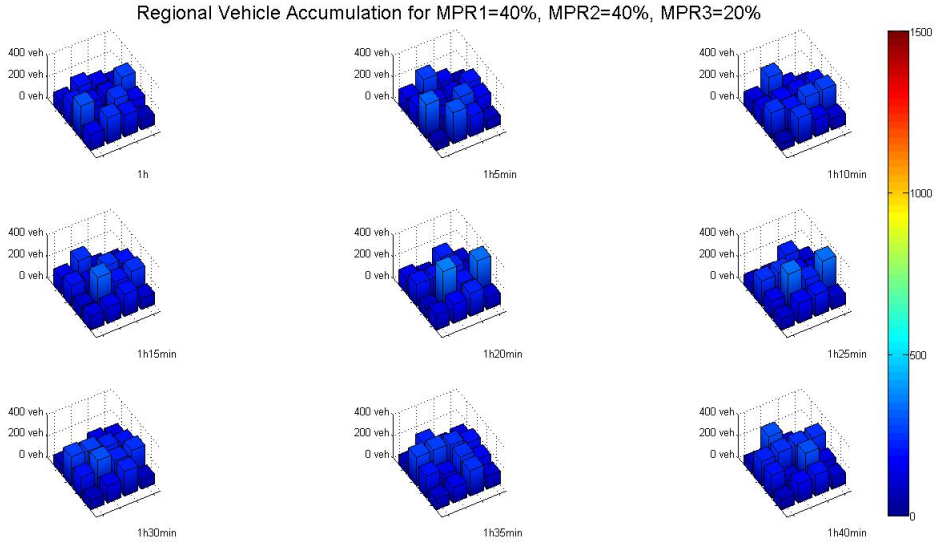


Figure 6.9: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

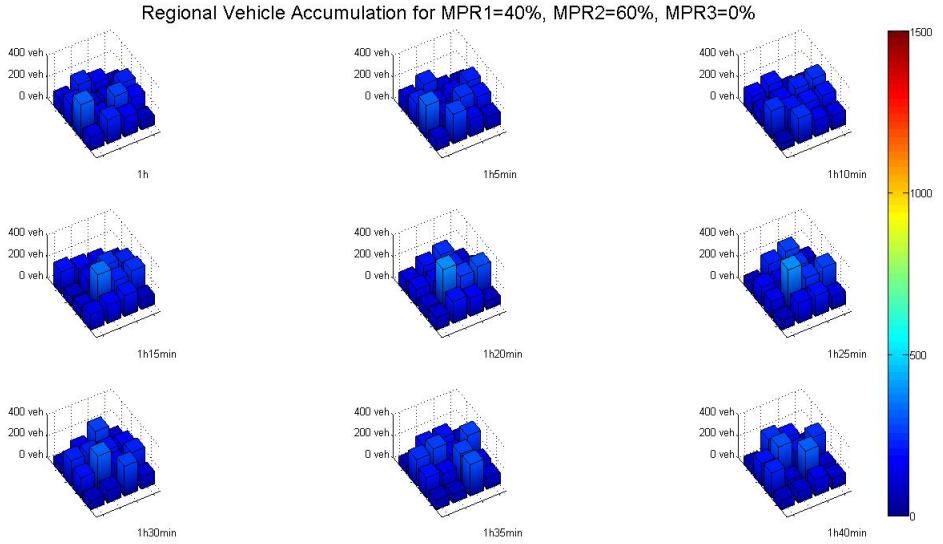


Figure 6.10: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

Table 6.6: Comparison of individual performance metrics for each routing method in mixed application of IRP, PRM and MLR with MPR1=10%, MPR2= 10(1-NC)%, MPR3=80(1+0.125NC)% with NC= 0%, 50%, 70% for scenarios (S1), (S2)

MPR1=10%, MPR2=10(1-NC)%, MPR3=80(1+0.125NC)%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	49.77	0.45	1284.7	57.94	0.42	1589.1
PRM	57.10	0.19	1022.2	48.43	0.84	2349.0
MLR	28.25	8.72	1584.3	17.75	17.20	2519.6
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	61.55	0.49	1073.1	61.22	0.04	1447.3
PRM	50.88	0.20	1175.4	49.55	0.42	2387.7
MLR	27.95	8.29	1443.0	19.14	19.68	2567.0
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	59.17	0.32	1595.4	60.78	0.07	1491.5
PRM	50.49	0.27	2454.7	51.58	0.24	2376.3
MLR	20.08	19.07	2580.0	19.60	19.42	2575.8

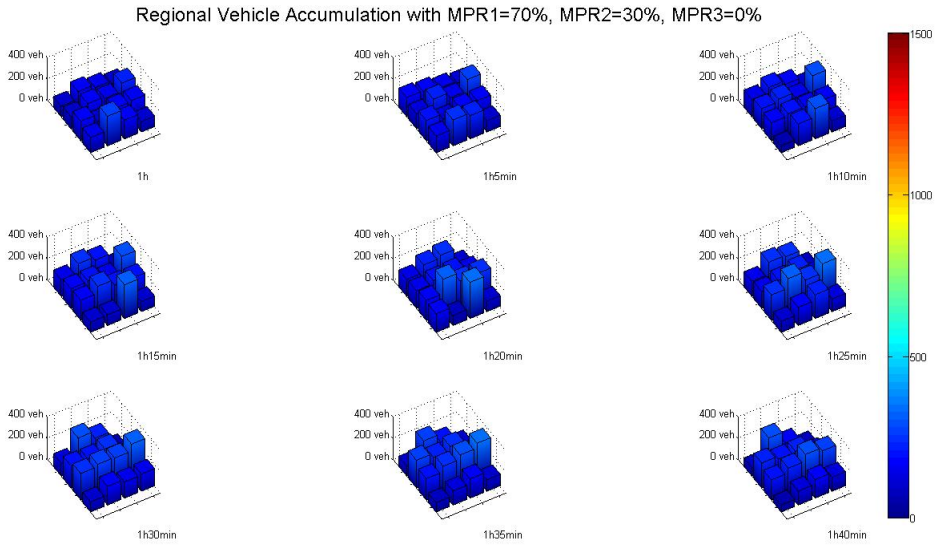


Figure 6.11: Regional Vehicle Accumulation for IRP\PRM\MLR from 1h-1h40min of the simulation, at 5 min intervals

We can visually confirm from Figures 6.5–6.11, that, when MPR1 exceeds 30%, regional vehicle accumulation levels never exceed 500 veh. Compared to an application of IRP with MPR=100%, however, as seen in Figure 6.4, vehicle distribution

Table 6.7: Comparison of individual performance metrics for each routing method for mixed application of IRP, PRM and MLR with MPR1=10%, MPR2=40(1-NC)%, MPR3=50(1+0.8NC)% with NC=
0%, 50%, 70% for scenarios (S1), (S2)

MPR1=10%, MPR2=40(1-NC)%, MPR3=50(1+0.8NC)%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	59.59	0.24	1563.8	45.73	0.26	1626.6
PRM	41.29	1.91	2051.9	38.67	2.65	2107.7
MLR	6.74	6.16	2146.3	6.31	6.55	2137.7
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	58.40	0.50	1622.1	55.48	0.06	1447.9
PRM	43.97	1.68	2248.0	45.15	1.28	2196.2
MLR	16.01	14.31	2454.3	15.29	12.91	2423.7
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	58.82	0.31	1596.4	59.37	0.06	1461.9
PRM	48.76	0.90	2319.5	48.20	0.95	2306.9
MLR	18.03	15.94	2527.4	16.86	15.12	2465.0

over all regions appears to be more uneven. As a general observation from Table 6.5, one can see that for most market penetration rates for (S2) and when MPR1= 40% for (S1), when non-compliance rate NC is higher, performance is the highest. This can be attributed to the fact, that, during the forecasting step of IRP, all virtual vehicles generated from the current step h onwards, are assigned a route choice based on MLR. High NC values mean that more travelers not belonging to 1^{st} class, are assigned routes based on MLR rather than PRM, improving travel time estimation accuracy for IRP during the mixed class application. This also means that PRM does not contribute as much as IRP to the overall network performance. Another observation is that when MPR1>30%, performance gains compared to MLR for when $\text{MPR}_{MLR} = 100\%$, the worst-performing routing method, almost always exceed 40%, for all non-compliance rate combinations. Even for MPR1\MPR2\MPR3 = 10% \10% \80%, with NC=70%, the case with least market penetration for both IRP and PRM, for the more realistic scenario (S1), there is still a performance gain of 13.60%. IRP for (S1), in combination with PRM, can potentially benefit

Table 6.8: Comparison of individual performance metrics for each routing method when it is for a mixed application of IRP, PRM and MLR with MPR1=10%, MPR2=70(1-NC)%, MPR3=20(1+3.5NC)%
with NC=
0%,50%,70% for scenarios (S1), (S2)

MPR1=10%, MPR2=70(1-NC)%, MPR3=20(1+3.5NC)%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	58.19	0.38	1711.3	39.66	0.13	1562.8
PRM	30.35	6.12	2158.8	29.99	0.17	2136.1
MLR	3.45	2.13	1972.3	3.38	1.66	1936.9
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	59.39	0.26	1583.4	47.96	0.17	1574.3
PRM	41.09	1.56	2031.5	42.37	1.84	2072.1
MLR	8.92	5.23	2107.1	7.75	0.66	2151.3
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	59.02	0.30	1629.7	54.89	0.08	1474.6
PRM	42.76	1.44	2192.3	45.01	1.61	2269.1
MLR	15.15	12.79	2417.4	15.10	12.60	2408.4

overall network performance, provided that 50%≥NC≤70%, even though MPR1 and MPR2 are as low as 10%. We also demonstrated the positive influence of autonomous vehicles in the overall network performance, evidenced by a performance gain $\mathcal{C}_{\text{RWA}}$=40.73% for (S1) and $\mathcal{C}_{\text{RWA}}$=46.25%, given equal market penetration rates MPR1=MPR2=MPR3=33.3% and non-compliance rate NC=0%. Individual performance metric results from Tables 6.6-6.12 provide a more detailed insight regarding the different effect the information provided to IRP in scenarios (S1) and (S2) has on Public Transit Diversion, Incomplete Trips Rate and Average Travel Time for each routing method individually. For MPR1=10%, MPR2=10(1-NC)%, MPR3=80(1+0.125NC)% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that, owing to the higher PTD rate values for IRP in all combinations but the one for NC=0% (S1), ATT for IRP never exceeds 1600s, whereas ATT for PRM and MLR exceed 2000s in the majority of combinations. For NC=0% (S1), ATT for PRM is shorter than ATT for IRP, which can be explained by the fact

Table 6.9: Comparison of individual performance metrics for each routing method when it is for a mixed application of IRP, PRM and MLR with MPR1=33.3%, MPR2=33.3(1-NC)%, MPR3=33.3(1+NC)% with NC= 0%,50%,70% for scenarios (S1),(S2)

MPR1=33.3%, MPR2=33.3(1-NC)%, MPR3=33.3(1+NC)%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	39.99	0.62	1504.5	28.26	0.0	1479.3
PRM	24.60	0.09	1876.8	27.48	0.14	1792.4
MLR	0.90	1.17	1713.9	0.09	0.0	1614.4
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	42.73	0.14	1432.9	32.79	0.05	1493.9
PRM	24.89	0.22	1831.8	28.61	0.14	1826.7
MLR	0.51	0.55	1662.6	0.31	0.59	1660.9
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	38.11	0.99	1558.9	35.36	0.02	1499.6
PRM	25.35	0.34	1915.3	29.00	0.12	1869.6
MLR	0.89	2.79	1801.8	0.51	0.40	1705.6

that this is the only case where the PTD value for PRM is higher than that of IRP. In all combinations, ITR for MLR never goes lower than 8%, whereas ITR values for IRP and PRM never exceed 0.50% and 0.90% respectively. While this could be partly attributed to IRP's and PRM's superior performance, the main factor is the low $1^{st}/3^{rd}$ class and $2^{nd}/3^{rd}$ class traveler ratio respectively. For MPR1=10%, MPR2=40(1-NC)%, MPR3=50(1+0.8NC)% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that, owing to the higher PTD rate values for IRP in all combinations, ATT for IRP never exceeds $1650s$, whereas ATT for PRM and MLR exceed $2000s$ in all combinations. In all combinations, ITR for MLR never goes lower than 6%, whereas ITR values for IRP never exceed 0.50%. ITR for PRM ranges from 0.90%-2.65%. This could be partly attributed to the $2^{nd}/3^{rd}$ class traveler ratio which ranges from 15%-80%. This can be confirmed by the fact that the highest ITR values for PRM occur for NC=0% (S1),(S2) corresponding to the highest ratio value of 80%. IRP outperforms both PRM and MLR while PRM finds itself consistently in the middle, in terms of performance. For

Table 6.10: Comparison of individual performance metrics for each routing method when it is for a mixed application of IRP, PRM and MLR with MPR1=40%, MPR2=40(1-NC)%,
MPR3=20(1+2NC)% with NC=
0%, 50%, 70% for scenarios (S1), (S2)

MPR1=40%, MPR2=40(1-NC)%, MPR3=20(1+2NC)%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	34.22	0.74	1529.4	25.33	0.01	1487.3
PRM	19.33	0.94	1828.8	20.21	0.38	1792.8
MLR	1.00	0.40	1647.3	0.08	0.0	1599.3
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	33.06	0.02	1453.2	29.53	0.0	1471.1
PRM	17.90	0.20	1775.4	18.68	0.17	1826.1
MLR	0.21	0.0	1600.5	0.31	0.0	1616.8
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	31.93	0.01	1469.4	29.65	0.0	1452.7
PRM	20.89	0.08	1817.2	18.53	0.10	1820.6
MLR	0.20	0.0	1611.8	0.51	0.0	1616.7

MPR1=10%, MPR2=70(1-NC)%, MPR3=20(1+3.5NC)% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that, owing to the higher PTD rate values for IRP in all combinations, ATT for IRP never exceeds 1750s, whereas ATT for PRM exceeds 2000s in all combinations. ATT for MLR exceeds 2000s for all combinations where NC>0%. Another observation that can be made as regards ATT for PRM is that it is consistently shorter than the ATT for MLR, except for the cases where NC=0%. Specifically, when PTD for PRM falls below 38%, ATT for PRM performs worse than MLR, regardless of the rate of non-compliance NC. This can be explained by the fact that, in this particular market penetration rates combination, MPR2>MPR3. This difference is highest for NC=0%. Especially for NC=0% (S1), it can also be observed that ITR for MLR is lower than ITR for PRM. ITR for PRM ranges from 0.17%-6.12%. In all combinations but for NC=0% (S1), NC=50% (S2), ITR for MLR performs the worst. An additional factor affecting ITR, which is the availability of route assignment information for the IRP forecast-

Table 6.11: Comparison of individual performance metrics for each routing method for mixed application of IRP, PRM and MLR with MPR1=40%, MPR2=60(1-NC)%, MPR3=60NC% with NC=0%,50%,70% for scenarios (S1),(S2)

MPR1=40%, MPR2=60(1-NC)%, MPR3=60NC%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	36.48	2.08	1595.8	25.18	0.03	1506.9
PRM	17.57	3.46	1909.1	19.03	0.74	1771.0
MLR	N/A	N/A	N/A	N/A	N/A	N/A
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	36.27	0.16	1458.7	27.88	0.01	1479.5
PRM	21.44	0.36	1769.6	19.73	0.18	1809.3
MLR	0.56	0.39	1622.2	0.17	0.0	1609.3
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	32.05	0.0	1443.9	28.05	0.0	1481.5
PRM	16.94	0.15	1785.6	21.59	0.09	1803.5
MLR	0.12	0.0	1593.7	0.09	0.0	1613.8

ing step, as is the case in (S2). For the case of NC=50% (S2), MLR outperforms PRM, as regards ITR, but lags slightly behind PRM, as regards ATT. This could be partly attributed to a combination of the positive influence of IRP with perfect information scenario (S2) route assignment and the fact that MPR3-MPR2>0. For NC=0% (S1), IRP with imperfect information scenario (S1) in combination with the low MPR3, seems to mostly benefit MLR with respect to both ITR and ATT. ITR for IRP and PRM achieve comparable performance for NC=0% (S2), and 10 times better than MLR, despite the fact that the MPR2-MPR3 difference is the highest possible. IRP consistently outperforms all other routing methods with respect to ITR and ATT. For MPR1=33.3%, MPR2=33.3(1-NC)%, MPR3=33.3(1+NC)% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that PTD values for IRP never exceed 45% which is less than the lowest PTD value for IRP for all market penetration rate combinations where MPR3<30%. As expected, the highest PTD rate for IRP is for (S1) with NC=0% and the lowest PTD rate for IRP is for (S2) with NC=0%. The reason for this is the fact that in (S1), IRP's forecasting step

Table 6.12: Comparison of individual performance metrics for each routing method for mixed application of IRP, PRM and MLR with MPR1=70%, MPR2=30(1-NC)%, MPR3=30NC% with NC=0%, 50%, 70% for scenarios (S1), (S2)

MPR1=70%, MPR2=30(1-NC)%, MPR3=30NC%						
Scenario	**(S1)**			**(S2)**		
NC= 0%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	22.48	1.18	1598.5	20.94	0.04	1554.0
PRM	10.98	0.58	1763.3	12.99	0.18	1738.5
MLR	N/A	N/A	N/A	N/A	N/A	N/A
NC=50%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	21.53	0.02	1555.8	21.64	0.01	1544.6
PRM	9.71	0.20	1734.0	11.93	0.09	1706.2
MLR	0.15	0.0	1550.6	0.07	0.0	1555.5
NC=70%	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$	$C_{PTD}(\%)$	$C_{ITR}(\%)$	$C_{ATT}(s)$
IRP	21.92	0.04	1578.7	22.05	0.03	1559.1
PRM	9.98	0.07	1727.6	11.32	0.05	1722.2
MLR	0.24	0.0	1559.7	0.06	0.0	1558.7

accuracy is diminished, due to information available only for the 1^{st} class travelers. All other traveler classes are assigned routes based on MLR. On the other hand, for (S2) and NC=0%, IRP's forecasting step accuracy is higher for this particular combination of market penetration rates, since information is available for both 1^{st} and 3^{rd} traveler classes. Due to it's overall high PTD rate, IRP presents with the lowest ATT in both (S1) and (S2). More specifically, it can be observed that when the individual PTD rate for IRP exceeds 40%, PTD contributes the most to the reduction of individual ATT for IRP, as is the case for (S1) and NC=50%. Another positive effect of IRP is the reduction of individual PTD rate for MLR, which drops below 1%. While the ITR for IRP for (S2) tends to be the lowest, for (S1), due to the limited information availability, prediction is not as accurate, leading to ITR almost half of that for MLR, which exhibits the highest ITR rate for all but one combinations. The ATT for PRM is consistently lagging behind the ATT for IRP and MLR. The best performing ATT for PRM is for (S2) and NC=0%. This means that an indirect observation can be made, of the effect market penetration rate has

on the individual performance of PRM and MLR, but not IRP, as 1^{st} class travelers are assumed to be fully compliant. ITR for PRM is at it's lowest for (S2), followed closely by (S1). Additionally, one must take into consideration the beneficial effect the 1^{st} traveler class, employing IRP, has on the 2^{nd} and especially the 3^{rd} traveler class. Compared to ATT results for PRM from Tables 6.2 and 6.4, for several degrees of penetration, ranges from 18%-30% improvement on the ATT for PRM in the mixed application. It should be noted that the ATT for MLR seems to benefit the most from the presence of 1^{st} class travelers, as can be seen from the improvement ranging between 40%-45%, compared to results from Table 6.2. Compared to MLR, PRM seems to benefit less from IRP, however, compared to performance results when IRP is absent, see 6.4, there is individual improvement of approximately 10% for ITR and 22% for ATT of PRM. For MPR1=40%, MPR2=40(1-NC)%, MPR3=20(1+2NC)% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that PTD values for IRP never exceed 35%. Similar to the previous market penetration rate combination, PTD values for MLR never exceed 1%, however, there is a 10% average decrease to PTD for PRM. This improvement comes about due to the 10% increase of MPR1, which, however, has adverse effects on ATT for PRM, which performs the worst in all cases. ITR for MLR is the lowest for all combinations, which can be partly attributed to the 10% increase of MPR1, but also to the fact that 3^{rd} class travelers comprise the least proportion of traveler population. IRP consistently outperforms all other routing methods with respect to ATT. For MPR1=40%, MPR2=60(1-NC)%, MPR3=60NC% with NC=0%,50%,70% for scenarios (S1),(S2), it can be observed that, similar to the previous market penetration rate combination, PTD values for IRP never exceed 35%. Similar to the previous market penetration rate combination, PTD values for MLR never exceed 1%. PTD results for PRM are comparable to the ones of the previous market penetration rate combination, never exceeding 22%. ITR for PTM is the worst for all combinations,

with the highest value presenting for NC=0% (S1), which also exhibits the highest ITR value for PRM. This can readily be explained by the fact that the forecasting step of IRP for (S1) makes use of imperfect route assignment information. Once again, IRP consistently outperforms all other routing methods with respect to ATT. For MPR1=70%, MPR2=30(1-NC)%, MPR3=30NC% with NC=0%,50%,70% for scenarios (S1),(S2), an observation can be made that, PTD values for IRP never exceed 25%. PTD values for PRM never exceed 15%. This improvement comes about due to the 30% increase of MPR1, compared to the previous market penetration rate combination, which leads to a decreased performance on ATT for PRM in all cases. ITR results for PRM are the worst performing in all cases but one, for NC=0% (S1). This can be explained by the fact that the forecasting step of IRP for (S1) makes use of imperfect route assignment information In all other cases ITR for MLR is the lowest, due to the fact that IRP tends to benefit the 3^{rd} traveler class the most, even in cases of high non-compliance (NC=70%), where MPR3>3×MPR2. Once again, IRP consistently outperforms all other routing methods with respect to ATT.

6.3 Chapter Overview

In this chapter, a market penetration scheme was introduced. Three distinct traveler classes are defined, the 1^{st} class of travelers equipped with autonomous vehicles, the 2^{nd} traveler class comprising of RGIS-equipped, conventional vehicles and the 3^{rd} class traveler class comprising of unequipped, conventional vehicles. Certain assumptions are made for each traveler class. The prescriptive pre-trip guidance approaches introduced in 5, IRP and PRM, a predictive and non-predictive routing method respectively, along with MLR, which gives the multinomial logit-based DSUE, are used to represent the 1^{st}, 2^{nd} and 3^{rd} traveler class respectively. First, 1^{st} and 2^{nd} traveler classes are assumed to have varied degrees of market

penetration, with the remaining population of travelers employing MLR. After the evaluation of performance for each individual traveler class, it is shown that MLR performs the worst in most cases. Then, a new performance metric is introduced, overall performance improvement $\mathcal{C}_{\text{RWA}}$, as the weighted average of improvement of all network performance metrics (Total Vehicle Time, Speed Variability, Public Transit Diversion rate, Incomplete Trips Rate, Average Travel Time), as compared with the 3^{rd} class of travelers. Subsequently, for varied combinations of market penetration rates corresponding to each class (MPR1, MPR2, MPR3), as well as various non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$ of travelers belonging to 2^{nd} class, the overall performance improvement $\mathcal{C}_{\text{RWA}}$ over the case of 3^{rd} traveler class employing MLR with 100% market penetration is evaluated. Individual performance results for each traveler class participating in the market penetration scheme are also evaluated. The performance robustness of IRP is verified. In the chapter that follows, a more detailed analysis of the conclusions regarding this work is presented as well as proposed future research, extending and building upon the current work.

Chapter 7

Conclusion and Future work

7.1 Conclusion

Congestion in urban traffic networks is a complex problem that is inextricably linked with quality of life in urban areas. It constitutes an important problem that affects people in explicit but also implicit ways. In a supply and demand scheme, where supply represents network resources (infrastructure) and demand represents travelers wishing to make use of the network (travel demand), the high cost of developing new infrastructure means that cities are only left with 2 options. To suppress demand through toll and parking pricing, or manage supply in an optimal manner by use of traffic signal control, route guidance. The latter of the two supply management applications, route guidance, is the main focus of this thesis.

Travelers, either through the use of smartphone applications or vehicle-integrated navigation equipment, are able to receive up-to-date information, including map location, prevailing traffic state and expected travel times. Route guidance can benefit drivers individually, however, through coordination with traffic control centers, system-wide benefits can be obtained. Congestion during peak periods can be alleviated through a suggestion mechanism that will enable the distribution of

vehicles along different paths with similar origin and destination locations. The application of this concept in large urban traffic networks is regrettably not a trivial problem, due to the large number of links, intersections and most importantly, travelers. Region-based route guidance can be considered as route guidance at an aggregated level. The advantages of the Urban-scale MFD concept become abundantly clear, when one considers the fact that, when partitioned appropriately into homogeneously congested regions, an urban traffic network can be modeled as a reservoir system with very well-defined operating points, critical accumulations derived from the Urban-scale MFDs that correspond to each homogeneously congested region. Subsequently a number of methodologies can be applied, such as region-based prescriptive route guidance with Public Transit Diversion integration, allowing for the distribution of travel demand so that congestion levels rarely exceed said operating points. In this thesis, weighted feature extensions of center-based clustering as well as normalized spectral clustering were applied to a network encapsulating Singapore's CBD to derive homogeneously congested regions. For this specific network setup, the weighted feature extension to k-harmonic means clustering outperformed all other clustering methods. Subsequently, a multi-class extension of the Network Transmission Model was introduced. A non-predictive strategy learning-based routing method employing Proxy Regret Matching (PRM) and a predictive simulation-based routing method, Incremental Route Planning (IRP), were developed as region-based prescriptive route guidance approaches with pre-trip information dissemination and subsequently applied on a regional network where the Network Transmission Model was used to simulate region-based traffic dynamics. Simulation Results for a diamond-shaped grid network comprising of 16 regions, with an area of 25 km^2 (5x5) for each region, where the Urban-scale MFDs for all regions are identical, have demonstrated the superiority of predictive routing methods, as IRP outperforms PRM and MLR. This, of course, is to be expected, since

IRP belongs to the predictive routing approach category, which means that travel time estimation accuracy is the highest amongst all methods compared. PRM, as expected from a non-predictive routing method, finds itself in the middle, performing better than MLR but worse than IRP. As expected, MLR under-performs in all defined performance metrics.

A potential issue that past Urban-scale MFD-related literature has failed to consider, is the fact that not all traveler features are uniform. Multi-class dynamic traffic models have become more appealing to investigators in recent years, however, due to the automotive industry's renewed interest in autonomous vehicle technology adoption. In this thesis, three distinct traveler classes were defined, the 1^{st} class employing a novel predictive routing method, IRP, to represent routing behavior of autonomous vehicles, the 2^{nd} class employing a non-predictive routing method, PRM, to represent RGIS-equipped, conventional vehicles, and the 3^{rd} class employing multinomial logit-based routing to represent unequipped, conventional vehicles. For the autonomous vehicles, 2 scenarios were considered, which make use IRP for fully compliant route guidance. In (S1), autonomous vehicles are assumed to be privately owned, which means that information provision is limited to other autonomous vehicles only. Hence, travel time prediction accuracy is negatively affected by the assumption that all other vehicles are routed as if unequipped with RGIS. In (S2), autonomous vehicles are assumed to be publicly owned, as part of a public one-way ridesharing service which means that information provision includes historical and real-time data about all vehicles in the urban traffic network. Hence, travel time prediction accuracy is positively affected. In essence, market penetration rates were varied explicitly through specification of MPR1, MPR2, MPR3 for 1^{st}, 2^{nd} and 3^{rd} traveler class respectively, as well as non-compliance rates NC for 2^{nd} class of travelers. In addition, respective market penetration rates for 1^{st}, 2^{nd} and 3^{rd} traveler classes were varied implicitly through the introduction of Public Transit

Diversion mechanisms to each routing method employed to represent each distinct traveler class. Subsequently, simulations were ran with a similar network setup as mentioned previously, however, the travel demand values were readjusted and free flow conditions speed was set to $v_{i,f} = 45$ km/h, to increase the level of realism for the simulation. After analyzing performance results for each routing method, MLR exhibits the worst performance for the majority of performance metrics. Then, a new performance metric is introduced, overall performance gain $\mathcal{C}_{\text{RWA}}$, as the weighted average of performance gains of all network performance metrics (Total Vehicle Time, Speed Variability, Public Transit Diversion rate, Incomplete Trips Rate, Average Travel Time), as compared with the 3^{rd} class of travelers. Subsequently, for varied combinations of market penetration rates corresponding to each class (MPR1, MPR2, MPR3), as well as various non-compliance rates $NC \in \{0\%, 50\%, 70\%\}$ of travelers belonging to 2^{nd} class, the overall performance gains $\mathcal{C}_{\text{RWA}}$ over the case of MLR with 100% market penetration is evaluated. Performance gains ranging from approximately 10%-50% could be seen, compared to individual application of MLR, a more realistic reference case. In the case of scenarios (S1),(S2) for IRP, performance gains seem to increase in tandem with the degree of market penetration, however it can be observed that for MPR$\geq$50%, performance gradually decreases until, at MPR=100%, it reaches a value of $\mathcal{C}_{\text{RWA}}$ $\approx$47% which lies in between values achieved for MPR=50% ($\mathcal{C}_{\text{RWA}}$ $\approx$50%) and MPR=30% ($\mathcal{C}_{\text{RWA}}$ $\approx$35%). These results demonstrate that performance gains are dependent on the forecasting step methodology. Nevertheless, performance robustness for both IRP (S1) and IRP (S2) is assured, as is evident from the consistent gain in performance for every degree of market penetration. An analysis of individual performance metric results for each market penetration rate combination was also conducted, in an effort to gain better insight regarding the effect MPR1, MPR2, MPR3, explicitly and implicitly, through the implementation of non-compliance NC, have on Public Transit Diversion, In-

complete Trips Rate and Average Travel Time for each routing method individually. A few highlights regarding individual performance that are worth mentioning follow:

- For the combination where MPR1=10% employing IRP, MPR2=10% employing PRM and MPR3=80% employing MLR, ITR for MLR never falls below 8%, whereas ITR values for IRP and PRM never exceed 0.50% and 0.90% respectively. While this could be partly attributed to IRP's and PRM's superior performance, the main factor is the very low $1^{st}/3^{rd}$ class and $2^{nd}/3^{rd}$ class traveler ratio respectively.

- When PTD for PRM falls below 38%, ATT for PRM performs worse than MLR, regardless of the rate of non-compliance NC.

- The 1^{st} traveler class, employing IRP, has beneficial effects on the 2^{nd} and especially the 3^{rd} traveler class. Compared to MLR, PRM seems to benefit less from IRP, however, compared to performance results when IRP is absent, there is individual improvement of approximately 10% for ITR and 22% for ATT of PRM.

- IRP was consistently providing 1^{st} class travelers with the shortest ATT, implicitly guaranteeing full compliance, even in the case full compliance is not a mandatory feature for 1^{st} class travelers. This indicates the positive potential impact of autonomous vehicle technology adoption.

Autonomous vehicle technology adoption is a prominent feature in news reports [20,21,130], becoming ubiquitous not only in the automotive industry, but also in technology companies like Apple Inc., Waymo (subsidiary of Alphabet Inc.) and many others. Smart city initiatives around the globe are also keen on integrating autonomous vehicle technology as part of their quest to improve urban traffic efficiency, reduce fuel consumption and upgrade their citizens' quality of life. Traffic

Control Centers can derive several benefits from the simulation of mixed application of various types of routing methods, representing distinct traveler classes, in a market penetration scheme. Regional vehicle accumulation as well as regional path travel time estimation accuracy increases, as the assumptions made regarding the routing behavior of each traveler class can have a significant impact on the resulting network traffic patterns. Due to the low computational effort required, region-based urban traffic management lends itself especially well for real-time traffic operations, minimizing total vehicle travel time during recurrent congestion, as well as providing updated route advisory in case of non-recurrent traffic congestion, e.g. incidents.

7.2 Future work

- This thesis proposes a multi-class extension of a regional dynamic traffic model, the Network Transmission Model. The bell-shaped function by [99] used to approximate the Urban-scale MFD has been shown to be suitable for describing region-based traffic dynamics, however, it is more frequent in recent Urban-scale MFD-related literature to represent the Urban-scale MFD with a 3^{rd} order polynomial function. The step forward would be, following the partitioning of a specific urban traffic network into homogeneously congested regions, to fit the Urban-scale MFD with a 3^{rd} order polynomial function, with coefficients calibrated according to the historical data provided.

- A public transit diversion mechanism, as integrated to each routing method presented in this dissertation, can be considered as an alternative route choice which does not compromise travelers' departure time preferences. For the next step forward, sophistication should be added to the rather simple Public Transit Model by introducing passenger transition delays and train frequency, with varying travel times.

- For the prescriptive, pre-trip route guidance provided to the 2^{nd} class of travelers, the compliance model originally suggested by Papageorgiou [125] was extended, whereby a certain fraction of 2^{nd} class travelers will comply to the advice provided and the remaining fraction will disregard it. As a next goal, a more sophisticated compliance model should be introduced, such that 2^{nd} class traveler compliance is dependent on the travelers' travel time information reliability, as well as their level of familiarity with the route assigned to them.

- Judging from the results in chapter 6, as regards ATT for PRM in a mixed-class application setting, one could argue that 2^{nd} class travelers would become non-compliant, eventually being subsumed by the 3^{rd} traveler class. Following the implementation of a sophisticated compliance model, as stated above, congestion pricing should be introduced as an additional traffic management measure, where departure time can be considered as an additional decision variable, so that the difference in individual performance between PRM and MLR can be eliminated, thus reaching equilibrium states, with respect to market penetration rate combinations.

Author's Publications

Main Author

1. Antonis F. Lentzakis, Simon I. Ware, Rong Su, Changyun Wen, "Region-based prescriptive route guidance for travelers of multiple classes", Transportation Research Part C: Emerging Technologies 87, 2018, pp. 138-158.

2. Antonis F. Lentzakis, Simon I. Ware, Rong Su, "Region-based dynamic forecast routing for autonomous vehicles", in 19^{th} IEEE Conference on Intelligent Transportation Systems (ITSC), 2016, pp. 1464-1469, 1-4 November 2016.

3. Antonis F. Lentzakis, Rong Su, Changyun Wen, "Strategic learning approach to region-based dynamic route guidance," in 12^{th} IEEE International Conference on Control & Automation (ICCA 2016), 2016, pp. 842-847, 1-3 June 2016.

4. Antonis F. Lentzakis, Rong Su, Changyun Wen, "Time-dependent partitioning of urban traffic network into homogeneous regions," in 13^{th} International Conference on Control, Automation Robotics & Vision (ICARCV),2014, pp. 535-540, 10-12 December 2014.

Co-author

1. Kaizhou Gao, Yicheng Zhang, Ali Sadollah, Antonios Lentzakis, Rong Su, "Jaya, harmony search and water cycle algorithms for solving large-scale real-life urban traffic light scheduling problem", Swarm and Evolutionary Computation, 17 May 2017, *In Press*.

2. Kaizhou Gao, Yicheng Zhang, Rong Su, Antonios Lentzakis, "Discrete harmony search algorithm for solving urban traffic light scheduling problem," in American Control Conference (ACC), 2016, pp. 6239-6244, 6-8 July 2016.

Bibliography

[1] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

[2] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012, pp. 141–154.

[3] V. L. Knoop and S. P. Hoogendoorn, "Empirics of a generalized macroscopic fundamental diagram for urban freeways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 133–141, 2013.

[4] "White paper - european transport policy for 2010: time to decide," European Commission, Tech. Rep., 2001.

[5] C. Wolf, "Traffic cost us $114.8 billion in 2009, texas institute says," 2011, online; accessed 19-December-2012. [Online]. Available: http://www.bloomberg.com/news/2011-01-20/traffic-jams-cost-us-114-8-billion-in-time-fuel-in-09-institute-says.html

[6] K. L. Head, P. B. Mirchandani, and D. Sheppard, *Hierarchical framework for real-time traffic control*, 1992, no. 1360.

[7] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Engineering Practice*, vol. 10, no. 2, pp. 183–195, 2002.

[8] L. Kuyer, S. Whiteson, B. Bakker, and N. Vlassis, "Multiagent reinforcement learning for urban traffic control using coordination graphs," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2008, pp. 656–671.

[9] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 163–174, 2009.

[10] J. Haddad, M. Ramezani, and N. Geroliminis, "Cooperative traffic control of a mixed network with two urban regions and a freeway," *Transportation Research Part B: Methodological*, vol. 54, pp. 17–36, 2013.

[11] A. Garcia, D. Reaume, and R. L. Smith, "Fictitious play for finding system optimal routings in dynamic traffic networks," *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 147–156, 2000.

[12] V. Knoop, S. Hoogendoorn, and J. W. Van Lint, "Routing strategies based on macroscopic fundamental diagram," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2315, no. 1, pp. 1–10, 2012.

[13] M. Hajiahmadi, V. L. Knoop, B. De Schutter, and H. Hellendoorn, "Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*. IEEE, 2013, pp. 1022–1028.

[14] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, "Equilibrium analysis and route guidance in large-scale networks with mfd dynamics," *Transportation Research Part C: Emerging Technologies*, 2015.

[15] N. Geroliminis and D. M. Levinson, "Cordon pricing consistent with the physics of overcrowding," in *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer, 2009, pp. 219–240.

[16] N. Zheng, R. A. Waraich, K. W. Axhausen, and N. Geroliminis, "A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1291–1303, 2012.

[17] M. Simoni, A. Pel, R. Waraich, and S. Hoogendoorn, "Marginal cost congestion pricing based on the network fundamental diagram," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 221–238, 2015.

[18] S. Zhong, M. Bushell, and W. Deng, "A reliability-based stochastic system optimum congestion pricing model under atis with endogenous market penetration and compliance rate," in *Transportation Research Board 91st Annual Meeting*, no. 12-1001, 2012.

[19] N. Zheng, G. Rérat, and N. Geroliminis, "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 133–148, 2016.

[20] IEEE, "Ieee news releases, look ma, no hands!" 2012, online; accessed 20-June-2016. [Online]. Available: http://www.ieee.org/about/news/2012/5september_2_2012.html

[21] J. Hauser, "Amerika schaltet auf autopilot," 2015, online; accessed 20-June-2016. [Online]. Available: http://www.faz.net/aktuell/wirtschaft/unternehmen/verkehrsminister-foxx-selbstfahrende-autos-in-10-jahren-standard-13811022.html

[22] A. Augustine and M. Nava, "Auto dealerships: Destined for disruption," 2016.

[23] A. Skabardonis, P. Varaiya, and K. Petty, "Measuring recurrent and nonrecurrent traffic congestion," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1856, pp. 118–124, 2003.

[24] S. P. Hoogendoorn and P. H. Bovy, *Optimal routing control using VMSs*. Delft University Press: Delft, The Netherlands, 1998.

[25] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.

[26] A. F. Lentzakis, R. Su, and C. Wen, "Time-dependent partitioning of urban traffic network into homogeneous regions," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 535–540.

[27] H. Etemadnia, K. Abdelghany, and A. Hassan, "A network partitioning methodology for distributed traffic management applications," *Transportmetrica A: Transport Science*, vol. 10, no. 6, pp. 518–532, 2014.

[28] T. Anwar, C. Liu, H. L. Vu, and C. Leckie, "Spatial partitioning of large urban road networks." in *EDBT*, 2014, pp. 343–354.

[29] M. Saeedmanesh and N. Geroliminis, "Clustering of heterogeneous networks with directional flows based on snake similarities," *Transportation Research Part B: Methodological*, vol. 91, pp. 250–269, 2016.

[30] A. F. Lentzakis, R. Su, and C. Wen, "Strategic learning approach to region-based dynamic route guidance," in *2016 12th IEEE International Conference on Control and Automation (ICCA)*. IEEE, 2016, pp. 842–847.

[31] K. A. Small and E. T. Verhoef, *The economics of urban transportation*. Routledge, 2007.

[32] M. van Essen, T. Thomas, E. van Berkum, and C. Chorus, "From user equilibrium to system optimum: a literature review on the role of travel information, bounded rationality and non-selfish behaviour at the network and individual levels," *Transport reviews*, vol. 36, no. 4, pp. 527–548, 2016.

[33] M. J. Lighthill and G. B. Whitham, "On kinematic waves. ii. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[34] R. Smeed, "The road capacity of city centers," *Highway Research Record*, 1967.

[35] J. M. Thomson, *Speeds and Flows of Traffic in Central London...* Printerhall, 1967.

[36] J. Wardrop, "Journey speed and flow in central urban areas," *Traffic Engineering & Control*, vol. 8, no. 8, 1968.

[37] Y. Zahavi, "Traffic performance evaluation of road networks by the α-relationship," *Traffic Engineering and Control*, vol. 14, no. 5-6, 1972.

[38] J. Godfrey, "The mechanism of a road network," *Traffic Engineering & Control*, vol. 8, no. 8, 1969.

[39] R. Herman and I. Prigogine, "A two-fluid approach to town traffic," *Science*, vol. 204, no. 4389, pp. 148–151, 1979.

[40] R. Herman and S. Ardekani, "Characterizing traffic conditions in urban areas," *Transportation Science*, vol. 18, no. 2, pp. 101–140, 1984.

[41] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007.

[42] N. Geroliminis and C. F. Daganzo, "Macroscopic modeling of traffic in cities," in *Transportation Research Board 86th Annual Meeting*, no. 07-0413, 2007.

[43] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, 2011.

[44] A. Mazloumian, N. Geroliminis, and D. Helbing, "The spatial variability of vehicle densities as determinant of urban network capacity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4627–4647, 2010.

[45] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 278–288, 2011.

[46] V. L. Knoop and S. P. Hoogendoorn, "Two-variable macroscopic fundamental diagrams for traffic networks," in *Traffic and Granular Flow'11*. Springer, 2013, pp. 351–360.

[47] J. Laval, "The effect of signal timing and network irregularities in the macroscopic fundamental diagram," in *Proceedings of the summer meeting of the TFTC TRB committee*, 2010, p. 5.

[48] L. Leclercq, N. Geroliminis *et al.*, "Estimating mfds in simple networks with route choice," *Transportation Research Part B: Methodological*, vol. 57, no. C, pp. 468–484, 2013.

[49] V. Knoop, D. De Jong, and S. Hoogendoorn, "Influence of road layout on network fundamental diagram," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2421, pp. 22–30, 2014.

[50] V. V. Gayah, X. S. Gao, and A. S. Nagle, "On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram," *Transportation Research Part B: Methodological*, vol. 70, pp. 255–268, 2014.

[51] L. Leclercq, C. Parzani, V. L. Knoop, J. Amourette, and S. P. Hoogendoorn, "Macroscopic traffic dynamics with heterogeneous route patterns," *Transportation Research Part C: Emerging Technologies*, 2015.

[52] M. Keyvan-Ekbatani, X. S. Gao, V. Gayah, and V. L. Knoop, "Combination of traffic-responsive and gating control in urban networks: Effective interactions," in *Transportation Research Board 95th Annual Meeting*, no. 16-5825, 2016.

[53] M. Simoni, "Congestion pricing schemes controlled by the gmfd: a comprehensive design and appraisal to bridge the engineering and economic perspective," Ph.D. dissertation, TU Delft, Delft University of Technology, 2013.

[54] M. Ramezani, J. Haddad, and N. Geroliminis, "Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control," *Transportation Research Part B: Methodological*, vol. 74, pp. 1–19, 2015.

[55] Y. Ji, J. Luo, and N. Geroliminis, "Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2422, no. 1, pp. 1–11, 2014.

[56] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transportation Research Part B: Methodological*, vol. 55, pp. 265–281, 2013.

[57] J. Haddad, M. Ramezani, and N. Geroliminis, "Model predictive perimeter control for two-region urban cities," in *Transportation Research Board 91st Annual Meeting*, no. 12-1445, 2012.

[58] M. Hajiahmadi, J. Haddad, B. De Schutter, and N. Geroliminis, "Optimal hybrid macroscopic traffic control for urban regions: Perimeter and switching signal plans controllers," in *Control Conference (ECC), 2013 European*. IEEE, 2013, pp. 3500–3505.

[59] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.

[60] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail, "Urban congestion gating control based on reduced operational network fundamental diagrams," *Transportation Research Part C: Emerging Technologies*, vol. 33, pp. 74–87, 2013.

[61] J. Haddad and B. Mirkin, "Adaptive perimeter traffic control of urban road networks based on mfd model with time delays," *International Journal of Robust and Nonlinear Control*, vol. 26, no. 6, pp. 1267–1285, 2016.

[62] A. Kouvelas, M. Saeedmanesh, and N. Geroliminis, "Enhancing model-based feedback perimeter control with data-driven online adaptive optimization," *Transportation Research Part B: Methodological*, vol. 96, pp. 26 – 45, 2017.

[63] M. Keyvan-Ekbatani, M. Yildirimoglu, N. Geroliminis, and M. Papageorgiou, "Multiple concentric gating traffic control in large-scale urban networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2141–2154, 2015.

[64] V. L. Knoop, J. van Lint, and S. P. Hoogendoorn, "Traffic network guidance using area accumulation and spatial variation in density," in *Proceedings of the 1st European Symposium on Quantative Methods in Transportation*, no. EPFL-ARTICLE-188493, 2012.

[65] C. Xiong, X. Chen, X. He, X. Lin, and L. Zhang, "Agent-based en-route diversion: Dynamic behavioral responses and network performance represented by macroscopic fundamental diagrams," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 148–163, 2016.

[66] M. Papageorgiou and A. Messmer, "Dynamic network traffic assignment and route guidance via feedback regulation," *Transportation Research Record*, vol. 1306, pp. 49–58, 1991.

[67] O. Chen and M. Ben-Akiva, "Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1617, pp. 179–188, 1998.

[68] H. Yang, "Evaluating the benefits of a combined route guidance and road pricing system in a traffic network with recurrent congestion," *Transportation*, vol. 26, no. 3, pp. 299–322, 1999.

[69] J. L. Adler, G. Satapathy, V. Manikonda, B. Bowles, and V. J. Blue, "A multi-agent approach to cooperative traffic management and route guidance," *Transportation Research Part B: Methodological*, vol. 39, no. 4, pp. 297–318, 2005.

[70] J. de Cea Ch, G. G. Valverde *et al.*, "Effect of advanced traveler information systems and road pricing in a network with non-recurrent congestion," *Transportation Research Part A: Policy and Practice*, vol. 43, no. 5, pp. 481–499, 2009.

[71] I. I. Sirmatel and N. Geroliminis, "Model predictive control of large-scale urban networks via perimeter control and route guidance actuation," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 6765–6770.

[72] K. Ampountolas, N. Zheng, and N. Geroliminis, "Perimeter flow control for bi-modal urban road networks," in *Transportation Research Board 93rd Annual Meeting*, no. 14-2846, 2014.

[73] N. Zheng, K. Aboudolas, and N. Geroliminis, "Investigation of a city-scale three-dimensional macroscopic fundamental diagram for bi-modal urban traffic," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*. IEEE, 2013, pp. 1029–1034.

[74] N. Geroliminis, N. Zheng, and K. Ampountolas, "A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks," *Transportation Research Part C: Emerging Technologies*, vol. 42, pp. 168–181, 2014.

[75] N. Chiabaut, "Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram," *Transportation Research Part B: Methodological*, vol. 81, pp. 410–420, 2015.

[76] J. Ortigosa, N. Zheng, M. Menendez, and N. Geroliminis, "Analysis of the 3d-vmfds of the urban networks of zurich and san francisco," in *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. IEEE, 2015, pp. 113–118.

[77] ——, "Traffic performance and road space allocation in multimodal urban networks with an mfd representation," Tech. Rep., 2017.

[78] A. Loder, L. Ambühl, M. Menendez, and K. W. Axhausen, "Empirics of multimodal traffic networks-using the 3d macroscopic fundamental diagram," *Arbeitsberichte Verkehrs-und Raumplanung*, vol. 1225, 2016.

[79] M. Yildirimoglu, A. Petit, N. Geroliminis, and Y. Ouyang, "Bus service design under demand diversion and dynamic roadway congestion based on aggregated network models," in *Transportation Research Board 95th Annual Meeting*, no. 16-0533, 2016.

[80] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, 2008.

[81] C. Buisson and C. Ladier, "Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2124, no. 1, pp. 127–136, 2009.

[82] Y. Ji, W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian, "Investigating the shape of the macroscopic fundamental diagram using sim-

ulation data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2161, no. 1, pp. 40–48, 2010.

[83] M. J. Cassidy, K. Jang, and C. F. Daganzo, "Macroscopic fundamental diagrams for freeway networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2260, no. 1, pp. 8–15, 2011.

[84] Y. Ji and N. Geroliminis, "Spatial and temporal analysis of congestion in urban transportation networks," in *Transportation Research Board Annual Meeting, Washington, DC*, 2011.

[85] X. Wu, H. X. Liu, and N. Geroliminis, "An empirical analysis on the arterial fundamental diagram," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 255–266, 2011.

[86] V. V. Gayah and C. F. Daganzo, "Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability," *Transportation Research Part B: Methodological*, vol. 45, no. 4, pp. 643–655, 2011.

[87] Y. Sheffi, "Urban transportation network," *Equilibrium analysis with mathematical programming methods, Prentice Hall*, 1985.

[88] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2013.

[89] M. Yildirimoglu and N. Geroliminis, "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams," *Transportation Research Part B: Methodological*, vol. 70, pp. 186–200, 2014.

[90] G. Hamerly and C. Elkan, "Alternatives to the k-means algorithm that find better clusterings," in *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, 2002, pp. 600–607.

[91] B. Zhang, M. Hsu, and U. Dayal, "K-harmonic means-a data clustering algorithm," *Hewlett-Packard Labs Technical Report HPL-1999-124*, 1999.

[92] A. Y. Ng, M. I. Jordan, Y. Weiss *et al.*, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.

[93] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[94] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[95] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.

[96] N. Geroliminis and B. Boyacı, "The effect of variability of urban systems characteristics in the network capacity," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1607–1623, 2012.

[97] K. Ampountolas and A. Kouvelas, "Real-time estimation of critical values of the macroscopic fundamental diagram for maximum network throughput," in *Transportation Research Board 94th Annual Meeting*, no. 15-1779, 2015.

[98] V. L. Knoop and S. P. Hoogendoorn, "Network transmission model: a dynamic traffic model at network level," in *Transportation Research Board 93rd Annual Meeting*, no. 14-1104, 2014.

[99] J. Drake, J. Schoffer, and A. May, "A statistical analysis of speed density hypothesis," *Highway Research Record*, vol. 154, no. 53-87, 1967.

[100] V. L. Knoop and S. P. Hoogendoorn, "Simulation model for traffic using network fundamental diagrams," in *Traffic and Granular Flow'13*. Springer, 2015, pp. 379–384.

[101] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[102] W. Jin and H. M. Zhang, "On the distribution schemes for determining flows through a merge," *Transportation Research Part B: Methodological*, vol. 37, no. 6, pp. 521–540, 2003.

[103] E. J. Gonzales, "Coordinated pricing for cars and transit in cities with hypercongestion," *Economics of Transportation*, vol. 4, no. 1, pp. 64–81, 2015.

[104] J. Haddad and A. Shraiber, "Robust perimeter control design for an urban region," *Transportation Research Part B: Methodological*, vol. 68, pp. 315–332, 2014.

[105] V. Knoop, G. Tamminga, and L. Leclercq, "Network transmission model: Application to a real world city," in *Transportation Research Board 95th Annual Meeting*, no. 16-1526, 2016.

[106] V. V. Dixit, E. A. Radwan, and T. R. Board, "Strategies to improve dissipation into destination networks using macroscopic network flow models," *the Fundamental Diagram for Traffic Flow Theory*, p. 212, 2007.

[107] C. M. Tampere, R. Corthout, D. Cattrysse, and L. H. Immers, "A generic class of first order node models for dynamic macroscopic simulation of traffic

flows," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 289 – 309, 2011.

[108] D. E. Kaufman, J. Nonis, and R. L. Smith, "A mixed integer linear programming model for dynamic route guidance," *Transportation Research Part B: Methodological*, vol. 32, no. 6, pp. 431–440, 1998.

[109] Y. Liu, X. Guo, and H. Yang, "Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks," *NETNOMICS: Economic Research and Electronic Networking*, vol. 10, no. 1, pp. 123–140, 2009.

[110] D. Fudenberg, *The theory of learning in games*. MIT press, 1998, vol. 2.

[111] L. S. Shapley *et al.*, "Some topics in two-person games," *Advances in game theory*, vol. 52, pp. 1–29, 1964.

[112] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.

[113] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning*. MIT Press, 1998.

[114] J. G. Wardrop, "Road paper. some theoretical aspects of road traffic research." in *ICE Proceedings: Engineering Divisions*, vol. 1, no. 3. Thomas Telford, 1952, pp. 325–362.

[115] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *STACS 99*. Springer, 1999, pp. 404–413.

[116] H. Yang, "System optimum, stochastic user equilibrium, and optimal link tolls," *Transportation science*, vol. 33, no. 4, pp. 354–360, 1999.

[117] H. Yang, Q. Meng, and M. G. Bell, "Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a

stochastic user equilibrium," *Transportation Science*, vol. 35, no. 2, pp. 107–123, 2001.

[118] Q. Meng, D.-H. Lee, R. L. Cheu, and H. Yang, "Logit-based stochastic user equilibrium problem for entry-exit toll schemes," *Journal of transportation engineering*, vol. 130, no. 6, pp. 805–813, 2004.

[119] R. D. Connors, A. Sumalee, and D. P. Watling, "Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes," *Transportation Research Part B: Methodological*, vol. 41, no. 6, pp. 593–615, 2007.

[120] K. Uchida, A. Sumalee, D. Watling, and R. Connors, "A study on network design problems for multi-modal networks by probit-based stochastic user equilibrium," *Networks and Spatial Economics*, vol. 7, no. 3, pp. 213–240, 2007.

[121] Q. Meng and H. L. Khoo, "A computational model for the probit-based dynamic stochastic user optimal traffic assignment problem," *Journal of Advanced Transportation*, vol. 46, no. 1, pp. 80–94, 2012.

[122] A. H. van der Weijde, V. A. van den Berg, and E. T. Verhoef, "Stochastic user equilibrium traffic assignment with price-sensitive demand: Do methods matter (much)?" Tinbergen Institute Discussion Paper, Tech. Rep., 2013.

[123] J. Y. Yen, "Finding the k shortest loopless paths in a network," *management Science*, vol. 17, no. 11, pp. 712–716, 1971.

[124] B. Kerner, H. Rehborn, M. Aleksić, and A. Haug, "Traffic prediction systems in vehicles," in *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*. IEEE, 2005, pp. 72–77.

[125] M. Papageorgiou, "Dynamic modeling, assignment, and route guidance in traffic networks," *Transportation Research Part B: Methodological*, vol. 24, no. 6, pp. 471–495, 1990.

[126] S. Hart and A. Mas-Colell, "A reinforcement procedure leading to correlated equilibrium," *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, vol. 4, p. 77, 2013.

[127] A. F. Lentzakis, S. I. Ware, and R. Su, "Region-based dynamic forecast routing for autonomous vehicles," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on.* IEEE, 2016, pp. 1464–1469.

[128] D. R. Ford and D. R. Fulkerson, *Flows in Networks.* Princeton, NJ, USA: Princeton University Press, 2010.

[129] U. Lauther, "An extremely fast, exact algorithm for finding shortest paths in static networks with geographical background," *Geoinformation und Mobilität-von der Forschung zur praktischen Anwendung*, vol. 22, pp. 219–230, 2004.

[130] A. Zhaki, "World's first driverless taxi trial kicks off in singapore," 2016, online; accessed 10-September-2016. [Online]. Available: http://www.straitstimes.com/singapore/transport/worlds-first-driverless-taxi-trial-kicks-off-in-singapore

[131] M. Saeedmanesh and N. Geroliminis, "Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks," *Transportation Research Procedia*, vol. 23, pp. 962–979, 2017.

[132] V. L. Knoop and W. Daamen, "Automatic fitting procedure for the fundamental diagram," *Transportmetrica B: Transport Dynamics*, vol. 5, no. 2, pp. 133–148, 2017.

[133] V. L. Knoop and S. P. Hoogendoorn, "An area-aggregated dynamic traffic simulation model," *European Journal of Transport and Infrastructure Research (EJTIR), 15 (2), 2015*, 2015.

[134] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, "A rolling horizon approach for route guidance in large-scale networks," in *Transportation Research Board 94th Annual Meeting*, no. 15-2073, 2015.

[135] M. Yildirimoglu and N. Geroliminis, "Observing and reconstructing aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks," in *Transportation Research Board 95th Annual Meeting*, no. 16-5186, 2016.

[136] R. Lamotte, A. de Palma, and N. Geroliminis, "On the use of reservation-based autonomous vehicles for demand management," *Transportation Research Part B: Methodological*, vol. 99, pp. 205–227, 2017.

[137] R. Lamotte, A. De Palma, and N. Geroliminis, "Sharing the road: the economics of autonomous vehicles," 2016.

[138] N. Geroliminis and J. Sun, "Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 9, pp. 966–979, 2011.

[139] H. S. Mahmassani, M. Saberi *et al.*, "Urban network gridlock: Theory, characteristics, and dynamics," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 79–98, 2013.

[140] X. Qian, J. Gregoire, F. Moutarde, and A. De La Fortelle, "Priority-based coordination of autonomous and legacy vehicles at intersection," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 1166–1171.

[141] A. J. Hawkins, "Voices clash at first public hearing on self-driving cars," 2016, online; accessed 10-April-2016. [Online]. Available: http://www.theverge.com/2016/4/8/11394454/self-driving-cars-nhtsa-dot-hearing-dc-regulations

[142] D. C. Gazis and R. B. Potts, "The oversaturated intersection," Tech. Rep., 1963.

[143] H. Etemadnia and K. Abdelghany, "On the network partitioning of large urban transportation networks."

[144] G. W. Brown, "Iterative solution of games by fictitious play," *Activity analysis of production and allocation*, vol. 13, no. 1, pp. 374–376, 1951.

[145] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *journal of economic theory*, vol. 68, no. 1, pp. 258–265, 1996.

[146] D. de Jong, V. L. Knoop, and S. P. Hoogendoorn, "The effect of signal settings on the macroscopic fundamental diagram and its applicability in traffic signal driven perimeter control strategies."

[147] V. L. Knoop, S. Hoogendoorn, and J. Van Lint, "Impact of traffic dynamics on macroscopic fundamental diagram," in *Transportation Research Board 92nd Annual Meeting*, no. 13-0595, 2013.

[148] J. Drake, J. Schofer, and A. May, "A statistical analysis of speed-density hypotheses," *Traffic Flow and Transportation*, 1965.

[149] D. I. Robertson, "Transyt: a traffic network study tool," 1969.

[150] J. D. Little, "The synchronization of traffic signals by mixed-integer linear programming," *Operations Research*, vol. 14, no. 4, pp. 568–594, 1966.

[151] N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou, "A multi-band approach to arterial traffic signal optimization," *Transportation Research Part B: Methodological*, vol. 25, no. 1, pp. 55–74, 1991.

[152] P. Koonce, L. Rodegerdts, K. Lee, S. Quayle, S. Beaird, C. Braud, J. Bonneson, P. Tarnoff, and T. Urbanik, "Traffic signal timing manual," Tech. Rep., 2008.

[153] R. Vincent and J. Peirce, "'mova': Traffic responsive, self-optimising signal control for isolated intersections," Tech. Rep., 1988.

[154] "MOVA, Microprocessor Optimised Vehicle Actuation, Controlling traffic light signals at isolated junctions."

[155] A. Hamilton, B. Waterson, T. Cherrett, A. Robinson, and I. Snell, "The evolution of urban traffic control: changing policy and technology," *Transportation planning and technology*, vol. 36, no. 1, pp. 24–43, 2013.

[156] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, "Scoot-a traffic responsive method of coordinating signals," Tech. Rep., 1981.

[157] P. Lowrie, "The sydney coordinated adaptive traffic system-principles, methodology, algorithms," in *International Conference on Road Traffic Signalling, 1982, London, United Kingdom*, no. 207, 1982.

[158] F. Donati, V. Mauro, G. Roncolini, and M. Vallauri, "A hierarchical decentralized traffic light control system. the first realisation" progetto torino"," 1984.

[159] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[160] M. Papageorgiou, M. Ben-Akiva, J. Bottom, P. H. Bovy, S. Hoogendoorn, N. B. Hounsell, A. Kotsialos, and M. McDonald, "Its and traffic management," *Handbooks in Operations Research and Management Science*, vol. 14, pp. 715–774, 2007.

[161] D. De Jong, "The effect of network structure and signal settings on the macroscopic fundamental diagram," Ph.D. dissertation, Masters thesis, 2012.

[162] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton university press, 2007.

[163] J. Nash, "Two-person cooperative games," *Strategy: Critical Perspectives on Business and Management*, vol. 4, no. l953, p. 247, 2002.

[164] T. Basar, G. J. Olsder, G. Clsder, T. Basar, T. Baser, and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1995, vol. 200.

[165] N. Dudebout and J. S. Shamma, "Empirical evidence equilibria in stochastic games." in *CDC*, 2012, pp. 5780–5785.

[166] R. J. Aumann, "Correlated equilibrium as an expression of bayesian rationality," *Econometrica: Journal of the Econometric Society*, pp. 1–18, 1987.

[167] Y. Shoham and K. Leyton-Brown, *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[168] C. H. Papadimitriou and T. Roughgarden, "Computing correlated equilibria in multi-player games," *Journal of the ACM (JACM)*, vol. 55, no. 3, p. 14, 2008.

[169] D. P. Foster and R. V. Vohra, "Calibrated learning and correlated equilibrium," *Games and Economic Behavior*, vol. 21, no. 1, pp. 40–55, 1997.

[170] J. F. Nash Jr, "The bargaining problem," *Econometrica: Journal of the Econometric Society*, pp. 155–162, 1950.

[171] L. S. Shapley, "A value for n-person games," DTIC Document, Tech. Rep., 1952.

[172] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.

[173] J. C. Harsanyi, "A simplified bargaining model for the n-person cooperative game," *International Economic Review*, vol. 4, no. 2, pp. 194–220, 1963.

[174] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press, 2007.

[175] F. A. Oliehoek, M. T. Spaan, and N. Vlassis, "Optimal and approximate q-value functions for decentralized pomdps," *Journal of Artificial Intelligence Research*, vol. 32, no. 1, pp. 289–353, 2008.

[176] J. C. Harsanyi, "Games with incomplete information played by bayesian players, i-iii part i. the basic model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967.

[177] M. Kearns, M. L. Littman, and S. Singh, "Graphical models for game theory," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 253–260.

[178] F. A. Oliehoek, M. T. Spaan, J. S. Dibangoye, and C. Amato, "Heuristic search for identical payoff bayesian games," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 1115–1122.

[179] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498–519, 2001.

[180] H.-A. Loeliger, "An introduction to factor graphs," *Signal Processing Magazine, IEEE*, vol. 21, no. 1, pp. 28–41, 2004.

[181] F. A. Oliehoek, S. Whiteson, and M. T. Spaan, "Exploiting agent and type independence in collaborative graphical bayesian games," *arXiv preprint arXiv:1108.0404*, 2011.

[182] U. Bertele and F. Brioschi, *Nonserial dynamic programming*. Elsevier, 1972.

[183] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[184] J. R. Kok, M. T. Spaan, and N. Vlassis, "Multi-robot decision making using coordination graphs," in *Proceedings of the 11th International Conference on Advanced Robotics, ICAR*, vol. 3, 2003, pp. 1124–1129.

[185] N. Vlassis, "Distributed decision making for robot teams," in *Advances in Intelligent and Distributed Computing*. Springer, 2008, pp. 35–40.

[186] K. Gao, Y. Zhang, R. Su, and A. Lentzakis, "Discrete harmony search algorithm for solving urban traffic light scheduling problem," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 6239–6244.

Appendices

Appendix A

Game Theory Fundamentals

A.1 Game Theory Fundamentals

Game theory, in its modern definition, was first explored in depth by John von Neumann in 1944 [162]. It is a multifaceted theoretical framework for decision-making that encompasses many classes of games representing a variety of real-life situations and setups. Two classes of games featured very prominently in literature include cooperative and competitive games. Competitive games can be used to represent partly or fully antagonistic players selecting actions independently in a strategic manner so as to maximize their perceived utility dependent on every player's decision. Players in this class of games lack communication capability and therefore take action in a distributed manner, optimizing their individual objective functions without any coordination dictated by an outside party. This does not mean that coordination does not occur in competitive games, but when it does, it is not due to outside influence. Cooperative games are augmented general-sum games, where two or more players are allowed to communicate and exchange information, including negotiation of binding agreements to reach a desirable outcome. General-sum refers to the fact that players' goals need not be completely misaligned [163].

Learning plays an important role in game theory. Through repeated iterations of the game under investigation, players are able to learn the optimal action selection that maximizes their respective perceived utilities [110].

A.2 Competitive games

The objective for every player i in a competitive game is to decide upon an action $a_i \in \mathcal{A}_i$ that maximizes their utility $u_i(a_i, a_{-i})$. As is evident by the notation for the utility function, player i's choice of action a_i is dependent on the set of actions a_{-i} of the remaining players in $\mathcal{K}\setminus\{i\}$. The actions in set $\mathcal{A}_i$ for player i that return maximum Game classes such as competitive games can have additional features such as time dependency or updated state information exchange which might alter the game outcome. Games without any time or information dependency are called static and dynamic otherwise. Dynamic games are characterized by time dependency, multiple iterations for players' decisions and partial information exchange among players. Static games represent games played in a single iteration, without any consideration for the utility gained from future game instances and can be defined in the following manner [164]:

Definition A.2.1.

A static competitive game can be described by the player set $\mathcal{K}$, the action sets $(\mathcal{A}_i)_{i\in\mathcal{K}}$ and the utility functions $(u_i)_{i\in\mathcal{K}}$, which describe some kind of payoff or cost for each player. utility for fixed actions of the remaining players in $\mathcal{K}\setminus\{i\}$ are called best responses.

We should mention that, in the dynamic class of competitive games, we must include additional information such as previous actions and time in the utility function structure. Besides deterministic action selection, each player may choose actions in

a probabilistic manner over the action sets $(\mathcal{A}_i)_{i\in\mathcal{K}}$. The former type of action selection is also called pure strategy, while the latter is called mixed strategy. Game theoretic description may cause one to draw comparisons to multi-objective optimization, but one should be aware of the marked difference between the two. Each objective belongs to an independent player whose action selection does not include any explicit form of coordination. One of the most important solution concepts for game theory in general and competitive games in particular is that of a Nash equilibrium. The Nash equilibrium characterizes a state in which no player i can improve its utility by changing unilaterally its strategy, given that the strategies of the other players are fixed. For a static game, the Nash equilibrium in pure strategies can be formally defined as follows:

Definition A.2.2.

A joint action that constitutes a best response for every player at the same time is also known as a pure-strategy Nash equilibrium. In the case of a static competitive game it is comprised of a combination of actions $a^\star \in \mathcal{A}$*, where* $\mathcal{A} = \prod_{i\in\mathcal{K}} A_i$*, where* $\forall i \in \mathcal{K}$*, the following inequality holds true:*

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \forall a_i \in \mathcal{A}_i. \tag{A.1}$$

The Nash equilibrium serves as the solution concept for the majority of game classes. Nash equilibria constitute self-enforcing action combinations. That means that unilateral deviation of player $i \in \mathcal{K}$ from the optimal action a_i^* reduces the utility of player i, provided that the remaining players' optimal actions a_{-i}^* remain fixed. Hence, in decentralized control settings, Nash equilibria facilitate coordination. Unfortunately, while there is a lot of merit in the use of Nash equilibrium as a solution concept, there are also demerits. Nash can merely provide guarantees for the existence of mixed strategies equilibria under the condition that the game

contains a finite number of actions. Mixed strategy Nash equilibria can be inefficient, with respect to individual player utility, for particular game setups, e.g. in the "battle of the sexes" game example [165]. There is an inherent inconsistency in the utility function structure selection. While different utility functions may represent the same kind of player preference, the same group of mixed strategies can have different probability values. Nash equilibria fail to describe situations where deviations are multilateral, i.e. more than one player deviates at the same time. Games may be characterized by multiple Nash equilibria, making coordination among players, i.e. selecting the same Nash equilibrium, problematic. In order to make some discrimination between Nash equilibria, we can consider Nash equilibria such that there is no other Nash equilibrium that is better for everyone. This intuition is formalized by the concept of Pareto optimality.

Definition A.2.3.

A joint action a is Pareto optimal if there is no other joint action a' that specifies at least the same utility value for every player and a higher utility value for at least one player, that is, there exists no a' such that

$$\forall a_i \in \mathcal{A}_i, u_i(a'_i, a'_{-i}) \geq u_i(a_i, a_{-i}) \wedge \exists i, u_i(a'_i, a'_{-i}) > u_i(a_i, a_{-i}) \tag{A.2}$$

If there exists a joint action a', such that equation A.2 is true, then joint action a is Pareto dominated by a'.

A Nash equilibrium a^* can be characterized as Pareto optimal if and only if there is no other Nash equilibrium $a^{*'}$ such that $a^{*'}$ is a Nash equilibrium and Pareto dominates a^*.

Aumann [166] proposed an alternative solution concept in the form of correlated equilibria. Correlated equilibria can be considered as a generalization of Nash equilibria, where mixed strategy probability distributions are drawn over *joint actions*.

Nash [172] assumes that all players are independent, thus, individual players' actions are not conditioned on the action distributions of other players. Correlated equilibria exhibit properties that make them a desirable alternative to Nash equilibria. The existence of correlated equilibria is guaranteed for finite games, since Nash equilibria are considered a degenerate form of correlated equilibria. Unlike Pareto optimal Nash equilibria calculation, an NP-hard problem [167], correlated equilibria can be calculated in polynomial time [168]. There also exists a number of learning algorithms that can achieve convergence to correlated equilibria [112,169], quite useful for online game implementation. An intuitive way to look at the correlated equilibrium solution concept is as follows. Assume that an impartial arbitrator has been added to the game setup. Drawing from the joint action distribution, this arbitrator then transmits to each player their individual actions only. The utility values for each player have to be high enough that every player is incentivized to comply with the suggested course of action. More formally, a joint action probability distribution vector σ is called a correlated equilibrium when, $\forall i \in \mathcal{K}, a_i, a_i' \in \mathcal{A}_i$

$$\sum_{a_{-i} \in \mathcal{A}_{-i}} \sigma(a_i, a_{-i}) u_i(a_i, a_{-i}) \geq \sum_{a_{-i} \in \mathcal{A}_{-i}} \sigma(a_i', a_{-i}) u_i(a_i', a_{-i}) \tag{A.3}$$

Correlated equilibria are more efficient than mixed strategy Nash equilibria, with respect to individual player utilities, while still preserving the fairness property of mixed strategy Nash equilibria.

The sum of the utilities of all players, also known as social welfare is another important quantity used in an alternative solution concept, where the primary objective is to incentivize players to take actions so as to maximize the social welfare, rather than their individual utilities.

A.3 Cooperative games

Unlike competitive games, in cooperative game setups, both coordination and communication among players are considered. By introducing the concept of transferable utility, i.e. assuming that players can trade part of their perceived utility with other players, one can impose desirable outcomes such as coordination among players.

Cooperative game theory extends competitive games by adding communication and information transmission among players. Another feature that further augments the game is the possibility to introduce formal agreements among players, representative of many real-life situations, such as legal contracts, marriage etc. It is commonly expected that the players will plan their actions independently. However, actions selected in such a manner may jointly be suboptimal. Thus the problem turns into a problem of coordination. We must ensure that the individual decisions made can be coordinated to achieve joint optimality. It is easy to see that determining the optimal joint action is a problem of Equilibrium selection.

In cooperative games and especially for a subclass of cooperative games called coalitional games, the main objectives can be summarized in determining what part of each players utility, if any, can be transferred to other players so as to ensure participation of players in coalitions, in essence preserving the fairness property for every coalition formed. If a player has nothing to gain from participation to a team, she would rather act independently. The solution concept developed to represent the situation where no player wishes to defect from their respective coalition is called the core. Another solution concept proposed by Shapley [171] is used to balance the transfers of excess utility due to the formation of coalitions, among individual players, so as to maintain fairness.

Another characteristic subclass of cooperative games are bargaining games, where

players have to reach an agreement that benefits all the players involved. It is evident that if players acquire higher utility values, when choosing to cooperate, a surplus of value is created. A well known solution concept implemented in this game class is the Nash Bargaining Solution originally proposed by Nash [170] for 2-player games and extended by Harsanyi [173] for n-player games. The Nash Bargaining Solution is unique and guaranteed to exist under the assumptions that players are allowed to communicate, no portion of the excess value remains undistributed, no change in outcome occurs from removal of dominated strategies or by linear scaling of utility functions and that all players benefit equally. After determination of all non-dominated strategy payoffs among all feasible strategy payoffs players must come to an agreement regarding the selection of the corresponding joint strategy that maximizes the global utility.

To summarize, a cooperative game framework describes situations where players are given the additional option to act in a concerted manner and form teams, through the fair distribution of excess utility that would not have been available to them had they acted independently.

A.4 Learning in Games

When discussing about learning in a game theoretical context, we refer to the algorithms used to teach players, over a number of iterations, to take actions that constitute equilibria for their particular game setup. Learning methods are also used to teach players selection of specific equilibria [174]. For the selection of a specific equilibrium, the generic iterative procedure is as follows:

1. Players select a joint action for a given iteration, depending on their respective strategies.

2. Each player observes the actions of all other players for the current iteration

and estimates their individual prospective utility.

3. Based on these observations, players inform their strategies accordingly and select a new joint action

The manner in which players update their strategy, based on past observations is also called the learning or update rule. The defining features of a learning algorithm are the acquisition of observations, the translation of observations into a cost function, subsequent optimization of the cost function and randomized action selection so as to promote exploration of the solution space and ensure convergence to global optima. One of the standard ways to introduce randomization is by using a mixed action concept, whereby the best response is assigned the highest probability value, while alternative actions receive smaller probability values.