Network-Wide Route Guidance with Consideration of Fairness: A Macroscopic Fundamental Diagram Approach

Hosseinzadeh, Fatemeh

http://hdl.handle.net/1880/114679

Downloaded from PRISM Repository, University of Calgary
Network-Wide Route Guidance with Consideration of Fairness: A Macroscopic Fundamental Diagram Approach

by

Fatemeh Hosseinzadeh

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING

CALGARY, ALBERTA

MAY, 2022

© Fatemeh Hosseinzadeh 2022
Abstract

This thesis introduces fair route guidance (RG) control schemes in a model predictive control (MPC) framework. The modelling approach used is based on the macroscopic fundamental diagram (MFD), which relates aggregated traffic variables, such as vehicle accumulation and trip completion rate. Earlier MFD-based RG schemes focus on improving network efficiency while overlooking fairness and equity issues. As a result, the controllers force some drivers to take longer paths for their trip to minimize total network travel time; thus, creating inequity issues. This problem motivated this thesis to develop a new control strategy that simultaneously addresses fairness and efficiency in RG control models. This is done by introducing various fairness-centered concepts, such as proportional fairness and anticipatory RG control in an MPC framework for online control.

The proportional fairness (PF) concept, which is rooted in economics and was successfully applied in wireless networks can address this issue by balancing the trade-off between network efficiency and fairness. This paper presents a two-level RG framework using MFD for a heterogeneous urban network divided into multiple pockets of congestion. The developed framework comprises an MPC-based RG optimization and an estimated route-choice model. Firstly, the optimized RG ratios are obtained from the optimization model with different objective functions including: proportional fairness of regional speed, path speed, and path travel time. These objective functions are examined and the results are compared. Then, to update the network traffic states, the drivers’ actual route choice is estimated based on the linear combination,
including the driver routing responses through a logit route-choice model and the optimized routing ratios, which is determined by using the given compliance rate.

Also, this thesis presents an MFD-based anticipative RG control approach by modelling a two-level optimization model in an online optimization framework and directly incorporates road users’ routing behaviour in the control model. The anticipatory RG controller is examined by replacing the basic objective function with the proportional fairness objective function. Based on the anticipatory control (AC) concept, incorporating user behaviour proactively as part of the control framework leads to a more optimum solution and more consistent routing schemes.

Intensive sensitivity analysis is conducted under high and low-demand profiles and for different compliance rates and MFD parameters. Compared with the basic control model, the results show that the fairness control models were more successful in reducing the variances of region accumulations and speeds. The results indicate that the proportionally fair RG model based on path time improves fairness in an urban network by increasing homogeneity while also maintaining a high level of efficiency. Having more homogenised traffic by FC models was consistent for all examined routing compliance levels, even when the compliance rates dropped to as low as 30%. However, because of integrating drivers routing decision directly with AC control models, the total travel time (TTT) efficiency of AC models was more than FC models in cases where the compliance rate was less than 70%. In addition, when examining the performance of the routing guidance for the scenarios with heterogeneous MFD parameters, the proportionally fair RG models exhibited a more homogenous traffic network by reducing the variance of speed, compared to other routing models.
Acknowledgments

First, thanks to God for his blessings and the love He has shown me throughout my life.

I would like to express my sincerest gratitude towards my supervisor, Prof. Lina Kattan, for her supervision, encouragement, guidance, and support during my M.Sc. program. Her ideas, feedback, and vision helped me shape my research career. I am sure without her wisdom, this thesis would not have been possible. She is very supportive in both my academic and my personal life. She is not only my supervisor but also a great friend. I feel extremely lucky to know her and work with her. I am very thankful for the support I received from my friends and colleagues in the transportation engineering group, especially from Nadia Moshahedi for her supervision and invaluable support during my M.Sc. studies.

My heartfelt thanks go to my wonderful family: my parents, Jamileh and Hossein, my sister Arezoo, and my brother Milad. I would like to thank them for being the best family they could ever be, for being an inspiration to how I understand caregiving, and for always having high hopes for me. Without their love, support, and prayers, I would not have been able to complete my M.Sc. Most importantly, I would like to express my gratitude to my love, my husband, Amir, for his understanding, patience, and support to follow my ambitions and goals.

This work is financially supported by NSERC Discovery Grant RT735236: RGPIN/03923-2014, NSERC Discovery Accelerator Grant (DAS): RGPAS/00099-2020, Alberta Innovate Strategic Grant on Integrated Urban Mobility 10024748: G2018000894, NSERC Collaborative Research and Training Experience (CREATE): CREATE/511060-2018, and Alberta Graduate Excellence Scholarship.

iv
Table of Contents

Abstract .......................................................................................................................... ii

Acknowledgments .......................................................................................................... iv

Table of Contents ............................................................................................................ v

List of Figures .................................................................................................................. viii

List of Tables .................................................................................................................. xii

List of Abbreviations ...................................................................................................... xiv

List of Symbols ............................................................................................................... xv

1. Introduction .................................................................................................................. 1

   1.1. Thesis Objectives and Contributions ................................................................ 3

   1.2. Thesis Roadmap .................................................................................................. 4

   1.3. Thesis Organization ............................................................................................ 5

2. Literature Review ......................................................................................................... 6

   2.1. Macroscopic Fundamental Diagram (MFD) ...................................................... 6

   2.2. Route Guidance (RG) ....................................................................................... 10
2.2.1. Overview of RG .................................................................................................................. 10

2.2.2. RG schemes based on MFD ............................................................................................... 11

2.3. Fairness issues in RG .............................................................................................................. 12

2.4. Anticipatory Control (AC) in RG ......................................................................................... 14

2.5. Proportional Fairness in Wireless Networks Routing ........................................................... 15

3. Methodology ............................................................................................................................. 18

3.1. An MFD-Based Modelling of a Multi-Region Urban Network ............................................. 18

3.1.1. Formulation of the first-level MFD-based RG control problem ...................................... 18

3.1.2. Formulation of the second-level MFD-based route choice problem and Plant model 22

3.2. Optimal Control of Urban Networks via Regional RG ....................................................... 24

3.2.1. Benchmark MPC controller with the objective of minimizing total travel time as a basic control ......................................................................................................................... 24

3.2.2. Proportionally fair RG control problem formulation ....................................................... 27

3.2.3. Framework for anticipatory control of urban networks .................................................. 34

3.3. Performance Measures ......................................................................................................... 39

4. Numerical Results and Discussion .......................................................................................... 41

4.1. Analysis of the Results Under Two Different Demand Profiles ........................................... 46

4.1.1. Analysis of the results in terms of regional accumulation for various RG schemes ... 50

4.1.2. Analysis of the results in terms of regional speed for various RG schemes ........... 54
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.3.</td>
<td>Analysis of the results in terms of regional RG ratios and path travel times for various RG schemes</td>
<td>58</td>
</tr>
<tr>
<td>4.2.</td>
<td>Compliance Rate Sensitivity Analysis</td>
<td>66</td>
</tr>
<tr>
<td>4.2.1.</td>
<td>Analysis of the results in terms of regional accumulation for various RG schemes</td>
<td>70</td>
</tr>
<tr>
<td>4.2.2.</td>
<td>Analysis of the results in terms of regional speed for various RG schemes</td>
<td>72</td>
</tr>
<tr>
<td>4.2.3.</td>
<td>Analysis of the results in terms of regional RG ratio and path travel time for various RG schemes</td>
<td>75</td>
</tr>
<tr>
<td>4.3.</td>
<td>MFD Parameters Sensitivity Analysis</td>
<td>84</td>
</tr>
<tr>
<td>4.3.1.</td>
<td>Analysis of the results in terms of regional accumulation for various RG schemes</td>
<td>86</td>
</tr>
<tr>
<td>4.3.2.</td>
<td>Analysis of the results in terms of regional speed for various RG schemes</td>
<td>88</td>
</tr>
<tr>
<td>4.3.3.</td>
<td>Analysis of the results in terms of regional RG ratio and path travel time for various RG schemes</td>
<td>90</td>
</tr>
<tr>
<td>5.</td>
<td>Conclusion</td>
<td>96</td>
</tr>
<tr>
<td>5.1.</td>
<td>Research Contributions and Findings</td>
<td>96</td>
</tr>
<tr>
<td>5.1.1.</td>
<td>Introducing proportionally fair RG controller for a large-size urban network</td>
<td>96</td>
</tr>
<tr>
<td>5.1.2.</td>
<td>Modelling an anticipatory RG controller for a large-size urban network</td>
<td>98</td>
</tr>
<tr>
<td>5.2.</td>
<td>Recommendations for Future Work</td>
<td>98</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Appendix A</td>
<td></td>
<td>110</td>
</tr>
</tbody>
</table>
# List of Figures

Chapter 1

Figure 1. 1. Thesis Roadmap ........................................................................................................................................ 5

Chapter 2

Figure 2. 1. Aggregated plot of average flow vs. average density derived from two sensors over two days (Geroliminis and Daganzo 2008) ........................................................................................................ 7

Figure 2. 2. Relationship between the average network flow and the average network density when the data are distinguished according to the standard deviation S of vehicles in the different road sections. (Mazloumian et al. 2010) ........................................................................................................ 8

Figure 2. 3. A simple illustration of a wireless network consisting of five wireless nodes and six wireless links where the objective of the nodes is to access the Internet services over Node C which acts as the gateway (SHI et al. 2014) ........................................................................................................ 16

Chapter 3

Figure 3. 1. Schematic view of the basic control in model predictive control framework ........ 25

Figure 3. 2. Schematic view of the proposed Proportionally Fair RG Control .................... 31

Figure 3. 3. Schematic view of the proposed Anticipatory Control ...................................... 36

Chapter 4

Figure 4. 1. The urban network which is divided into seven regions ................................. 41
Figure 4. 2. Demand profile ................................................................. 44

Figure 4. 3. Regional accumulations in high-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs ................................................................. 51

Figure 4. 4. Regional accumulations in low-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs ................................................................. 52

Figure 4. 5. Regional speed in high-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs ................................................................. 55

Figure 4. 6. Regional speed in low-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs ................................................................. 56

Figure 4. 7. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 5 to destination 2 with paths {{[5 3 2]}; {[5 4 2]}; {[5 6 7 1 2]}} ........................................................................................................ 59

Figure 4. 8. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 3 to destination 1 with paths {{[3 2 1]}; {[3 4 1]}; {[3 5 6 7 1]}} ........................................................................................................ 60

Figure 4. 9. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 6 to destination 3 with paths {{[6 5 3]}; {[6 4 3]}; {[6 7 1 2 3]}} ........................................................................................................ 61

Figure 4. 10. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 5 to destination 2, paths {{[5 3 2]}; {[5 4 2]}; {[5 6 7 1 2]}} ........................................................................................................ 62

Figure 4. 11. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 3 to destination 1, paths {{[3 2 1]}; {[3 4 1]}; {[3 5 6 7 1]}} ........................................................................................................ 63
Figure 4. 12. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 6 to destination 3, paths \{[[6 5 3]];[[6 4 3]];[[6 7 1 2 3]]\} ................................................................. 64

Figure 4. 13. Regional accumulations under various compliance rate scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt .......... 71

Figure 4. 14. Regional speed under various compliance rate scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt .............. 73

Figure 4. 15. Regional speed under various compliance rate scenario: A) 20%, B) 10%, C) 5% and D) 0%, by various control models: a) NC, b) BC, e) FCpt ................................. 75

Figure 4. 16. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 5 to destination 2 with paths \{[[5 3 2]];[[5 4 2]];[[5 6 7 1 2]]\} ......................................................... 77

Figure 4. 17. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 5 to destination 2 with paths \{[[5 3 2]];[[5 4 2]];[[5 6 7 1 2]]\} ......................................................... 78

Figure 4. 18. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 3 to destination 1 with paths \{[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]\} ......................................................... 79

Figure 4. 19. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 3 to destination 1 with paths \{[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]\} ......................................................... 80

Figure 4. 20. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 6 to destination 3 with paths \{[[6 5 3]];[[6 4 3]];[[6 7 1 2 3]]\} ......................................................... 81
Figure 4. 21. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 6 to destination 3 with paths \{ {{[6 5 3]}; {[6 4 3]}; {[6 7 1 3]}} \} ................................................................. 82

Figure 4. 22. Regional accumulations under various regions’ MFD parameters: A) same regions’ MFD parameter, B) different regions’ MFD parameter, by various control models: a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt ........................................................................................................... 87

Figure 4. 23. Regional speed under various regions’ MFD parameters: A) same regions’ MFD parameter, B) different regions’ MFD parameter, by various control models: a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt ........................................................................................................... 89

Figure 4. 24. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 5 to destination 2 with paths \{ {{[5 3 2]}; {[5 4 2]}; {[5 6 7 2]}} \} .................................................. 91

Figure 4. 25. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 3 to destination 1 with paths \{ {{[3 2 1]}; {[3 4 1]}; {[3 5 6 7 1]}} \} ......................... 92

Figure 4. 26. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 6 to destination 3 with paths \{ {{[6 5 3]}; {[6 4 3]}; {[6 7 1 3]}} \} ......................... 93
List of Tables

Chapter 3
Table 3. 1. Summary of the measures of performance ................................................................. 39

Chapter 4
Table 4. 1. Summary of the examined scenarios and their assumptions ................................. 42
Table 4. 2. Simulation results of control models under high-demand scenario (the percentage improvements in each measure compared to NC is shown in parentheses) ........................................ 47
Table 4. 3. Simulation results of control models under low-demand scenario (the percentage improvements in each measure compared to NC is shown in parentheses) ................................. 47
Table 4. 4. Simulation results of control models under 100% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses) .......... 66
Table 4. 5. Simulation results of control models under 85% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses) .......... 67
Table 4. 6. Simulation results of control models under 70% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses) .......... 67
Table 4. 7. Simulation results of control models under 30% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses) .......... 68
Table 4.8. Simulation results of control models for the network with same region MFD parameters (the percentage improvements in each measure compared to NC is shown in parentheses)........ 84

Table 4.9. Simulation results of control models for the network with different region MFD parameters (the percentage improvements in each measure compared to NC is shown in parentheses)................................................................. 85
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>anticipatory control</td>
</tr>
<tr>
<td>ABC</td>
<td>anticipatory basic control</td>
</tr>
<tr>
<td>AFCrs</td>
<td>anticipatory fair control based on regional speed</td>
</tr>
<tr>
<td>AFCpt</td>
<td>anticipatory fair control based on path time</td>
</tr>
<tr>
<td>BC</td>
<td>basic control</td>
</tr>
<tr>
<td>CBD</td>
<td>central business district</td>
</tr>
<tr>
<td>FC</td>
<td>fair control</td>
</tr>
<tr>
<td>FCrS</td>
<td>fair control based on regional speed</td>
</tr>
<tr>
<td>FCps</td>
<td>fair control based on path speed</td>
</tr>
<tr>
<td>FCpt</td>
<td>fair control based on path time</td>
</tr>
<tr>
<td>ITS</td>
<td>intelligent transportation systems</td>
</tr>
<tr>
<td>MFD</td>
<td>macroscopic fundamental diagram</td>
</tr>
<tr>
<td>MPC</td>
<td>model predictive control</td>
</tr>
<tr>
<td>NC</td>
<td>no control</td>
</tr>
<tr>
<td>NMPC</td>
<td>nonlinear model predictive control</td>
</tr>
<tr>
<td>OCP</td>
<td>optimum control problem</td>
</tr>
<tr>
<td>OD</td>
<td>origin-destination</td>
</tr>
<tr>
<td>PF</td>
<td>proportionally fair</td>
</tr>
<tr>
<td>RG</td>
<td>route guidance</td>
</tr>
<tr>
<td>TMC</td>
<td>traffic management center</td>
</tr>
<tr>
<td>TTS</td>
<td>total time spent</td>
</tr>
<tr>
<td>TT_{path}</td>
<td>paths’ travel time</td>
</tr>
<tr>
<td>TTT</td>
<td>total travel time</td>
</tr>
<tr>
<td>UE</td>
<td>user equilibrium</td>
</tr>
</tbody>
</table>
# List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>region number</td>
</tr>
<tr>
<td>$R$</td>
<td>set of homogeneous regions</td>
</tr>
<tr>
<td>$W$</td>
<td>set of OD region pairs</td>
</tr>
<tr>
<td>$P_w$</td>
<td>set of routes between OD region pairs $w$</td>
</tr>
<tr>
<td>$P$</td>
<td>set of all routes</td>
</tr>
<tr>
<td>$R_p$</td>
<td>set of regions which belong to the path $p$</td>
</tr>
<tr>
<td>$G$</td>
<td>MFD function (veh/min)</td>
</tr>
<tr>
<td>$G_i$</td>
<td>MFD function for region $I$ (veh/min)</td>
</tr>
<tr>
<td>$N_I$</td>
<td>cumulative accumulation for region $I$ (veh)</td>
</tr>
<tr>
<td>$N_I^{\text{remaining}}$</td>
<td>remaining capacity of region $I$ (veh)</td>
</tr>
<tr>
<td>$N_I$</td>
<td>set of regions neighboring region $I$</td>
</tr>
<tr>
<td>$N_{II}$</td>
<td>accumulation of region $I$ with destination in the same region (internal trip) (veh)</td>
</tr>
<tr>
<td>$N_{IJ}$</td>
<td>accumulation of region $I$ with the final destination in region $J$ (external trip) (veh)</td>
</tr>
<tr>
<td>$L_I$</td>
<td>average trip length inside region $I$(Km)</td>
</tr>
<tr>
<td>$v_{\text{free},I}$</td>
<td>free-flow speed at region $I$. (Km/hr)</td>
</tr>
<tr>
<td>$\tau_I$</td>
<td>regional travel time for region $I$. (min)</td>
</tr>
<tr>
<td>$\tau_{ij}^p$</td>
<td>path total travel time for path $p$ that begins from region $I$ and end at region $J$. (min)</td>
</tr>
<tr>
<td>$N_I^{\text{crit}}$</td>
<td>critical accumulation at region $I$. (veh)</td>
</tr>
<tr>
<td>$Q_{II}$</td>
<td>internal demand at region $I$ (veh/min)</td>
</tr>
<tr>
<td>$Q_{IJ}$</td>
<td>exogenous inflow demand generated in region $I$ with destination $J$ (veh/min)</td>
</tr>
<tr>
<td>$\theta_{IJ}$</td>
<td>regional route guidance ratio (origin region $I$, destination region $J$, neighbor region $H$)</td>
</tr>
<tr>
<td>$\hat{\theta}_{IJ}$</td>
<td>optimized route guidance ratio</td>
</tr>
<tr>
<td>$\theta'_{IJ}$</td>
<td>driver route choice</td>
</tr>
<tr>
<td>$\theta''_{IJ}$</td>
<td>estimated real route choice</td>
</tr>
</tbody>
</table>
\( \theta''_{IJ} \) consistent route split ratios
\( \theta^0 \) initial route guidance
\( \theta_{\text{min}} \) minimum route guidance
\( \theta_{\text{max}} \) maximum route guidance
\( M_{II} \) flow from region I to destination I, internal trip completion (veh/min)
\( M_{I|H|J} \) transfer flow from region I to destination J through the next immediate region H (veh/min)
\( \hat{M}_{I|H|J} \) capacity-restricted transfer flow (origin region I, destination region J, neighbor region H) (veh/min)
\( C_{|I|H} (N_{H}) \) boundary capacity between regions I and H (veh/min)
\( C_{|I|H}^{\text{max}} \) maximum boundary capacity between regions I and H (veh/min)
\( N_{H}^{\text{jam}} \) jam accumulation of the receiving region H (veh)
\( N(k) \) vector containing all \( N_{H}(k) \)
\( Q(k) \) vector containing all \( Q_{IJ}(k) \) (veh/min)
\( \theta(k) \) vector containing all \( \theta_{IJ}(k) \)
\( \hat{\theta}(t) \) vector containing all \( \hat{\theta}_{IJ}(t) \)
\( \theta'(t) \) vector containing all \( \theta'_{IJ}(t) \)
\( \hat{N}(t_c) \) Measurement states taken at \( t_c \)
\( k \) control interval counter
\( t \) time (min)
\( t_c \) current control time step (min)
\( T_c \) control sampling time (min)
\( T \) total simulation time (min)
\( m \) iteration number
\( n_p \) prediction horizons
\( n_c \) control horizons
\( f \) time discretized version of dynamics equations
\( \gamma \) compliance rate
\( w \) weight parameter
$V_r$ average speed of region $r$ (Km/hr)

$V_p$ average speed of path $p$ (Km/hr)

$T_p$ travel time of path $p$ (min)

$\alpha_I, \beta_I$ MFD scaling parameters

$\beta$ gridlock parameter

$\alpha$ boundary capacity parameter

$U_i(x_i)$ utility function of $x_i$

$x_i^*$ proportional fair allocation of resource $i$
1. Introduction

Transportation infrastructure is the backbone of urban societies that facilitate the movement of road users (i.e., vehicles, bikers, pedestrians) from one part of the urban network to another. In recent years, the ever-increasing trend of urbanization and vehicle ownership have led to congestion in urban networks. In this sense, proper management of road networks is crucial to maintain urban mobility and sustainability. Due to fiscal, land and environmental constraints, building more roads is often not a viable solution. Improving the operational efficiency of urban transportation networks by using the existing unused capacity is a sustainable solution. Intelligent Transportation Systems (ITS) rely on real-time control, management, and information dissemination to reduce network congestion.

The local traffic controllers fall short of coordinating with the rest of the urban road controllers, so developing a global controller that addresses this issue at a network-wide level is needed. The initial step toward improving urban mobility is modelling of urban networks, which can be done at a microscopic, mesoscopic or macroscopic level. Due to the existence of short links, traffic signals, and unpredictable driver responses in a large-size urban network, microscopic traffic modelling suffers from computational burden and is deemed impractical for managing large-size networks in real time. In the last decade, the macroscopic fundamental diagram (MFD) was developed as an aggregated and efficient traffic model that facilitates network-level modelling and control of large-size urban networks (Daganzo 2007; Daganzo and Geroliminis 2008; Geroliminis and Daganzo 2008).

While MFD provides an elegant and reliable tool for describing homogeneously congested areas of urban networks, this approach should not be used universally since it is not fit for some types of urban networks. Geroliminis and Daganzo (2008) proved that only networks with a
regionally homogenous distribution of density may have a well-defined, low-scatter MFD. So, a heterogeneous network can be further partitioned into multiple smaller homogeneous sub-regions, each with a well-defined MFD, to make modelling and control task of large-size urban networks more manageable (Buisson and Ladier 2009; Mahmassani and Saberi 2013; Mazloumian et al. 2010). Due to its elegance and simplicity, MFD is now widely adopted as the foundation for various control methods such as route guidance, perimeter control, and congestion pricing.

Route Guidance (RG) is an advisory traffic control that recommends a set of routes to be traversed by road users to travel from their origin to their destination zone. Jangshan (2012) believed that routing responds to the query, "where should I go?", while determining a set of acceptable routes. On the other hand, RG responds to the query, "what should I do next?", by routing and guiding a vehicle through monitoring and controlling its movements along a predefined path. MFD-based RG models were developed as advisory control systems and were proven to be effective in managing congestion at homogeneously congested pockets of urban networks (Sirmatel and Geroliminis 2018).

So far, the majority of existing MFD-based RG control models are designed to improve network efficiency by minimizing total network travel time or maximizing the network’s outflow (Sirmatel and Geroliminis 2018, Wei and Yang 2019). In doing so, often the congestion and delays are not fairly distributed among road users in various subregions of the network. In addition, to meet the objective of minimizing total travel time, controllers advise some drivers to take longer paths for their trip to minimize total network travel time; thus, creating inequity issues. This gap in literature motivated this thesis to develop a new control strategy that simultaneously addresses fairness and efficiency in RG control models. The research is done through introducing various fairness–aware concepts such as proportional fairness and anticipatory RG control in an MPC framework for online control.

The concept of proportional fairness, which creates a balanced trade-off between efficiency and fairness, was initially introduced for allocating resources in wireless networks (Kelly 1997). It was later adapted for transportation application, in the context of transit emergency evacuation (Aalami and Kattan 2018, 2021) and transport market flow allocation (Aalami and Kattan 2022).
A proportional fair allocation is shown to be equivalent to maximizing the summation of the proportions of the users’ utilities. In other words, if we remove a piece of resource from one user and allocate it to another user, and this “move” reduces the utility of the first user by p%, but adds to the utility of the other user by p% or more; we will do this move because it increases the total proportion of users’ utilities. The proportional fairness concept can thus address the above described RG equity issues by balancing the trade-off between network efficiency and fairness.

Additionally, previous control models only considered a preset rate of behavioural compliance response of road users when updating the networks’ traffic state. Not all drivers comply with such routing advice; therefore, endogenously incorporating driver responses into the control optimization framework is important to achieve an optimal and consistent routing solution. Because the AC model is more realistic and integrates driver responses directly in the control model without requiring predefined assumptions on routing, it creates a more fair distribution of traffic compared to basic models with a realistic compliance rate of RG instructions. The anticipatory control concept aims to provide a consistent RG instruction by anticipating road user reactions to the optimized RG and thus resulting in more consistent RG schemes (Ben-Akiva et al. 2001; Bottom 2000).

### 1.1. Thesis Objectives and Contributions

This research aims to advise drivers to select efficient and fair roads to reach their destination. More specifically, the objective of this research is to study fair RG control schemes in an MPC framework with a focus on proportional fairness and anticipatory control concepts.

The emphasis of this research is to address network “fairness” issues as defined in terms of user utilities while balancing the trade-off between network efficiency and fairness. The proportionally fair RG offers an equitable allocation of shared resources without sacrificing efficiency by considering the improvement in the proportion of user utilities.

Next, the anticipatory RG control is modelled as a Stackelberg game two-level optimization model with two players: a leader and a follower (Brückner and Scheffer 2011). At the first level, the traffic controller (i.e., the leader) optimizes the RG policies while endogenously incorporating
the anticipated road users’ (i.e., the follower) response to changes in RG control as obtained from the second level. In addition, the anticipatory control (AC) of RG is developed by examining different objective functions including total travel time minimization and several variants of proportional fairness objective functions.

Generally, in this thesis, we address the research question, “how can we fairly recommend a set of routes to be traversed by road users without much compromise to the efficiency of the network?”

This thesis makes the following contributions:

- Developing a novel two-level proportional fair RG control scheme for a large-size network-based on MFD.
- Developing several routing utility measures (path time and path speed) to accommodate the fair-aware RG controller.
- Developing a two-level network-wide anticipatory RG control scheme in the context of MFD by integrating the drivers routing decisions into the RG control framework for more consistent, and thus fair RG schemes.
- Comprehensively evaluating the performance of the developed routing algorithms through new sets of measures that address the different perspectives (i.e., users and non-users).

1.2. Thesis Roadmap

Figure 1. 1 illustrates the thesis roadmap. This dissertation has been organized into five chapters.
1.3. Thesis Organization

This thesis is organized as follows: In Chapter 2, an overview is presented of the literature pertaining to the MFD-based RG problem, proportional fairness, and anticipatory control concepts. Chapter 3 introduces the basic foundations for MFD-based modelling of a multi-regional network. Included are descriptions of different routing control schemes such as basic MPC controller, with the objective of minimizing total travel time; a novel proportionally fair RG control scheme in an MPC framework; and an anticipatory control scheme with traditional and novel objective functions. In Chapter 4, the results of various control models are presented and discussed for three sensitivity analysis cases pertaining to: (a) the demand, (b) compliance rate to RG instructions and (c) MFD parameters variation. Finally, Chapter 5 concludes this thesis and provides some directions for future works.
2. Literature Review

This chapter presents a review of the literature pertaining to the MFD-based RG problem. The MFD concept is first briefly presented in section 2.1 and then extended to the MPC-based control in the context of MFD. A review of route-guidance schemes in general and in the MFD context is provided in section 2.2. Particular attention is then paid in section 2.3 to fairness issues pertaining to RG, which is the focus of this dissertation. Anticipatory RG control, which is considered one of the first attempts to address equity issues in RG, is then reviewed in section 2.4 for general traffic networks, and will be applied in an MFD context. Finally, section 2.5 reviews routing strategies in the context of a wireless network based on the proportional fairness concept.

2.1. Macroscopic Fundamental Diagram (MFD)

The unique structure of an urban network with short links, traffic signals, and the unpredictable nature of road users' response, makes it impractical and inefficient to model a large-size urban network based on a microscopic traffic model. The macroscopic fundamental diagram (MFD) was developed as an aggregated traffic model that illustrates the traffic states of a network based on a single state variable. MFD provides a reliable and elegant tool for parsimonious and efficient network-level modelling and for control of large-size networks (Daganzo 2007; Sirmatel and Geroliminis 2018).

The general concept of MFD was initially proposed by Godfrey (1969) and its theoretical concept was later introduced by Daganzo (2007). In this work, the dynamics of a single region with a homogeneous distribution of vehicles were modelled by mapping accumulation (the total number of vehicles in the system) to outflow (trip completion rate). In homogeneous urban networks, MFD provides demand insensitive, unimodal, and low-scatter relationships between
accumulation and network outflow. Due to this simple and elegant relationship between aggregated traffic variables, MFD is widely adopted for large-size traffic management and control. Daganzo and Geroliminis (2008) used a variational theory approach based on Daganzo (2005a; b) to estimate MFD analytically for a single arterial consisting of sequential links separated by traffic signals. This analytical-based MFD presentation was later extended by Geroliminis and Boyac (2012), and later by Leclercq and Geroliminis (2013).

Geroliminis and Daganzo (2008) confirmed the empirical existence of MFD using data from fixed detectors and GPS-equipped taxis in the Yokohama (Japan) traffic network. Figure 2.1 illustrates an aggregated plot of average flow vs. average density derived from two sensors over two days (labels A1-D2 denote different time periods in two days) (Geroliminis and Daganzo 2008). Figure 2.1 demonstrates that although origin-destination (OD) demand values differ significantly during the two days, MFD is independent of O-D (i.e., demand-insensitive). The resulting plot indicates that a well-defined low-scatter relationship between average flow and average density exists for Yokohama traffic network.

Figure 2.1. Aggregated plot of average flow vs. average density derived from two sensors over two days (Geroliminis and Daganzo 2008)
These findings supported the assumption that the average trip length within a region is relatively constant throughout a day. While the assumption of constant trip length is a conventional assumption in the accumulation-based MFD method, more recent MFD literature partitions the network into sub-regional MFD to incorporate their unique characteristics and dynamics for more rigorous modelling effort (Ramezani et al. 2015; Yildirimoglu et al. 2015, 2018). Geroliminis and Daganzo (2008) attribute the high scatter noticed in the right branch of the MFD networks to the prevailing heterogeneity of flow under very congested conditions. In other words, under severe congestion occurring in the region, the gridlock phenomenon occurs; accordingly preventing flows reaching the downstream links, which exhibit zero or very low flow. Thus, while congested links exhibit high density, other links might be almost empty; thereby explaining the occurrence of high variance/high scatter in the right branch of the MFD. Hence, only networks with a regionally homogenous distribution of density may have a well-defined, low-scatter MFD.

Since heterogeneity undermines the existence of a well-defined MFD, a heterogeneous network can be further partitioned into multiple smaller homogeneous sub-regions with a well-defined MFD to make modelling and control more manageable (Buisson and Ladier 2009; Mahmassani and Saberi 2013; Mazloumian et al. 2010). Figure 2.2 from Mazloumian et al. (2010) illustrates how, for the same number of vehicles in the sub-network, if the system’s standard deviation of link density is larger, then the maximum outflow would be lower and vice-versa; which necessitates network partitioning.

![Figure 2.2](image-url)

**Figure 2.2.** Relationship between the average network flow and the average network density when the data are distinguished according to the standard deviation S of vehicles in the different road sections. (Mazloumian et al. 2010)
Today, MFD is widely used as the foundation for various control methods such as RG, perimeter control, and congestion pricing. A large body of literature has developed MFD-based control schemes for urban single-region networks (Daganzo 2007; Gayah et al. 2014; Haddad 2017a; Haddad and Shraiber 2014; Keyvan-Ekbatani et al. 2012) and multi-region networks (Aboudolas and Geroliminis 2013; Haddad 2017b; Haddad and Geroliminis 2012; Kouvelas et al. 2017; Ramezani et al. 2015; Sirmatel and Geroliminis 2018). For in-depth literature reviews on MFD-based modelling, refer to Saberi and Mahmassani 2012; Yildirimoglu et al. 2015 and more recently Johari et al. 2021.

According to Sirmatel and Geroliminis 2018, the following factors must be considered when designing network-level controllers for urban networks using MFD-based modelling:

- a) constraints on traffic states and control inputs,
- b) nonlinear dynamics of the MFD-based network model, and
- c) the possibility of having access to future information such as demand profile.

These points strongly facilitate model predictive control (MPC), an advanced control methodology based on real-time repeated rolling-horizon optimization. MPC’s most essential benefit over conventional control methods is its capability of handling constraints systematically (Aalipour et al. 2018; Haddad et al. 2013; Hajiahmadi et al. 2014; Ramezani et al. 2015; Sirmatel and Geroliminis 2018; Yildirimoglu et al. 2015). MPC is, thus, a computationally fast approach for resolving infinite horizon in the form of a constrained optimum control problem (OCP). The finite horizon OCP is solved at each sampling time instant using the current states of network as the initial state to yield a series of optimum controls, while only the control corresponding to the first-time step is applied to the model. This process is repeated in a rolling horizon fashion over all upcoming time steps until the simulation is terminated. For discussions on major MPC concerns, refer to Garcia et al. (1989); and for an outline of theoretical elements of MPC, refer to Mayne et al. (2000).

The application of MPC in traffic control is abundant in the context of arterial, freeway, logistics, and MFD control literature. Instances of freeway MPC-based control schemes include the work of Gomes and Horowitz (2006), Papamichail et al. (2010), Hajiahmadi et al. (2016) who
studied ramp metering; and Karimi et al. (2004) who integrated ramp metering with RG. Lin et al. (2011) and Zhou et al. (2015) studied signal control on urban networks, while Van den Berg et al. (2007) studied signal control for mixed urban and freeway networks. In addition, control of railways and logistics systems are other examples of MPC applications (Kersbergen et al. 2016; Li et al. 2016).

MPC approaches for urban networks using MFD-based prediction models have recently appeared in the literature. A nonlinear MPC for a two-region urban network with perimeter control actuation was first introduced by Geroliminis et al. (2013), followed by Haddad et al. (2013) who studied a nonlinear MPC for a mixed transportation network including two urban regions and a freeway. Hajiahmadi et al. (2014) proposed a hybrid MPC for an urban network that includes switching both signal timing plans and perimeter control systems. Ramezani et al. (2015) developed a model that captures the dynamics of heterogeneity, as well as a hierarchical control based on MPC. Hajiahmadi et al. (2013) and Yildirimoglu et al. (2015) developed a route guiding actuation in combination with perimeter control using MPC.

2.2. Route Guidance (RG)

2.2.1. Overview of RG

RG is an advisory traffic control system which recommends a set of routes to be traversed by road users to travel from their origin to their destination points. From a system perspective, RG redistributes the flow in a traffic network to optimize a pre-defined control measure, such as minimizing total system travel time. Route finding schemes have long been a research priority in intelligent transportation systems (Ben-Akiva et al. 1991). Traditional RG systems were designed to find the shortest route based on trip time, distance, travel cost, or a combination of these criteria. Earlier work focused on static situations without considering changing traffic conditions in the road network. These static RG schemes were unsuitable for real-time management of congested traffic where RG is mostly needed. In dynamic RG, travellers are guided based on the fluctuation of traffic conditions. For instance, if a traffic warning indicates an event occurrence such as delays
or congestion, the navigation system examines whether the affected region can be avoided and guides the driver on a newly computed alternative route (Dong et al. 2011).

The concept of route concept is tightly intertwined with the traffic assignment problem which describes how traffic is distributed on various routes and links of a given traffic network. Wardrop introduced the concepts of user equilibrium and system optimum as “Wardrop's first principle” and “Wardrop's second principle” respectively, to formalize different notions of equilibrium given the assumption of perfect travel time information (Wardrop 1952; Wardrop and Whitehead 1952). To relax this assumption of perfect information, stochastic user equilibrium is introduced later as more realistic version of user equilibrium (Daganzo and Sheffi 1977). In addition, the Price of Anarchy (PoA) is introduced as a related concept in economics and game theory to measure how a system's efficiency degrades due to the selfish behaviour of travellers (Koutsoupias and Papadimitriou 2009). PoA measures the deviation of the total system travel time between user equilibrium and system optimum. In other words, the potential fall in efficiency from social to selfish equilibria is an example of PoA.

2.2.2. RG schemes based on MFD

As advisory control systems, MFD-based RG models were developed and proven to be effective in managing congestion in homogeneously congested pockets of urban networks. In some models, if necessary, a second level of control was applied to further control each region locally (Hajiahmadi et al. 2013; Sirmatel and Geroliminis 2018; Yildirimoglu et al. 2015, 2018). Yildirimoglu et al. (2015) examined a network divided into three homogeneously congested regions, where each was further divided into six or seven sub-regions with their own well-defined MFD. They developed a region-based RG control system based on the aggregated model that uses aggregated traffic states, while only partial sub-regional information is available. In more recent work, Yildirimoglu et al. (2018) developed a hierarchical traffic control approach with a centralized upper-level regional RG and localized lower-level sub-regional path assignment. The network was divided to seven regions and each region divided to seven subregions. The model establishes the state in the region-based model and accordingly provides drivers with RG information that meets the same constraints as in the subregion-based model. RG is generated
through a rolling horizon framework. This means that the model sets the state in the region-based model at each time step (e.g. 1 min), then this model defines a sequence of routing decisions over a rolling horizon. However, only the first phase of routing decisions is used in the subregion-based model. This process is repeated with a shifted horizon until the end of simulation time.

To improve mobility in urban networks, Sirmatel and Geroliminis (2018) developed a network-level economic MPC method which integrates perimeter control with regional route guiding. They stated that, “economic MPC involves objective functions that express economically optimal plant operation (e.g., maximising profits or minimising time spent)” (Sirmatel and Geroliminis 2018). At first, a novel MFD-based model with cyclic behaviour avoidance is developed. Then, the challenge of calculating the route guiding and perimeter control inputs for a multi-region urban network with the goal of minimising total time spent (TTS) is framed as an economic MPC problem. They used a logit route choice model to express road user routing in the network, while also taking into account vehicle compliance with their issued RG advice. Thus, they offered a hybrid formulation to calculate real route choice based on the given compliance rate.

Wei and Yang (2019) developed a bi-level RG approach for urban road networks based on MFD, which is a combination of central and distributed RG. Their approach considers the perspectives of both traffic management and drivers to balance between user optimum and system optimum. Thus, a system optimum traffic assignment approach is followed at upper-level to find the optimum RG, while the lower-level estimates the resulting user routing responses that follow a stochastic user equilibrium (Abdulhaq and Qasim 2017) in sub-regions (Wei and Yang 2019).

2.3. Fairness issues in RG

In the literature, MFD-based control models are typically developed with the objective of minimizing total network travel time or maximizing network outflow; however, congestion and delay are often not fairly distributed among users in various regions of the network. Only a few studies tried to solve this issue and considered fair distribution. In order to have more efficient and fair RG control models, various fairness-aware concepts such as proportional fairness and anticipatory control can be combined with RG in an MPC framework.
Hajiahmadi et al. (2013) modelled a regional RG problem as a multi-objective optimization problem. He minimised both the difference between the average speed of regions and total network travel time. They discuss how minimising overall network travel time without paying attention to each individual region would result in long travel times in some regions, thus creating low average regional speed in these areas, while other regions experience short travel times and higher average speeds. Aboudolas and Geroliminis (2013) used a simple multivariable feedback controller to control the flow transfer between neighboring areas after dividing a heterogeneous network into many homogeneous sections. They concluded that the model could consider both mobility and equity factors by trying to decrease heterogeneity within the urban network.

Recently, Essen (2018) developed the novel concept of “social routing advice” also known as “social navigation”. Under the social routing concept, travellers might accept to follow rerouting advice of following slightly longer routes to compensate for the selfish behaviour of other travellers, thereby benefiting the total system travel time. In a more recent work, Essen et al. (2020) conducted an empirical behavioural analysis that examined determinants of traveller compliance with social routing advice. The findings revealed a considerable difference in compliance behaviour across different information frames, difference in additional travel time, social goals, and characteristics of travellers. An intrinsic motivation to contribute to increased throughput is the primary motivation for revealed compliance, while the key motivation for non-compliance relates to perceived traffic conditions.

Then, Eikenbroek et al. (2021) examined the impact of the concept of the “social routing” approach for improving the performance of the traffic system and drawing the system closer to the optimal state while considering equity at the final state. They believed that the system optimum is efficient but unfair, whereas the user equilibrium is inefficient but completely fair. As a result, they demonstrate how to find the best path by solving a bi-level algorithm with drivers having to follow a limited detour in their trip; clearly accounting for traveller utility. Also, they demonstrate fair traffic distribution in the social routing method with a numerical example of the travel time difference between the advised path for drivers who were advised to take the longer path at each method compared to the selfish method. They numerically show that the travel time difference of the selected path in the social routing method with a shorter possible path (based on the selfish
method) is less than the travel time difference of selected path in the system optimum model with the shorter path.

2.4. Anticipatory Control (AC) in RG

Previous models control the traffic state of the network by minimizing disabilities or maximizing utilities. Those models only considered assumed behavioural responses (e.g., compliance rates) when updating models, rather than incorporating these responses as an integral step in optimizing the control. Since some drivers may not comply with such routing advice; thus, incorporating their behaviour proactively, as part of the optimizing step, is important to achieve a more optimal solution and more consistent routing schemes. This proactive incorporation of expected traveler behavioural responses to routing is at the core of the anticipatory RG concept which provides routing recommendations based on the prediction of future traffic demand patterns (Kaufman et al. 1991).

Ben-Akiva et al. (2001) and Bottom (2000) stated that anticipatory control (AC) inherently aims to provide a consistent RG ratio by anticipating how users react to the optimized RG instructions, therefore, resulting in more consistent RG schemes. In other words, the design of anticipatory control policies needs to endogenously incorporate user reactions to these RG, which in turn depend on the issued RG. A circular dependency entails the formulation and solving of a fixed-point problem as suggested by Ben-Akiva et al. (2001) and Bottom (2000). At the fixed-point problem, RG ratios are determined based on the traffic state patterns received from the route choice model, while routing decisions are made based on the route travel time estimated at the RG model. Crittin and Bierlaire (2001) suggest a heuristic strategy based on an estimated objective function to reduce the computing burden of Bottom's framework (2000). Additionally, Huang et al. (2017) stated that anticipatory control with the goal of achieving global optimality, predicts drivers’ reactions to control (e.g. based on a logit route choice model) and considers the predicted responses into the control decision. The work of Han et al. (2015), Rinaldi et al. (2018), Rinaldi and Tampère (2015), Taale (2008), Zhang and Yang (2004), and Zhou et al. (2015a) have proven that integrating driver routing responses directly into control models improves controller efficacy, while also enhancing mobility and alleviating congestion; specifically in congested traffic.
conditions. Also, because the AC model is more realistic and integrates driver responses directly into the control model without needing predefined assumptions on routing, it seems to have a more fair distribution of traffic compared to basic models with a realistic compliance rate of RG instructions.

In this thesis, the anticipatory RG control is modelled as a two-level optimization model to incorporate the driver route choice behaviour within the control model. Additionally, in this thesis, the anticipatory control (AC) of RG is developed by considering proportional fairness objective functions as the objective function of the optimization model, then integrating drivers behavioural responses as an integral part of the control problem. The details are explained in the Methodology Section.

2.5. Proportional Fairness in Wireless Networks Routing

The proportional fairness concept originated in computer science studies of wireless networks, with the seminal works of Kelly (1997) and Kelly et al. (1998). The focus was on the fair distribution of available network resources (i.e., links, bandwidth capacities, etc.) among competing streams, users or devices.

Figure 2. 3 illustrates a simple wireless network from SHI et al. (2014) with five wireless nodes, users, or devices; with six wireless links. The defined network comprised devices A, B, D, and E which communicate with each other via wireless links L1 to L5; while device C is the gateway/router device to access internet services. This is illustrated by the following example which shows computing resources and memory being shared among different applications within an individual node or device (A to E), while needing to cooperate at the system level to achieve successful communication.
Figure 2. 3. A simple illustration of a wireless network consisting of five wireless nodes and six wireless links where the objective of the nodes is to access the Internet services over Node C which acts as the gateway (SHI et al. 2014)

Similar to road traffic networks, a wireless network has limited shared network resources (i.e., link capacities) that are subject to delays and queue formation (i.e., externalities). To achieve fairness in resource utilization, both resources and externalities must be equitably distributed. Proportional fairness is developed as a regulatory control mechanism that can mitigate the issue by preventing the network from becoming congested; thereby, losing its efficiency when it is most needed. Proportional fairness is a utility-based optimization formulation, where Kelly et al. (1998) introduces the concept that users (i.e., devices) have different needs/satisfaction-levels regarding the shared resources.

The proportional fairness is rooted in economics and is a generalization of the Nash bargaining game (Nash and John 1950). The main concept of proportional fairness suggests that an equitable allocation of shared resources can be achieved by raising the proportion of user utilities without sacrificing efficiency. Based on this concept, if we remove a piece of resource from one user and allocate it to another user, and this “move” reduces the utility of the first user by p% but adds to the utility of the other user by p% or more, we will do this move because it increases the total proportion of user utilities.
Recently, Aalami and Kattan adapted this fairness concept for transportation application in the context of transit emergency evacuation (2018, 2021), and more recently for transport market flow allocation in traffic networks (2022). Moshahedi (2021) developed a perimeter control strategy based on the proportional fairness concept to address issues of fairness and efficiency when applying perimeter rates in a multi-region urban network.

Earlier MFD-based RG models focus on improving network efficiency while overlooking fairness and equity issues. As a result, the controllers advise some drivers to take longer paths for their trip to minimize total network travel time; thus, creating inequity issues. The proportional fairness concept can address this issue by balancing the trade-off between network efficiency and fairness. The concept is adopted in this thesis to develop a proportionally fair RG control scheme.
3. Methodology

3.1. An MFD-Based Modelling of a Multi-Region Urban Network

In this thesis, a two-level optimization model based on the concept of the macroscopic fundamental diagram (MFD) for a multi-region urban network is presented. While the first-level optimizes the RG rates, the second-level estimates the resulting route choice by incorporating driver responses to the optimized RG instructions obtained from the first-level model. The optimized RG rates are estimated in the first-level optimization model based on a model predictive control (MPC) scheme. In this MPC framework, various routing schemes, including a proportionally fair and a two-level anticipatory RG controller, are developed and compared to a benchmark routing scheme in the literature that optimizes total system travel time. These different routing schemes developed in this thesis are described in this chapter.

3.1.1. Formulation of the first-level MFD-based RG control problem

During the formulation of an MFD-based model, an urban network is considered \( G = (\mathcal{R}, W) \) with heterogeneous distribution of accumulation partitioned into several smaller homogeneous regions; where \( \mathcal{R} = \{1, ..., r_m\} \) is a set of homogeneous regions; \( W \) is a set of Origin-Destination (OD) region pairs, \( P_w \) is a set of routes \( w \) between any given OD pairs \( w \in W \); and \( P \) is the set of all routes. The dynamic of a given homogeneous region \( I \in \mathcal{R} \) is described by its outflow MFD = \( G_I(N_I) \) (veh/min), expressing the number of trips completed within a region as a function of its accumulation \( N_I \) (veh).

The functional form of MFD function, as suggested by Hajiahmadi et al. (2013), can be approximated by an exponential function as follows:
\[ G_i(N_i(t)) = \frac{v_{\text{free},i}}{L_i} \cdot N_i(t) \cdot \exp\left(-0.5\left(\frac{N_i(t)}{N_i^{\text{crit}}}ight)^2\right) \]

where \( v_{\text{free},i} \) is the free-flow speed at region I, and \( L_i \) is the average trip length inside region I. Accumulation-based MFD models conventionally assume that the average trip lengths inside a region are independent of the origin and destination location, and are, thus, considered constant. This assumption of constant trip length is supported by Geroliminis and Daganzo (2008) who found that the value of production per outflow of a region to be a relatively constant value throughout a given day.

To model the MFD dynamics, accumulation of any region \( N_i \) within region I are disaggregated into internal and external accumulations according to the final trip destinations \( J \in R \).

\[ N_i(t) = N_{ii}(t) + \sum_{J \in R; I \neq J} N_{ij}(t) \]

\( N_{ij} \) denotes the accumulation of region I with the final trip destination in another region J (external trip), while \( N_{ii} \) is the accumulation of region I with the final trip destination within the same region (internal trip).

In this research, each region is assumed to be equipped with regional RG control rates \( \theta_{iHJ}(t) \) (I \( \in R \), J \( \in R \setminus \{I\} \), H \( \in N_i \)), where \( N_i \) is the set of regions neighboring region I, and J is the final destination region. \( \theta_{iHJ}(t) \), thus, distributes the flows exiting region I and destined to region J over initial region’s neighboring regions H. Thus, if \( M_{iHJ}(t) \) is the transfer flow (veh/min) from region I to final destination J traversing through the next immediate region H, it is expressed as follows:

\[ M_{iHJ}(t) = \theta_{iHJ}(t) \frac{N_{ij}(t)}{N_i(t)} G_i(N_i(t)) \]

The dynamical equation states for a multi-region MFDs network are from (Geroliminis and Daganzo 2008; Yildirimoglu et al. 2015, 2018):
\[ \dot{N}_{II}(t) = Q_{II}(t) - M_{II}(t) + \sum_{H \in \mathcal{N}_I} \tilde{M}_{HI}(t) \]  
\[ \dot{N}_{IJ}(t) = Q_{IJ}(t) - \sum_{H \in \mathcal{N}_I} \tilde{M}_{HIJ}(t) + \sum_{H \in \mathcal{N}_I; H \neq J} \tilde{M}_{HIJ}(t) \]

where \( \dot{N}_{II}(t) \) is the rate of accumulation of internal trips inside region I (veh/min); \( \dot{N}_{IJ}(t) \) is the rate of accumulation of external trips from region I to another region J; \( Q_{II}(t) \) is internal demand (veh/min) in region I; and \( Q_{IJ}(t) \) is the exogenous inflow demand (veh/min) generated in region I with final destination J.

\( M_{II}(t) \) is the trip completion flow rate (veh/min) from region I to destination I (internal trip completion). It is expressed as follows:

\[ M_{II}(t) = \frac{N_{II}(t)}{N_{I}(t)} G_I(N_I(t)) \]  

The summation of receiving inflows from the boundary region can also be restricted by existing high accumulations in neighboring regions \( H \); thus, justifying the definition of the capacity-restricted transfer flow \( \tilde{M}_{HIJ}(t) \) (veh/min) from region I to destination J through the immediate neighboring region \( H \) that is restricted by the boundary capacity \( C_{IH}(N_H(t)) \) between regions I and H. Thus, exiting flow can be defined as follows (Ramezani et al. 2015; Sirmatel and Geroliminis 2018; Yildirimoglu et al. 2015):

\[ \tilde{M}_{HIJ}(t) = \min\left(M_{HIJ}(t), C_{IH}(N_H(t)) \frac{M_{HIJ}(t)}{\sum_{K \in \mathcal{R}} M_{IKH}(t)}, (\beta N_H^{jam} - N_H(t))/dt\right) \]

where \( N_H^{jam} \) is jam accumulation (the maximum accumulation at which a network will reach gridlock) of the receiving region \( H \), and \( \beta \) is the gridlock parameter that prevents occurrence of jam accumulation in the next immediate region \( H \); therefore, \( (\beta N_H^{jam} - N_H(t))/dt \) restricts the transfer flow to be less or equal to the remaining capacity of the immediate neighboring region \( H \). \( C_{IH}(N_H(t)) \) (veh/min) is the exiting boundary capacity between regions \( I \) and \( H \) that depends on the remaining accumulation of neighbor region \( N_H \) at time \( t \). Ramezani et al. (2015) considered the boundary capacity as a piecewise function expressed in terms of \( N_H(t) \) taking the form of
either a constant capacity value \( C_{IH}^{\text{max}} \) or a decreasing function as indicated in the formula below (Mariotte et al. 2019; Ramezani et al. 2015; Sirmatel and Geroliminis 2018):

\[
C_{IH}(N_H(t)) = \begin{cases} 
C_{IH}^{\text{max}} & \text{if } 0 \leq N_H(t) < \alpha N_H^{jam} \\
C_{IH}^{\text{max}} \left( 1 - \frac{N_H(t)}{N_H^{jam}} \right) & \text{if } \alpha N_H^{jam} \leq N_H(t) \leq N_H^{jam}
\end{cases}
\]

where \( C_{IH}^{\text{max}} \) (veh/min) denotes the maximum boundary capacity; \( \alpha N_H^{jam} \) (with \( 0 < \alpha < 1 \)) determines the point where boundary capacity starts to decrease from the constant value \( C_{IH}^{\text{max}} \) with increasing accumulation, and drops to zero when accumulation reaches to jam accumulation \( N_H^{jam} \). Haddad et al. (2013) stated that the boundary capacity constraint can be omitted in the MPC-based models for computational advantage and for two other reasons. First, the boundary capacity decreases for accumulations much larger than the critical accumulation; and secondly, the controller should initially prevent the regions from having accumulations reaching gridlock. While Knoop and Hoogendoorn (2014); Sirmatel and Geroliminis (2018); and Hajiahmadi et al. (2014) considered this point equivalent to the MFD critical accumulation, Mariotte et al. (2019) identified this assumption to be too restrictive and claimed that the turning point of receiving capacity is higher than the critical accumulation value.

Also, \( \bar{M}_{HIJ}(t) \) and \( \bar{M}_{HII}(t) \) are defined similarly as follows:

\[
\bar{M}_{HIJ}(t) = \min \left( M_{HIJ}(t), C_{HI}(N_J(t)) \right) \frac{M_{HIJ}(t)}{\sum_{K \in \mathcal{R}} M_{HKI}(t)}, \frac{\beta N_J^{jam} - N_J(t)}{dt} 
\]

\[
\bar{M}_{HII}(t) = \min \left( M_{HII}(t), C_{HI}(N_J(t)) \right) \frac{M_{HII}(t)}{\sum_{K \in \mathcal{R}} M_{HKI}(t)}, \frac{\beta N_J^{jam} - N_J(t)}{dt} 
\]

The first-level RG control problem is constructed in an MPC fashion with optimized RG \( \hat{\theta} \) as described next. In an MPC framework, at each time step, the RG ratios are optimized over a given prediction horizon; but only the results of the first time of prediction horizon are deployed (Garcia et al. 1989; Mayne et al. 2000). Then, the horizon is rolled forward, and RG ratios are optimized again using the newly updated states and the newly injected RG based on the assumed
drivers’ compliance rates and routing responses at the second-level. The whole process is repeated continuously until the end of the simulation time.

The advantage of MPC compared to other control methods such as bang-bang controller (Daganzo 2007) and PI-type controller (Aboudolas and Geroliminis 2013; Keyvan-Ekbatani et al. 2012) is that MPC can effectively handle the situation in congested and saturated traffic network conditions (Moshahedi 2021; Sirmatel and Geroliminis 2018). In addition, MPC can handle disturbances, uncertainty, and model mismatch errors that exist in the model as a function of data collection with noisy measurements and demand with prediction error. Furthermore, it can easily handle constraints and the nonlinear dynamics of traffic problems (Geroliminis et al. 2013; Haddad et al. 2013; Moshahedi 2021; Ramezani et al. 2013; Sirmatel and Geroliminis 2018). Also, the type of solver affects computation time (Diehl et al. 2009), so MPC is an appropriate basic framework for the RG control problem. Its performance is strongly influenced by the value of the control and prediction horizon \( n_p \).

This research develops several MPC-based RG models such as the proportionally fair RG control and the anticipatory RG control, which are compared to a benchmark as illustrated in detail in the next sections.

3.1.2. Formulation of the second-level MFD-based route choice problem and Plant model

As explained previously, real-time behavioural responses to route advice are not often accounted in the first-level control model. In other words, some drivers who are advised to take longer paths for the system’s benefit, may not comply with such routing advice. So, at the second-level model, it is assumed that the actual route choice ratio, which is used for updating the traffic states in the plant model that represents the “synthesised real” traffic network, is a linear combination of optimized RG instructions and driver route-choice decision as follows:

\[
\theta''(t) = \gamma \hat{\theta}(t) + (1 - \gamma)\theta'(t)
\]

where \( \gamma \) is the assumed compliance rate to the optimized RG and is considered a constant that indicates the percentage of drivers following the RG recommendations of the RG control scheme.
(Le et al. 2017; Sirmatel and Geroliminis 2018). \( \hat{\theta}(t) \) is a vector containing all \( \hat{\theta}_{IJ}(k) \) which is the optimized RG, and \( \theta'(t) \) is a vector containing all \( \theta'_{IJ}(k) \) which is the driver route choice decision obtained through a stochastic route choice model which will be explained later in this section. According to Eikenbroek et al. (2021), such formulation which also incorporates drivers behavioural responses results in a more fair routing model because it uses actual drivers routing decisions.

This thesis assumes a region-based model, determining the sequence of regions taken for a trip from an origin to a destination region, since the link-level traffic and tracked vehicle information is unavailable. So, to establish a stochastic user equilibrium that represents the impacts of RG on how traffic is distributed in the network to obtain a more realistic routing ratio, an MFD-based regional route choice model is used in the second-level model.

At each simulation step, the total accumulation for each region is calculated based on the prevailing traffic states. Then, for any region \( I \), regional travel time is estimated as \( \tau_I(N_I(t)) = \frac{N_I(t)}{G_I(N_I(t))} \) with the assumption that regional trip length has a constant value.

The summation of travel time for the regions which belong to each path, estimates that path’s total travel time as:

\[
\tau^p_{ij}(t) = \sum_{r \in R_p} \tau_r(N_r(t)) \quad p \in P_{ij}
\]

where \( \tau^p_{ij} \) is the travel time of path \( p \) which starts from the origin in region I and heads towards the final destination region J, and IJ pairs are named \( w \ (w \in W) \); \( R_p \) is the set of regions which belong to the path \( p \), and \( P_{ij} \) is a set of paths between IJ pairs \( \in W \). Accordingly, the driver route choice decision without control is calculated from the following logit assignment formula (Prashker and Bekhor 2004):
\[ \theta'_{IJ}^{p} = \frac{e^{-\phi \cdot t_{IJ}^{p}}}{\sum_{p \in P_{IJ}} e^{-\phi \cdot t_{IJ}^{p}}} \quad p \in P_{IJ} \]

where \( \theta'_{IJ}^{p} \) is the drivers’ stochastic user equilibrium routing decision based on existing region states, and \( \phi \) is the logit model parameter that indicates the drivers’ level of awareness of regions travel time; thus, higher \( \phi \) corresponds to higher level of knowledge of the network travel time and vice versa.

At each time step, once the optimized RG instructions \( \hat{\theta} \) are obtained from the first-level model and the actual routing ratio is obtained from the second-level model, the traffic state of the network is updated using the plant model before starting the next step. It is assumed that the plant model has access to real values of demand while the prediction model has only access to the average values (Moshahedi 2021). The observed state of the plant model is assumed to have noisy measurement. Thus, to update the traffic states based on real demand, an error term which has a normal distribution with the mean equal to 0 and a standard deviation equal to 0.2 is used to represent demand uncertainty. So, using current states, real values of demand and \( \theta''(t) \) route choice ratio as obtained in Eq. (3.11); the updated states are obtained to be used as initial states at the next step of the optimization program. Also, the routing ratio \( \theta''(t) \) is used as the updated new initial value of RG ratios for the next step of the RG optimization problem.

### 3.2. Optimal Control of Urban Networks via Regional RG

#### 3.2.1. Benchmark MPC controller with the objective of minimizing total travel time as a basic control

Figure 3.1 provides an overview of the benchmark method similar to Sirmatel’s and Geroliminis (2018) and Moshahedi et al. (2021) papers. Figure 3.1 shows that at each current simulation time, \( t_c \), MPC RG controller advise RG ratio by minimizing total travel time at the first-level optimization model, based on formulation presented below. Since some drivers may not comply with such routing advice, the real route choice is estimated by a combination of driver routing decisions by the logit model and the optimized RG result at the second-level of model.
Then, the plant model estimates states. These new states and the estimated route choice will be injected into the optimization model as the network’s states and initial values of RG variables before starting optimization at the next step.

Figure 3.1. Schematic view of the basic control in model predictive control framework

In the first-level MPC-based optimizing RG, the total travel time of the network is considered as the objective function. At each time step (t), the RG ratios are optimized over a prediction horizon \([k = 0, \ldots, n_p - 1]\); but only the results of the first-time steps are used at the next second-level model. The first-level MPC optimization model is formulated as follows:

\[
\underset{\theta}{\text{minimize}} \quad T_c \cdot \sum_{k=0}^{n_p-1} \sum_{l \in \mathcal{R}} N_l[k]
\]

subject to:
\[ N[0] = \hat{N}(t_c) = \hat{N}(t_c - 1) + dt \times \dot{N}(t_c - 1) \]
\[ \theta^0 = \theta[0] = \theta''(t_c - 1) = \gamma \dot{\theta}(t_c - 1) + (1 - \gamma)\theta'(t_c - 1) \]

for \( k = 0, ..., n_p - 1:\)
\[ N_i[k + 1] = f(N[k], Q[k], \theta[k]) \]
\[ 0 \leq \sum_{j \in R} N_{ij}[k] \leq \beta N_i^l \text{jam}, \quad \forall I \in R \]
\[ \theta_{\text{min}} \leq \theta_{IJ}[k] \leq \theta_{\text{max}}, \quad \forall I, J \in R, I \neq J, H \in N_i \]
\[ \sum_{H \in N_i} \theta_{IHJ}[k] = 1, \quad \forall I, J \in R, I \neq J \]
\[ if \ k \geq n_c: \ \theta[k] = \theta[k - 1] \]

where \( T_c \) is the control sampling time; \( k \) is the control interval counter; \( t_c \) is the current control time step; \( \hat{N}(t_c) \) is the measurement taken at \( t_c \); \( \dot{\theta}(t_c - 1) \) is the optimization result from the previous optimization step; \( n_p \) and \( n_c \) are the prediction horizons and control horizons; \( f \) is the time discretized version of the dynamic equations; \( \gamma \) is the assumed compliance rate; \( N[k], Q[k], \theta[k] \) are the vectors containing all \( N_{ij}[k], Q_{ij}[k], \theta_{IJ}[k] \) terms respectively; \( \theta' \) is the drivers routing decision at current state based on their response to current state as obtained from the logit model; and \( \theta''(t_c - 1) \) is the estimated route choice ratio at the previous simulation time step.

Eq. (3.14) minimizes the total travel time of the network by minimizing total accumulation of the network (Keyvan-Ekbatani et al. 2012), with respect to RG ratio value, over a prediction horizon \( n_p \), with \( T_c \) as a sampling time. To solve the optimization problem, the newly estimated RG, \( \theta''(t_c - 1) \), which is estimated based on Eq. (3.11) from the previous step, is used as the initial value of RG which is needed at each step.

Eq. (3.18) limits the region accumulation to jam/gridlock accumulation level. Eq. (3.19) constrains the RG ratio values between a lower limit \( \theta_{\text{min}} = 0 \), preventing the flow from transferring to other neighbor regions; and an upper limit \( \theta_{\text{max}} = 1 \), permitting the whole flow from origin
region to transfer to a specific neighbor region (Sirmatel and Geroliminis 2018). Eq. (3. 20) is a flow conservation equation which forces the summation of RG ratios over each region’s neighboring regions to be equal to 1; where all the flows that exit from a region are distributed only in neighboring regions; indicating that new flow cannot be created or erased.

3.2.2. Proportionally fair RG control problem formulation

This thesis introduces a novel proportionally fair RG control scheme in an MPC framework. Most previous models of MFD-based RG manage the traffic state of network while mainly focusing on improving network efficiency by minimizing total travel time (Sirmatel and Geroliminis 2018; Wei and Yang 2019). In doing so, some drivers might be directed to follow longer paths instead of shorter ones; thereby creating an inequity issue. Proportional fairness concept can address this issue by balancing the trade-off between network efficiency and fairness.

As not above, the concept of proportional fairness was initially proposed by Kelly (1997) for allocating resources in wireless networks and was later adapted for transportation application by Aalami and Kattan, in the context of transit emergency evacuation (2018, 2021) and transport market (2022). The underlying concept of proportional fairness is offering an equitable allocation of shared resources by considering the improvement in the proportion of user utilities without sacrificing efficiency. Thus, user utility, which expresses user benefits in terms of network resource allocation is at the core of the proportional fairness concept. The utility $U_i(x_i)$ expresses the benefits that user i gains because of allocating resource $x_i$; while $U_i(x_i)$ is a positive, non-decreasing function of $x_i$.

A proportional fair allocation is shown to be equivalent to maximizing the summation of the proportions of the user utilities (Kelly 1997; Aalami and Kattan 2018, 2021, 2022). In other words, if we remove a piece of resource from one user and allocate it to another user, and this “move” reduces the utility of the first user by p% but adds to the utility of the other user by p% or more, we will do this move because it increases the total proportion of user utilities.

A resource allocation $A^* = \{x_1^*, ..., x_n^*\}$ is a feasible, proportional fair allocation of resources if it satisfies the following conditions defined as:
Definition 1: Given a set of utility functions \( \{U_1(x_1), U_2(x_2), \ldots, U_n(x_n)\} \), and a set of weights \( \{w_1, w_2, \ldots, w_n\} \), a set of feasible solutions \( \{x_1^*, \ldots, x_i^*\} \) is a weighted proportional fair allocation of resources, if it satisfies the following condition:

\[
\sum_{i=1}^{n} w_i \frac{U_i(x_i) - U_i(x_i^*)}{U_i(x_i^*)} \leq 0 \quad \forall A^* = \{x_1^*, \ldots, x_i^*\}
\]

Where \( w_i \) denotes the weight of utility function \( U_i(x_i) \). Kelly (1997), and recently Aalami and Kattan (2018, 2021, 2022) proved Definition 1.

Definition 2: Eq. (3.22) is achieved by maximizing the weighted sum of the logarithm of the utility functions:

\[
\sum_{i} w_i \ln U_i(x_i)
\]

Definition 3: For a utility function to satisfy Eq. (3.22), \( \sum_i w_i \ln U_i(x_i) \) must be a non-negative, strictly concave function.

In the context of MFD, Moshahedi (2021) explained that the utility function \( U_i(x_i) \) indicates the benefit that the achieving resource \( x_i \) gives to user \( i \). It is a positive, non-decreasing function of \( x_i \). A set of solutions \( A^* = \{x_1^*, \ldots, x_i^*\} \) is feasible if it does not violate the capacity constraints.

To model the fair RG control problem, firstly, a utility function must be developed, and secondly the objective function must be constructed while satisfying the above utility functions characteristics. Depending on the focus of the traffic management center (TMC), a user can be defined as an entire region to represent the perspective of all transportation networks affected by the routing (i.e., both road and non-road user perspectives) in a given region. Due to the complexity of the transportation system, the RG benefits and externalities can vary significantly among travellers, road users as well as non-users; and over space or time. For instance, while a routing guidance can improve the travel time of the drivers receiving it, externalities of increased congestion, pollution, noise and probability of collisions are projected on non-drivers such as
pedestrians, cyclists and users of adjacent lands in a given region. Such externalities are often overlooked in the RG literature. According to Banister (2018), those who receive less benefits and more externalities are known to experience double inequality. It is, thus, the share of distribution of these benefits and externalities that the proportional fairness concept is concerned about when defining users in a region. Thus, if the goal is to address equity as an aggregate measure of travellers, roads and non-road users alike, a utility focusing on the region is more appropriate, which can be expressed in terms of region speed or region accumulation. Alternatively, if the focus is on travellers of different paths, a user can be defined as a group of travellers using a single path in the network; accordingly, the user utility can be expressed as the user path travel times or path travel speed. A more disaggregate formulation of the utility can be also formulated by distinguishing between the valuation of travel time between commuters and non-commuters, but that is outside the scope of this thesis. The weight \( w_i \) in equations (3.22) and (3.23) gives the TMC the flexibility of prioritizing one user over another (e.g., downtown or a path travelled by emergency vehicles).

At the RG control problem, the utility function \( U_I(N_I) \) should be expressed as a function of regional accumulation \( N_I \) for each region. RG ratios thus can influence the allocation of network resources (i.e., network capacities) by manipulating the divergence of vehicles to the next regions. Hajiahmadi et al. (2013) proposed a speed-accumulation relationship for MFD as an exponential function. The speed function for region I is defined as follows:

\[
V_I = v_{free} \cdot \exp (-0.5 \left( \frac{N_I}{N_{cr}} \right)^2) \tag{3.24}
\]

where \( v_{free} \) (m/sec) is the free-flow speed, and \( N_{cr} \) (veh) is the critical accumulation in region I. To express utility as a non-decreasing function, Moshahedi (2021) used regional speed as an indicator of a region’s utility level and manipulated Eq. (3.24) by expressing region speed as a function of available (remaining) accumulation; thus, substituting \( N_I \) with \( (N_{jam}^i - N_{remaining}^i) \). This substitution results in a non-decreasing function of available accumulation \( (N_{remaining}^i) \) which is expressed as the following utility function:
\[ U_I(N_I^{remaining}) = v_{free} \cdot \exp \left( -0.5 \left( \frac{(N_I^{jam} - N_I^{remaining})}{N_I^{cert}} \right)^2 \right) \quad \forall I \in \mathcal{R} \]

Also, Moshahedi proved that definition 3 is satisfied for the proposed function by showing that the second derivative of \( \Sigma_{I \in \mathcal{R}} \ln U_I(N_I) \) is always a negative number.

\[
\frac{\partial^2 \ln(U_I(x_i))}{\partial x_i^2} = \frac{\partial^2 \ln (v_{free} \cdot \exp \left( -0.5 \left( \frac{(N_I^{jam} - N_I^{remaining})}{N_I^{cert}} \right)^2 \right))}{\partial N_I^{remaining}^2} = -1 \frac{1}{N_I^{cert}^2} < 0
\]

Given that Eq. (3.26) is always satisfied, \( \ln U_I(N_I) \) is a strictly concave function. Given a set of positive weight parameters and concave function \( \ln U_I(N_I) \), the weighted summation of concave functions is concave as well; thereby proofing the existence of a unique equilibrium solution.

As part of the sensitivity analysis of the proportionally fair RG control scheme, besides the regional average speed being the utility of each region, two other utilities based on path average speed and path travel time were also examined. So, three proportional Fair RG Controls were developed: Fair Control based on regional speed (FCrs) – an RG scheme that represents the perspectives of road users and non users within region; and Fair Control based on path speed (FCps) and Fair Control based on path time (FCpt) – two variants of RG schemes that represent the perspective of the travellers on a given path.

While the average regional speed is obtained as the ratio of production and accumulation (Eq. (3.32)), the regional average travelling time is estimated from \( \left( \frac{L_I}{V_I} \right) \), (i.e., the average trip length inside a region divided by average regional speed). In other words, based on Eq. (3.33), the average speed of the regions which belong to each path is used as the estimate of that path’s speed. Also, the summation of travel time of regions which belong to each path is the estimate of that path’s travel time (Eq. (3.34)).
Figure 3.2. Schematic view of the proposed Proportionally Fair RG Control

Figure 3.2 provides an overview of the proportionally fair RG control method. At first-level, using the utility functions which are discussed and based on the MFD dynamical traffic states, a proportionally fair RG control problem is defined. Since some drivers may not comply with such routing advice, the actual route choice $\theta''(t)$ is estimated at the second-level as described in Eq. (3.11). Then, the plant model estimates the traffic states for the next step, based on these newly updated states and the estimated route choices, the optimization model at the first-level restarts the optimization for the next step.

The formulations of the three variants of the proportional Fair RG Control are as follows:

1) Fair Control based on regional speed (FCrs),

$$FCrs \quad maximize_{\theta} \sum_{k=0}^{n_p-1} \sum_{l \in \mathbb{R}} w_l \ln(V_l[k])$$

2) Fair Control based on path speed (FCps),

$$FCps \quad maximize_{\theta}$$

3) Fair Control based on travel time (FCtt),

$$FCtt \quad minimize_{\theta}$$
FCps \[\text{maximize}_\theta \sum_{k=0}^{n_p-1} \sum_{p \in P} w_p \ln(V_p[k])\]

and

3) Fair Control based on path time (FCpt),

FCptime \[\text{minimize}_\theta \sum_{k=0}^{n_p-1} \sum_{p \in P} w_p \ln(T_p[k])\]

All of the above are subject to:

\[N[0] = \hat{N}(t_c) = \hat{N}(t_c - 1) + dt \times \hat{N}(t_c - 1)\]

\[\theta^0 = \theta[0] = \theta''(t_c - 1) = \gamma \hat{\theta}(t_c - 1) + (1 - \gamma) \theta'(t_c - 1)\]

For \(k = 0, ..., n_p - 1:\)

\[V_I[k] = v_{free} \cdot \exp(-0.5(N_I[k] / N_{I,crit})^2) = v_{free} \cdot \exp(-0.5(N_I^{lam} - N_I^{remaining}[k]) / N_{I,crit}^2)\]

\[V_p[k] = \text{ave}(V_I[k]) \quad I \in R_p\]

\[T_p[k] = \sum_{I \in R_p} \left( \frac{L_I}{V_I[k]} \right)\]

\[N_I[k + 1] = f(N[k], Q[k], \theta[k])\]

\[0 \leq \sum_{j \in \mathcal{R}} N_{IJ}[k] \leq \beta N_I^{lam}, \quad \forall I \in \mathcal{R}\]

\[0 \leq \theta_{IJH}[k] \leq 1, \quad \forall I, J \in \mathcal{R}, I \neq J, H \in \mathcal{N}_I\]

\[\sum_{H \in \mathcal{N}_I} \theta_{IJH}[k] = 1, \quad \forall I, J \in \mathcal{R}, I \neq J\]

\[\text{if } k \geq n_c: \quad \theta[k] = \theta[k - 1]\]

where \(T_c\) is the control sampling time; \(k\) is the control interval counter; \(t_c\) is the current control time step; \(\hat{N}(t_c)\) is the measurement taken at \(t_c\); \(\hat{\theta}(t_c - 1)\) is the optimization result from the previous optimization step; \(\theta''(t_c - 1)\) is the estimated route choice ratio at the previous simulation time step; \(n_p\) and \(n_c\) are the prediction horizons and control horizons; \(f\) is the time discretized version of dynamics equations; \(\gamma\) is the assumed compliance rate; \(N[k], Q[k], \theta[k]\) are
vectors containing all $N_{IJ}[k]$, $Q_{IJ}[k]$, $\theta_{HI}[k]$ terms, respectively; $\theta'$ is the drivers' routing decision at the current state based on the logit mode; $N_{I}^{crt}$ is critical accumulation region $I$; $v_{free}$ is the free flow speed; $w$ is a weight parameter corresponding to each utility; $V_r$ is average speed of region $r$; $L_r$ is average length of region $r$; $V_p$ is average speed of path $p$; $T_p$ is travel time of path $p$; and $R_p$ is the set of regions which passed by path $p$.

The objective functions introduced in Eq. (3.27), Eq. (3.28), and Eq. (3.29) are defined in terms of the weighted sum of the logarithm of a region’s utility, path’s utility, and path’s disutility, respectively. These objective functions maximize the total utility Eq. (3.27), Eq. (3.28) or minimize the disutility Eq. (3.29) of a network with respect to RG ratios. Weight parameters $w_I$ and $w_p$ are included to prioritize important regions or paths of a network by assigning higher relative values. For example, high density downtown areas can be given higher priority in the form of a larger weight to alleviate its congestion by further discouraging routing through it; thus, supporting more sustainable travel modes (i.e., transit, biking and pedestrian). Alternatively, an individual weighing scheme can be assigned to each region to reflect a specific region's characteristics and/or related transport policies. Instances of such weighing can be designed to incorporate considerations such as regions with high access to public transit, regions with high density, regions experiencing high congestion, etc.

Constraint equation (3.36) limits the region accumulation to reach jam/gridlock accumulation level. Constraint equation (3.37) constrains the RG ratio values between a lower limit $\theta_{min} = 0$ and an upper limit $\theta_{max}=1$. Constraint equation (3.38) is a form of flow conservation constraint that restricts the summation to 1 of RG ratios over each region’s neighboring regions, because all the flows that exist from a region are distributed only to neighboring regions, since flow cannot be created or eliminated. Equation (3.39) keeps the RG fixed for the simulation steps than are greater than the control horizon $n_c$.

The initial RG ratio value at each step used to solve the optimization problem comes from the estimated route choice ratio, $\theta''(t_c - 1)$ of Eq. (3.11), that is a combination of drivers’ routing decisions by the logit model and the optimization result at the previous step.
At each time step, a cost function over a prediction horizon \((n_p)\) is optimized while satisfying the constraints, yet only the results of the first steps are deployed. The horizon is shifted forward, while RG ratios are optimized again using the newly updated states and the resulting route choice (which is based on the assumed drivers compliance rate and responses to the optimized RG obtained from the first-level). The whole process is repeated continuously until the end of the simulation time. The cost function is a weighted sum of the logarithm of the utility functions.

3.2.3. Framework for anticipatory control of urban networks

In this thesis, the anticipatory RG control is modelled as a two-level optimization model to involve the drivers route choice behaviour in the control model where the user reactions to changes in RG control is endogenously incorporated as part of the design of the control policies. It can be named consistent routing schemes as defined by Ben-Akiva et al. (2001) and Bottom (2000) who stated that anticipatory control inherently aims to provide consistent RG ratio by anticipating how users react to the optimized RG and, thus, resulting in more consistent RG schemes.

Previous models control the traffic state of the network by minimizing dis-utilities or maximizing utilities. Those models only considered behavioural response when updating models using a assumed compliance rate, rather than an integral step of optimizing the control. Since some drivers may not comply with such routing advice, incorporating their behaviour proactively as part of the optimizing step is important to achieve a more optimum solution and more consistent routing schemes. So, in AC models, the assumption of compliance rate is not needed and then, this model is more realistic.

The above circular dependencies aiming to reach consistent guidance entail the formulation and solving of a fixed-point problem as suggested by Ben-Akiva et al. (2001) and Bottom (2000). At the fixed-point problem, RG ratios are determined based on the traffic states’ patterns received from the route choice model, and routing decisions are made based on the route travel time estimated at the RG model.

In this thesis, the anticipatory control of RG is presented to integrate driver attempts in minimising the travel costs (i.e., user equilibrium) and the drivers behavioural response as part of
the control problem. Three objective functions are used to optimize the RG ratio, and introduce three AC models as follow:

Anticipatory basic RG control (ABC): minimize the total travel time of network

Anticipatory fair RG control based on path time (AFCpt): minimize the proportional fairness of path travel time

Anticipatory fair RG control based on regional speed (AFCrs): maximize the proportional fairness of region’s speed

Anticipatory control considers drivers’ routing decisions by incorporating their anticipated responses to control, and including these likely responses in the control decision, with an aim to reach global optimality (Huang et al. 2017b). Figure 3. 3 provides an overview of the two-level anticipatory RG control scheme, in which the MPC-based RG models are at the first level and the driver route choice model at the second level mutually interacting and influencing each other. Thus, at the first level, MPC optimizes the RG ratio to meet either the objective function of minimizing the total travel time or those of the proportional fairness formulations; while at the second level, the model checks traveller response to this RG advice by choosing their path, which is estimated using the logit assignment formula. The current state is updated based on the driver route choice at the second level which is then input to the first level and the process continues iteratively. If the convergence criterion between the RG ratio and the drivers’ behavioural response is not satisfied, then the loop continues by using the recently updated traffic state and the drivers routing decision as the initial values of the variables to be optimized (i.e., RG) at the next iteration. Once convergence is achieved or the stopping criterion is satisfied, route split ratios are sent to the plant model to update the traffic states for the next time step. Then, both updated traffic states and the route split ratios are injected into the optimization model at the first level to reflect the current network’s traffic state and initial value of RG when starting optimization at next time step. The algorithm continues over all simulation time.
In the first-level MPC-based optimizing RG, each of total travel times of the network, proportional fairness based on path travel time, and proportional fairness based on region speed, are considered as the objective function. At each time step \((t)\), the RG ratios are optimized over a prediction horizon \([k = 0, \ldots, n_p - 1]\); but only the results of the first time of prediction horizon are used at the next second-level model. The first-level MPC optimization model is formulated as follows:
\[
ABC: \quad \text{minimize}_{\theta} \quad T_c. \sum_{k=0}^{n_p-1} \sum_{I \in \mathcal{R}} N_I[k]
\]

\[
AF Cpt: \quad \text{minimize}_{\theta} \quad \sum_{k=0}^{n_p-1} \sum_{p \in P} w_p. \ln(T_p[k])
\]

\[
AF Crs: \quad \text{maximize}_{\theta} \quad \sum_{k=0}^{n_p-1} \sum_{I \in \mathcal{R}} w_I. \ln(V_I[k])
\]

subject to:
\[
N[0] = \tilde{N}(t_c) = \tilde{N}(t_c - 1) + dt \times \dot{\tilde{N}}(t_c - 1)
\]
\[
\theta^0 = \theta[0] = \theta'''(t_c - 1) = \hat{\theta}(t_c - 1)
\]

for \( k = 0, ..., n_p - 1 \):
\[
N_I[k + 1] = f(N[k], Q[k], \theta[k])
\]
\[
0 \leq \sum_{J \in \mathcal{R}} N_{IJ}[k] \leq \beta N_{IJ}^{jam}, \quad \forall I \in \mathcal{R}
\]
\[
\theta_{\min} \leq \theta_{IJ}[k] \leq \theta_{\max}, \quad \forall I, J \in \mathcal{R}, I \neq J, H \in \mathcal{N}_I
\]
\[
\sum_{H \in \mathcal{N}_I} \theta_{IJ}[k] = 1, \quad \forall I, J \in \mathcal{R}, I \neq J
\]
\[
if \ k \geq n_c: \quad \theta[k] = \theta[k - 1]
\]
\[
V_I[k] = v_{free}. \exp(-0.5(\frac{N_I^k}{N_I^{crit}})^2) = v_{free}. \exp(-0.5(\frac{N_I^{jam}-N_I^{remaining}[k]}{N_I^{crit}})^2)
\]
\[
T_p[k] = \sum_{I \in R_p} \left( \frac{L_I}{V_I[k]} \right)
\]

where \( T_c \) is the control sampling time; \( k \) is the control interval counter; \( t_c \) is the current control time step; \( \tilde{N}(t_c) \) is the measurement taken at \( t_c \); \( n_p \) and \( n_c \) are the prediction horizons and control horizons; \( f \) is the time-discretized version of dynamic equations; \( \gamma \) is the assumed compliance rate; \( N[k], Q[k], \theta[k] \) are the vectors containing all \( N_{IJ}[k], Q_{IJ}[k], \theta_{IJ}[k] \) terms, respectively; \( \theta'''(t_c - 1) \) is the converged route split ratios at previous time step that will be the initial value of \( \text{RG} \) at next step; \( N_{I}^{crit} \) is critical accumulation region \( I \); \( w \) is a weight parameter corresponding to
each utility; $V_r$ is average speed of region $r$; $L_r$ is average length of region $r$; $T_p$ is travel time of path $p$; and $R_p$ is the set of regions which passed by path $p$. The following algorithm is used to demonstrate the iteration loop, which occurs between upper- and second-level models at each time step, which leads to the converged route split ratios ($\theta'''$).

Step 1:

- Set iteration number $m = 1$.
- Calculate optimized RG ($\hat{\theta}$) from Eq. (3.40) at first-level MPC controller.

Step 2:

- Update current traffic states $\hat{N}(t_c)$ based on optimized RG ratios by using Eq. (3.4) and Eq (3.5).
- Estimate path travel time ($\tau_{IJ}^P$) by using Eq. (3.12).
- Estimate drivers route decision ($\theta'$) using logit assignment formula at Eq. (3.13).
- Update current traffic states $\hat{N}(t_c)$ based on drivers route choice ratios by using Eq. (3.4) and Eq (3.5).

Step 3:

- If the difference between optimized RG ($\hat{\theta}$) and drivers route decision ($\theta'$) is less than the convergence criterion, convergence is achieved, and the value of route split will be named consistent route split ratios ($\theta'''$) and will be sent to the plant model. Otherwise, increase the iteration number to $m = m+1$, and return to Step 1 to run first-level optimization model by using the recently updated traffic states and drivers routing decisions as the initial value of optimizing variables (RG) at the next iteration.

Eq. (3.40-a), which is the objective function of ABC, minimizes the total travel time of the network by minimizing total accumulation of the network (Keyvan-Ekbatani et al. 2012), which
in turn is expressed as a function of RG ratio value, over a prediction horizon $n_p$, with $T_c$ as a sampling time. Eq. (3. 40-b), which is the objective function of AFCpt, is defined in terms of the weighted sum of the logarithm of path disutility, which minimizes the assigned user disutility with respect to RG ratios. Eq. (3. 40-c), which is the objective function of AFCrs, is defined in terms of the weighted sum of the logarithm of the region utility, which maximizes the assigned user utility with respect to RG ratios. The Constraint Eq. (3. 44) limits the region accumulation to jam/gridlock accumulation level. Eq. (3. 45) constrains the RG ratio values between a lower limit $\theta_{min} = 0$, which means flow does not transfer to that neighboring region; and an upper limit $\theta_{max} = 1$, which means all flow from origin region transfers to that specific neighboring region (Sirmatel and Geroliminis 2018). Next, Eq. (3. 46) is a flow conservation equation that restricts the summation to 1 of RG ratios over each region’s neighboring regions, because, as mentioned previously, all of the flows that exit from a region are distributed only in neighboring regions (i.e., new flow cannot be created or eliminated).

3.3. Performance Measures

To assess the effectiveness of the developed RG in terms of fairness and efficiency, several measures are analyzed and compared. While efficiency is often measured in terms of Total Travel time (TTT), fairness cannot be easily measured since it is different from various perspectives. Due to the complexity of the transportation system, RG benefits and externalities can significantly vary over space and time among road users as well as non-users. For instance, if we evaluate equity from the perspective of road and non-road users alike, measures focusing on a specific region rather than path are more appropriate and this can include measures such as region accumulation, region speed and speed variance. Table 3. 1 explains a summary of the measures of performance examined.

<table>
<thead>
<tr>
<th>Measured Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (TTT) (min)</td>
<td>Lower Total travel time indicates a higher efficiency of the network.</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Regional Accumulation (veh)</strong></td>
<td>Accumulation of each region is directly proportional to TTT of the region, and thus, is considered as a measure of efficiency. A lower average accumulation reflects a higher level of efficiency of the network. Also, the accumulation graph shows which regions are subject to higher accumulation. Using accumulation graphs, regional accumulation states as under various controlled and uncontrolled scenarios are compared. The closer is the accumulation level of the regions, the fairer is the distribution of the traffic congestion from the regional perspective.</td>
</tr>
<tr>
<td><strong>∑ std. of accumulation</strong></td>
<td>Lower summation of standard deviation (∑ std.) of accumulation shows a higher level of homogeneity of traffic distribution among the regions. This is a measure of fairness from the regional perspective.</td>
</tr>
<tr>
<td><strong>Speed (Km/hr)</strong></td>
<td>A higher average region’s speed shows a higher performance of the network. Like the accumulation graph, the speed plot shows which regions benefit more from the RG in terms of higher average speed. The closer is the regions’ speed plot, the fairer is the distribution of the traffic congestion among the regions.</td>
</tr>
<tr>
<td><strong>∑ std. of regional speed</strong></td>
<td>A low ∑ std. of regional speed is a measure of fairness as it portrays the ability of the RG in homogenizing traffic speed among the multi-region network.</td>
</tr>
<tr>
<td><strong>Paths’ travel time</strong></td>
<td>The Path’ travel time indicates how long it takes to travel between each OD pair using various paths. Thus, the closer the value of path travel time among the various paths with same OD, the higher is the fairness level from the travellers’ perspective.</td>
</tr>
</tbody>
</table>

TT(path) (min)
4. Numerical Results and Discussion

The performance of the proposed models is examined using a numerical example on a hypothetical seven-region urban network as shown in Figure 4.1. The blue arrows show possible flow distributions in the neighbouring regions 3, 4, and 6 using the RG ratios. The orange lines are examples of the resulting paths to be followed from origin region 5 to destination region 1 over the listed three neighbouring regions. It is assumed that up three shortest paths exist between any OD pair in the network. As the real network’ paths is built before, this hypothetical network’s path is assumed to be an input for the model. Path enumeration is avoided by considering the three shortest routes found using Dijkstra’s algorithm between any OD region pair in the network. The assumed paths between any OD pair are listed in Appendix A.

Figure 4.1. The urban network which is divided into seven regions

In this network, a specific MFD in the form of an exponential function similar to Hajiahmadi et al. (2013) is defined for each region that relates accumulation to travel production:
\[ G_I(N_I(t)) = \frac{v_{\text{free}}}{L_I} \cdot N_I(t) \cdot \exp\left(-0.5\left(\frac{N_I(t)}{N_I^{\text{crm}}}ight)^2\right) \]

In the base scenarios, it is assumed that all regions have the same MFD parameters with jam accumulation \( N_I^{\text{jam}} = 10000 \) (veh), critical accumulation \( N_I^{\text{crm}} = 2500 \) (veh), and free-flow speed \( v_{\text{free}} = 70 \) (Km/hr). In addition, it is assumed that the average trip length of vehicles inside a region is independent of the origin and destination locations and is constant for all regions with average trip length \( L_I = 3.6 \) (Km). Also, the summation of the average trip length for the regions which belong to each path, estimates that path’s travel distance. So, the path that passes from more regions, will have longer travel distance.

The values are assumed for each region’s boundary receiving capacity \( c_{IH}^{\text{max}} = 192 \) (veh/min) and \( \alpha = 0.64 \).

The shortest path between any two peripheral non-neighbor regions is usually accessed through the central region (i.e., region 4 in this hypothetical traffic network). Consequently, it is expected that more vehicles travel in region 4 compared to other regions.

Table 4.1. Summary of the examined scenarios and their assumptions illustrates the summary of the scenarios examined in this chapter.

<table>
<thead>
<tr>
<th>No Control</th>
<th>Control Type formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>Predefined Routing Compliance rate ( \gamma )</td>
</tr>
<tr>
<td></td>
<td>Anticipatory Routing Scheme</td>
</tr>
</tbody>
</table>

In addition, in the next chapter, rigorous sensitivity analysis is conducted to examine the performance of the model under high and low-demand profiles, various levels of region heterogeneity, and different road user compliance rates.

In all two-level RG formulations with predefined routing compliance rate, the “synthesised real” route choice ratio used for updating the traffic states in the plant model is a linear combination...
of the optimized RG and drivers route choice decisions estimated as $\gamma \hat{\theta} + (1 - \gamma)\theta'$; where $\gamma$ denotes the predefined drivers compliance rates; $\hat{\theta}$ is the optimized route ratio estimated using the control model; and $\theta'$ is driver route choice decision obtained through the logit route choice model. The use of the logit model allows the driver route behavioural choice to be incorporated in the model formulation.

In the two-level anticipatory RG control model, in order to consider drivers behaviour directly in the control model, the MPC at the first-level and driver route choice model at the second level mutually interact with each other. In other words, driver routing behaviour is obtained from a logit route choice model.

The lower control bound for RG ratio is $\theta_{min}=0$ (i.e., no routing to a neighboring region) and the upper control bound is $\theta_{max}=1$ (i.e., all flow is routed to the next region).

The demand profile follows a triangular distribution for 110 minutes with distinct parameters for different regions. To discharge the traffic network at the end of the simulation, the value of demand is considered zero for the last 70 minutes of the simulation. Because central region 4 is the most traversed region in the network, it is assumed that it has less internal demand than other regions. Demand uncertainty is considered for both low and high-demand scenarios. The error terms, which have a normal distribution with a mean equal to 0 and a standard deviation equal to 0.2, are assumed to describe demand uncertainty. The real value of demand which is presented in Figure 4. 2 for some ODs, is used to update states at the plant model, while the control model has only access to the average values of the demand. For example, Q62 show the demand profile of trips between origin region 6 to destination region 2.
The network is simulated for 180 minutes (3 hours) using a control sampling time $T_c = 1\text{(min)}$. The prediction and control horizons are chosen as $N_p = 7$ and $N_c = 3$ for the MPC model (Sirmatel and Geroliminis 2018). As demonstrated in previous work, the model’s performance is strongly influenced by the value of the prediction horizon $N_p$ (Sirmatel and Geroliminis 2018). If we use lower $N_p$, the model’s performance reduces.

For solving the MPC problem, the CasADi MATLAB solver was used. CasADi is an open-source nonlinear optimization tool that facilitates rapid and efficient implementation of different methods for numerical optimal control for nonlinear model predictive control (NMPC) (“CasADi’s documentation”, Andersson et al. 2019). For all simulations, MATLAB R2021a is used on a 64-bit Windows PC with 2.59-GHZ Intel Core-i7 and 16-GB of RAM.

As performance measures, data on regional accumulation, regional speed, regional RG ratio, and path travel time were collected and compared for multiple controlled and uncontrolled scenarios. The type of control scenarios investigated in this thesis are:

1. No Control (NC) scenario, wherein the flow ratios exiting a region and distributing over its neighboring regions are not controlled. Thus, drivers choose their routes according to the stochastic UE model (i.e., logit assignment).

2. Basic control (BC) RG scenario, constructed in a two-level model predictive control (MPC) framework, wherein RG rates are optimized using an MPC controller over the prediction horizon. Real routing is only updated after each control time step based on a linear
combination of the optimized RG ratio and the driver route choice using a stochastic UE model (i.e., logit assignment) with various compliance rates.

3. Proportionally fair control based on a regional speed (FCrs) RG scenario, wherein similar to the two-level MPC BC scenario, RG rates are optimized using the MPC controller over a prediction horizon using the proportional fairness concept, based on which an objective function on average regional speed is defined. Real routing is calculated similar to the BC scenario.

4. Proportionally fair control based on path speed (FCps) RG scenario, wherein similar to the two-level MPC BC scenario, RG rates are optimized using the MPC controller over a prediction horizon using the proportional fairness concept, based on which an objective function on average path speed is defined. Real routing is calculated similarly to the BC model.

5. Proportionally fair control on path time (FCpt) RG scenario, wherein, similar to the two-level MPC BC scenario, RG rates are optimized using the MPC controller over a prediction horizon using the proportional fairness concept, based on which an objective function on average path time is defined. Real routing is calculated similarly to the BC model.

6. Anticipatory basic control (ABC) RG scenario, in which the RG rates are optimized using a two-level optimization framework that integrates driver routing responses (i.e., logit assignment) directly into the control model. The two levels of the model actively interact and influence each other. The objective function is the same as BC and minimizes the total travel time of the network.

7. Anticipatory fair control based on regional speed (AFCrs) RG scenario – same as in ABC scenario, the RG rates are optimized using a two-level optimization framework that integrates driver routing responses (i.e., logit assignment) directly into the control model. The two levels of the model actively interacted and influenced each other. The objective function is the same as FCrs and maximizes a function on average regional speed.
8. Anticipatory fair control based on path time (AFCpt) RG scenario – same as in the ABC scenario, the RG rates are optimized using a two-level optimization framework that integrates driver routing responses (i.e., logit assignment) directly into the control model. The two levels of the model actively interacted and influenced each other. The objective function, which is the same as FCpt, minimizes a function on average path time.

The simulation was conducted for three sensitivity analysis cases as listed below:

1. Low and high-demand scenarios; 100% compliance rate and same MFD for all regions are considered.

2. Various levels of driver compliance to RG instructions (i.e., 100%, 85%, 70%, and 30% are considered); high-demand profile and same MFD for regions are considered.

3. Various regional MFDs to represent regions with variable size and traffic features. In the region heterogeneity sensitivity analysis, it is assumed that each region has a different MFD that is within ±5% of the original MFD. It is assumed that the jam accumulation is $N^{jam}_I = 10000, \beta_I$ (veh); the critical accumulation is $N^{crit}_I = 2500, \beta_I$ (veh); and the free-flow speed is $v_{free_I} = 70, \alpha_I$ (Km/hr); where $\alpha_I = [1.02 0.99 1.04 1 0.98 0.97 0.96]$ and $\beta_I = [0.96 1.02 0.98 1 1 0.97 1.03]$ are considered (Moshahedi 2021).

The sensitivity analysis results with respect to demand, drivers compliance rate, and various MFD parameters are presented in Sections 4.1–4.3. For each sensitivity analysis scenario, analysis of efficiency and fairness are conducted. Fairness is defined as the summation of variation of average regional speed within the multi-region network. Efficiency is defined as total travel time (TTT) of the network. Also, for each sensitivity analysis case, results are further analyzed in terms of regional accumulation, regional average speed, yielded RG ratios, and path travel time analysis.

4.1. Analysis of the Results Under Two Different Demand Profiles

The hypothetical network simulation results under different control scenarios during a 3-hour control period are summarized in Table 4.2 and Table 4.3.
Table 4.2. Simulation results of control models under high-demand scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>100% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>ABC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
<th>AFCpt</th>
<th>AFCrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time(min)</td>
<td>1324124</td>
<td>887342</td>
<td>898891</td>
<td>881562</td>
<td>878450</td>
<td>880009</td>
<td>911553</td>
<td>910174</td>
</tr>
</tbody>
</table>
|                        | (--)     | (-32.99%-32.11%(-33.42%(-33.66%(-33.54%(-31.16%(-31.26%))|)
| Accumulation (veh)     | 7356.24  | 4929.68  | 4993.84  | 4897.57  | 4880.28  | 4888.94  | 5064.18  | 5056.52  |
|                        | (--)     | (-32.99%-32.11%(-33.42%(-33.66%(-33.54%(-31.16%(-31.26%))|)
| Avg. (veh)             | 7789.47  | 1557.44  | 2053.50  | 801.48   | 1007.77  | 7747.48  | 1922.05  | 1925.75  |
|                        | (--)     | (-80.01%-73.64%(-89.09%(-87.07%(-90.05%(-75.32%(-75.28%))|)
| Speed (Km/hr)          | 60.05    | 64.68    | 64.48    | 64.77    | 64.80    | 64.79    | 64.34    | 64.35    |
| Avg. (Km/hr)           | (--)     | (7.71%)(7.36%)(7.85%)(7.90%)(7.88%)(7.14%)(7.16%)|)
| Sum std.              | 1126.95  | 224.63   | 285.85   | 107.60   | 135.44   | 99.33    | 270.28   | 272.58   |
|                        | (--)     | (-80.07%-74.64%(-90.45%(-87.98%(-91.19%(-76.02%(-75.81%))|)
| Loop CPU time(s)       | 0.73     | 25.92    | 110.82   | 20.80    | 20.38    | 27.39    | 707.67   | 668.93   |

Table 4.3. Simulation results of control models under low-demand scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>100% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>ABC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
<th>AFCpt</th>
<th>AFCrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time(min)</td>
<td>921797</td>
<td>778204</td>
<td>791178</td>
<td>778333</td>
<td>776118</td>
<td>776929</td>
<td>797556</td>
<td>797748</td>
</tr>
</tbody>
</table>
|                        | (--)     | (-15.58%(-14.17%(-15.56%(-15.80%(-15.72%(-13.48%(-13.46%))|)
| Accumulation (veh)     | 5121.09  | 4323.35  | 4395.43  | 4324.07  | 4311.77  | 4316.27  | 4430.87  | 4431.93  |
|                        | (--)     | (-15.58%(-14.17%(-15.56%(-15.80%(-15.72%(-13.48%(-13.46%))|)
| Avg. (veh)             | 25139.40 | 9575.55  | 16962.28 | 7756.81  | 7795.87  | 6888.84  | 15684.94 | 15441.07 |
| Sum std.              | (--)     | (-61.91%(-32.53%(-69.14%(-68.99%(-72.60%(-37.61%(-38.58%))|)
| Speed (Km/hr)          | 64.34    | 65.87    | 65.67    | 65.88    | 65.90    | 65.90    | 65.60    | 65.60    |
| Avg. (Km/hr)           | (--)     | (2.37%)(2.05%)(2.38%)(2.41%)(2.41%)(1.96%)(1.95%)|)
| Sum std.              | 366.27   | 124.31   | 220.43   | 92.48    | 96.42    | 83.28    | 204.68   | 203.36   |
|                        | (--)     | (-66.06%-39.82%(-74.75%(-73.68%(-77.26%(-44.12%(-44.48%))|)
| Loop CPU time(s)       | 0.68     | 22.73    | 810.08   | 23.07    | 19.35    | 21.77    | 831.32   | 661.79   |

In the high-demand scenarios with the 100% compliance rate assumption, the efficiency measure (i.e., TTT) has improved compared to NC by 31.16%, 31.26%, 32.11%, 32.99%, 33.42%, 33.54% and 33.66%, under AFCpt, AFCrs, ABC, BC, FCrs, FCpt and FCps, respectively. Based on these results, the difference in terms of improving efficiency measure is not significantly different among the various examined control strategies and they are all successful in improving
the network TTT. Furthermore, all these control schemes lead the multi-region network to have more homogeneous traffic distribution as indicated in the resulting standard deviations of region accumulation and region speed (i.e., $\sum \text{std.}$) in Table 4. 2. The control model with a lower value of $\sum \text{std.}$ for speed, is more capable of homogenizing traffic speed in the multi-region network, which has improved compared to NC by 74.64%, 75.81%, 76.02%, 80.07%, 87.98%, 90.45%, and 91.19%, under ABC, AFCrs, AFCpt, BC, FCps, FCrs, and FCpt, respectively. While ABC obtains the lowest, FCpt obtains the highest rank in fairness. In other words, FC-based control scenarios outperform BC and AC-based scenarios. More specifically, compared to other control strategies, FCpt shows superior performance in improving both TTT efficiency and fairness in the distribution of speed and accumulations, followed by FCrs and FCps which yield the second-best performances in terms of balancing both efficiency and fairness. The reason behind FCpt’s higher performance is that it directly optimizes path travel time by manipulating the number of vehicles using that path and, thereby, indirectly manipulating the traffic state of the regions. Therefore, it ensures fairness in the distribution of travel time among the different paths and regions in the network.

As explained earlier, due to the complexity of the transportation system, RG benefits and impacts can vary significantly among travellers, road users as well as non-users; and over space or time. The proportional fairness concept is concerned about the fair distribution of these benefits and externalities. If the goal is to address equity as an aggregate measure of travellers, road and non-road users alike, a utility focusing on the region is more appropriate, such as FCrs that is expressed in terms of region speed. On the other hand, if the focus is on travellers of different paths, a user can be defined as a group of travellers using a single path in the network; accordingly, the user utility function for the control model can be expressed as the user path travel times or path travel speed, as done in FCpt and FCps.

FCpt, which is based on the path travel time, ensures a level of fairness among the various paths within the network, and thus, yields a better outcome for drivers by improving efficiency and reducing the variance of the region accumulation and speed. Also, FCps is based on path speed that is estimated by summing the average speed of regions traversed by a given path. Since multiple paths use the regions at the same time, in FCps, the regions that are traversed more often by the paths are incorporated multiple times in the objective function that maximized the path utilities. In
contrast, in FCrs the objective function is defined based on region; thus, each region is accounted for once. Thus, in FCrs, reasonable weights can be assigned to each region to prioritize it over other regions in the network. Herein, the weight of region 4 (i.e., CBD) is considered 20% higher than other regions, to alleviate its congestion by further discouraging routing through it.

In the low-demand scenario with 100% compliance rate assumption, the efficiency measure (i.e., TTT) improved compared to NC by 13.46%, 13.48%, 14.17%, 15.58%, 15.56%, 15.72% and 15.80%; under AFCrs, AFCpt, ABC, BC, FCrs, FCpt and FCps, respectively. According to these results, while all AC-based models are more realistic as they do not need an assumption of compliance rate, they do not perform comparably well to non-AC scenarios with the assumption of 100% compliance rate, which is non-realistic in practice. Furthermore, BC and FC models lead the multi-region network to have more homogeneous traffic distribution as indicated in the resulting standard deviations of region accumulation and region speed (i.e., $\Sigma \text{std.}$) in Table 4. 3. A lower value of $\Sigma \text{std.}$ for speed, indicates that the model is more capable of homogenizing traffic speed in the multi-region network. Providing a more fair distribution in regional speed is an indication of fairness of the distribution of both benefits and impacts of RG. Table 4. 3 shows that $\Sigma \text{std.}$ of speed improved compared to NC by 66.06%, 73.68%, 74.75%, and 77.26% under BC, FCps, FCrs, and FCpt control models, while ABC, AFCpt, and AFCrs improved it by only 39.82%, 44.12%, and 44.48%, respectively. Therefore, AC-based RG results in the lowest performance in terms of the fair distribution of region speed. Yet, FC-based RG consistently outperforms their BC and AC-based counterparts in terms of fair distribution of speed among the different regions and paths. More specifically, among FC-based control scenarios, FCpt shows a superior performance not only in improving efficiency (TTS) but also in enhancing fairness. FCpt does that by directly optimizing path travel time while also manipulating the number of vehicles transferring through that path to other regions. For example compared to NC, in the FCps scenario with a high-demand profile, the efficiency improved by 33.66% and fairness improved by 87.98% with a reduction of the $\Sigma \text{std.}$ for speed. On the other hand, FCpt had a 0.12% efficiency improvement less than FCps, while it could improve fairness around 3.21% more than FCps. The results indicate the ability of FC-based RG in increasing the homogeneity of traffic and regional speed in an urban network while maintaining a high level of efficiency improvement.
Thus, in summary, a balanced trade-off between efficiency and fairness can be achieved under the proportional fairness RG strategies under the two examined demand levels. Among the different examined proportional fairness-based strategies, FCpt is found to be outperforming the other strategies in finding a more balanced trade-off by meeting the fairness objective without much compromise in efficiency.

4.1.1. Analysis of the results in terms of regional accumulation for various RG schemes

Figure 4. 3 and Figure 4. 4 show the accumulation states of regions under different demand scenarios. In both scenarios, all regions are emptied at the same time at t=120 min.
Figure 4. 3. Regional accumulations in high-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs
Figure 4. Regional accumulations in low-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs.
As shown in Figure 4. 3(a), under the high-demand scenario and NC, the CBD’s maximum accumulation reaches 5,300 veh (53% of jam accumulation) approximately 21% higher compared to the low-demand scenario (3,220 veh) presented by Figure 4. 4(a). Further, while regions 2, 4, and 6 are subject to high congestion conditions that lasts up to time interval 60 min, the remaining regions experience low accumulation during the whole simulation period. Under low-demand scenario, only regions 2 and 4 experience congested conditions with the longest one lasting less than 40 min.

As illustrated in the accumulation graphs of Figure 4. 3 (b, c, g, h) and Figure 4. 4 (b, c, g, h); BC, ABC, AFCpt, and AFCrs are successful in managing the traffic conditions in CBD (i.e., region 4) by manipulating its maximum accumulation level. In the low-demand scenario, ABC, AFCpt, AFCrs and BC results are somehow similar with CBD experiencing higher accumulation than the other regions in the network; however, all region accumulation is still maintained at less than the critical accumulation. Under the high-demand scenario, although the maximum accumulation of CBD under ABC (2,853 veh) is lower compared to that of BC (2,915 veh), under ABC, AFCpt, and AFCrs; region 2 experiences higher accumulations around the critical accumulation compared to that of BC.

According to the accumulation graphs of Figure 4. 3 (d, e, f) and Figure 4. 4 (d, e, f), in both demand scenarios, in all FC-based control scenarios, no region experiences congested conditions, and CBD’s maximum accumulation never exceeds 2,400 vehicles. In the low-demand scenario, the FC results are nearly similar; however, in the high-demand scenario, the results in Figure 4. 3 (d, e, f) show that fair routing schemes, based on regional speed and path time, give better results than fair control model on path speed. The path speeds are estimated based on the average speed of the regions existing on that path. FCps work similarly to FCrs; however, in the case of FCrs, reasonable weights can be assigned to each region to prioritize it over other regions in the network. Herein, the weight of region 4 (i.e., CBD) is considered 20% higher than other regions to alleviate its congestion and discourage routing through it; \( w = [1, 1, 1, 1.2, 1, 1, 1] \). The maximum regional accumulation does not exceed critical accumulation in FCrs, FCpt, and FCps models.
In all FC-based control models, while the accumulations reduced in regions 2, 4, and 6, the accumulations of the other regions increased (i.e., regions 1, 3, 5, 7), leading to a more even distribution of the accumulation in the various regions during the simulation time in FC models. On the other hand, based on Table 4.2 and Table 4.3 results, the efficiency of the network does not degrade much. Therefore, the FC-based control schemes proposed in this thesis create a balanced trade-off between efficiency and fairness, while in BC model efficiency, improvement is usually achieved at the cost of fairness.

4.1.2. Analysis of the results in terms of regional speed for various RG schemes

Figure 4.5 and Figure 4.6 demonstrate regional speed under the high- and low-demand scenarios, respectively.
Figure 4.5. Regional speed in high-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs
Figure 4.6. Regional speed in low-demand scenario, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs
In the NC analysis results for the high-demand scenario shown by Figure 4. 5(a), the CBD’s speed dropped to less than 8(Km/hr) at t=75min. Further, regions 2 and 6 experience regional speeds below 25 Km/hr reflecting the congested conditions in this scenario. Under the low-demand scenario, CBD’s speed dropped to 30 Km/hr as presented in Figure 4. 6 and Figure 4. 4(a). Average speeds of regions 2 and 6 are maintained around 40 Km/hr and above.

As shown in the accumulation graphs of Figure 4. 5(b, c, g, h) and Figure 4. 6(b, c, g, h); BC, ABC, AFCpt, and AFRs successfully control traffic conditions in the CBD by diverting traffic away from this region through the RG, resulting in an increase in CBD regional speed. In the low-demand scenario, the performance of BC and ACs are not much different, and the minimum resulting speed varies between 40.9 and 45 Km/hr. These findings are expected as the competition for available shared capacity is more pronounced when the demand is increased. In congested conditions of high-demand scenario, however, ABC is successful in slightly increasing the minimum speed in CBD to 37.77 Km/hr from 35.48 Km/hr under BC model. However, as shown in Table 4. 1, the average speed in the entire network was reported to be slightly higher in BC compared to that of ACs through the simulation time.

As shown in the speed graphs (Figure 4. 5), in addition to CBD, regions 2 and 6 experience higher congestion under AC compared to BC with their minimum speed reaching 40 Km/hr compared to 50 Km/hr under BC. These findings can be attributed to the fact that under AC, the drivers routing choice is integrated into the main control scheme; thus, to reduce their total path travel time, rather than traveling through the CBD, drivers prefer to use less congested regions to reach their destination. While AC is more successful in reducing CBD accumulation and increasing its minimum speed compared to the BC model, AC leads to increased accumulation and thus reduced minimum speed in regions 2 and 6, where traffic is diverted.

According to regional speed graphs of Figure 4. 5 (d, e, f) and Figure 4. 6 (d, e, f), the speeds of regions in FC models are maintained to be more harmonised in all the regions during the entire simulation time. The speed harmonisation results among the regions are especially obtained for the FC with the objective function focusing on path travel time (FCpt) and then region speed (FCRs). These findings are attributed to the core concept of proportional fairness that focuses on
the marginal gain in utilities of each region, rather than directly maximizing their utilities; thereby resulting in more fair, yet efficient, control schemes. The reasons are explained earlier in section 4.1 and subsection 4.1.1, that FCpt directly optimizes path travel time, which is associated with the number of vehicles transferring through that path, and thus, ensures fairness between different paths in the network. Also, because FCrs has reasonable weights that are assigned to each region to prioritize it over other regions in the network, its result is more fair than FCps that works similarly to FCrs with various weights.

In the low-demand scenario, various FC model results are close to each other; the minimum regional speeds are 50.48, 50.60, and 49.78 Km/hr in FCrs, FCpt, and FCps models, respectively. With increasing demand, in the high-demand scenario, the difference between results is more pronounced, and the minimum regional speeds are 45.91, 45.88, and 44.26 Km/hr in FCrs, FCpt, and FCps models, respectively. While the speeds of regions 2, 4, and 6 increased, those of the other regions decreased. As expected from the fairness concept, a balanced trade-off between efficiency and fairness happens. In other words, with the decrease in speed of regions 1, 3, 5, and 7; while keeping the efficiency of TTT and an acceptable level of average speed on the whole network; the speeds are more harmonised in all the regions, and fairness is achieved.

4.1.3. Analysis of the results in terms of regional RG ratios and path travel times for various RG schemes

Figure 4. 7, Figure 4. 8, and Figure 4. 9 show regional RG ratios and path travel times under the high-demand profile for the no control and control scenarios (a) NC, b) BC, c) AC, d) FCrs, e) FCps, and f) FCpt) for OD pairs (5, 2), (3, 1), and (6, 3), respectively. Also, Figure 4. 10, Figure 4. 11, and Figure 4. 12 show the regional RG ratio and path travel time under the low-demand profiles for the controlled and uncontrolled scenarios for the same ODs. In each case, three routes are considered from a given origin region to reach the destination: one passing through CBD, one traversing only one neighboring region, and another one traversing three intermediate regions.
Figure 4. 7. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 5 to destination 2 with paths \{\{[5\ 3\ 2]\};\{[5\ 4\ 2]\};\{[5\ 6\ 7\ 1\ 2]\}\}
Figure 4. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 3 to destination 1 with paths \{[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]\}
Figure 4.9. Regional RG ratios and path travel times in high-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 6 to destination 3 with paths {
\{[6 5 3]\};\{[6 4 3]\};\{[6 7 1 2 3]\}\}

61
Figure 4.10. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 5 to destination 2, paths \{[[5 3 2]],[[5 4 2]],[[5 6 7 1 2]]\}
Figure 4.11. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 3 to destination 1, paths \{[3 2 1];[3 4 1];[3 5 6 7 1]\}
Figure 4.12. Regional RG ratios and path travel times in low-demand scenarios, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt, g) AFCpt, h) AFCrs, origin 6 to destination 3, paths \{[[6 5 3]];[[6 4 3]];[[6 7 1 2 3]]\}
The results in all figures show that RG ratio graphs and path travel times are consistent. In other words, with a decrease in travel time on a given path, its guidance ratio increases and vice versa.

Based on the results, paths consisting of only one intermediate region (e.g., 532 and 542), have lower travel time compared to paths consisting of three intermediate regions (e.g., 56712). When no control is applied to the network, the CBD gets congested quickly as demand is loaded during the simulation time. Thus, to avoid passing through the CBD, vehicles are advised to follow longer routes; accordingly the route ratio increases for paths consisting of five regions. Conversely, in controlled scenarios, due to the proper network management, there is no need for RGs to advise vehicles to divert on long routes.

The results of BC and ACs shown in Figure 4. 7, Figure 4. 8, and Figure 4. 9 (for high-demand scenarios); and in Figure 4. 10, Figure 4. 11, and Figure 4. 12 (for low-demand scenarios) show that the paths with the same length and consisting of only three regions have a fairly similar travel time; thus, avoiding the use of paths that contain five regions. The results in Figure 4. 7 and Figure 4. 8 show that for the first hour of simulation drivers are advised to route through the CBD (i.e., region 4) when they have to select one of regions 2 or 4 as the next region of travel, while they are advised to route through region 3 when they have to select one of regions 3 or 4 as the next region of travel. These results are attributed to the fact that region 2 has more internal demand compared to region 3. Then, around t= 60 min, once CBD becomes congested, BC and AC start guiding drivers through the peripheral regions (i.e., regions 2 or 3 at each case). Only after t=110 min, the network starts to discharge, and the model advises drivers to also select the next region as the CBD. Accordingly, while the AC-based RGs stop routing along the longest path, they only direct travelers on two paths which have the same travel time result and the same RG ratio in the optimization solver.

The notable advantage of AC-based over the non-AC-based controllers is their lower RG fluctuation results as depicted in Figure 4. 7 and Figure 4. 12(parts c, h, g). Unlike the non-AC-based strategies, AC-based RG incorporates drivers choice directly in its model formulation; thus, the resulting consistency in routing choice is manifested as smoother RG ratios, and a sudden
rerouting between alternative routes is observed less than the one detected in BC and FCs scenarios (Moshahedi et al. 2021).

According to the FC model results in Figure 4.7- Figure 4.12 (parts d, e, f), the travel time of the paths with the same length is approximately similar. These findings are attributed to the controller logic that focuses on manipulating either regional speed or path speed or path travel time to be more harmonized; thereby leading to similar path travel time as shown in Figure 4.5 and Figure 4.6. So, because the region speeds in all FC scenarios are closer to each other than those in BC and AC scenarios, travel times for paths with the same length are closer. Under FC, when selecting the CBD region or the peripheral region as the next region, more frequent fluctuation in RG ratios is observed because path travel time is approximately the same, and thus, both paths have an equal chance to be advised by the controller. After t=120 min, when the network is discharged, all regions are in free-flow traffic condition; therefore, all paths can have the same RG ratio. Since path travel time is calculated as the converse of speed multiplied by a constant length value, a similar condition exists under FCpt as well.

4.2. Compliance Rate Sensitivity Analysis

The hypothetical network simulation results under different control scenarios with different compliance rates during a 3-hour control period are summarized in Table 4.4, Table 4.5, Table 4.6, and

Table 4.7. All simulations are performed under the high-demand scenario.

Table 4.4. Simulation results of control models under 100% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>100% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>1324124</td>
<td>887342</td>
<td>881562</td>
<td>878450</td>
<td>88009</td>
</tr>
<tr>
<td>(-)</td>
<td>(-32.99%)</td>
<td>(-33.42%)</td>
<td>(-33.66%)</td>
<td>(-33.54%)</td>
<td></td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>7356.24</td>
<td>4929.68</td>
<td>4897.57</td>
<td>4880.28</td>
<td>4888.94</td>
</tr>
<tr>
<td>Avg. (veh)</td>
<td>(-)</td>
<td>(-32.99%)</td>
<td>(-33.42%)</td>
<td>(-33.66%)</td>
<td>(-33.54%)</td>
</tr>
<tr>
<td>Σ std.</td>
<td>77899.47</td>
<td>15575.44</td>
<td>8501.48</td>
<td>10073.77</td>
<td>7747.48</td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>(-)</td>
<td>(-80.01%)</td>
<td>(-89.09%)</td>
<td>(-87.07%)</td>
<td>(-90.05%)</td>
</tr>
</tbody>
</table>
### Table 4.5. Simulation results of control models under 85% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>85% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>1324124</td>
<td>908962</td>
<td>896965</td>
<td>895931</td>
<td>896862</td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>7356.24</td>
<td>5049.78</td>
<td>4983.13</td>
<td>4977.39</td>
<td>4982.56</td>
</tr>
<tr>
<td>Avg. (veh)</td>
<td>(-)</td>
<td>(-31.35%)</td>
<td>(-32.26%)</td>
<td>(-32.34%)</td>
<td>(-32.27%)</td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>60.05</td>
<td>64.44</td>
<td>64.60</td>
<td>64.61</td>
<td>64.60</td>
</tr>
<tr>
<td>Avg. (Km/hr)</td>
<td>(-)</td>
<td>(7.30%)</td>
<td>(7.57%)</td>
<td>(7.58%)</td>
<td>(7.57%)</td>
</tr>
<tr>
<td>Σ std.</td>
<td>77899.47</td>
<td>17500.58</td>
<td>8932.06</td>
<td>10719.40</td>
<td>8257.47</td>
</tr>
<tr>
<td></td>
<td>(-)</td>
<td>(-77.53%)</td>
<td>(-88.53%)</td>
<td>(-86.24%)</td>
<td>(-89.40%)</td>
</tr>
<tr>
<td>Loop CPU time (s)</td>
<td>0.73</td>
<td>26.03</td>
<td>21.30</td>
<td>19.77</td>
<td>18.52</td>
</tr>
</tbody>
</table>

### Table 4.6. Simulation results of control models under 70% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>70% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>1324124</td>
<td>933264</td>
<td>916458</td>
<td>914305</td>
<td>916137</td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>7356.24</td>
<td>5184.80</td>
<td>5091.43</td>
<td>5079.47</td>
<td>5089.65</td>
</tr>
<tr>
<td>Avg. (veh)</td>
<td>(-)</td>
<td>(-29.52%)</td>
<td>(-30.79%)</td>
<td>(-30.95%)</td>
<td>(-30.81%)</td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>60.05</td>
<td>64.17</td>
<td>64.39</td>
<td>64.41</td>
<td>64.39</td>
</tr>
<tr>
<td>Avg. (Km/hr)</td>
<td>(-)</td>
<td>(6.85%)</td>
<td>(7.22%)</td>
<td>(7.25%)</td>
<td>(7.22%)</td>
</tr>
<tr>
<td>Σ std.</td>
<td>77899.47</td>
<td>20166.43</td>
<td>9385.27</td>
<td>11419.34</td>
<td>8811.18</td>
</tr>
<tr>
<td></td>
<td>(-)</td>
<td>(-74.11%)</td>
<td>(-87.95%)</td>
<td>(-85.34%)</td>
<td>(-88.69%)</td>
</tr>
<tr>
<td>Loop CPU time (s)</td>
<td>0.73</td>
<td>28.76</td>
<td>20.67</td>
<td>19.15</td>
<td>20.20</td>
</tr>
</tbody>
</table>
Table 4.7 Simulation results of control models under 30% compliance rate scenario (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>30% compliance rate</th>
<th>NC</th>
<th>BC</th>
<th>FCRs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>1324124</td>
<td>1130200</td>
<td>997337</td>
<td>993950</td>
<td>997287</td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>7356.24</td>
<td>6278.89</td>
<td>5540.76</td>
<td>5521.94</td>
<td>5540.48</td>
</tr>
<tr>
<td>Avg. (veh)</td>
<td>(-)</td>
<td>(-14.65%)</td>
<td>(-24.68%)</td>
<td>(-24.94%)</td>
<td>(-24.68%)</td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>60.05</td>
<td>62.18</td>
<td>63.47</td>
<td>63.51</td>
<td>63.47</td>
</tr>
<tr>
<td>Avg. (Km/hr)</td>
<td>(-)</td>
<td>(-3.54%)</td>
<td>(5.70%)</td>
<td>(5.75%)</td>
<td>(5.69%)</td>
</tr>
<tr>
<td>Loop CPU time (s)</td>
<td>0.73</td>
<td>33.34</td>
<td>20.64</td>
<td>51.44</td>
<td>21.98</td>
</tr>
</tbody>
</table>

In BC, the TTT or efficiency improvement compared to NC has dropped from 32.99% to 31.35%, 29.52%, and 14.65%, with a decrease in the compliance rate from 100% to 85%, 70%, and 30%. Also, compared to the NC scenario, the average regional speed improvement changed from 7.71% to 7.30%, 6.85%, and 3.54%; with a decrease in the compliance rate from 100% to 85%, 70%, and 30%, respectively. These results show that with a decrease in the compliance rate from 100% to 30%, 55.59% and 54.08%, reduction in efficiency improvements and in average regional speed are observed, respectively.

Compared to NC, in FC models, with a decrease in the compliance rate from 100% to 85%, 70%, and 30%; efficiency improvement decreased from 33.42% to 32.26%, 30.79%, and 24.68% using FCRs; from 33.66% to 32.34%, 30.95%, and 24.94% using FCps; and from 33.54% to 32.27%, 30.81%, and 24.68% using the FCpt model, respectively. Also, the average regional speed improvement compared to NC decreased from 7.85% to 7.57%, 7.22%, and 5.70% using FCRs; from 7.90% to 7.58%, 7.25%, and 5.75% using FCps; and from 7.88% to 7.57%, 7.22%, and 5.69% using the FCpt model. These results show that with a decrease in driver compliance to RG instructions from 100% to 30%, the various FC models sustain similar levels of efficiency improvements and lose only less than 30% of their TTT and speed improvements; hence, there are significantly less reductions compared to those under BC. These findings are due to the fact that
FC-based RG models consistently seek harmonizing traffic condition in the network while incorporating travellers utility as the core of its formulation. Consequently, FC-based RGs inherently induce higher compliance rate and are, therefore, more resilient to lower assumptions of the compliance rate and initial RG values compared to BC.

Furthermore, the effect of the compliance rate on network homogeneity is investigated. This issue is presented by \( \sum \text{std.} \) of accumulation values and \( \sum \text{std.} \) of speed values in Table 4.4, Table 4.5, Table 4.6, and Table 4.7. In the BC model, by decreasing the compliance rate from 100% to 30%, the improvement of \( \sum \text{std.} \) of accumulation compared to NC is decreased from 80.01% to 77.53%, 74.11%, and 25.38%; while the improvement of \( \sum \text{std.} \) of speed compared to NC is decreased from 80.07% to 77.36%, 73.60%, and 25.17%; with a decrease in the compliance rate from 100% to 85%, 70%, and 30%, respectively. These results indicate that the BC model is highly sensitive to the decrease in the drivers compliance rate and thus loses its ability in homogenizing the network (i.e., 68.27% drop in \( \sum \text{std.} \) of accumulation improvement, and 68.58% drop in \( \sum \text{std.} \) of speed improvement).

On the other hand, in the FC-based model, with a decrease in the compliance rate from 100% to 30%, the reduction in \( \sum \text{std.} \) of accumulation compared to NC is dropped from 89.09% to 80.33% using FCrs, from 87.07% to 76.96% using FCps, and from 90.05% to 81.20% using the FCpt model. Also, the reduction in \( \sum \text{std.} \) of speed compared to NC dropped from 90.45% to 80.51% using FCrs, from 87.98% to 76.81% using FCps, and from 91.19% to 81.54% using the FCpt model. These results show that FC models lost only around 10% of their resilience in homogenizing traffic network even with a reduced compliance rate from 100% to 30%. One plausible interpretation of these findings is that FC strategies are utility-fairness centered that focus on balancing the difference in travel time or speed among the paths or regions. Thus, in incorporating driver and/or user perspective in the problem formulation in the form of utilities, FC inherently induces a higher compliance rate; therefore, making it more resilient to the assumption of a low compliance rate and initial value of RG ratios. On the other hand, BC strategies consider only the network perspective in minimizing total travel time, with the assumption of a predefined compliance rate, making it highly vulnerable when the compliance rates are low.
Finally, in the comparison of models in FC strategies under various compliance rates with that of AC strategies, which are independent of compliance rate, it is observed that with decreasing the compliance rate to around 70% and below (while the efficiency improvement of FC-based models deteriorates more than that of AC-based models), FC models still maintain a more fair and homogeneous traffic irrespective of the assumed compliance rate. This finding is important in highlighting the insensitivity of FC-strategies to an assumed compliance rate. On the other hand, we cannot ignore the advantage of AC models with a more realistic compliance rate (70% and less than that) in keeping high levels of TTT efficiency compared to all other models. Also, because the AC model is more realistic without needing predefined assumptions on routing, and integrates drivers responses directly in the control model, it results in a more fair distribution of traffic compared to BC with a realistic compliance rate of RG instructions.

4.2.1. Analysis of the results in terms of regional accumulation for various RG schemes

Compliance rate sensitivity analyses are conducted for the high-demand scenario in the previous section. In BC and FC scenarios, the real route choice ratio is a combination of the optimized RG and drivers route choice decisions estimated as $\gamma \hat{\theta} + (1 - \gamma)\theta'$, where $\gamma$ is the drivers compliance rate to the optimized route ratios instructed by the RG controller. In this sensitivity analysis, the value of 100%, 85%, 70%, and 30% are considered for the drivers compliance rate. Figure 4. 13 demonstrates the accumulation states of regions under different drivers compliance rates. In all scenarios, all regions are emptied at the same time at $t=120$ min.
Figure 4.13. Regional accumulations under various compliance rate scenarios: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt.
In Figure 4. 13 (a), the accumulation values of regions do not change for the case of NC for different compliance rates, because there is no controlled and optimized RG in the NC model. In the BC and FC models, the regional accumulation values increase as the compliance rate decreases from 100% to 30% such that some regions get congested. Especially under BC, the maximum accumulation of CBD has increased from 2,915 vehicles to 3,108; 3,357; and 5,016 vehicles as the drivers compliance rate decreases from 100% to 85%, 70%, and 30%, respectively. However, the fairness-based control was shown to be more resilient to a low compliance rate. Accordingly, the maximum accumulation is increased from 2,297 vehicles to 2,336; 2,384; and 2,864 vehicles under FCrs; from 2,298 vehicles to 2,357; 2,434; and 2,909 vehicles under FCpt; and from 2,394 vehicles to 2,468; 2,547; and 3,148 vehicles under FCps; with a decrease in compliance rate from 100% to 85%, 70%, and 30%, respectively. Similarly, the results show that the maximum accumulation increases by 25%, 27%, and 32% under FCrs, FCpt, and FCps models, respectively, by reducing the compliance rate from 100% to 30%, while the increase was as much as 72% under BC. It is clear that fairness models, especially FCpt, alleviate the traffic conditions in CBD well by diverting traffic from CBD and reducing the maximum accumulation level even with a low compliance rate. Fair models outperform BC even when the compliance rate is low since these models consider driver goals in routing through use of utility functions.

4.2.2. Analysis of the results in terms of regional speed for various RG schemes

Figure 4. 14 demonstrates the regional speed of the network under different compliance rates of RG optimization results.
Figure 4.14. Regional speed under various compliance rate scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt
In the NC results shown in Figure 4. 14 (a), the region speeds do not change by changing the compliance rate, because there is no controlled and optimized RG in the no-control model. In BC and FC models, the values of regional speed decrease by the decrease in the drivers compliance rate from 100% to 30%. Also, the $\sum std$ of speed increase (presented in detail in Table 4. 4, Table 4. 5, Table 4. 6, and Table 4. 7.) by the decrease in the compliance rate. Under BC, the minimum speed belonging to CBD decreased from 35.48 Km/hr to 32.33, 28.41, and 9.35 Km/hr by a decrease in the compliance rate from 100% to 85%, 70%, and 30%, respectively. In fairness models, the minimum speed is decreased from 45.91 Km/hr to 45.23, 44.43, and 36.31 Km/hr in the FCrs model; from 45.88 Km/hr to 44.89, 43.57, and 35.58 Km/hr in the FCpt model; and from 44.26 Km/hr to 43.01, 41.66, and 31.68 Km/hr in the FCps model; by a decrease in compliance rate from 100% to 85%, 70%, and 30%, respectively. These results show that the minimum speed belonging to the CBD decreased by 21%, 23%, and 29% under FCrs, FCpt, and FCps models, respectively, by reducing the compliance rate from 100% to 30%. However, this decrease was 74% under BC. It is clear that fair control scenarios, specially FCrs and FCpt, manage the traffic conditions well by maintaining the regional speed at the acceptable level even under low drivers compliance rates.

Using RG to provide advice in a network with drivers who do have very low level of compliance rate, is not a logical symptom. Thus, we assumed the lowest possible compliance rate of 30% and ran the sensitivity analysis for the compliance range between 100% to 30%. However, to see what theoretically can happen for regional speed under a lower compliance rate, the results are presented for lower compliance rates that are further decreased from 30% to 0% and the results are depicted in Figure 4. 15 for BC and NC models and only for FCpt as the performing FC-based model.
The results show that with a compliance rate of 0%, RG models work similar to NC model, which is intuitive as drivers do not follow the RG advice. While it is visible from Figure 4. 14 and Figure 4. 15 that BC model starts to lose its ability to homogenize traffic when the compliance rate falls under 70%, the FCpt model can still sustain its ability to homogenize traffic until the compliance rate falls below 20%. These results depict the superior performance of the FC-based model as it is less sensitive to the assumption of compliance rate and initial value of RG compared to the BC.

4.2.3. Analysis of the results in terms of regional RG ratio and path travel time for various RG schemes

Figure 4. 15, Figure 4. 16, Figure 4. 18, and Figure 4. 20 show regional RG ratio under various compliance rate scenarios for control models a) NC, b) BC, c) FCrs, d) FCps, and e) FCpt) for OD pairs (5, 2), (3, 1), and (6, 3), respectively. Also, Figure 4. 17, Figure 4. 19, and Figure 4. 21 show path travel time.
time under various compliance rates for three ODs. These figures confirm that RG ratio graphs and path travel time are consistent. It is observed that with decreasing one path’s travel time, the routing ratio for that path increases, and the controller advises drivers to choose that path for reaching their destination.

Figure 4. 16 and Figure 4. 17 show the results of the compliance rate sensitivity analysis for OD pair (5, 2). Other sets of Figure 4. 18 and Figure 4. 19 show the results for OD pair (3, 1), and Figure 4. 20 and Figure 4. 21 for those of OD pair (6, 3).
Figure 4.16. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 5 to destination 2 with paths \{[5 3 2];[5 4 2];[5 6 7 1 2]\}
A) 100% compliance rate  B) 85% compliance rate  C) 70% compliance rate  D) 30% compliance rate

Figure 4. 17. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 5 to destination 2 with paths \{[[5 3 2]];[[5 4 2]];[[5 6 7 1 2]]\}.
A) 100% compliance rate  B) 85% compliance rate  C) 70% compliance rate  D) 30% compliance rate

Figure 4.18. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 3 to destination 1 with paths \{[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]\}
Figure 4. 19. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 3 to destination 1 with paths {[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]}
Figure 4. 20. Regional RG ratios under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 6 to destination 3 with paths \{[[6 5 3]]:[[6 4 3]]:[[6 7 1 2 3]]\}
Figure 4.21. Path travel time under various compliance rates scenario: A) 100%, B) 85%, C) 70% and D) 30%, by various control models: a) NC, b) BC, c) FCrs, d) FCps, e) FCpt for origin 6 to destination 3 with paths \{\{6 5 3\};\{6 4 3\};\{6 7 1 2 3\}\}
Figure 4. 16 shows that RG ratios under BC- and FC-based models result in less fluctuation in the network with low drivers compliance rates compared to the network with a high compliance rate. These findings are attributed to the fact that with a decrease in compliance rate, drivers prefer to choose CBD as their next region; therefore, travel time of paths traversing through CBD increases, and travel time graphs would be spaced apart. On the other hand, the controllers consistently advise drivers to divert away from the CBD and choose the peripheral region as the next region, which explains the decreased fluctuation of RG. Additionally, under BC corresponding to a low compliance rate, CBD gets congested, and as a result, the path that passes through CBD (e.g., path 542) has a higher travel time than the longer path that contains five regions (e.g., path 56712) for a short period of simulation time. During that period of time, BC guides drivers through a path that contains five regions (i.e., the green line in part (D, b) in Figure 4. 16, Figure 4. 18, and Figure 4. 20). On the other hand, in Figure 4. 17, it is observed that FC-based schemes address this issue by stabilizing the travel time for paths of the same length by keeping them close to each other. It shows that FC control models lead to more fair and more homogeneous traffic in the network even when the compliance rate is low. The reason is related to the objective function of models. In BC, the model considers only the network benefit with minimizing total travel time, but fair models consider driver aims when modelling by viewing the problem from the perspective of the driver and other users. That makes the fair models more successful compared to BC when the compliance rate is low.

After t= 110 min, the network becomes empty, and the controllers advise drivers to select their path through the CBD. After around t= 120 to 130 min, when the network is completely empty, the BC model that focuses only on minimizing the summation of accumulation gives a value to each path RG ratio based on their initial value of RG for the optimization model at that time, which is equal to the RG in the previous time when compliance rate is 100%. Hence, RG ratios are fixed on the value at the time the network is completely discharged. On the other hand, in FC models, which work based on speed and travel time (i.e., inverse of speed), all paths have the same RG ratio because all regions have free-flow speeds, and there is no restriction for maximizing speed until it reaches the free-flow speed. When reducing the compliance rate to 30%, both BC and FC models tend to suggest the longest path that contains five regions (after t= 120–130 min). These results are attributed to the fact that with decreasing compliance rate, the real
route choice is achieved from a combination of the previous time optimized RG and drivers routing choices based on the logit model. Because the logit model assigns a value for each path based on their travel time, all paths (even the paths which contain five regions) have a non-zero value as their RG ratio. Also, graphs tend to be more like the NC model after \( t = 120 \text{ min} \).

### 4.3. MFD Parameters Sensitivity Analysis

Sensitivity analysis of MFD parameters is conducted under the same demand profile as Figure 4. 2 and with assumed 100% compliance rate. In this sensitivity analysis, the values of \( \alpha_I = [1.02, 0.99, 1.04, 1, 0.98, 0.97, 0.96] \) and \( \beta_I = [0.96, 1.02, 0.98, 1, 1, 0.97, 1.03] \) are considered for MFD parameters of regions [1, 2, 3, 4, 5, 6, and 7] in this network which is within \( \pm 5\% \) of the network observed in downtown Yokohama (Geroliminis and Daganzo 2008).

To compare the various models’ performances under the assumption of same and different MFD parameters scenarios, the \( \sum std. \) of accumulation and \( \sum std. \) of speed were calculated for normalized accumulation and normalized speed. Since in different MFD parameters scenarios the maximum speed and jam accumulation is different for regions, average regional speed and accumulation are normalized between 0 and 1 by dividing them by their maximum values.

Table 4. 8. Simulation results of control models for the network with same region MFD parameters (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>Same MFD Parameters</th>
<th>NC</th>
<th>BC</th>
<th>ABC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>921797</td>
<td>778204 (-15.58%)</td>
<td>791178 (-14.17%)</td>
<td>778333 (-15.56%)</td>
<td>776118 (-15.80%)</td>
<td>776929 (-15.72%)</td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>5121.09</td>
<td>4323.35 (-15.58%)</td>
<td>4395.43 (-14.17%)</td>
<td>4324.07 (-15.56%)</td>
<td>4311.77 (-15.80%)</td>
<td>4316.27 (-15.72%)</td>
</tr>
<tr>
<td>Avg. of accumulation (veh)</td>
<td>2.51</td>
<td>0.96 (-61.91%)</td>
<td>1.70 (-32.53%)</td>
<td>0.78 (-69.14%)</td>
<td>0.78 (-68.99%)</td>
<td>0.69 (-72.60%)</td>
</tr>
<tr>
<td>( \sum std. ) of normalized accumulation</td>
<td>5.23</td>
<td>1.78</td>
<td>3.15</td>
<td>1.32</td>
<td>1.38</td>
<td>1.19</td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>64.34</td>
<td>65.87</td>
<td>65.67</td>
<td>65.88</td>
<td>65.90</td>
<td>65.90</td>
</tr>
<tr>
<td>Avg. of speed (Km/hr)</td>
<td>5.84</td>
<td>65.87 (2.37%)</td>
<td>65.67 (2.05%)</td>
<td>65.88 (2.38%)</td>
<td>65.90 (2.42%)</td>
<td>65.90 (2.41%)</td>
</tr>
<tr>
<td>( \sum std. ) of normalized speed</td>
<td>5.23</td>
<td>1.78 (66.06%)</td>
<td>3.15 (39.82%)</td>
<td>1.32 (74.75%)</td>
<td>1.38 (73.68%)</td>
<td>1.19 (77.26%)</td>
</tr>
<tr>
<td>Loop CPU time (s)</td>
<td>0.68</td>
<td>22.73</td>
<td>810.08</td>
<td>23.07</td>
<td>19.35</td>
<td>21.77</td>
</tr>
</tbody>
</table>
### Table 4.9. Simulation results of control models for the network with different region MFD parameters (the percentage improvements in each measure compared to NC is shown in parentheses)

<table>
<thead>
<tr>
<th>Different MFD Parameters</th>
<th>NC</th>
<th>BC</th>
<th>ABC</th>
<th>FCrs</th>
<th>FCps</th>
<th>FCpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total travel time (min)</td>
<td>942384</td>
<td>790782</td>
<td>803297</td>
<td>787068</td>
<td>785683</td>
<td>786806</td>
</tr>
<tr>
<td>Accumulation (veh)</td>
<td>5235.47</td>
<td>4393.23</td>
<td>4462.76</td>
<td>4372.60</td>
<td>4364.90</td>
<td>4371.14</td>
</tr>
<tr>
<td>Avg. of accumulation (veh)</td>
<td>(--)</td>
<td>(-16.09%)</td>
<td>(-14.76%)</td>
<td>(-16.48%)</td>
<td>(-16.63%)</td>
<td>(-16.51%)</td>
</tr>
<tr>
<td>(\sum) std. of normalized accumulation 2.68</td>
<td>(1.15)</td>
<td>(1.81)</td>
<td>0.76</td>
<td>0.83</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Speed (Km/hr)</td>
<td>63.68</td>
<td>65.31</td>
<td>65.11</td>
<td>65.37</td>
<td>65.39</td>
<td>65.37</td>
</tr>
<tr>
<td>Avg. of speed (Km/hr)</td>
<td>(--)</td>
<td>(2.56%)</td>
<td>(2.25%)</td>
<td>(2.65%)</td>
<td>(2.67%)</td>
<td>(2.66%)</td>
</tr>
<tr>
<td>(\sum) std. of normalized speed 5.83</td>
<td>(2.07)</td>
<td>3.47</td>
<td>1.26</td>
<td>1.43</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>Loop CPU time (s)</td>
<td>0.57</td>
<td>30.10</td>
<td>989.89</td>
<td>18.02</td>
<td>20.10</td>
<td>19.84</td>
</tr>
</tbody>
</table>

As illustrated in the low-demand scenario with the same MFD parameters in Table 4.3, the FCpt model was more successful than other models in improving the efficiency and average regional speed while also homogenizing the distribution of traffic by resulting in more even regional average speeds and decreasing the \(\sum\) std. of normalized speeds at the same time.

According to the results of Table 4.8 and Table 4.9 under NC scenario, the multi-region network with different MFDs has a higher average accumulation and lower average speed compared to the same network with the same MFDs. All control models (i.e., BCs, ACs, and FCs) improve over the no control results in both scenarios. All control models improve efficiency in the network with different MFDs compared to the network with the same regional MFDs. Improvement in the \(\sum\) std of normalized speed compared to NC was reduced by 1.61% using BC while it was increased by 0.6% using ABC, 3.58% using FCrs, 1.82% using FCps, and 2.53% using FCpt in the network with different MFD parameters. Thus, in a multi-region network with various MFDs, BC performs poorly in terms of reducing the \(\sum\) std of normalized speed while ABC, and especially FC-based models, suppress the \(\sum\) std of normalized speed, and thus, result in more homogeneous traffic.

In conclusion, fair-based models are more appropriate to implement in a network with various MFDs as the utility-based formulation of the proportional fairness is tailored to the
individual needs of each region. Similar to the previous sensitivity analysis, the FCpt outperforms other control models in enhancing homogeneity of speeds and accumulation.

4.3.1. Analysis of the results in terms of regional accumulation for various RG schemes

Figure 4. 22 shows the accumulation states of regions under the same and different MFD parameters. In all scenarios, all regions are emptied simultaneously at t= 120 min.
Figure 4. 22. Regional accumulations under various regions’ MFD parameters: A) same regions’ MFD parameter, B) different regions’ MFD parameter, by various control models: a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt
In the NC model of Figure 4.22(a), more regions experience congested conditions when they have different MFDs compared to the network with the same MFDs. When MFD parameters are the same, only regions 2 and 5 experienced congested conditions, while with changing MFD parameters, regions 2, 5, and 6 experienced congested conditions because both parameters for region 6 were less than 1 as explained in what follows. Based on the definition of outflow \((G_t(N_t(t)))\) at Eq. (4.1), and definition of critical accumulation and free-flow speed for regions with various MFD parameters \((N_t^{crt}=2500, \beta_t(veh), v_{free_t} = 70, \alpha_t (Km/hr))\), it is observed that \(G_t\) is directly proportional to the value of \(\alpha_t\), and inversely proportional to the square of \(\beta_t\). When \((\alpha_t, \beta_t)\) is set to \((0.97,0.97)\) for region 6, then \(N_6^{crt}\) decreases and \(G_6\) increases, leading to an accumulation increase and, thus, a congested condition. These results show that the maximum region accumulation increased by 5.02%, 1.34%, 4.30%, 3.20% and 2.88% under BC, ABC, FCrs, FCps, and FCpt models, respectively, with changing regions MFD parameters from the same value to different values.

4.3.2. Analysis of the results in terms of regional speed for various RG schemes

Figure 4.23 shows the regional average speed values under the same and different MFD parameters.
Figure 4.23. Regional speed under various regions’ MFD parameters: A) same regions’ MFD parameter, B) different regions’ MFD parameter, by various control models: a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt
In Figure 4.23, it is observed that in cases with different MFDs, the fluctuation of the average speed among regions increases. In region 6, it is observed that the average normalized regional speed decreases to a value close to that of CBD; this confirms the accumulation result earlier observed in Figure 4.22 and explained based on the formulation for region 6’s MFD parameters. The detailed results of the mean and summation of the standard deviation of speed for various models are presented in Table 4.8 and Table 4.9. These results show that minimum regional normalized speeds decreased by 4.48%, 1.38%, 1.48%, 1.98% and 1.88% under BC, ABC, FCrs, FCps, and FCpt models, respectively, with changing region MFD parameters from the same value to different values.

Also, the table results show insignificant changes in the average of normalized speeds, while the $\sum \text{std}$ of normalized speed changes with changing region MFD parameters from the same value to different values. Compared to NC, the improvement of the $\sum \text{std}$ of normalized speed has reduced 1.61% using BC, but it has increased 0.6% using ABC, 3.58% using FCrs, 1.82% using FCps, and 2.53% using FCpt in the network with different MFD parameters compared with networks with same MFDs.

4.3.3. Analysis of the results in terms of regional RG ratio and path travel time for various RG schemes

Figure 4.24, Figure 4.25, and Figure 4.26 show regional RG ratios and path travel times for two types of networks: A) a multi-region network with similar MFDs and B) a multi-region network with different MFDs. These figures show regional RG ratios and path travel times for uncontrolled and controlled scenarios [a) NC, b) BC, c) AC, d) FCrs, e) FCps, and f) FCpt] for various OD pairs from 5 to 2, from 3 to 1, and from 6 to 3, respectively.
Figure 4. 24. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 5 to destination 2 with paths $\{[5\ 3\ 2];([5\ 4\ 2]);([5\ 6\ 7\ 1\ 2])\}$
Figure 4. 25. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 3 to destination 1 with paths \{[[3 2 1]];[[3 4 1]];[[3 5 6 7 1]]\}
Figure 4. 26. Regional RG ratios and path travel times for both the “same” and “different” regions’ MFD parameters scenarios by using various models, a) NC, b) BC, c) ABC, d) FCrs, e) FCps, f) FCpt for origin 6 to destination 3 with paths {{[6 5 3]}; {[6 4 3]}; {[6 7 1 2 3]}}
These figures confirm that RG ratio graphs and path travel times are consistent. In other words, with a decrease in travel time on a given path, its guidance ratio increases and vice versa.

In Figure 4.24, it is observed that RG ratios in the network with different regional MFD parameters result in less fluctuations compared to the network with the same regional MFD parameters, especially in FC models. Based on the definition of speed \( V_I\left(N_I(t)\right) = \frac{G_I(N_I(t))}{N_I(t)} \cdot L_I(t) \), MFD at Eq. (4.1), critical accumulation \( N_I^{crit} = 2500 \beta_I \) (veh), and free flow speed \( v_{free_I} = 70 \alpha_I \) (Km/hr) for regions with various MFD parameters, it is observed that speed has a direct relation with \( \alpha_I \) and an inverse relation with the second order of \( \beta_I \). In a multi-region network with dissimilar MFDs, average regional speed values are dramatically different when compared to a network with similar MFDs. In such networks, the variation of speeds increases as well. Then, the controllers advise drivers to select the neighboring region which has a higher speed than other neighboring regions. In the network with various MFD parameters, because the difference of region speed increase compared to same regions network, optimization models advise drivers to drive in a region with higher speed, rather than dividing vehicles between both paths with the same situation. For instance, Figure 4.23 shows that the difference of speeds in regions 3 and 4 increases in the network with different MFD parameters, compared to the difference of speeds in regions 3 and 4 in the network with the same MFD parameters. So, Figure 4.24 shows that both of the next regions 3 and 4 are selected in the “same MFD parameters” scenario, while in the “different MFD parameters” scenario, the fairness controllers advise drivers to choose a path in which driving with higher speed is possible, and the RG ratio of path 532 is more than 542 the entire time, and fluctuation is decreased.

In Figure 4.26, it is observed that only the path with the next region 5 is selected for the first 50 min; then, both short paths are advised to be selected. As depicted in Figure 4.23, in which at both same and different MFD parameters scenarios in FC models, the region 5’s speed graph shows higher value than CBD’s speed curve, during the 50 min of simulation, while after this time the two curves fall in a closer range. So, in the network with different MFD parameters, rerouting on paths with neighboring regions 4 and 5 starts after time = 50 min; thus causing more fluctuation in RG graphs. All the figures’ graphs match and approve each other.
5. Conclusion

This final chapter summarizes the main achievements of the thesis and presents an outlook on issues in MFD-based RG control of large-size urban networks that need to be addressed.

5.1. Research Contributions and Findings

While a large body of research exists on MFD-based RG control models aimed at maximizing total network outflow, not enough effort has been put into the development of fair RG control schemes. This study aims to develop such a scheme that simultaneously addresses both efficiency and fairness issues. This thesis develops new control schemes that create a balanced trade-off between efficiency and fairness when devising RG control strategies by using various fairness-centered concepts such as proportional fairness and anticipatory control in an MPC framework. All control schemes were modelled in the MPC framework and compared to a benchmark RG model. For solving the MPC problem, the CasADi solver, an open-source nonlinear optimization tool that facilitates rapid and efficient implementation of different methods for numerical optimal control, was implemented in MATLAB. Intensive numerical analysis was conducted to examine the performance of the developed algorithms under different demand profiles, MFD parameters and route compliance rates.

5.1.1. Introducing proportionally fair RG controller for a large-size urban network

This thesis introduces the novel proportionally fair RG control scheme for a multi-region urban network based on MFD. The emphasis of this study is on user utilities while creating a balanced trade-off between network efficiency and fairness. The proportionally fair RG control schemes are modelled in an MPC framework with three different objective functions based on
either maximizing average regional speed, maximizing average path speed, or minimizing average path travel time, with the aim of fair distribution of benefits and externalities. The performance of the developed models aligns with the main concept of proportionally fair RG, without sacrificing efficiency, and exhibits an increased level of fairness in the proposed solutions.

Among FC-based control scenarios, FCpt shows an outstanding performance since it not only improves efficiency, but also enhances fairness more than the other control scenarios. The reason behind FCpt’s outstanding performance is that this controller is modelled from the driver perspective of reducing the path trip time. FCpt directly optimizes path travel time by managing the number of vehicles transferring through that path and other alternative paths, and thus, ensures fairness between different paths in the network. FCr and FCps yields the second-best performances in terms of efficiency and fairness. FCps is based on path speed that is estimated from the average speed of regions belonging to a particular path, and it works very similarly to the FCr control scenario. In other word, in FCps, the regions that are traversed more often by the paths are incorporated multiple times in the objective function. In contrast to the FCpt and FCps scenarios, in which objective function is defined based on path, the FCr objective function is defined based on region. Therefore, one of the key features of FCr is that more weights can be assigned to a given region to prioritize it over other regions in the network. Herein, the weight of region 4 (i.e., the CBD) is considered to be 20% higher than other regions to further alleviate its congestion by discouraging routing through it.

Generally, it is concluded that FC models are superior to BC models in terms of fairness, even when the drivers compliance rate is relatively low. Since fairness is measured in terms of variation of speed in regions, these models are expected to enhance network homogeneity. One plausible interpretation of these findings is that FC strategies are users’ utility centered that focus on balancing the difference in travel time or speed among the paths or regions. Thus, in incorporating driver and/or user perspectives to the problem formulation, FC inherently induces higher compliance rate; thus, leading it to be more resilient to the assumption of low compliance rate. Also, FC models are more appropriate for use in a multi-region network with various MFDs, while BC falls short in homogenizing in such network.
5.1.2. Modelling an anticipatory RG controller for a large-size urban network

This thesis developed an MFD-based anticipative RG control approach in MPC framework for a network-wide control of an urban network. The anticipatory RG control is modelled as a two-level optimization model in an online optimization framework and directly incorporates road user routing behaviour in the control model. The anticipatory RG controller is then examined by replacing the basic objective function with the proportional fairness objective functions introduced earlier. At the heart of the anticipatory control concept is the incorporation of user behaviour indigenously as part of the control framework to lead to a more optimum solution and more consistent routing schemes.

With the exception of an assumption of 100% compliance rate for non-AC models, the efficiency of the network measured in terms of TTT is consistently improved under AC compared to BC, but not to FC. When the compliance rate decreases to a realistic value around 70% and below, the efficiency of the AC model is more than the efficiency of FC models, because the AC model considers drivers behaviour directly into the control model. However, FC models have more fair and more homogeneous traffic at any compliance rate. Another advantage of anticipatory controllers over the non-AC controllers is their low RG fluctuation, because AC models compromise between controller and drivers. Therefore, under ACs, which account for drivers routing decisions directly in the control model, a smoother pattern for RG ratios is achieved, and sudden rerouting between alternative routes detected in BC and FCs scenarios is less observed.

An additional observation is that the effect of instruction of AC on the results is not highly dependent on the optimization objective function. So, the ABC, AFCpt, and AFCrs results are somehow similar.

5.2. Recommendations for Future Work

Several possible future directions are listed as follows:

- The proportionally fair RG scheme can be compared to other routing schemes such as social routing in terms of both efficiency and fairness performance.
• The proportional fairness concept can be adopted for other MFD-based demand management strategies, such as congestion pricing or for integrated traffic management strategies such as simultaneous RG and perimeter control.

• MFD-based multi-level traffic management schemes can be developed that target fairness locally for sub-regions and globally at the regional level.

• Real-world traffic networks can be used to assess the applicability of the fair RG control schemes.


diagram approach.” University of Calgary.


Transportation Research Board, 2315, 153–163.


The list of assumed path for hypothetical network is presented at next OD matrix:

<table>
<thead>
<tr>
<th>Destination Region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>12</td>
<td>123</td>
<td>143</td>
<td>16753</td>
<td>14</td>
<td>145</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>*</td>
<td>23</td>
<td>24</td>
<td>245</td>
<td>235</td>
<td>2356</td>
</tr>
<tr>
<td>3</td>
<td>321</td>
<td>34</td>
<td>35671</td>
<td>32</td>
<td>*</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>*</td>
<td>45</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>541</td>
<td>5671</td>
<td>5321</td>
<td>532</td>
<td>542</td>
<td>56712</td>
<td>53</td>
</tr>
<tr>
<td>6</td>
<td>641</td>
<td>671</td>
<td>65321</td>
<td>642</td>
<td>6712</td>
<td>6532</td>
<td>653</td>
</tr>
<tr>
<td>7</td>
<td>71</td>
<td>742</td>
<td>76532</td>
<td>743</td>
<td>7123</td>
<td>7653</td>
<td>74</td>
</tr>
</tbody>
</table>

Appendix A