

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

EXPLORING THE ROLES OF AUTOMATED VEHICLES IN
SYSTEM OPTIMAL FOR TRAFFIC ASSIGNMENT
PROBLEMS

JINXIAO DU

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering

Exploring the Roles of Automated Vehicles in System
Optimal for Traffic Assignment Problems

Jinxiao Du

A thesis submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

March, 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Jinxiao Du)

Abstract

The rise of automated vehicles (AVs) presents significant opportunities to improve transportation efficiency, safety, and sustainability. By leveraging advanced sensing, computing, and communication technologies, AVs have the potential to reduce congestion and improve the network-wide performance. Besides, traffic assignment model plays a crucial role in routing AVs efficiently in a traffic network. System optimal (SO) is one of the important routing principles in traffic assignment problems, which quantifies the best possible performance of a traffic network and provides policy implications and operation strategies for decision-makers. The rapid development of automated technology reveals the possibility to control all AVs to achieve SO in a fully automated environment. Therefore, this research aims to explore the roles of AVs in SO for traffic assignment problems in different aspects: vehicles, platform and network. Besides, the traffic assignment problem can be classified into two categories: static traffic assignment (STA), which assumes constant travel demand and network conditions, and dynamic traffic assignment (DTA). It is suitable to explore the equilibrium when AV control rules and routing decisions are fixed over the analysis horizon. On the other hand, DTA models capture the relationship between dynamic route choice behaviors and transportation network states. Therefore, depending on the nature of the strategies of AVs to achieve SO, different types of traffic assignment models are employed to ensure optimal routing and system performance. Furthermore, the non-uniqueness of the solution of SO in traffic assignment problems implies that the SO could be achieved by multiple strategies in each aspect and it inspires us to develop the second criterion of optimal strategy except SO in each aspect.

In the vehicles aspect, this study develops the headway control framework in a fully automated road network, as we believe headway of AVs is another influencing factor to traffic dynamics in addition to conventional traveler behaviors (*e.g.* route and departure time choices).

Specifically, this study aims to search for the optimal time headway between AVs on each link that achieves the network-wide system optimal dynamic traffic assignment (SO-DTA). To this end, the headway-dependent fundamental diagram (HFD) and headway-dependent double queue model (HDQ) are developed to model the effect of dynamic headway on roads, and a dynamic network model is built. It is rigorously proven that the minimum headway could always achieve SO-DTA, yet the optimal headway is non-unique. Motivated by these two findings, this study defines a novel concept of maximin headway, which is the largest headway that still achieves SO-DTA in the network. Mathematical properties regarding maximin headway are analyzed and an efficient solution algorithm is developed. Numerical experiments on both a small and large network verify the effectiveness of the maximin headway control framework as well as the properties of maximin headway. This study sheds light on deriving the desired solution among the non-unique solutions in SO-DTA and provides implications regarding the safety margin of AVs under SO-DTA.

In the platform aspect, this thesis proposes a publicly-owned centralized platform (POCP) for shared automated vehicles (SAVs) that connects travelers, traffic network companies (TNCs), and parking spaces. It provides a novel insight for the regulation of government to save social costs by adjusting travelers' behaviors with spatial-temporal trip subsidies. In the proposed POCP, (i) government operates this centralized platform with the objective of minimizing social cost, collecting and distributing necessary information (*i.e.*, trip orders, trip fees, discounts, parking vacancies) to corresponding users; (ii) travelers request and pay the trip order in the centralized platform and their trip order choices may be associated with the trip fee and discount; (iii) TNCs receive trip orders from the platform and provide mobility services by routing SAVs to maximize the profit. This study formulates the operation of POCP as a bi-level programming, where the government aims at minimizing the social cost by providing travelers with spatial-temporal discounts in the upper-level problem, and TNC aims at maximizing the profit given travel demand from the upper level as a DTA in lower level problem. This study proves that the proposed bi-level programming could be equivalently decomposed into two sub-problems with linear programming to solve for the minimum social cost and minimum subsidy amount, respectively. To validate our proposed model and solution algorithms, this study conducted numerical experiments in both small and real-world large networks. The results reveal substantial savings in social costs, equivalent to 110.38% of the subsidies spent in the small network and 121.96% in

the large network. These findings underscore the efficacy of the proposed POCP model, emphasizing the potential of spatial-temporal subsidies in reducing traffic congestion and improving social welfare.

In the network aspect, considering the route choice for AVs referring to the routes selected for the vehicle (e.g., via app, voice command, or a central controller), this study propose the concept of mixed sub-network equilibrium (MSNE) that divides a traffic network into two sub-networks: the user equilibrium sub-network (UESN), where AVs are controlled by users to follow the user equilibrium (UE) principle for seeking the route with minimum travel cost, and the system optimal sub-network (SOSN), where AVs are centrally controlled to minimize total travel cost by following the SO principle. This hybrid network setting addresses the challenge of applying system-wide control across large-scale networks by focusing centralized control only on critical regions. A key contribution of this research is the introduction of the critical size of the system optimal sub-network (CS-SOSN), which quantifies the minimum size of the SOSN required to achieve system-wide optimal performance. To solve for the CS-SOSN, this study propose a novel method of adaptive variable neighborhoods search (AVNS) algorithm. Through theoretical analysis and numerical simulation in three networks, this study demonstrate that the proposed framework effectively reduces total travel cost while minimizing the need for extensive centralized control. This study further validate the proposed CS-SOSN model on the real Hong Kong network and the results show that, in real-world traffic operations, only 44.17% of the network needs to be controlled to bring the whole network to the state of SO, without requiring full control of the whole network. The CS-SOSN model provides an effective method for managing the traffic network, reducing congestion, and improving performance in fully AV environment. This research offers practical insights for managing AVs in traffic networks by optimizing the balance between decentralized and centralized control strategies.

Publications arising from the thesis

Peer-reviewed journal papers

- **Du, J.**, & Ma, W. (2024).^{*} Maximin Headway Control of Automated Vehicles for System Optimal Dynamic Traffic Assignment in General Networks. *Transportation Research Part E: Logistics and Transportation Review*, 188, 103628.
- **Du, J.**, Yu, S., Yang, B., He, J., & Ma, W. (2025). Mixed Sub-Network Equilibrium of Automated Vehicles in General Traffic Networks. *Transportation Research Part B: Methodological (Under Review)*.
- **Du, J.**, Ma, W., Ge, Q.^{*}, & Fukuda, D.^{*} (2025). Optimal Dynamic Subsidy Strategy Considering the Competition between the Privately-Owned and Publicly-Owned Ride-Hailing Platform. *Transportation Research Part C: Emerging Technologies (Under Review)*.
- Yu S., **Du, J.**^{*}, Cruz, A., Schmöcker, Jan., & Ma, W. (2025). Multi-Class Traffic Mass Evacuation Considering Non-Compliance Behavior. *Transportation Research Part A: Policy and Practice (Minor Revision)*.
- **Du, J.**, Yu, S., He, J., Yang, B., Sun, X., & Ma, W.^{*} (2025). Designing the Publicly-Owned Centralized Platform for Ride-Hailing Services with Shared Automated Vehicles: A System Optimal Dynamic Traffic Assignment Approach. *Transportation Research Part E: Logistics and Transportation Review (Under Review)*.
- Yang, B., **Du, J.**, He, J., An, S., & Ma, W.^{*} (2025). Estimating Real Demand Using a Flipped Queueing Model: Insights from Shared Micro-Mobility Services. *Transportation Research Part E: Logistics and Transportation Review (Under Review)*.

- Yang, B., He, J., **Du, J.**, An, S., & Ma, W.* (2025). Real Demand Estimation for Bike-Sharing Systems: A Variational Inference Approach Incorporating Spatiotemporal and Migration Correlations. *IEEE Transactions on Intelligent Transportation Systems (Under Review)*.

Conferences proceedings

- He J., **Du, J.**, & Ma, W.* (2024). Preventing Dimensional Collapse in Self-Supervised Learning via Orthogonality Regularization. *Advances in Neural Information Processing Systems*, 37, 95579–95606.
- He J., **Du, J.**, Xu, S., & Ma, W.* (2024). Preventing Model Collapse in Deep Canonical Correlation Analysis by Noise Regularization. *Advances in Neural Information Processing Systems*, 37, 82375–82410.
- Wang X., **Du, J.**, Peng, Y., & Ma, W.* (2025). From Sequential to Recursive: Enhancing Decision-Focused Learning with Bidirectional Feedback. *The Fortieth AAAI Conference on Artificial Intelligence (Rebuttal in Phase 2)*.

Conferences presentations

- **Du, J.**, Chen, Z., He, Jun., & Ma, W.* (2024). Sub-Network Control of Automated Vehicles in General Traffic Networks, *The World Transportation Convention*, 26-29 June 2024, Qingdao, China.
- **Du, J.**, He, J., & Ma, W.* (2024). Designing the Publicly-Owned Centralized Platform for Ride-Hailing Services with Shared Automated Vehicles, *The World Transportation Convention*, 26-29 June 2024, Qingdao, China.
- **Du, J.**, Wan, Z., He, Jun., & Ma, W.* (2024). Optimal Subsidy Strategies to Reduce Pollution in the Adoption of Electric Vehicles, *The Service Operations and Logistics, and Informatics*, 22-24 June 2024, Macau, China.
- **Du, J.**, & Ma, W.* (2024). Designing the Publicly-Owned Centralized Platform for Ride-Hailing Services with Shared Automated Vehicles, *103rd Transportation Research Board (TRB) Annual Meeting*, 7-11 January 2024, Washington, D.C..

- **Du, J.,** He, J., & Ma, W.* (2023). Designing the Publicly-Owned Centralized Platform for Ride-Hailing Services with Shared Automated Vehicles, *The 27th International Conference of Hong Kong Society for Transportation Studies*, 11-12 December 2023, Hong Kong, China.
- **Du, J.,** & Ma, W.* (2022). Maximin Headway Control of Automated Vehicles for System Optimal Dynamic Traffic Assignment in General Network, *The 26th International Conference of Hong Kong Society for Transportation Studies*, 12-13 December 2022, Hong Kong, China.

Acknowledgements

I would like to express my sincere gratitude to Dr. Ma, my Ph.D. supervisor, for his steadfast support and mentorship throughout my research journey. I feel incredibly privileged to have been part of the MobilityAI Lab and to have conducted my studies under his guidance. When I first joined the group, I had little background in the field, having transitioned into Transportation Engineering. However, thanks to Dr. Ma's patience, encouragement, and insightful advice, I grew from a beginner into a proficient researcher. His expertise and support helped me navigate challenges, and under his supervision, I have developed strong skills in mathematical modeling, academic writing, and presentations over the past four years. I am also deeply appreciative of Prof. Anthony Chen, my co-supervisor, for his valuable insights and guidance. His thought-provoking lectures and stimulating discussions helped me identify key research gaps, while his meticulous and constructive feedback on my presentations significantly contributed to refining my work. Additionally, I was fortunate to gain international academic exposure and industry insights through the exchange program at The University of Tokyo (UTokyo), an opportunity made possible by Dr. Ma's open-minded support and generosity. I would also like to extend my thanks to Prof. Fukuda from UTokyo for his guidance during my time there, which broadened my academic perspective and provided me with invaluable research experience. To conclude, their collective mentorship has been instrumental in shaping this dissertation, and I am truly grateful for their support and encouragement.

I would also extend my appreciation to all members in MobilityAI Lab in PolyU for all the insightful discussions and suggestions. Especially, I would express my deep thanks to Binyu Yang, Junlin He and Shuangyuan Yu for all the joyful time we spent together which eased the pain of research.

Finally, I would like to extend my deepest gratitude to my parents for their unconditional love, unwavering support, and endless encouragement throughout my academic journey. Their

belief in me has been a constant source of motivation, and their sacrifices have provided me with the opportunities that made this achievement possible. No matter how far I was from home, their support always gave me the strength to persevere through challenges.

Contents

1	Introduction	2
1.1	Background	2
1.2	Problem statement	4
1.3	Research objectives	7
1.4	Dissertation organization	8
2	Literature review	11
2.1	Overview of DTA modeling	11
2.1.1	Dynamic network loading models	11
2.1.2	DTA with AVs	13
2.1.3	Non-uniqueness of SO-DTA	15
2.2	Ride-hailing platform of AVs	16
2.2.1	Ride-hailing platform regulation and ownership	16
2.2.2	Subsidy	18
2.3	Roles of AVs in traffic assignment problem	19
2.3.1	Mixed equilibrium with AVs	19
2.3.2	Control strategies of AVs in traffic networks	20
3	Maximin headway control for system optimal dynamic traffic assignment	23
3.1	Introduction	23
3.2	Modeling headway-dependent flow dynamics	27
3.2.1	Notations	27
3.2.2	Headway-dependent fundamental diagram	28
3.2.3	Headway-dependent double queue model	30

3.2.4	Dynamic network traffic model	33
3.3	Optimal headway control	35
3.3.1	Continuous-time formulation	36
3.3.2	Discretized-time formulation	36
3.4	Maximin headway control	41
3.4.1	Maximin headway control formulation	41
3.4.2	Maximin headway control solution algorithm	43
3.5	Numerical experiments	45
3.5.1	A small network	45
3.5.1.1	Settings	46
3.5.1.2	Results for maximin headway control in small network	46
3.5.1.3	Sensitivity analysis for the maximin headway control	49
3.5.2	The Hong Kong network	51
3.5.2.1	Settings	51
3.5.2.2	Experimental results	52
4	The publicly-owned centralized platform for ride-hailing services	54
4.1	Introduction	54
4.2	Problem settings	59
4.3	Model	60
4.3.1	Notations	61
4.3.2	Dynamic network modeling for POCP	63
4.3.2.1	Traffic flow dynamics	64
4.3.2.2	Travel demand and government subsidy	69
4.3.3	Cost for different stakeholders	71
4.3.3.1	Travelers cost	71
4.3.3.2	TNC profit	72
4.3.3.3	Government cost	72
4.3.3.4	Social cost	73
4.3.4	Formulation	73
4.4	Solution framework	75

4.4.1	Strong duality of lower level	75
4.4.2	Solution of bi-level programming	78
4.4.2.1	Linearity in DTA	78
4.4.2.2	Linearity in subsidy	84
4.4.3	Solution algorithm	85
4.5	Numerical experiments	87
4.5.1	Small Network	87
4.5.1.1	Computation performance	88
4.5.1.2	Comparison of different strategies	90
4.5.2	Hong Kong Network	92
4.5.2.1	Experimental setting	92
4.5.2.2	Results	95
5	Mixed sub-network equilibrium of automated vehicles in general traffic networks	97
5.1	Introduction	97
5.2	Problem Settings	102
5.3	Model	105
5.3.1	Formulation of mixed sub-network equilibrium	105
5.3.2	Properties of mixed sub-network equilibrium	112
5.3.3	Critical size of SOSN	115
5.4	Solution Method	119
5.4.1	Sub-network-augmented method of successive averages (SNA-MSA) for MSNE	120
5.4.2	Adaptive variable neighborhoods search algorithm (AVNS) algorithm for CS-SOSN	120
5.5	Numerical Results	124
5.5.1	Small network	125
5.5.2	Large network	127
5.5.3	Hong Kong network	129
6	Summary and future directions	132
6.1	Research findings and contributions	132

6.2	Future direction	134
References		136
Appendix A Appendix in maximin headway control		154
A.1	Notations	155
A.2	Proof of Proposition 1	156
A.3	Proof of Proposition 2	158
A.4	Proof of Proposition 3	159
A.5	Sensitivity analysis of the headway control for SO-DTA	162
A.5.1	LP formulation	162
A.5.2	Sensitivity-based optimal headway control solution algorithm	164
A.6	Proof of Proposition 4	167
A.7	Proof of Proposition 5	173
A.8	Proof of Proposition 6	173
A.9	Settings of the small network	174
A.10	Settings of the Hong Kong network	174
Appendix B Appendix in publicly-owned centralized platform		180
B.1	Proof of Feasibility of POCP Problem	180
B.2	Proof of Proposition 11	184
B.3	Proof of Lemma 1	184
B.4	Proof of Proposition 12	187
B.5	Proof of Proposition 13	189
B.6	Proof of Proposition 14	190
B.7	Parameters of POCP	191
Appendix C Appendix in mixed sub-network equilibrium		193
C.1	Proof of Proposition 18	193
C.2	Proof of Proposition 19	194
C.3	Proof of Proposition 20	195
C.4	Proof of Proposition 22	196
C.5	Proof of Proposition 23	199

List of Figures

1.1	Non-uniqueness of AVs strategies to achieve SO.	7
1.2	Thesis outline.	9
3.1	Illustration of headway control around an intersection.	25
3.2	Illustrative relationship between the probability of accident and headway. . . .	26
3.3	Illustration of the maximin headway for SO-DTA.	26
3.4	Fundamental diagram of AVs with different headways.	29
3.5	Flow area and buffer area in a link.	31
3.6	Small network.	46
3.7	Inflow, outflow, and maximin headway of each link under SO-DTA.	47
3.8	Convergence curve in terms of TTT by Algorithm 6.	48
3.9	The ratio of maximin headway under different lengths of time intervals.	49
3.10	The ratio of maximin headway under different traffic demands.	50
3.11	The ratio of maximin headway under different minimum headway settings. . . .	50
3.12	Overview of the Hong Kong network.	51
3.13	The distribution of maximin headway and link flow under SO-DTA.	52
4.1	Comparisons of the structures of different platforms under government subsidy.	58
4.2	Overview of the double flow with POCP in link level.	65
4.3	Overview of the double flow with POCP in node level.	66
4.4	Structure of small network.	88
4.5	Overview of the Hong Kong network.	93
4.6	Overview of the link flow of POCP in Hong Kong network.	96
4.7	Dynamics of subsidy in POCP.	96

5.1	Illustration of SOSN and UESN.	99
5.2	Critical size of system optimal sub-network.	101
5.3	An example of SOSN and UESN.	103
5.4	Workflow of AVNS algorithm.	122
5.5	Illustration of small network.	125
5.6	Illustration of large network.	127
5.7	The variation of the total travel cost for different size of SOSN in large network.	128
5.8	The location of CS-SOSN in large network.	129
5.9	The Hong Kong network.	130
5.10	The variation of the total travel cost for different size of SOSN in the Hong Kong network.	130
5.11	The location of CS-SOSN in the Hong Kong network.	131
A.1	The hour O-D demand profile in the Hong Kong network. A darker color indicates more trips between the corresponding O-D pair.	176

List of Tables

2.1	Literature summary of traffic assignment with AVs.	15
2.2	Literature of the subsidy in the ride-hailing market.	17
2.3	Literature of the control strategies of AVs in traffic networks.	21
3.1	Notations in the discretized formulation.	37
3.2	Maximin and minimum headway in the small network.	49
4.1	Notation list.	63
4.2	The small network link properties.	88
4.3	Network parameters and model coefficients for a small network.	89
4.4	Computation performance of Algorithm 3 and heuristic method.	90
4.5	Flexible demand setting in the small network.	90
4.6	Results of POCP on the small network.	92
4.7	Network parameters and model coefficients for Hong Kong network.	94
4.8	Results of POCP on Hong Kong network.	95
5.1	Notations table.	103
5.2	Small network parameters.	125
5.3	Link flow under MSNE in small network.	126
5.4	MSNE of path in small network.	126
5.5	O-D demand.	128
A.1	Notation table.	155
A.2	Parameters settings in the small network.	174
A.3	Settings of minimum headway.	175
A.4	Results of the maximum headway.	175

A.5	Link parameters in the Hong Kong network.	179
B.1	Parking parameters for Hong Kong network.	192

Chapter 1

Introduction

1.1 Background

The rapid advancement of automated vehicle (AV) technology has opened up new possibilities for improving transportation efficiency, safety, and sustainability. AVs are motor vehicles equipped with automated sensing, perception, decision-making and control subsystems that enable them to operate with reduced or no real-time human steering inputs under specified operational conditions (Litman, 2017). AVs typically integrate multiple sensors (e.g., cameras, radar, lidar), onboard localization, artificial intelligence, and vehicle-to-everything (V2X) technology. The characteristics of AVs include: (1) automated trajectory generation and short reaction times compared to human drivers; (2) the ability to execute centralized or coordinated routing instructions when connected to a traffic management center; (3) potential for applying AVs strategies to improve system performance. The introduction of AVs allows for a shift from traditional, human-centric transportation systems to more controlled and optimized traffic management solutions. As the penetration rate of AVs increases, their impact on traffic dynamics and routing strategies becomes a key area of research. A fundamental challenge in integrating AVs into road networks lies in their routing and control mechanisms, which directly influence the overall system efficiency.

Traffic assignment models play a critical role in determining how AVs navigate through transportation networks. The traffic assignment problem seeks to allocate vehicles to available routes to optimize network performance (Sheffi, 1985). One of the key routing principles in traffic assignment problems is the system optimal (SO), which seeks to minimize total travel costs across

the entire network rather than optimizing individual user experiences (Beckmann et al., 1956). The SO principle quantifies the best possible traffic network performance and provides policy-makers with policy implications and operational strategies to improve system-wide performance. Achieving SO in transportation networks under human-driven environment is challenging due to individualistic driving behavior, but the emergence of AVs offers a unique opportunity to control vehicle movements more effectively and enforce SO routing strategies.

With advancements in automation and connectivity, the potential for a fully automated transportation environment where all AVs adhere to SO principles is becoming increasingly feasible (Levin and Boyles, 2016a). The ability to control all AVs enables traffic managers to implement system-wide optimization, reducing congestion and improving travel reliability. However, the realization of SO under fully automated environment requires sophisticated control strategies to manage vehicle interactions effectively. This research aims to develop strategies for AVs routing to achieve SO in traffic assignment problems from three key perspectives: vehicle level, platforms level, and network level.

The traffic assignment problem can be broadly classified into two categories: static traffic assignment (STA) and dynamic traffic assignment (DTA). STA models assume that travel demand and network flows remain constant over time, making them suitable for long-term planning and equilibrium analysis (Patriksson, 2015). These models typically focus on optimizing steady-state traffic flow without considering temporal variations. In contrast, DTA models account for time-dependent changes in traffic conditions, congestion, and real-time vehicle interactions (Peeta and Ziliaskopoulos, 2001). DTA provides a more realistic framework for managing AV movements, especially in highly dynamic urban environments. Depending on the characteristics of the strategy of routing AVs, different types of traffic assignment models are employed to ensure optimal routing and overall system performance.

An important challenge in SO-based traffic assignment is the non-uniqueness of solutions. Since multiple route configurations can achieve SO, different strategies can be designed to reach system-wide optimization in each aspect: vehicle, platform, and network. This flexibility introduces an additional dimension in the development of AV control mechanisms, inspiring the exploration of a secondary optimization criterion beyond SO. In other words, given multiple feasible strategies for achieving SO, additional factors such as safety, subsidy, or minimum government intervention may serve as secondary optimization objectives.

To conclude, a comprehensive understanding of the network-wide impacts of AVs in a fully automated environment is crucial for both public agencies and private sectors in developing effective management and routing strategies. This research formulates the routing strategies of AVs for SO in a traffic network in three key dimensions: vehicle level, platform level, and network level. Additionally, given the non-uniqueness of solutions in SO problems, this study explores the selection of optimal control strategies by incorporating other transportation-related factors, such as safety, financial incentives, and minimal governmental intervention.

1.2 Problem statement

This study aims to explore the roles of AVs in system optimal (SO) for traffic assignment problems. The SO problem is important to quantify the optimal performance of the system, then provide valuable insights of designing strategies of policies. The emergence of AVs provides a feasible and effective tool for the government to achieve the SO state in the whole network, then reducing congestion and saving social welfare. However, how to incorporate AVs into the SO problem and develop effective strategies to achieve SO are still a challenge. This study aims to explore the roles of AVs in SO for traffic assignment problems in three key dimensions: vehicles level, platform level, and network level. Additionally, given the non-uniqueness of solutions in SO problems, this research investigates a secondary criterion for determining the optimal strategy of routing AVs. The specific research challenges associated with each aspect are outlined as follows.

In the vehicle aspect, we focus on the headway of AVs. Time headway is the time gap between two successive vehicles as measured at a point on the road. AV headway plays a crucial role in alleviating congestion and ensuring safety within a traffic network. Besides, it is feasible to dynamically control the AVs headway, then influencing the traffic states and congestion. Therefore, the dynamic traffic assignment (DTA) framework is more suitable for application in exploring the effects of controlling AVs headway. Previous studies have demonstrated that varying AV headway settings influence the fundamental diagrams of roads (Zhou and Zhu, 2020a, Gong et al., 2016a), thereby affecting the solutions of the system optimal dynamic traffic assignment (SO-DTA). Reducing AV headway at the network level can increase road capacity and throughput, which contributes to optimizing system performance and achieving the SO-DTA. However,

the feasible range of AV headway is constrained by minimum safety requirements, AV specifications, and the maximum capacity of individual roads (Li and Chen, 2017). Furthermore, similar to existing SO-DTA models (Shen and Zhang, 2014a), it is reasonable to expect that multiple AV headway settings may yield the same SO-DTA solution across general networks. This observation motivates the exploration of a secondary criterion in the SO-DTA problem. Safety remains a critical consideration for AV operations. For instance, a trade-off has been established a trade-off between the time lag gap and the safety buffer, both of which contribute to AV headway (Li, 2022). If the time lag gap is fixed—being determined by vehicle specifications and mechanical properties—then increasing headway can enhance safety. Under identical traffic conditions, a larger headway between AVs is preferable. This study analytically validates these intuitions in general networks and introduces the concept of maximin headway control, defined as the "largest" headway setting that satisfies SO-DTA requirements. This approach ensures that the selected headway setting is the "safest" under SO-DTA conditions. The mathematical properties of maximin headway are examined, proving its uniqueness, and it is shown that this optimal headway setting can be determined analytically using Linear Programming (LP).

In the platform aspect, ride-hailing platforms provide convenience for both users and drivers but may cause an increase in congestion in traffic networks. For its convenience and efficiency, the ride-hailing platform has been a popular choice for travelers but the rapid growth of ride-hailing vehicles causes a significant increase in the number of idle vehicles cruising on streets and a reduction in traffic speed. Hence the popularity of the current ride-hailing platform pattern may cause an increase in congestion in traffic networks (Schaller, 2017). Most of the studies regarding ride-hailing platform design to reduce congestion assume that the ride-hailing platforms are operated by third parties for maximal profits, while the management strategies by governments mainly include: 1) a cap on the number of drivers or vehicles in the ride-hailing platform, and 2) a congestion tax for each ride-hailing trip (Li et al., 2019, Liu et al., 2022, Vignon et al., 2021, Zhang and Nie, 2018a). Both of the above strategies are effective in reducing congestion but with the cost of actually impeding the development of the ride-hailing industry for the reason that the revenue of drivers in the ride-hailing platform would decrease if imposing cap because the platform would reap most of the benefits and platform revenue would decrease if imposing congestion tax (Li et al., 2019). In this study, we propose a novel concept of publicly-owned centralized platform (POCP), which receives order requests from travelers and sends the orders

to the TNCs. The key feature of the POCP is that it is owned by the government, and it aims to achieve the SO-DTA using a limited budget. Furthermore, shared automated vehicles (SAVs) have been shown and proven the possibility to replace personal vehicles representing the great potential for wide application in future mobility services (Chen, Kockelman, and Hanna, 2016, Spieser et al., 2014). Therefore, we formulate the operation of the POCP as nonlinear bi-level programming where the government aims at minimizing the social cost by subsidy to adjust travelers' behaviors in the upper level and TNC aims at maximizing its revenue by providing mobility service of SAVs in the lower level. The bi-level problem can be converted and solved exactly with analytical guarantees.

In the network aspect, sub-network control of AVs is an effective method to indirectly improve the traffic condition and mitigate congestion in the whole traffic network given the limited resources. Although it would be practical to control all the AVs in the whole traffic network to achieve SO under fully automated environments, travelers always desire to choose the shortest path from the origin to destination in real traffic assignment following user equilibrium (UE) principle and Fraedrich et al., 2019 pointed out that the excessive intervention of government is harmful to promote the popularity of AVs as it would aggravate the financial burden of government and reduce the trust of travelers in it. Therefore, in this study, we propose the sub-network control of AVs, where AVs would be controlled by different agents in corresponding sub-networks following different principles. We divide the whole traffic network into many sub-networks, categorized as system optimal sub-network (SOSN) and user equilibrium sub-network (UESN). In SOSN, such as CBD and downtown areas, the system would have the right of controlling AVs to minimize congestion in the traffic network following the SO principle. In UESN, travelers would have the right of controlling AVs to minimize their travel cost following the UE principle. Existing studies mainly focus on the control scheme of AVs applied homogeneously in traffic networks, while the heterogeneous control scheme and the right of control AVs in the sub-network level is still an open question. Besides, it is obvious that the total cost would be minimized if the size of SOSN equals the size of the whole network. Inspired by saving resources and reducing intervention to find a “smaller” size of SOSN with the minimum total cost, we propose a novel concept of the critical size of SOSN as the minimum size of SOSN to achieve the minimum total cost in the whole network. Mathematical properties regarding SOSN and UESN are also analyzed. Besides, due to the fact that the government policy can not be changed frequently, STA

is more suitable than DTA for sub-network control of AVs. In reality, this work aims to leverage the sub-network control of AVs to achieve the optimal system performance with minimum government intervention in the network level.

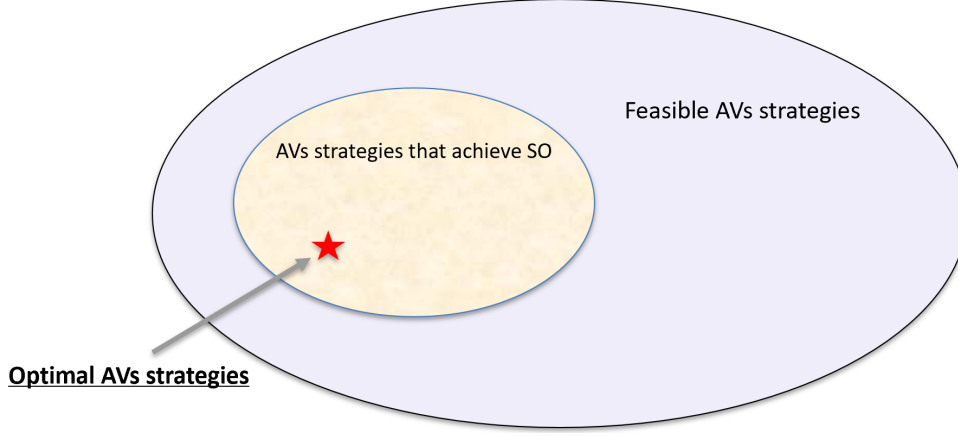


Figure 1.1: Non-uniqueness of AVs strategies to achieve SO.

To conclude, as is shown in Figure 1.1, the set of AV strategies to achieve SO is non-unique. This naturally raises a key question: If multiple strategies could achieve SO (or SO-DTA), how to find the “best” strategies? To fill this research gap, this research proposes the second criterion to find the optimal AVs strategies. Specifically, this research explores the roles of AVs in SO for traffic assignment problems in different aspects: vehicles, platform and network. Although many previous works have considered the effects of routing AVs under different routing principles, they have always ignored the non-uniqueness of the solution of SO. It means that the SO could be achieved by multiple strategies in each aspect and it inspires us to develop the second criterion of the optimal routing strategy of AVs in each aspect. Therefore, this research proposes the concept and framework of maximin headway, publicly-owned ride-hailing platform (POCP), and the critical size of system optimal sub-network (CS-SOSN), each representing the safest, most economical, and minimally invasive approach to achieving the SO, respectively.

1.3 Research objectives

This research aims to achieve an integral framework for exploring the roles of AVs in SO in different aspects: vehicles, platform and network. Furthermore, this study explores the secondary criteria for determining the optimal strategies for achieving SO within each of these components,

ensuring that the proposed strategies of AVs not only achieve primary system performance objectives but also address additional considerations such as safety, subsidy, or minimum government intervention. Specific objectives in each component are listed as follows.

- In the vehicles aspect, we aim to explore the effect of headway control on the fundamental diagram of AVs and traffic states in traffic networks. Headway dependent fundamental diagram (HFD) and headway-dependent double queue (HDQ) model are developed to capture of effects of dynamic AV headway control. We prove that SO-DTA could be achieved under minimum headway setting and SO-DTA may be achieved by multiple headway control. The maximin headway control is defined as the “largest” headway setting that still achieves SO-DTA and we propose an efficient solution algorithm to derive the maximin headway.
- In the platform aspect, we aim to develop a publicly-owned centralized platform (POCP), where the government is to minimize total social cost by changing travelers’ behavior in the upper level and the TNC is to maximize its revenue by providing ride-hailing services. An analytical solution algorithm is proposed to solve the minimum subsidy to achieve the minimum social cost.
- In the network aspect, we aim to find out the heterogeneous effect of the control scheme of AVs and divide the whole traffic network into SOSN and UESN. We propose the control scheme of AVs in SOSN and UESN. We analyze the equilibrium in the whole traffic network and discuss the conditions of whether to enter or leave SOSN, respectively. A novel concept of critical size of SOSN is proposed as using minimum resources or the least involvement of government to minimize total travel cost in the whole network.

1.4 Dissertation organization

This dissertation presents the strategies of AVs to achieve SO in traffic assignment problems. The main structure of other chapters are presented in Figure. 1.2 below. Chapter 1 introduces the background for this research, states its objectives, and summarizes the main contributions.

Chapter 2 provides a comprehensive review of the application of AVs in traffic assignment, examining the strengths and limitations of various models while establishing the theoretical

foundation for the proposed approach in this study. The review begins with an analysis of existing dynamic network loading (DNL) models, focusing on the impact of headway on traffic flow dynamics and safety. Additionally, we explore different ride-hailing platforms, considering variations in ownership structures and modeling approaches. A discussion is included on how subsidies are implemented in previous studies, covering aspects such as the provider, recipient, subsidy type, and intended objective. Furthermore, we examine existing models that analyze the mixed equilibrium of AVs and investigate various strategies for AV control within traffic networks. These strategies include lane-based control, ratio-based control, and sub-network control.

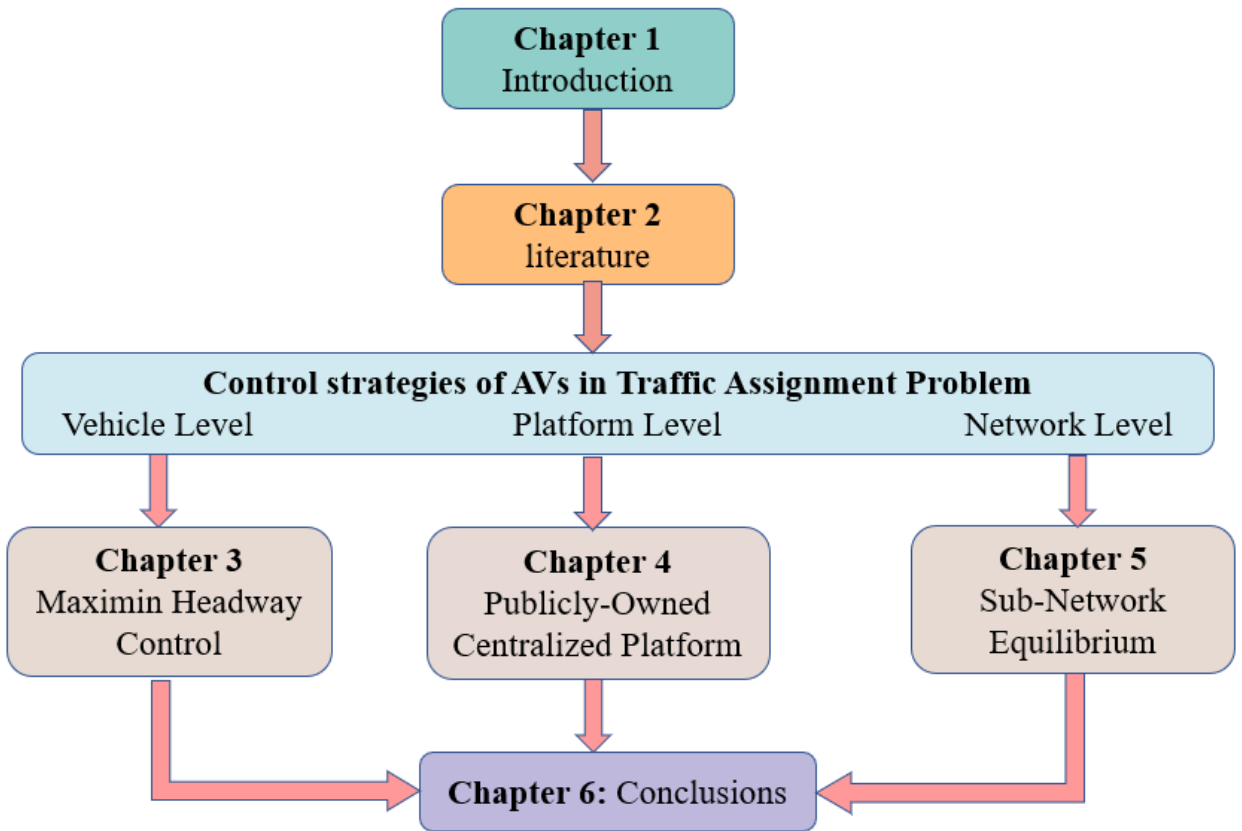


Figure 1.2: Thesis outline.

In Chapter 3, we develop a headway control framework for fully automated road networks, recognizing headway as a key factor in traffic dynamics. The goal is to determine the optimal time headway for AVs on each link to achieve SO-DTA. To model the headway effects, we introduce the headway-dependent fundamental diagram (HFD) and headway-dependent double queue model (HDQ), integrated into a dynamic network model. It is proven that while the minimum headway always ensures SO-DTA, the optimal headway is not unique. Based on this, we

define the novel concept of maximin headway—the largest headway that still achieves SO-DTA. We analyze its mathematical properties and develop an efficient solution algorithm.

In Chapter 4, we introduce a publicly-owned centralized platform (POCP) for shared automated vehicles (SAVs), linking travelers, traffic network companies (TNCs), and parking spaces. It offers a novel approach for government regulation to reduce social costs by influencing traveler behavior through spatial-temporal trip subsidies. The POCP operation is formulated as a bi-level programming problem, where the government optimizes subsidies to minimize social costs in the upper level, while TNCs solve a DTA problem to maximize profits in the lower level. We prove that this bi-level problem can be decomposed into two linear programming sub-problems to determine the minimum social cost and subsidy amount efficiently.

In Chapter 5, we introduce the mixed sub-network equilibrium (MSNE), which divides a traffic network into two parts: (i) the user equilibrium sub-network (UESN), where AVs independently select routes to minimize their own travel costs, and (ii) the system optimal sub-network (SOSN), where AVs are centrally controlled to minimize overall travel costs. This hybrid approach enables targeted control of critical areas rather than system-wide regulation. A key contribution is the introduction of the critical size of the system optimal sub-network (CS-SOSN), which determines the minimum SOSN size needed for system-wide optimal performance. We develop an adaptive variable neighborhood search (AVNS) algorithm to solve for CS-SOSN and validate its effectiveness through theoretical analysis and simulations on three networks.

Finally, Chapter 6 summarizes the findings and conclusion of this thesis.

Chapter 2

Literature review

In this section of the literature review, the focus is on summarizing the related work about the role of AVs in the traffic assignment problem. The section begins with an overview of the current state of DTA models, providing insight into the foundational work and models used in DTA for traffic systems. Following this, the literature addresses two critical issues in traffic assignment: the ride-hailing platform and the mixed equilibrium problem. These two topics represent key challenges in modern traffic systems, and the review further explores how AVs influence and interact within these areas. Specifically, the literature highlights the roles of AVs in determining the properties of traffic assignment models, shaping the outcomes and strategies for both ride-hailing platforms and mixed traffic equilibrium.

2.1 Overview of DTA modeling

In this section, we first summarize the literature regarding the dynamic network loading network models for AVs in Section 2.1.1. In particular, we compare with existing studies on dynamic link models, network models, and SO-DTA, to further highlight the contributions of this work. Then in Section 2.1.2, we discuss the application of AVs in DTA in the current related research works. Lastly, we discuss the non-uniqueness of the solution of SO-DTA problem in Section 2.1.3.

2.1.1 Dynamic network loading models

Dynamic link models are an essential component that models the flow dynamics on road segments, and it is necessary to develop an appropriate link model to depict the effects of AV

headway control. The Lighthill, Whitham, and Richards (LWR) model forms the basis of the dynamic network loading model (DNL) by capturing key congested patterns on road segments (Lighthill and Whitham, 1955; Richards, 1956). The LWR model is a system of partial differential equations (PDE), which consists of three parts: static hydrodynamics relation for flow, density, and speed; fundamental diagram; and dynamic continuity equation. Many famous DNL models are designed to solve the PDE system of LWR in different methods.

The cell transmission model (CTM) provides an approximate solution to the PDE system of the LWR model through the discretization in time and space (Daganzo, 1995). CTM is formulated based on the density in each cell and the flow rates over the boundaries of cells. While the CTM is capable of accurately capturing spillback phenomena and determining detailed traffic states within a link, it results in high computational costs in our problem due to the potentially larger dimension of a DTA problem when using CTM compared to other link models (Nie and Zhang, 2005). Compared with CTM which approximately solves LWR by discretizing in space and time, the link transmission model (LTM) directly solves the LWR system using the Newell-Daganzo method (Yperman, 2007). LTM only tracks the cumulative number of vehicles at the entrance and exit of a link with a smaller computation cost, and it could also capture the spillbacks by incorporating the receiving constraints for congested states and the sending constraints for free-flow states. Recently, Osorio et al., 2011 first proposed the concept of double queue, where the dynamics of a link are captured by the downstream queue and upstream queue. The downstream queue models the delay in free-flow time, the queuing process, and the outflow at the downstream end of a link. On the other hand, the upstream queue captures the delay caused by backward shockwave and the congestion propagation process, as well as the inflow at the upstream end of a link. The continuous form of DQ is developed by Ma et al., 2014 and it is proved that DQ could be directly derived by LTM but with the simplified mathematical representation compared with LTM. Besides, the spillback in DQ could be tracked by the traffic states at either the entrance or the exit of a link (Ma et al., 2014). Furthermore, Ngoduy et al., 2021 proposed a multi-class two regime transmission model (MTTM) to capture the evolution of the queue lengths, where a link is divided into two areas: free-flow area and congested area. But MTTM is a nonlinear model for traffic states and it would add complexity and difficulty in solving the traffic states and analyzing the properties in DTA problems.

The abovementioned models are all capable of capturing the spillback phenomenon repre-

senting the time when the congestion shockwave reaches the entrance of a link. Besides, they all require a node model to satisfy the invariance principle (Lebacque, 2005). To conclude, all these models are also designed to solve the PDE system of the LWR model, which actually depicts the dynamics of traffic states. Technically, the choice of DNL models would not significantly influence the modeling of the dynamic networks, while we prefer a simpler model that has more analytical capabilities. Compared with other DNL models, DQ has the advantages of low computation cost, simplified mathematical formulation, and linear representations for traffic state variables. Therefore, we choose the DQ model as the basis to incorporate the effects of headway in flow dynamics in this thesis.

Besides, in addition to the link models, the node model is also an important component of the DNL process. It captures how flows on different links interact at nodes. By studying the sending flow and receiving flow, the DNL process is divided into two parts: capturing how the traffic flow propagates within a single link and the interactions between link flows at nodes. Daganzo, 1994 proposed simple demand–supply based junction rules that couple CTM cells at merges and diverges. It is built on upstream “sending” and downstream “receiving” concepts and is proved as the first order discrete Godunov approximation. Lebacque, 1996 clarified how demand–supply interpretations provide a unifying treatment of boundary and junction conditions for first-order macroscopic traffic models and explained how junction constraints should be imposed in Godunov-type DNL models. Besides, Tampère et al., 2011 proposed a generic class of first-order node models and stated a set of desirable axioms, including conservation, non-negativity and respect of split ratios. In summary, DNL process is an important part in DTA problems and consists of link models and node models to capture different flow dynamics. In this study, we focus on exploring the effects of headway in the propagation of traffic flow within a link.

2.1.2 DTA with AVs

Dynamic traffic assignment (DTA) models are designed to explore the time-dependent traffic state pattern given a time-varying travel demand on general traffic networks (Szeto and Lo, 2006). There are two representative types of DTA: dynamic user equilibrium (DUE) and system optimal dynamic traffic assignment (SO-DTA). SO-DTA quantifies the best possible performance of a traffic network based on the dynamic extension of Wardrop’s second principle

(Wardrop, 1952). For its advantage of exploring the spatial-temporal traffic dynamics to achieve the best performance of traffic networks, the research of the application of SO-DTA problem focuses on many different areas, such as emission reduction (Long et al., 2018; Lu et al., 2016; Ma et al., 2017; Tan et al., 2021), congestion mitigation (Levin, 2017; Liu et al., 2020; Samaranyake et al., 2018), parking services (Levin, 2019; Qian and Rajagopal, 2014), traffic signal control (Han, Liu, et al., 2016; Lo, 2001), emergency evacuation (Chiu et al., 2007; Liu et al., 2006) and network design (Waller and Ziliaskopoulos, 2001; Waller et al., 2006).

Different from human-driven vehicles (HVs), AVs show great potential in improving control efficiency and may play an important role in achieving SO-DTA (Wang et al., 2018). The current research on traffic assignment with AVs is summarized in Table 2.1. For the static traffic assignment problem of AVs, Chen, He, Zhang, and Yin, 2016a proposed a mathematical model to derive an optimal AV lane control framework for mixed traffic, where AV lines are only permitted for AVs and it is necessary to specify the deployment plan. Bagloee et al., 2017 focused on a static traffic assignment problem on mixed traffic where AVs are seeking the system optimal (SO) and HVs are seeking for user equilibrium (UE), and considers the influence of many realistic features in the solution, such as road capacity and elastic demand. Chen et al., 2017a presented an optimal AV zone design framework to improve the performance of the whole traffic network, where the government may make a plan of AVs zones only allowing AVs to enter to promote the adoption of AVs in the future. Zhang and Nie, 2018b proposed a route control scheme to balance the system efficiency and the control intensity by controlling a relatively small portion of all vehicles. Wang, Peeta, and He, 2019a addressed the multi-class traffic assignment problem by assuming AVs follow SO and HVs follow cross-nested logit (CNL) stochastic user equilibrium (SUE) to capture HV drivers' uncertainty due to their limited knowledge of traffic conditions in the whole network. Wang et al., 2020 developed a mixed behavioral equilibrium model in mixed traffic considering different mode choices for HVs and AVs, and Chen, Lin, et al., 2020 determined the minimum control ratio of AVs to achieve SO by path controlling. Levin and Boyles, 2016b proposed a multiclass cell transmission model for HVs and AVs and proves the consistency with the hydrodynamic theory. A collision avoidance car-following model incorporating drivers' reaction time is developed to explore the effect of reaction time on link capacity and backward wave speed. Their works provide valuable insights into the wide applications of AVs in avoiding collision by a smaller reaction time compared with human drivers.

Reference	Type	HV	AV	Description
Levin and Boyles, 2016b	Dynamic	DUE	DUE	A multi-class CTM model for mixed traffic
Chen, He, Zhang, and Yin, 2016a	Static	UE	UE	An optimal AV lanes control framework for mixed traffic
Bagloee et al., 2017	Static	UE	SO	The mixed traffic network with elastic demand
Chen et al., 2017a	Static	UE	SO	Optimal design of AV zones for mixed traffic
Zhang and Nie, 2018b	Static	UE	SO	A route control scheme to balance efficiency and control
Wang, Peeta, and He, 2019a	Static	CNL	UE	A multi-class traffic assignment model for mixed traffic
Wang et al., 2020	Static	SUE	SO	A mixed equilibrium model with mode choice
Chen, Lin, et al., 2020	Static	UE	SO	Minimum control ratio of AVs for SO in mixed traffic
Ngoduy et al., 2021	Dynamic	SO-DTA	SO-DTA	Impact of AVs in traffic network by multi-class SO-DTA
Seo and Asakura, 2022	Dynamic	\	SO-DTA	A unified optimization strategy framework for shared AVs
This work	Dynamic	\	SO-DTA	Maximin AV headway control for SO-DTA on general networks

Table 2.1: Literature summary of traffic assignment with AVs.

In summary, the work related to the traffic assignment problem of AVs is mainly static and there is few work exploring how AV headway control would affect the traffic states in whole general networks as a SO-DTA problem. There are few studies discussing the roles of headway in the control of AVs. Therefore, this work is designed to propose an optimal AV headway control framework for SO-DTA in general networks.

2.1.3 Non-uniqueness of SO-DTA

The solution of SO-DTA provides valuable insights for decision-makers to develop traffic management strategies (van Essen et al., 2016). Most of the related works focus on the solution algorithm of SO-DTA (Chow, 2009; Long and Szeto, 2019; Qian et al., 2012; Zhang and Qian, 2020), and there are a few works discussing the non-uniqueness and properties of the solutions of SO-DTA. For example, Shen and Zhang, 2014b proved that SO-DTA may have multiple solutions with the same objective but different queue locations in networks. To explore the optimal queue distribution for the solution of SO-DTA, Ma et al., 2014 proposed the solution algorithm and discusses the existence of the free-flow optimal solution of SO-DTA that all vehicles are only allowed to wait in the origin nodes, and Ngoduy et al., 2016 optimized the heterogeneity of the queue lengths and the total queue lengths in traffic networks for all solutions of SO-DTA. Though both of the above studies discuss the non-uniqueness of the solutions of SO-DTA and propose the criteria to “pick up” the desired solution of SO-DTA, the criteria of both works are mainly related to congestion patterns. In reality, decision-makers could develop traffic management strategies considering different perspectives. Therefore, this thesis discusses the “safest” solution among the non-unique solutions of SO-DTA and proposes the maximin headway in this problem as the largest headway solution in SO-DTA.

2.2 Ride-hailing platform of AVs

In this section, we are focusing on the current research studies on the design of ride-hailing platform and the application of subsidy in platform. Specifically, we first present the regulation of ride-hailing platforms and its ownership structure in Section 2.1.1. Besides, we also discuss the role of subsidies in changing travelers' behavior and reducing congestion in Section 2.1.2.

2.2.1 Ride-hailing platform regulation and ownership

The drastic growth of ride-hailing services has attracted extensive research interest. Most research efforts related to ride-hailing platforms primarily revolve around various key stakeholders, including but not limited to travelers, drivers/service providers, platform operators, and government/policy makers (Wang and Yang, 2019). These stakeholders collectively influence and shape the ride-hailing ecosystem, with a particular emphasis on the experiences of passengers and the engagement of drivers—central pillars that underlie the demand-supply equilibrium of these services (Chen et al., 2023; Hu et al., 2022; Ke et al., 2020; Lavieri and Bhat, 2019; Sun et al., 2019; Wang, Liu, et al., 2019). Platform companies, typically managed by external third-party entities, are primarily driven by the pursuit of maximizing their profits (Bao et al., 2023; Jiao and Ramezani, 2022; Li and Liu, 2021; Lyu et al., 2023). However, ride-hailing services have led to a significant increase in idle vehicles on the road, contributing to heightened traffic congestion (Beojone and Geroliminis, 2021; Diao et al., 2021; Erhardt et al., 2019; Tarduno, 2021). Consequently, regulatory authorities have initiated the development of policy frameworks aimed at promoting public welfare. These regulations seek to mitigate the adverse impacts of ride-hailing services by implementing measures such as subsidies (Mo et al., 2020; Zhao et al., 2023; Zhu, Xu, et al., 2021), congestion tolls (Li et al., 2021; Vignon et al., 2021), and service caps (Ke et al., 2021; Li et al., 2019; Liu et al., 2022). A thorough comprehension of these measures is critical for policymakers tasked with crafting regulations that balance the operational efficiencies of ride-hailing companies, improve system-wide performance, and safeguard consumer interests. Such a comprehensive perspective is indispensable in enhancing the quality of mobility services and strengthening a sustainable urban transportation ecosystem.

Literature	Platform				Subsidy												
	Ownership		Model		Provider		Receiver			Type			Objective				
	Private	Public	AM ¹	STA ¹	DTA ¹	P ²	G ²	P ³	D ³	T ³	C ⁴	S ⁴	T ⁴	S-T ⁴	SW ⁵	PP ⁵	SM ⁵
Mo et al., 2020	✓		✓				✓	✓			✓				✓		
Pantelidis et al., 2020		✓		✓			✓		✓			✓			✓		
Tafreshian and Masoud, 2020		✓		✓			✓			✓		✓					✓
Zhu, Xu, et al., 2021	✓		✓				✓			✓	✓				✓		
Zhu, Ke, and Wang, 2021	✓		✓			✓				✓				✓		✓	
Guo and Huang, 2022	✓		✓			✓				✓	✓					✓	
Zhao et al., 2023	✓		✓				✓	✓			✓				✓		
Zhang, Liu, and Fan, 2023	✓		✓			✓		✓	✓		✓					✓	
Liu and Chow, 2023		✓		✓			✓			✓		✓					✓
Tang et al., 2023	✓		✓			✓				✓	✓				✓		
Zhou et al., 2024	✓		✓			✓		✓			✓					✓	
This work		✓			✓		✓			✓				✓		✓	
The dynamic network-wide effects of applying a spatial-temporal subsidy in publicly-owned platform																	

The dynamic network-wide effects of applying a spatial-temporal subsidy in publicly-owned platform

¹ AM: aggregated model; STA: Static traffic assignment; DTA: Dynamic traffic assignment

² P: platform; G: government.

³ P: platform; D: Driver/Mobility operator; T: traveler.

⁴ C: constant; S: spatial; T:temporal; S-T: spatial-temporal.

⁵ SW: Social welfare; PP: Platform profit; SM: Stable matching

Table 2.2: Literature of the subsidy in the ride-hailing market.

Furthermore, the ownership structure of the ride-hailing platform plays a significant role in the effective implementation of spatial-temporal subsidies to enhance ride-hailing services and promote social welfare in real world. The limitations in information sharing, as highlighted in Table 2.2.1, indicate the challenges faced by the government in providing subsidies to ride-hailing travelers for enhancing social welfare in privately-owned platforms (Guo and Huang, 2022; Mo et al., 2020; Tang et al., 2023; Zhao et al., 2023; Zhu, Ke, and Wang, 2021; Zhu, Xu, et al., 2021). Zhong et al., 2022 pointed out that there exists competition between government and ride-hailing platforms in promotion and market regulation, which leads to information asymmetry such that it is difficult for the government to provide subsidies directly to travelers in real-time online ride-hailing orders. In privately-owned platforms, the government can only directly provide subsidies to the platform operators to promote social welfare rather than to travelers themselves (Mo et al., 2020; Zhao et al., 2023). Zhu, Xu, et al., 2021 assumed that the government would offer static subsidies to travelers utilizing public transit operated by the government, excluding those using ride-hailing services from privately-owned platforms.

2.2.2 Subsidy

Subsidies serve as effective and commonly implemented regulatory tools within the ride-hailing market (Mo et al., 2020; Zhao et al., 2023; Zhu, Xu, et al., 2021). The diverse range of subsidy types and models of the ride-hailing platform is concisely cataloged in Table 2.2.1. The primary goals of applying subsidies to the ride-hailing market consist of enhancing social welfare (Mo et al., 2020; Pantelidis et al., 2020; Zhao et al., 2023; Zhu, Xu, et al., 2021), obtaining more platform profit (Guo and Huang, 2022; Tang et al., 2023; Zhu, Ke, and Wang, 2021), and guaranteeing the existence of stable matching (Liu and Chow, 2023; Tafreshian and Masoud, 2020). Yet, the literature presents a scant exploration of the dynamic and network-wide implications resulting from the application of spatial-temporal subsidies within ride-hailing platforms. Notably, Liu and Chow, 2023 put forth a subsidy mechanism originating from the platform aiming to support the local stability of the matching process. The prescribed subsidy in the said study is path-based by following a generalized form of user equilibrium. However, the study did not sufficiently explore the dynamic impacts of this subsidy (Liu and Chow, 2023). It is worth noting that in practical traffic situations, the departure time selections made by commuters significantly impact congestion levels, and thus social welfare. Moreover, Zhu, Ke, and

Wang, 2021 present a spatial-temporal subsidy framework that utilizes a mean-field Markov decision process to characterize system dynamics, which diverges from a traffic network modeling approach. Nonetheless, in large-scale traffic networks, it becomes imperative to integrate and articulate the network-wide implications of subsidies. This ensures that the effect of subsidy strategies is more neatly aligned with real-world traffic conditions. To bridge the identified gaps in research, this study introduces a spatial-temporal subsidy approach that capitalizes on the dynamic traffic assignment model to promote social welfare. Additionally, it provides an in-depth exploration of diverse subsidy types tailored for various ride-hailing/ride-sharing behaviors.

2.3 Roles of AVs in traffic assignment problem

In this section, we mainly focus on finding the roles of AVs in traffic assignment. The existence of AVs provides the feasibility of different routing principles, then leading to the mixed equilibrium in the whole network. Besides, different control strategies of AVs would make different effects in the whole traffic network. Therefore, in this section, we summarize the literature regarding the mixed equilibrium and control strategies of AVs in traffic assignment problems. The related work regarding the mixed equilibrium with AVs is presented in Section 2.3.1. Then we compare with existing studies on control strategies of AVs to further highlight the contributions of this work in Section 2.3.2.

2.3.1 Mixed equilibrium with AVs

The concept of mixed equilibrium, which simultaneously encompasses UE and SO principles, has attracted significant attention in traffic networks with the rise of AVs. Traditional traffic management typically focuses on one of these models. However, with the integration of AVs, mixed equilibrium becomes feasible, allowing AVs to follow different routing principles. For instance, Zhang et al., 2020 studied the mixed equilibrium of human-driven vehicles (HVs) and AVs under a tradable credit scheme, where both HVs and AVs follow the stochastic user equilibrium (SUE) principle, but CAV users have much lower perception variances than HV users on travel costs. Chen et al., 2017b developed an AV zone design framework, which only allows AVs to enter. In the AV zone, AVs follow the SO principle. In areas outside the AV zone, HVs follow the UE principle. Jiang, 2017 proposed an optimal signal design framework for mixed

equilibrium, where AVs follow the Cournot-Nash (CN) principle and RVs follow the UE principle. Zhang and Nie, 2018a developed an optimal-ratio control scheme for AVs, where some AVs follow the UE principle and other parts of AVs follow the SO principle. Sharon et al., 2018 introduced a linear programming model to determine the minimum number of AVs required to achieve SO, given the presence of AVs and CVs. Wang, Peeta, and He, 2019b explored a mixed equilibrium model, where the routing principle of HVs is characterized by the cross-nested logit (CNL) model. Zhang, Lu, Hu, and Meng, 2023 explored the mixed equilibrium under the deployment of dedicated lane and roadside unit, where HVs and AVs follow the stochastic network stochastic user equilibrium (SN-SUE) and stochastic network user equilibrium (SN-UE) principle, respectively. In summary, the mixed equilibrium models offer a promising approach to optimizing the performance of complex traffic networks by leveraging the advantages of both UE and SO, particularly as the adoption of AVs continues to rise.

2.3.2 Control strategies of AVs in traffic networks

The control strategies of AVs in traffic networks mainly consist of lane, ratio, and sub-network control, as summarized in Table 2.3.2.

For lane control of AVs, the approaches are mainly divided into dedicated AV lanes and automated vehicle/toll (AVT) lanes. Dedicated AV lanes are specialized traffic lanes exclusively reserved for the use of AVs, prohibiting access to HVs (Chen, He, Zhang, and Yin, 2016b; Liu and Song, 2019; Madadi et al., 2021; Zhang, Lu, Hu, and Meng, 2023). Another common method for AV lane control is the use of AVT lanes, where both AVs and HVs can pass, but AVs are exempt from tolls while HVs are charged (Liu and Song, 2019; Wang et al., 2021). This approach optimizes traffic flow by regulating access through tolls. The AV ratio control typically refers to controlling the proportion of vehicles that follow the SO principle, while the remaining vehicles follow the UE principle (Chen, Lin, et al., 2020; Zhang and Nie, 2022). Besides, it is also possible to control a certain ratio of AVs to follow other routing choice principles. For example, Battifarano and Qian, 2023 considered the fleet optimal (FO) principle for the controlled AVs and analyzes their effects in the traffic networks. Zhang et al., 2022 took environmental management factors into account and incorporates the least-emission (LE) principle, in addition to the SO principle, as one of the control principles for AV routing choices.

Literature	AV			Control			“Best” Control	Remark
	SO	UE	SUE	Lane	Ratio	Sub-network		
Chen, He, Zhang, and Yin, 2016b		✓		✓			×	AV lines with endogenous market penetration
Chen et al., 2017b	✓	✓				✓	×	Optimal design of AV zones ¹
Liu and Song, 2019		✓		✓			✓	Planning of AV lines and AV toll lines
Chen, Lin, et al., 2020	✓	✓			✓		✓	Minimum control ratio of AVs to achieve SO
Madadi et al., 2020			✓			✓	×	Sub-network for automated driving of AVs ²
Wang et al., 2021		✓		✓			✓	Autonomous vehicle/toll (AVT) lanes design
Madadi et al., 2021			✓	✓		✓	×	AVs lines and sub-network for AD of AVs ³
Zhang and Nie, 2022	✓	✓			✓		✓	Optimal-ratio control scheme of AVs
Zhang et al., 2022	✓ ⁴	✓			✓		×	Path controlling and subsidy for AVs
Zhang, Lu, Hu, and Meng, 2023			✓	✓			×	Deployment of AV lanes and roadside unit
Battifarano and Qian, 2023	✓	✓			✓		✓	Impact of optimized fleets of AVs on network
Kashniri et al., 2025	✓	✓			✓		×	Non-subscribers follow UE and others follow SO
This work	✓	✓				✓	✓	Divide the network into UESN and SOSN Propose critical size of SOSN (CS-SOSN)

¹ AV zones in Chen et al., 2017b only allows AVs to enter.

² Sub-network in Madadi et al., 2020 refers to the sub-network where the AD of AVs are allowed. In other areas, AVs are manual driving.

³ AV lanes in Madadi et al., 2021 are dedicated for AVs. Besides, the setting of the sub-network in Madadi et al., 2021 are same as it in Madadi et al., 2020.

⁴ In addition to UE and SO, the routing principles of AVs in Zhang et al., 2022 also include least-emission (LE).

Table 2.3: Literature of the control strategies of AVs in traffic networks.

Sub-network control is also an effective strategy to control AVs to improve system performance. Besides, Kashmiri et al., 2025 presented a transportation management center (TMC) framework that minimizes the total system cost (OTSC), but non-subscribers, who opt not to subscribe to the TMC, follow their user equilibrium (UE) routing patterns and compete with travelers adhering to the OTSC strategy. Sub-network control for automated driving (AD) of AVs is a common measure to improve transportation systems by influencing the capacity, the value of time (VoT), and fuel efficiency (Madadi et al., 2020, 2021). Both AVs and CVs are permitted to enter these sub-networks, but AD for AVs is restricted to operation within these sub-networks. In other words, manual driving of AVs will be required in the remaining areas of the network. Besides, Chen et al., 2017b designed the designated AV zones, where AVs have the option to enter or bypass, while CVs are prohibited from accessing these zones. It is assumed that AVs in the AV zones are forced to minimize the total travel cost in the AV zones, which equals choosing the path with minimum marginal travel time from the entrance to the exit of AV zones. For CVs and AVs in other areas of the network, they follow the UE principle. However, it overlooks the impact of controlling AVs within the AV zones on the overall performance of the entire traffic network. Both the routing choices within AV zones and the choices of entrance and exit influence the system performance of the traffic network.

Besides, the selection of the “best” control strategy is also an important consideration. The non-uniqueness of the AVs control problem implies that there might be multiple control strategies to achieve the SO (Battifarano and Qian, 2023). For example, Battifarano and Qian, 2023 proposed the critical fleet size for system optimum (CFS-SO), representing the smallest fleet to induce SO. However, current research on sub-network control primarily focuses on identifying the optimal strategies to achieve SO, without investigating the “best” strategy among those that achieve SO (Chen et al., 2017b; Madadi et al., 2020, 2021). To address the above two problems, this thesis divides the entire network into UESN and SOSN. In SOSN, both the routing choices in SOSN and the choices of entrance and exit of SOSN are optimized to minimize the total travel cost in the entire network. Besides, this thesis also proposes the framework of the critical size of SOSN (CS-SOSN) as the minimum size of SOSN to achieve SO, which could be regarded as the minimum government intervention to achieve the maximum system performance.

Chapter 3

Maximin headway control for system optimal dynamic traffic assignment

3.1 Introduction

With the rapid development of advanced sensors, wireless communication, and artificial intelligence, it is envisioned that automated vehicles (AVs) will play a significant role in future intelligent transportation systems. Ultimately, the fully automated environment would be realized to significantly mitigate congestion and ensure safety in general urban networks (Wadud et al., 2016). Therefore, it is essential for public agencies (*e.g.*, Transport Department) and private sectors (*e.g.*, Transportation Network Companies) to understand the network-wide effects of AVs in the fully automated environment and to develop effective management and control schemes for AVs. However, recent studies mainly focus on the microscopic behaviors of AVs, while studies on the spatio-temporal effects of AVs in networked traffic dynamics are still lacking (Yu et al., 2021).

Among various traffic network models, system optimal dynamic traffic assignment (SO-DTA) quantifies the best possible performance of a traffic network given the dynamic network demand and supply. Both the optimal objective function (*i.e.*, the total travel time) and the optimal solution (*i.e.*, path flow and link flow) could benchmark different traffic operation and management strategies and provide policy implications for decision-makers. Additionally, it is practical to control all the AVs to achieve SO-DTA under fully automated environments (Nguyen et al., 2021).

In this work, we consider the headway control of AVs, in addition to the travelers' behaviors (*e.g.*, route and departure time choices) for SO-DTA. That is, AVs can decide the headway between each other according to the mobility and safety requirements of each road segment (Hatipoglu et al., 1996). In particular, headway refers to the time headway throughout the thesis.

Existing studies have shown that different headway settings of AVs could affect the fundamental diagrams of each road (Gong et al., 2016b; Zhou and Zhu, 2020b), hence affecting the solutions of SO-DTA. However, most current studies are at a microscopic level, while how the AV headway control could affect the network-wide SO-DTA is still an open question.

Many studies have explored the applicability and practicality of AV headway control in future traffic scenarios Elmorshedy et al., 2023; Xiao et al., 2023. The concept and control principles of headway control are first proposed as an effective longitudinal control strategy of vehicles Chiu et al., 1977; Shladover et al., 1991. The development of automated technology has attracted interest in studying the application of headway control in fully automated vehicles. Li et al., 2017 and Xiao and Gao, 2011 propose a practical constant time headway (CTH) control policy for AVs. Further, Middleton and Braslavsky, 2010 considers the control strategy under a heterogeneous headway. Moreover, a variable time headway (VTH) control method was developed and proved as a more flexible, effective, and practical method compared with the CTH control policy (Chen, Liang, et al., 2020). A brief illustration of the headway control around an intersection is shown in Figure 3.1. All the AVs are connected through Vehicle-to-Vehicle (V2V) communications, and their headway can be set differently on different roads. The cooperative area is at the intersection (de La Fortelle and Qian, 2015), where AVs can communicate with the Roadside Unit (RSU) to receive the headway setting for each road and adjust their headway accordingly. The cooperative area should be designed to have strong communication capacities between AVs and RSUs and enough space for the adjustment of headway.

Intuitively, a smaller AV headway yields higher road capacities and throughput, leading to shorter travel time, and hence SO-DTA can be achieved with the smallest AV headway. To verify this intuition, later this work *analytically* proves that Observation 1 holds in general networks.

Observation 1. Given a range of the AV headway, SO-DTA is always achieved when the AV headway is set as the smallest value in the range.

Note that the range of AV headway depends on the minimum safety requirements of AVs,

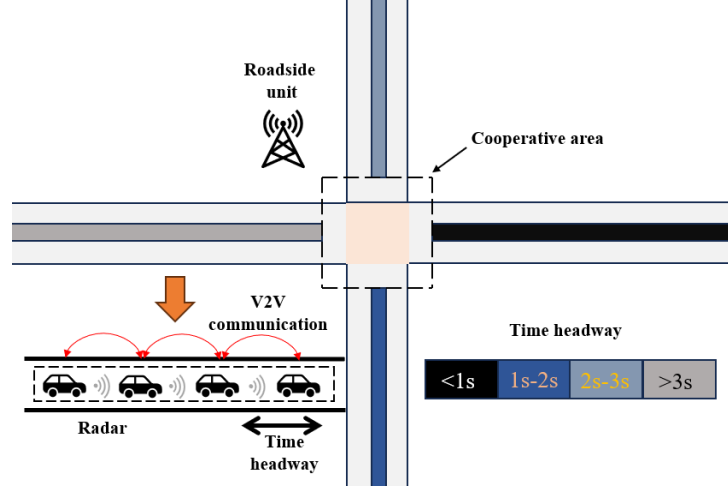


Figure 3.1: Illustration of headway control around an intersection.

specifications of AVs, and the maximum capacity of the specific road (Li and Chen, 2017). Furthermore, similar to the existing SO-DTA models (Shen and Zhang, 2014b), we have Observation 2 hold, indicating that SO-DTA may be achieved under different headway settings.

Observation 2. There can be multiple AV headway settings yielding the SO-DTA in general networks.

Observation 2 inspires us to search for the second criterion in the SO-DTA problem. For AVs, safety is always an indispensable factor. The impact of headway on traffic safety is widely studied in the literature (Ayres et al., 2001; Fenton, 1979). For example, Vogel, 2003 demonstrates that headway is a suitable indicator of safety and suggests using headway to measure safety for enforcement purposes given that the small headway has the potential to create dangerous driving situations. Moreover, Biswas, 2022 states that the small headway corresponds to a high risk of crash due to the inadequate response time, and in contrast, the large headway allows drivers to adjust their vehicles to decrease the risk of crash. Therefore, we present Figure 3.2 to briefly illustrate a general relationship between the probability of accident and headway. The headway on each road should be between the minimum and maximum headway, and within the range, a larger headway could yield a lower probability of an accident, thus leading to safer traffic conditions. Although traffic safety is a complex issue in reality, often influenced by multiple factors including headway, this study will only consider the impact of headway on safety. Future work will expand on the influence of more factors.

In view of this, this work proposes a novel concept of the maximin headway of AVs for

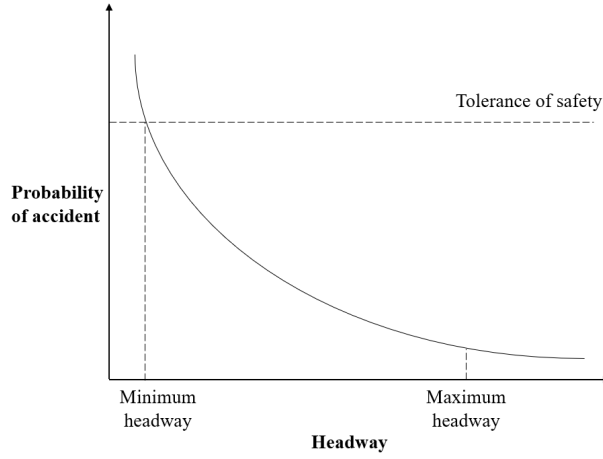


Figure 3.2: Illustrative relationship between the probability of accident and headway.

SO-DTA, as defined in Definition 3.1.1.

Definition 3.1.1 (Maximin Headway - Layman’s version). The largest headway setting for AVs that still achieves the SO-DTA is referred to as the maximin headway of AVs.

One can see that the maximin headway is the “largest” headway setting that achieves SO-DTA, meaning that such a headway setting is the “safest” under the SO-DTA. An illustration of the maximin headway is also presented in Figure 3.3. Existing literature mainly focuses on the solutions to SO-DTA (Qian et al., 2012; Zhang and Qian, 2020), while this work discusses the “best” solution among the non-unique solutions of SO-DTA for AVs, which we believe is interesting and unique. In particular, differences between minimum headway and maximin headway provide policy implications for road segment design and AV deployment.

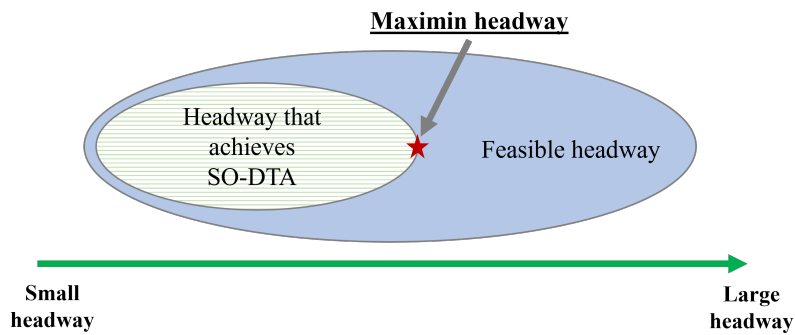


Figure 3.3: Illustration of the maximin headway for SO-DTA.

To summarize, this work proposes a maximin headway control framework that achieves SO-DTA. Specifically, we first explore how the headway of AVs would affect the road fundamental

diagram and develop a headway-dependent double queue model (HDQ), which is proved to be a generalized form of the double queue model (Ma et al., 2014). Then we formulate the network-wide traffic dynamics with SO-DTA for AVs, followed by the rigorous proof that Observation 1 and Observation 2 hold. Then we formally define the maximin headway control for SO-DTA, propose a provable approach to solve it analytically, and show the uniqueness of the maximin headway. Lastly, numerical experiments are conducted on the small and large networks to verify our proofs.

The contributions of this work are summarized as follows:

- The headway-dependent double queue (HDQ) model is developed to depict the effects of AV headway on network-wide traffic dynamics.
- It is analytically proved that the SO-DTA can be achieved under minimum headway.
- It first time proposes and solves the maximin headway control of AVs for SO-DTA in general dynamic networks.

3.2 Modeling headway-dependent flow dynamics

In this section, we propose the continuous-time form of the headway-dependent dynamic traffic flow. The continuous-time formulation is generally used in DNL, such as the continuous-time form of LTM (Han, Piccoli, and Szeto, 2016; Jin, 2015), continuous DQ (Ma et al., 2014) continuous-time point queue (PQ) model (Ban et al., 2012). To this end, we first present the headway-dependent fundamental diagram (HFD) to capture the effects of time headway on FD. Then we develop the headway-dependent double queue (HDQ) model and prove that HDQ is a generalized form of continuous-time DQ (Ma et al., 2014). Lastly, the network-wide flow dynamics are modeled through a series of linear constraints.

3.2.1 Notations

The road network is represented as a directed graph $G(V, E)$, where V is the set of all nodes including dummy nodes and E is the set of all links including O-D connectors. We denote $\tilde{R}$ as the set of dummy origins and $\tilde{S}$ as the set of dummy destinations. Naturally, we have $\tilde{R} \subset V$ and $\tilde{S} \subset V$. The frequently used notations are stated in Table A.1 in Appendix A.1.

3.2.2 Headway-dependent fundamental diagram

This section explores how different headways would affect the fundamental diagram. Without loss of generality, it is assumed that the headway and length of AV are homogeneous. To be specific, headway can be categorized into space and time headway. The space headway refers to the distance between the rear bumper of one vehicle and the front bumper of the vehicle following it; time headway refers to the time interval that elapses from the moment the rear of one vehicle passes a fixed point until the front bumper of following vehicle passes the same point (Guo and Yue, 2012; Zhu and Zhu, 2018). The space headway is not sufficient to represent the safety between AVs, while the time headway has been widely used in the literature regarding the AV control (Becker et al., 2022; Chen, Liang, et al., 2020). The time headway in this study is capable of measuring the probability of accidents by considering both the spacing and speed of AVs, which serves as a desirable control variable that trades off efficiency and safety.

Then we present the fundamental diagram for AVs based on the results by Zhou and Zhu, 2020b. Considering link (i, j) , we denote $\Delta_{x_{i,j}}$ as the spacing of AVs in link (i, j) , $v_{i,j}$ as the speed of AVs on link (i, j) , $h_{i,j}$ as the headway of AVs in link (i, j) and L as the length of each AV. The relationship between the headway $h_{i,j}$ and spacing $\Delta_{x_{i,j}}$ can be described as follows:

$$\Delta_{x_{i,j}} = h_{i,j}v_{i,j} + L.$$

Besides, the relationship between the spacing $\Delta_{x_{i,j}}$ and density $\rho_{i,j}$ is shown as follows:

$$\Delta_{x_{i,j}} = \frac{1}{\rho_{i,j}}.$$

The speed $v_{i,j}$ should be limited within free-flow speed $v_{i,j}^f$ as follows:

$$0 \leq v_{i,j} \leq v_{i,j}^f.$$

Combining the above equations, we have Equation 3.1 to represent the headway-dependent fundamental diagram (HFD) of AVs for $(i, j) \in E \setminus (L_R \cup L_S)$ and $t \in [0, T]$, where L_R and L_S represent the set of origin and destination connectors, respectively.

$$\begin{aligned}
f_{i,j}(t) &= \begin{cases} v_{i,j}^f \rho_{i,j}(t), & 0 \leq \rho_{i,j}(t) \leq \frac{1}{h_{i,j}(t)v_{i,j}^f + L} \\ \frac{1 - \rho_{i,j}(t)L}{h_{i,j}(t)}, & \frac{1}{h_{i,j}(t)v_{i,j}^f + L} \leq \rho_{i,j}(t) \leq \frac{1}{L} \end{cases} \\
&= \min \left\{ v_{i,j}^f \rho_{i,j}(t), \frac{1 - \rho_{i,j}(t)L}{h_{i,j}(t)} \right\}, \quad 0 \leq \rho_{i,j}(t) \leq \frac{1}{L}
\end{aligned} \tag{3.1}$$

where $\rho_{i,j}(t)$ and $h_{i,j}(t)$ are the density of the flow area and headway of link (i, j) at time t , respectively. Figure 3.4 presents the effect of headway on flow by HFD based on Equation 3.1, where $h^1 < h^2 < h^3 < h^4 < h^5$. This figure is also validated by the experimental results from Shi and Li, 2021, which empirically validates the fundamental diagram by collecting GPS trajectories of a platoon of AVs.

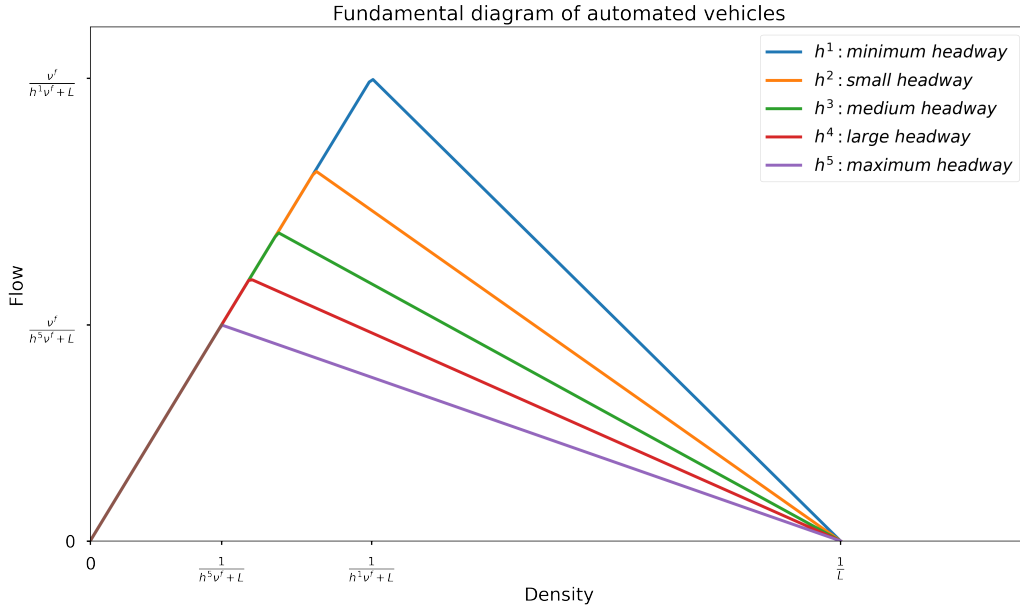


Figure 3.4: Fundamental diagram of AVs with different headways.

As shown in Figure 3.4, headway could affect the HFD of AVs in the following aspects:

- Scope of the free-flow region: a smaller headway corresponds to a larger scope of the free-flow region.
- Upper bound of flow: a smaller headway corresponds to a larger upper bound of flow.

- Wave travel speed in congested states: a smaller headway corresponds to a higher wave travel speed in congested states.

The wave travel time in congested states is to capture how the traffic conditions in the downstream link boundary would influence traffic conditions in the upstream link boundary (Yperman, 2007). Suppose $\tau_{i,j}^w(t)$ denotes the wave travel time in congested states on link (i, j) at time t , representing the used time traveling from the exit to entrance when reaching the entrance at time t . If the fundamental diagram is fixed, we would have a constant wave travel time in congested states $\tau_{i,j}^w = \frac{L_{i,j}}{v_{i,j}^w}$, where $v_{i,j}^w$ is the wave travel speed in congested states on link (i, j) and $L_{i,j}$ is the length of link (i, j) . However, in HFD, Equation 3.1 indicates that the wave travel speed in congested states $v_{i,j}^w(t)$ in HFD is time-dependent given a dynamic headway $h_{i,j}(t)$ and we formulate $v_{i,j}^w(t)$ in the following equation as the slope of HFD in congested states.

$$v_{i,j}^w(t) = \frac{L}{h_{i,j}(t)}$$

Then Equation 3.2 indicates that the wave travel time in congested states should satisfy Equation 3.2 for $(i, j) \in E \setminus (L_R \cup L_S)$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$\int_{t-\tau_{i,j}^w(t)}^t v_{i,j}^w(\hat{t}) d\hat{t} = \int_{t-\tau_{i,j}^w(t)}^t \frac{L}{h_{i,j}(\hat{t})} d\hat{t} = L_{i,j}, \quad (3.2)$$

where $L_{i,j}$ is the length of link (i, j) .

3.2.3 Headway-dependent double queue model

This section proposes a dynamic link model considering time-dependent AV headway, namely the headway-dependent double queue (HDQ) model. We further prove that HDQ is a generalized form of the double queue (DQ) model.

In HDQ, each link is divided into two areas: flow area and buffer area. The flow area depicts the traffic flow dynamics behavior and captures the upstream queue in the link, and we suppose the traffic flow would propagate based on HFD. The buffer area captures the downstream queue in the link and the maximum number of queued vehicles is limited based on the holding capacity in the buffer area (Guo and Ban, 2020). An illustration of the HDQ is shown in Figure 3.5. One can see that the HDQ is inspired by the MTTM and DQ model, and it shares similarities with

the Spatial Queue model (Gawron, 1998).

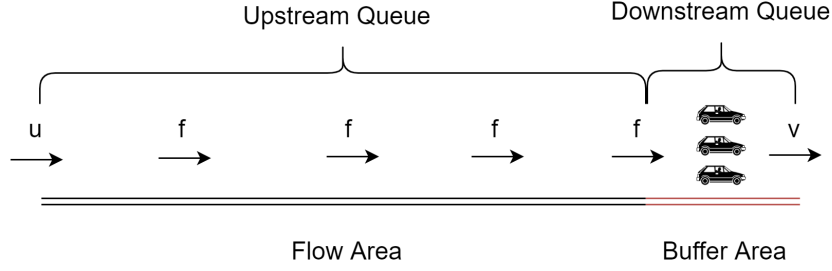


Figure 3.5: Flow area and buffer area in a link.

The HDQ has the following characteristics that are different from the DQ and MTTM:

- The MTTM overlooks the possible congestion in the flow area, while the HDQ considers both upstream and downstream queues.
- Comparing to the DQ, the HDQ makes use of the variable f to represent the effect of headway control on the traffic states and network conditions.

The cumulative flow for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$ is defined as follows:

$$U_{i,j}^{s'}(t) = \int_0^t u_{i,j}^{s'}(\hat{t}) d\hat{t}, \quad V_{i,j}^{s'}(t) = \int_0^t v_{i,j}^{s'}(\hat{t}) d\hat{t}, \quad F_{i,j}^{s'}(t) = \int_0^t f_{i,j}^{s'}(\hat{t}) d\hat{t},$$

where $U_{i,j}^{s'}(t)$, $V_{i,j}^{s'}(t)$ and $F_{i,j}^{s'}(t)$ are the cumulative inflow of the flow area, the cumulative outflow of the buffer area and cumulative inflow at the boundary of the flow area and buffer area of link (i, j) with destination s' at time t , respectively. $u_{i,j}^{s'}(t)$, $v_{i,j}^{s'}(t)$ and $f_{i,j}^{s'}(t)$ are the inflow of the flow area, outflow of the buffer area and inflow at the boundary of the flow area and buffer area with destination s' of link (i, j) at time t , respectively.

Then the downstream queue and upstream queue in HDQ for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$ are presented as follows:

$$q_{i,j}^{\mathcal{D},s'}(t) = \int_0^t f_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^t v_{i,j}^{s'}(\hat{t}) d\hat{t} + q_{i,j}^{\mathcal{D},s'}(0) = F_{i,j}^{s'}(t) - V_{i,j}^{s'}(t), \quad (3.3)$$

$$q_{i,j}^{\mathcal{U},s'}(t) = \int_0^t u_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^{t-\tau_{i,j}^w(t)} f_{i,j}^{s'}(\hat{t}) d\hat{t} + q_{i,j}^{\mathcal{U},s'}(0) = U_{i,j}^{s'}(t) - F_{i,j}^{s'}(t - \tau_{i,j}^w(t)) \quad (3.4)$$

where $q_{i,j}^{\mathcal{D},s'}(t)$ and $q_{i,j}^{\mathcal{U},s'}(t)$ are the downstream queue and upstream queue with destination s' of link (i, j) at time t , respectively.

Proposition 1 presents the relationship between the DQ and HDQ, indicating that HDQ could be reduced to DQ under its assumption.

Proposition 1 (Reduction of HDQ to DQ). HDQ is a generalized form of DQ by incorporating the influence of headway; HDQ could be reduced to the DQ under the assumption of DQ.

Proof. See Appendix A.2. □

We further have Equations 3.5 and 3.6 hold, meaning that traffic flow dynamics variables are the summation of their destination-specified variables for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$q_{i,j}^{\mathcal{D}}(t) = \sum_{s'} q_{i,j}^{\mathcal{D},s'}(t), \quad q_{i,j}^{\mathcal{U}}(t) = \sum_{s'} q_{i,j}^{\mathcal{U},s'}(t), \quad \rho_{i,j}(t) = \sum_{s'} \rho_{i,j}^{s'}(t), \quad (3.5)$$

$$u_{i,j}(t) = \sum_{s'} u_{i,j}^{s'}(t), \quad f_{i,j}(t) = \sum_{s'} f_{i,j}^{s'}(t), \quad v_{i,j}(t) = \sum_{s'} v_{i,j}^{s'}(t), \quad (3.6)$$

where $q_{i,j}^{\mathcal{D}}(t)$ and $q_{i,j}^{\mathcal{U}}(t)$ are the downstream queue and upstream queue of link (i, j) at time t , respectively. $u_{i,j}(t)$, $v_{i,j}(t)$ and $f_{i,j}(t)$ are the inflow of the flow area, outflow of the buffer area and inflow at the boundary of the flow area and buffer area of link (i, j) at time t , respectively. $\rho_{i,j}(t)$ and $\rho_{i,j}^{s'}(t)$ are the density of the flow area and the density of the flow area with destination s' of link (i, j) at time t , respectively.

Assuming the density is homogeneous in flow area, the change rate of density is presented in Equation 3.7 for $(i, j) \in E \setminus (L_R \cup L_S)$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$\dot{\rho}_{i,j}^{s'}(t) = \frac{u_{i,j}^{s'}(t) - f_{i,j}^{s'}(t)}{L_{i,j}} \quad (3.7)$$

Additionally, Equation 3.8 and 3.9 indicate that the downstream queue, upstream queue, inflow rate, and outflow rate are bounded by their respective capacity for $(i, j) \in E \setminus (L_R \cup L_S)$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$q_{i,j}^{\mathcal{D}}(t) \leq \bar{Q}_{i,j}^{\mathcal{D}}, \quad q_{i,j}^{\mathcal{U}}(t) \leq \bar{Q}_{i,j}^{\mathcal{U}}, \quad (3.8)$$

$$u_{i,j}(t) \leq \bar{C}_{i,j}^{\mu}, \quad v_{i,j}(t) \leq \bar{C}_{i,j}^{\nu}, \quad (3.9)$$

where $\bar{Q}_{i,j}^{\mathcal{D}}$ and $\bar{Q}_{i,j}^{\mathcal{U}}$ are the downstream queue capacity and upstream queue capacity, respectively; $\bar{C}_{i,j}^{\mu}$ and $\bar{C}_{i,j}^{\nu}$ are the inflow and outflow capacity of link (i, j) at time t , respectively. We further have Proposition 2 indicating that the impact to the maximal flow rate of the link is dependent on the HDQ.

Proposition 2 (Headway-dependent flow rate bounds). The upper bounds of the inflow rate $u_{i,j}(t)$ and outflow rate $v_{i,j}(t)$ are dependent on the upstream queue $q_{i,j}^{\mathcal{U}}(t)$ and downstream queue $q_{i,j}^{\mathcal{D}}(t)$, respectively, and their relation is shown as follows.

$$\begin{aligned} u_{i,j}(t) &\leq \begin{cases} \bar{C}_{i,j}^{\mu}, & q_{i,j}^{\mathcal{U}}(t) < \bar{Q}_{i,j}^{\mathcal{U}} \\ \min \{f_{i,j}(t - \tau_{i,j}^w(t)), \bar{C}_{i,j}^{\mu}\}, & q_{i,j}^{\mathcal{U}}(t) = \bar{Q}_{i,j}^{\mathcal{U}} \end{cases} \\ v_{i,j}(t) &\leq \begin{cases} \bar{C}_{i,j}^{\nu}, & q_{i,j}^{\mathcal{D}}(t) > 0 \\ \min \{f_{i,j}(t), \bar{C}_{i,j}^{\nu}\}, & q_{i,j}^{\mathcal{D}}(t) = 0 \end{cases} \end{aligned}$$

Proof. See Appendix A.3. □

3.2.4 Dynamic network traffic model

This section models the network-wide effects of AV headway control by developing a dynamic network model.

The flow conservation at nodes is stated in Equation 3.10 for $i \in N \setminus (\tilde{R} \cup \tilde{S})$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$\sum_p v_{p,i}^{s'}(t) = \sum_j u_{i,j}^{s'}(t) \quad (3.10)$$

As the node model is not the main contribution of this work, we adopt the simplified model that only considers the conservation law, while other advanced node models can be also incorporated if necessary. For example, node models could be developed in macroscopic levels to reduce congestion combined with AVs (Li, Cao, et al., 2023), prevent the conflicting movement at intersections (Zhu and Ukkusuri, 2015) and eliminate the traffic-holding phenomenon (Jabari, 2016). Similarly, our proposed framework could also incorporate the node model and apply cor-

responding strategies at intersections to avoid conflicts, such as traffic signal control (Yu et al., 2018).

The link conservation is ensured through the HDQ model. Equation 3.11 presents the non-negative constraint of all the variables involved for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$.

$$\rho_{i,j}^{s'}(t), u_{i,j}^{s'}(t), f_{i,j}^{s'}(t), v_{i,j}^{s'}(t), q_{i,j}^{\mathcal{D},s'}(t), q_{i,j}^{\mathcal{U},s'}(t) \geq 0 \quad (3.11)$$

Besides, the number of AVs in link (i, j) with destination s' at time t could be represented as the sum of the number of AVs in the flow area and buffer area, as shown in Equation 3.12.

$$\begin{aligned} N_{i,j}^{s'}(t) &= \int_0^t u_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^t v_{i,j}^{s'}(\hat{t}) d\hat{t} \\ &= \int_0^t u_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^t f_{i,j}^{s'}(\hat{t}) d\hat{t} + \int_0^t f_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^t v_{i,j}^{s'}(\hat{t}) d\hat{t} \\ &= L_{i,j} \rho_{i,j}^{s'}(t) + q_{i,j}^{\mathcal{D},s'}(t) \end{aligned} \quad (3.12)$$

The initial condition of the network is stated in Equation 3.13 for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$, which guarantees the whole traffic network is empty before loading.

$$\rho_{i,j}^{s'}(0), q_{i,j}^{\mathcal{D},s'}(0) = 0 \quad (3.13)$$

The end condition constraint is presented in Equation 3.14 for $(i, j) \in E$, $s' \in \tilde{S}$ and $t \in [0, T]$, which guarantees the whole traffic network is empty in the end.

$$q_{i,j}^{\mathcal{D},s'}(T) = 0, \rho_{i,j}(T) \leq \frac{1}{L_{i,j}} \quad (3.14)$$

Importantly, the change rate of link density can be represented by $\dot{\rho}_{i,j}(t) = -\frac{v_{i,j}^f \rho_{i,j}(t)}{L_{i,j}}$, $t \geq t_0$ based on Equations 3.5 to 3.7 when $u_{i,j}(t) = 0$. This represents the scenario in which no vehicles will enter the link when $\rho_{i,j}(t_0) > 0$. $\rho_{i,j}(t)$ has the shape of the negative exponential function, and hence it will take infinite time to make $\rho_{i,j}(t) = 0$. In order to avoid this situation, we set the end constraint of the flow area in link (i, j) as $\rho_{i,j}(T) \leq \frac{1}{L_{i,j}}$, which means that we suppose the flow area is empty when the number of vehicles in flow area is smaller than 1 at the end time T . Furthermore, we could also set a large positive constant c as $\rho_{i,j}(T) \leq \frac{1}{cL_{i,j}}$ to ensure the vehicles in the total network is close to 0 at end time T .

Equation 3.15 depicts the number of AVs waiting at origins for $r' \in \tilde{R}$, $s' \in \tilde{S}$ and $t \in [0, T]$:

$$w_{r',r}^{s'}(t) = \int_0^t d_{r',s'}(\hat{t}) d\hat{t} - \int_0^t v_{r',r}^{s'}(\hat{t}) d\hat{t} \quad (3.15)$$

where $w_{r',r}^{s'}(t)$ represents the number of AVs waiting at origins in the origin connectors (r', r) with destination s' at time t . $v_{r',r}^{s'}(t)$ is the outflow of origin connectors (r', r) with destination s' at time t . $d_{r',s'}(t)$ represents the travel demand of O-D pair (r', s') at time t . The demand is restricted in the time interval $[0, T_1]$, so we have $d_{r',s'}(t) = 0$ for $T_1 < t \leq T$.

Additionally, Equation 3.16 limits the minimum and maximum headway for $(i, j) \in E \setminus (L_R \cup L_S)$, $s' \in \widetilde{S}$ and $t \in [0, T]$.

$$h_{i,j;t}^{min} \leq h_{i,j}(t) \leq h_{i,j;t}^{max} \quad (3.16)$$

where $h_{i,j;t}^{min}$ and $h_{i,j;t}^{max}$ are the minimum and maximum headway of link (i, j) at time t . $h_{i,j;t}^{min}$ represents the tolerance of safety. If the headway $h_{i,j}(t)$ is less than $h_{i,j;t}^{min}$, the driving condition in a link would be highly dangerous. $h_{i,j;t}^{max}$ depicts the bound of V2V communication. If the headway $h_{i,j}(t)$ is larger than $h_{i,j;t}^{max}$, the V2V communication may be damaged.

3.3 Optimal headway control

In this section, we first formulate the continuous and discretized headway control for SO-DTA. Then we prove that the minimum headway control could achieve SO-DTA, and it is not the unique solution.

To search for the SO-DTA, there are also two types of formulations in the community of dynamic traffic assignment (DTA): path-based formulation and link-based formulation. The two formulations are actually equivalent, and we can solve either one to obtain the SO-DTA conditions. Specifically, the link flow and path flow can be converted to each other through the path-link incidence matrix or Dynamic Assignment Ratio (DAR) matrix (Ma and Qian, 2018). Both formulations consider the utility changes due to the change of link-specific headway. In this work, we adopt the link-based formulation as it can facilitate analytical analysis and solution algorithms, where the headway control is directly applied in each link, then the link flow/cost would be influenced, thus each path flow/cost is also impacted under the link-specific headway control. Therefore, the network-wide effect of link-specific time headway control could be ana-

lyzed.

3.3.1 Continuous-time formulation

The objective of the SO-DTA aims to minimize total travel time (TTT), which equals to the sum of the travel time on road networks and the waiting time at origins, as represented by Equation 3.17. The decision variables include traffic dynamics $\mathbf{x}$ and headway $\mathbf{h}$.

$$\min_{\mathbf{x}, \mathbf{h}} TTT = \int_0^T \left(\sum_{r'} \int_0^t v_{r',r}(\hat{t}) d\hat{t} - \sum_{s'} \int_0^t u_{s,s'}(\hat{t}) d\hat{t} \right) dt + \int_0^T \sum_{(r',s')} w_{r',r}^{s'}(t) dt \quad (3.17)$$

subject to:

- Headway-dependent fundamental diagram constraints: *Eqs. (3.1)-(3.2)*
- Headway-dependent double queue constraints: *Eqs. (3.3)-(3.9)*
- Flow conservation constraints: *Eq. (3.10)*
- Non-negative constraints: *Eq. (3.11)*
- Initial constraints: *Eq. (3.13)*
- End constraints: *Eq. (3.14)*
- Other constraints: *Eqs. (3.15)-(3.16)*

3.3.2 Discretized-time formulation

This section discretizes the formulation proposed in Section 3.3.1. We set $T = N\Delta_t$, where N represents the number of time intervals and Δ_t represents the length of a time interval. A summary of the notations is shown in Table 3.1.

Notation	Description
$\rho_{i,j}(k)$	Density of the flow area of link (i, j) in the k th time interval
$\rho_{i,j}^{s'}(k)$	Density of the flow area of link (i, j) with destination s' in the k th time interval
$q_{i,j}^D(k)$	Downstream queue of link (i, j) in the k th time interval
$q_{i,j}^{D,s'}(k)$	Downstream queue of link (i, j) with destination s' in the k th time interval
$q_{i,j}^U(k)$	Upstream queue of link (i, j) in the k th time interval
$q_{i,j}^{U,s'}(k)$	Upstream queue of link (i, j) with destination s' in the k th time interval
$u_{i,j}(k)$	Inflow of the flow area of link (i, j) in the k th time interval
$u_{i,j}^{s'}(k)$	Inflow of the flow area of link (i, j) with destination s' in the k th time interval
$f_{i,j}(k)$	Inflow at the boundary of the flow area and buffer area of link (i, j) in the k th time interval
$f_{i,j}^{s'}(k)$	Inflow at the boundary of the flow area and buffer area of link (i, j) with destination s' in the k th time interval
$v_{i,j}(k)$	Outflow of the buffer area of link (i, j) in the k th time interval
$v_{i,j}^{s'}(k)$	Outflow of the buffer area of link (i, j) with destination s' in the k th time interval
$h_{i,j}(k)$	Headway of automated vehicles on link (i, j) in the k th time interval
$w_{r',r}^{s'}(k)$	Number of automated vehicles waiting in the origin connectors (r', r) with destination s' in the k th time interval
$n_{i,j}^w(k)$	Wave travel time in congested states on link (i, j) in the k th time interval

Table 3.1: Notations in the discretized formulation.

We are now ready to present the discretized version of the headway-dependent SO-DTA formulation as follows:

$$\min_{\mathbf{x}, \mathbf{h}} TTT = \sum_{r'} \sum_{k=1}^N \Delta_t \left(\sum_{l=1}^k \Delta_t v_{r',r}(l) \right) - \sum_{s'} \sum_{k=1}^N \Delta_t \left(\sum_{l=1}^k \Delta_t u_{s,s'}(l) \right) + \sum_{(r',s')} \sum_{k=1}^N \Delta_t w_{r',r}^{s'}(k) \quad (3.18)$$

subject to

1. Headway-dependent fundamental diagram constraints for $(i, j) \in E \setminus (L_R \cup L_S)$, $s' \in \tilde{S}$ and $1 \leq k \leq N$.

$$f_{i,j}(k) = \min \left\{ v_{i,j}^f \rho_{i,j}(k), \frac{1 - \rho_{i,j}(k)L}{h_{i,j}(k)} \right\}, \quad 0 \leq \rho_{i,j}(k) \leq \frac{1}{L} \quad (3.19)$$

The discretized form of Equation 3.2 representing the wave travel time in congested states is shown as follows.

$$\sum_{l=k-n_{i,j}^w(k)}^k \Delta_t \frac{L}{h_{i,j}(l)} \leq L_{i,j} \quad (3.20)$$

$$\sum_{l=k-n_{i,j}^w(k)-1}^k \Delta_t \frac{L}{h_{i,j}(l)} \geq L_{i,j} \quad (3.21)$$

For the convenience of searching the maximin headway defined in Section 3.4, we assume that wave travel time in congested states $n_{i,j}^w(k)$ is only determined by the headway $h_{i,j}(k)$ in Equation 3.22 and 3.23. This relaxation would not influence the properties and definition of both optimal headway and maximin headway. In the detailed proof, Propositions 3 to 6 could also be achieved under Equations 3.20 and 3.21. The relaxation is only designed to convert the proposed SO-DTA problem into two LPs to solve the maximin headway in Algorithm 1.

$$n_{i,j}^w(k) \Delta_t \frac{L}{h_{i,j}(k)} \leq L_{i,j} \quad (3.22)$$

$$(n_{i,j}^w(k) + 1) \Delta_t \frac{L}{h_{i,j}(k)} \geq L_{i,j} \quad (3.23)$$

2. Headway-dependent double queue constraints constraints for $s' \in \widetilde{S}$ and $1 \leq k \leq N$. We limit $(i, j) \in E$ in Equations 3.24 to 3.27 and $(i, j) \in E \setminus (L_R \cup L_S)$ in Equations 3.29 to 3.30.

$$q_{i,j}^{\mathcal{D},s'}(k) = \sum_{l=0}^k \Delta_t f_{i,j}^{s'}(l) - \sum_{l=0}^k \Delta_t v_{i,j}^{s'}(l) + q_{i,j}^{\mathcal{D},s'}(0) \quad (3.24)$$

$$q_{i,j}^{\mathcal{U},s'}(k) = \sum_{l=0}^k \Delta_t u_{i,j}^{s'}(l) - \sum_{l=0}^{k-n_{i,j}^w(k)} \Delta_t f_{i,j}^{s'}(l) + q_{i,j}^{\mathcal{U},s'}(0) \quad (3.25)$$

$$q_{i,j}^{\mathcal{D}}(k) = \sum_{s'} q_{i,j}^{\mathcal{D},s'}(k); \quad q_{i,j}^{\mathcal{U}}(k) = \sum_{s'} q_{i,j}^{\mathcal{U},s'}(k); \quad \rho_{i,j}(k) = \sum_{s'} \rho_{i,j}^{s'}(k) \quad (3.26)$$

$$u_{i,j}(k) = \sum_{s'} u_{i,j}^{s'}(k); \quad f_{i,j}(k) = \sum_{s'} f_{i,j}^{s'}(k); \quad v_{i,j}(k) = \sum_{s'} v_{i,j}^{s'}(k) \quad (3.27)$$

$$\rho_{i,j}^{s'}(k) = \rho_{i,j}^{s'}(k-1) + \Delta_t \frac{u_{i,j}^{s'}(k) - f_{i,j}^{s'}(k)}{L_{i,j}} \quad (3.28)$$

$$q_{i,j}^{\mathcal{D}}(k) \leq \bar{Q}_{i,j}^{\mathcal{D}}; \quad q_{i,j}^{\mathcal{U}}(k) \leq \bar{Q}_{i,j}^{\mathcal{U}} \quad (3.29)$$

$$u_{i,j}(k) \leq \bar{C}_{i,j}^{\mu}; \quad v_{i,j}(k) \leq \bar{C}_{i,j}^{\nu} \quad (3.30)$$

3. Flow conservation constraints for $i \in N \setminus (\widetilde{R} \cup \widetilde{S})$, $s' \in \widetilde{S}$ and $1 \leq k \leq N$.

$$\sum_p v_{p,i}^{s'}(k) = \sum_j u_{i,j}^{s'}(k) \quad (3.31)$$

4. Non-negative Constraint for $(i, j) \in E$, $s' \in \widetilde{S}$ and $1 \leq k \leq N$.

$$\rho_{i,j}^{s'}(k), u_{i,j}^{s'}(k), f_{i,j}^{s'}(k), v_{i,j}^{s'}(k), q_{i,j}^{\mathcal{D},s'}(k), q_{i,j}^{\mathcal{U},s'}(k) \geq 0 \quad (3.32)$$

4. Initial condition constraint for $(i, j) \in E$ and $s' \in \widetilde{S}$.

$$\rho_{i,j}(0), q_{i,j}^{\mathcal{D},s'}(0) = 0 \quad (3.33)$$

5. End condition constraint for $(i, j) \in E$ and $s' \in \widetilde{S}$.

$$q_{i,j}^{\mathcal{D},s'}(N) = 0, \rho_{i,j}(N) \leq \frac{1}{L_{i,j}} \quad (3.34)$$

6. Other constraints for $(i, j) \in E \setminus (L_R \cup L_S)$, $r' \in \widetilde{R}$, $s' \in \widetilde{S}$ and $1 \leq k \leq N$, where $d_{r',s'}(k)$ represents the travel demand of O-D pair (r', s') in the k th time interval and $N_1 = \frac{T_1}{\Delta_t}$.

$$w_{r',r}^{s'}(t) = \sum_{l=0}^k \Delta_t d_{r',s'}(l) - \sum_{l=0}^k \Delta_t v_{r',r}^{s'}(l) \quad (3.35)$$

$$h_{i,j;k}^{min} \leq h_{i,j}(k) \leq h_{i,j;k}^{max} \quad (3.36)$$

Proposition 3 presents the feasibility of the proposed optimal headway control problem in Equation 3.18, where N_I , $N^{r',s'}$, $P^{r',s'}$, $N_{i,j}^{r',s'}$ and $\delta_{i,j}^{Pr',s'}$ are defined in Appendix A.4.

Proposition 3 (Feasibility). If $T > T_I + \sum_{(r',s')} n^{r',s'} \sum_{(i,j)} T_{i,j}^{r',s'} \delta_{i,j}^{Pr',s'}$, the optimal headway control problem in Equation 3.18 is always feasible.

Proof Sketch. It is obvious that if we choose a large enough time horizon N , the problem is always feasible, which means all AVs would arrive at their destinations before N . So the basic idea for the proof is to find a large enough N and construct a feasible solution under N . Details are presented in Appendix A.4.

We define the minimum total travel time and optimal headway of the discretized form of

SO-DTA for Equations 3.18 to 3.36 in Definition 3.3.1. $\Omega_{\mathbf{h}}$ represents the set of flow dynamics $\mathbf{x}$ such that $\mathbf{x}$ and $\mathbf{h}$ constitute a feasible solution of SO-DTA. $TTT_{\mathbf{h}}$ represents the minimum TTT if we set the headway variable $\mathbf{h}$ as an exogenous input.

Definition 3.3.1. Suppose $\Omega_{\mathbf{h}} = \{\mathbf{x} \mid \mathbf{x} \text{ and } \mathbf{h} \text{ satisfy Equations 3.19; 3.22 to 3.36}\}$, and $TTT(\mathbf{x})$ represents the total travel time under $\mathbf{x}$ as Equation 3.18 is only related with flow dynamics $\mathbf{x}$. Then we denote $TTT_{\mathbf{h}} = \min\{TTT(\mathbf{x}) \mid \mathbf{x} \in \Omega_{\mathbf{h}}\}$ and define the minimum total travel time and optimal headway as follows:

- Minimum total travel time (minimum TTT). TTT^* is the minimum total travel time if $TTT^* = \min\{TTT_{\mathbf{h}} \mid h_{i,j;k}^{min} \leq h_{i,j}(k) \leq h_{i,j;k}^{max}; (i, j) \in E \setminus (L_R \cup L_S); 1 \leq k \leq N\}$.
- Optimal headway. $\mathbf{h}$ is the optimal headway if $TTT_{\mathbf{h}} = TTT^*$; $\mathbf{h}$ and $\mathbf{x}$ is the optimal solution of SO-DTA problem if $\mathbf{x} \in \{\mathbf{x} \mid TTT(\mathbf{x}) = TTT_{\mathbf{h}} = TTT^*; \mathbf{x} \in \Omega_{\mathbf{h}}\}$.

The proposed SO-DTA problem is nonlinear programming due to Equations 3.19 and Equations 3.22 to 3.23. However, if we set the headway variable $\mathbf{h}$ as an exogenous input, the problem would become a linear programming (LP) in Appendix A.5.1. The corresponding sensitivity-based solution algorithm is presented in Appendix A.5.2.

However, for the SO-DTA problem, the sensitivity analysis approach has a large computation cost and it is hard to guarantee the convergence to the global optimal solution. Proposition 4 states that the minimum TTT can be achieved under the minimum headway setting, providing us with a direct way to reach the state of SO-DTA. On top of that, Proposition 5 states that it is possible to have multiple AV headway settings that yield the same SO-DTA conditions.

Proposition 4 (Minimum headway achieves SO-DTA). Supposing that $\mathbf{h}^{min} = \{h_{i,j}(k) = h_{i,j;k}^{min} \mid (i, j) \in E \setminus (L_R \cup L_S); 1 \leq k \leq N\}$, for $\forall \mathbf{h}$ satisfies Equation 3.36, we have

$$TTT_{\mathbf{h}^{min}} = TTT^* \leq TTT_{\mathbf{h}}.$$

Proof Sketch. For any feasible solution $\mathbf{x}$ and $\mathbf{h}$ of the proposed SO-DTA formulation, we only change a smaller $h_{i,j}(k)$ in $\mathbf{h}$ and keep all other headway in $\mathbf{h}$ unchanged. The new headway is denoted as $\mathbf{h}^*$. Then we prove that we could always find a feasible $\mathbf{x}^*$ under $\mathbf{h}^*$ and the total

travel time of $\mathbf{x}^*$ is less than or equal to the total travel time of $\mathbf{x}$ under original headway $\mathbf{h}$. When we traverse all links and time intervals, it is proved that SO-DTA could be achieved under the minimum headway. Details are presented in Appendix A.6.

Proposition 5 (Non-uniqueness).

$\forall \mathbf{h}$ s.t. $TTT_{\mathbf{h}} = TTT^*$, if $\exists \mathbf{x} \in \Omega_{\mathbf{h}}$ s.t. $TTT(\mathbf{x}) = TTT_{\mathbf{h}} = TTT^*$ and $\exists (i, j) \in E \setminus (L_R \cup L_S)$ and $k \in [1, N]$ s.t. $v_{i,j}^f \rho_{i,j}(k) < \frac{1-\rho_{i,j}(k)L}{h_{i,j}(k)}$, then $\exists \mathbf{h}^1 \neq \mathbf{h}$ s.t. $\mathbf{x} \in \Omega_{\mathbf{h}^1}$ and $TTT_{\mathbf{h}^1} = TTT(\mathbf{x}) = TTT_{\mathbf{h}} = TTT^*$

Proof. See Appendix A.7. □

Proposition 4 aligns with Observation 1, and Proposition 5 proves Observation 2. Then we have two natural questions arise:

- If we decide to control AVs with minimum headway in reality, would it cause other concerns (e.g., safety)?
- Is there a “best” headway among all the optimal headway given the non-uniqueness of SO-DTA solutions?

Both questions motivate us to search for a better headway control setting that still achieves the SO-DTA, which is presented in the following section.

3.4 Maximin headway control

In this section, we define the maximin headway as the “largest” headway setting among all the optimal headway settings that achieve the SO-DTA. The problem of maximin headway control is formulated and solved analytically.

3.4.1 Maximin headway control formulation

We first rigorously define the maximin headway in Definition 3.4.1, where $\mathbf{h}^*$ is a vector consisting of the time-dependent headway in different links. For the purpose of further converting the

proposed SO-DTA problem into a LP, we use the 1-norm to represent the largest headway. The definition of maximin headway could also be extended by using other norms for future study.

Definition 3.4.1 (Maximin Headway). $\mathbf{h}^*$ is called the maximin headway if it satisfies the following two conditions:

- SO-DTA: $TTT_{\mathbf{h}^*} = TTT^*$, meaning that $\mathbf{h}^*$ is the optimal headway for SO-DTA.
- Largest headway: $\forall \mathbf{h} \text{ s.t. } TTT_{\mathbf{h}} = TTT^*$, we have $\|\mathbf{h}\|_1 \leq \|\mathbf{h}^*\|_1$, where $\|\cdot\|_1$ is the 1-norm.

Definition 3.4.1 aligns with Definition 3.1.1 with more rigor and mathematical details. We formulate a nonlinear optimization to search for the maximin headway in Equation 3.37, where $TTT^* = TTT_{\mathbf{h}^{min}}$ represents the optimal value of TTT under minimum headway as denoted in Proposition 4.

$$\begin{aligned}
& \max_{\mathbf{h}} \quad \mathbf{1}^T \mathbf{h} \\
& \text{s.t.} \quad TTT = TTT^* \\
& \quad \text{Headway-dependent fundamental diagram constraints (3.19); (3.22) – (3.23)} \\
& \quad \text{Headway-dependent double queue constraints (3.24) – (3.30)} \\
& \quad \text{Flow conservation constraints (3.31)} \\
& \quad \text{Non-negative Constraints (3.32)} \\
& \quad \text{Initial condition constraints (3.33)} \\
& \quad \text{Initial condition constraints (3.34)} \\
& \quad \text{Other constraints (3.35) – (3.36)}
\end{aligned} \tag{3.37}$$

We note it is a direct extension to replace $\mathbf{1}^T$ with a weight vector $\mathbf{w}^T$ to represent the importance of headway on different roads. Proposition 6 discusses the condition of the uniqueness of maximin headway of Formulation 3.37.

Proposition 6 (Uniqueness of maximin headway). Suppose $\Phi = \bigcup_{\mathbf{h}} \Omega_{\mathbf{h}}$ is the set of feasible

traffic dynamics $\mathbf{x}$,

If $\exists \mathbf{x}^* \in \Phi$ s.t. $TTT(\mathbf{x}^*) < TTT(\mathbf{x})$, for $\forall \mathbf{x} \in \Phi$ and $\mathbf{x} \neq \mathbf{x}^*$, then $\exists \mathbf{h}^*$ satisfying $TTT_{\mathbf{h}^*} = TTT(\mathbf{x}^*)$, we have $\|\mathbf{h}^*\|_1 > \|\mathbf{h}\|_1$, $\forall \mathbf{h}$ satisfying $TTT_{\mathbf{h}} = TTT(\mathbf{x}^*)$ and $\mathbf{h}^* \neq \mathbf{h}$.

$\iff$ *If $\mathbf{x}^*$ with $TTT(\mathbf{x}^*) = TTT^*$ is unique, the optimal solution of Formulation 3.37 is unique.*

Proof. See Appendix A.8. □

3.4.2 Maximin headway control solution algorithm

This section proposes a maximin headway control solving algorithm with the idea of exploring the maximum increase of headway to maintain the states of SO-DTA achieved by minimum headway setting.

Proposition 7 provides us a method to derive a “larger” optimal headway given the current optimal headway.

Proposition 7 (Larger headway for SO-DTA). Suppose $\mathbf{h}^*$ satisfies Equation 3.36 and $\mathbf{x}^* \in \Omega_{\mathbf{h}^*}$ s.t. $TTT(\mathbf{x}^*) = TTT_{\mathbf{h}^*} = TTT^*$. If $\mathbf{h}$ satisfies following constraints:

$$\rho_{i,j}^*(k) v_{i,j}^f h_{i,j}(k) \leq 1 - \rho_{i,j}^*(k)L \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \quad (3.38)$$

$$L_{i,j} h_{i,j}(k) \geq \Delta_t L n_{i,j}^{w,*}(k) \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \quad (3.39)$$

$$L_{i,j} h_{i,j}(k) \leq \Delta_t L n_{i,j}^{w,*}(k) + \Delta_t L \quad (3.40)$$

$$h_{i,j}^*(k) \leq h_{i,j}(k) \leq h_{i,j;k}^{max} \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \quad (3.41)$$

$$h_{i,j}(k) = h_{i,j}^*(k) \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \geq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \quad (3.42)$$

where $\rho_{i,j}^*(k)$ and $n_{i,j}^{w,*}(k) \in \mathbf{x}^*$, then we have $\mathbf{x}^* \in \Omega_{\mathbf{h}}$, $TTT_{\mathbf{h}} = TTT(\mathbf{x}^*) = TTT^*$ and $\|\mathbf{h}^*\|_1 < \|\mathbf{h}\|_1$.

Proof. It is obvious that if headway variables $\mathbf{h}$ satisfies the Equations 3.38 to 3.42, then $\mathbf{h}$ and $\mathbf{x}^*$ would constitute the optimal solution of SO-DTA problem in Equation 3.18 to 3.36, so we have $TTT_{\mathbf{h}} = TTT(\mathbf{x}^*) = TTT^*$ and $\|\mathbf{h}^*\|_1 < \|\mathbf{h}\|_1$. $\square$

Propositions 4 and 7 intuitively inspire us a way to derive the maximin headway control and it mainly consists of the following two steps: 1) we first derive the flow dynamics $\mathbf{x}^*$ for SO-DTA under the minimum headway $\mathbf{h}^{min}$; 2) then we explore the “largest” headway setting based on $\mathbf{x}^*$ by remaining the same TTT, which implies the “largest” headway to still achieve SO-DTA and also represents the “safest” headway among optimal headway.

To this end, we propose an algorithm to solve the proposed maximin headway control in Algorithm 1 as follows and prove the correctness of Algorithm 1 under the condition of the existence of unique optimal flow dynamics in Proposition 8, which states that the output of Algorithm 1 is the optimal solution of Formulation 3.37 if $\exists \mathbf{x}^* \in \Phi$ s.t. $TTT(\mathbf{x}^*) < TTT(\mathbf{x}), \forall \mathbf{x} \in \Phi$ and $\forall \mathbf{x} \neq \mathbf{x}^*$. One can see from Algorithm 1 that, once the traffic condition $\mathbf{x}^*$ is determined, we can run a linear programming (LP) to solve for the maximin headway, and hence the solution process is computationally efficient.

Algorithm 1 Maximin Headway Control Solving Algorithm

Input: parameters in Table A.1.

- 1: Solve the discretized SO-DTA formulation (Equations 3.18 to 3.19; Equations 3.22 to 3.36) under minimum headway $\mathbf{h}^{min}$ to derive the optimal flow dynamics $\mathbf{x}^*$ with $TTT(\mathbf{x}^*) = TTT_{\mathbf{h}^{min}} = TTT^*$. To be specific, we can solve Formulation 3.43 to obtain $\mathbf{x}^*$, and details of the formulation is presented in Appendix A.5.1.

$$\begin{aligned}
 \mathbf{x}^* &= \arg \min_{\mathbf{x}} TTT = \mathbf{P}^T \mathbf{x} \\
 \text{s.t. } & \mathbf{A}(\mathbf{h}^{min})\mathbf{x} = \mathbf{B}(\mathbf{h}^{min}) \\
 & \mathbf{C}(\mathbf{h}^{min})\mathbf{x} \leq \mathbf{D}(\mathbf{h}^{min})
 \end{aligned} \tag{3.43}$$

- 2: Solve Formulation 3.44 under the flow dynamics $\mathbf{x}^*$ and derive the maximin headway $\mathbf{h}^*$, where $\rho_{i,j}^*(k)$ and $n_{i,j}^{w,*}(k) \in \mathbf{x}^*$.

$$\begin{aligned}
& \max_{\mathbf{h}} \quad \mathbf{1}^T \mathbf{h} \tag{3.44} \\
& s.t. \quad \rho_{i,j}^*(k) v_{i,j}^f h_{i,j}(k) \leq 1 - \rho_{i,j}^*(k)L \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \\
& \quad L_{i,j} h_{i,j}(k) \geq \Delta_t L n_{i,j}^{w,*}(k) \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \\
& \quad L_{i,j} h_{i,j}(k) \leq \Delta_t L n_{i,j}^{w,*}(k) + \Delta_t L \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \\
& \quad h_{i,j}^*(k) \leq h_{i,j}(k) \leq h_{i,j;k}^{max} \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \leq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\} \\
& \quad h_{i,j}(k) = h_{i,j}^*(k) \quad i, j, k \in \left\{ (i, j; k) \mid v_{i,j}^f \rho_{i,j}^*(k) \geq \frac{1 - \rho_{i,j}^*(k)L}{h_{i,j}^*(k)} \right\}
\end{aligned}$$

Output: maximin headway $\mathbf{h}^*$ and optimal flow dynamics $\mathbf{x}^*$

Proposition 8 (Correctness of Algorithm 1). The output $\mathbf{h}^*$ of Algorithm 1 is the optimal solution of Formulation 3.37 if $\exists \mathbf{x}^* \in \Phi$ s.t. $TTT(\mathbf{x}^*) < TTT(\mathbf{x}), \forall \mathbf{x} \in \Phi$ and $\forall \mathbf{x} \neq \mathbf{x}^*$.

Proof. Proposition 8 can be directly obtained by combining Proposition 4, 6, and 7. □

3.5 Numerical experiments

The proposed model and propositions are validated on both a small and a large-scale Hong Kong network in this section.

3.5.1 A small network

In the small network, we first solve for the optimal headway control and maximin headway control to demonstrate that the maximin headway control achieves the SO-DTA with large headway settings. Additionally, the sensitivity analysis for the maximin headway control is also conducted.

3.5.1.1 Settings

The small network consists of 5 nodes and 6 links, Nodes 1 and 2 are connected to the origins, and Node 5 is connected with the destination. We further set the total time horizon T as 90 minutes and the length of each time interval Δ_t as 5 minutes. It is worth noting that, in simulation-based microscopic models, small time steps (e.g., 5 seconds) could be used (Ben-Akiva et al., 2012; Sumalee et al., 2011; Zhang and Qian, 2020). However, for the DTA models using dynamic link models, numerous studies opt for a time interval of five minutes or greater (Bellei et al., 2005; Chan et al., 2021; Chiu et al., 2011; Janson, 1991). Besides, we would conduct the sensitivity analysis for Δ_t on the proposed maximin headway control framework in 3.5.1.2. The travel demand for each O-D pair is 50 *vehicles/mins* and the time horizon of demand T_1 is 40 minutes. Other parameters are illustrated in Appendix A.9.

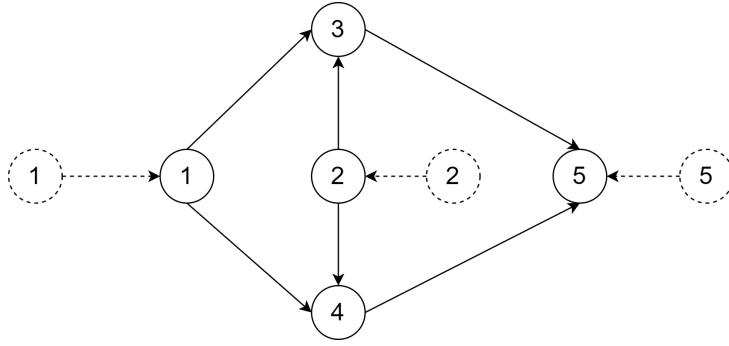


Figure 3.6: Small network.

3.5.1.2 Results for maximin headway control in small network

The three different scenarios are defined for comparison as follows:

- **Min-HW**: the headway of each link is directly set as the minimum headway.
- **SO-HW**: we directly solve for the SO-DTA by searching for the optimal headway, and the solution algorithm is based on the sensitivity-based algorithm (Algorithm 6) described in Appendix A.5.2.
- **Maximin-HW**: we solve the maximin headway by Algorithm 1.

Based on Proposition 4, we already know that the minimum headway setting would yield the SO-DTA, and the total travel time $TTT(\mathbf{h}^{min}) = 25,210$. Additionally, the dynamic inflow and

outflow of each link in the SO-DTA are presented in Figure 3.7. The path-based and link-based formulations are equivalent to SO-DTA. Utilizing the path-link incidence matrix or the Dynamic Assignment Ratio (DAR) matrix, we can obtain the path flow under SO-DTA from the optimal link flow as determined by Algorithm 1.

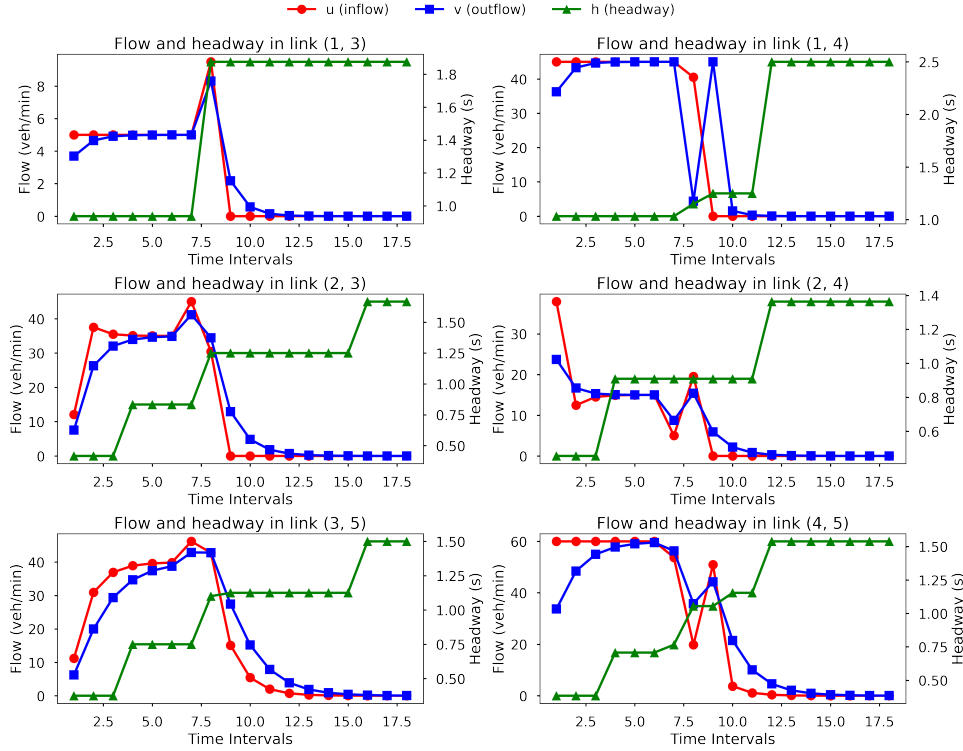


Figure 3.7: Inflow, outflow, and maximin headway of each link under SO-DTA.

The convergence curve of Algorithm 6 in **SO-HW** is shown in Figure 3.8, where the convergence criterion is that the difference of TTT in the latest two iterations is less than 1 minute. The sensitivity-based Algorithm 6 may converge to the local optimal and it does not guarantee to find the global optimal solution by Algorithm 6. However, it is proved that our proposed Algorithm 1 could analytically solve the SO-DTA problem to find the global optimal solution. Therefore, one can see that the sensitivity-based algorithm converges to $TTT = 31,342$ minutes, which is larger than 25,210 derived by Algorithm 1 in **Maximin-HW**. This verifies that the sensitivity-based method might not converge to the optimal solution of SO-DTA and demonstrates the merits of the proposed analytical algorithm in Algorithm 1.

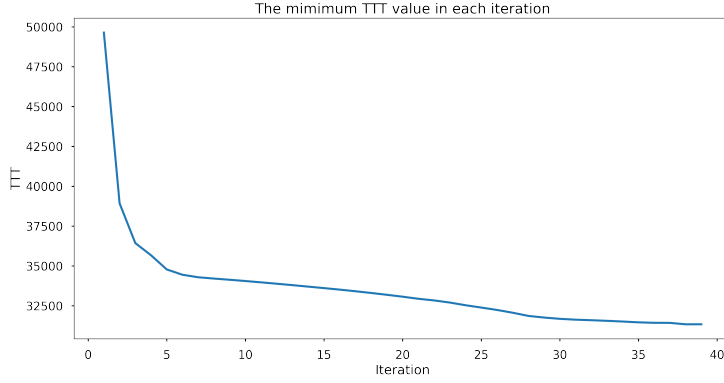


Figure 3.8: Convergence curve in terms of TTT by Algorithm 6.

As for the **Maximin-HW**, the TTT remains 25,210 minutes, meaning that the maximin headway still achieves the SO-DTA. Besides, to measure the improvement of maximin headway compared with minimum headway when considering the safety aspect, we define the ratio of maximin headway **rmh** in Equation 3.45, where $\mathbf{h}^* = \{h_{i,j}^*(k) \mid (i, j) \in E \setminus (L_R \cup L_S); 1 \leq k \leq N\}$ is the vector of maximin headway solved by Algorithm 1 and $\mathbf{h}^{min} = \{h_{i,j;k}^{min} \mid (i, j) \in E \setminus (L_R \cup L_S); 1 \leq k \leq N\}$ is the vector of minimum headway.

$$\mathbf{rmh} = \frac{\|\mathbf{h}^*\|_1}{\|\mathbf{h}^{min}\|_1} \quad (3.45)$$

The ratio of maximin headway **rmh** in the small network is 1.43, which indicates that the maximin headway is about 1.43 times as large as the minimum headway on average, demonstrating the merits of the maximin headway control. Table 3.2 presents the average maximin headway and minimum headway of each link over time. One can see that the gap between the maximin and minimum headway is significant, and the gap could also reflect the safety margin of each link. The time-dependent maximin headway is illustrated in Figure 3.7 and it could infer a rough relationship between the optimal flow dynamics and maximin headway. The link flow under SO-DTA is usually large when the maximin headway is small and vice versa. This result aligns with the real-world traffic phenomena. When the flow in a road or highway is small, it is reasonable to set a relatively large headway.

Link	Average Minimum Headway	Average Maximin Headway
1 $\rightarrow$ 3	0.969s	1.510s
1 $\rightarrow$ 4	0.997s	1.646s
2 $\rightarrow$ 3	0.783s	1.088s
2 $\rightarrow$ 4	0.742s	1.088s
3 $\rightarrow$ 5	0.753s	0.978s
4 $\rightarrow$ 5	0.828s	1.068s

Table 3.2: Maximin and minimum headway in the small network.

3.5.1.3 Sensitivity analysis for the maximin headway control

This section explores how different network configurations would affect the ratio of maximin headway. In particular, we study the sensitivity regarding the length of time interval, travel demand, and minimum headway.

Time interval. We first explore the influence of the length of time interval on the ratio of maximin headway. The length of time interval Δ_t varies from 2 to 12 minutes.

Figure 3.9 illustrates that the ratio of maximin headway remains stable when the length of time interval increases, implying that the current length of time interval is sufficient for the case study.

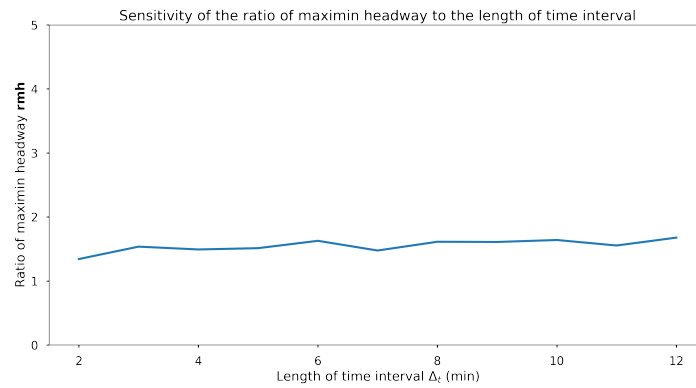


Figure 3.9: The ratio of maximin headway under different lengths of time intervals.

Travel demand. We conduct the sensitivity analysis regarding travel demand under different minimum headway requirements. We set a constant minimum headway for all links and vary the demand of each O-D pair from 30 to 70 veh/min under different minimum headway require-

ments, then we solve Algorithm 1 and derive the ratio of maximin headway, and the results are shown in Figure 3.10.

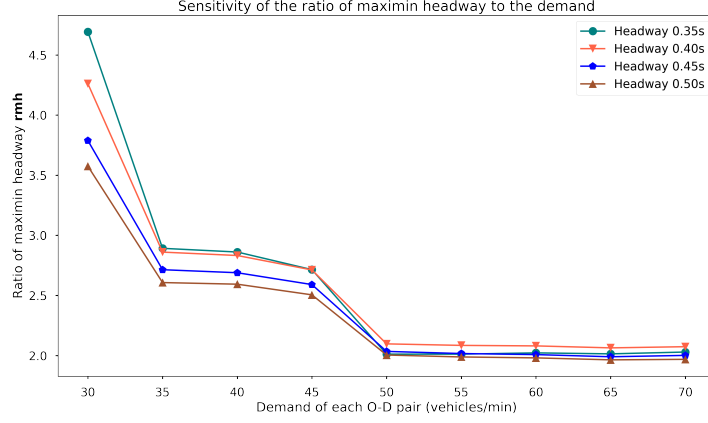


Figure 3.10: The ratio of maximin headway under different traffic demands.

One can see that the ratio of maximin headway decreases when the demand increases, indicating that the safety margin for SO-DTA will become marginal when more people travel.

Minimum headway. Then the sensitivity analysis for the minimum headway value is conducted. The minimum headway is constant for all links and we vary the value of minimum headway from 0.20s to 1.10s under different O-D demands and keep other parameters unchanged. Then we solve the Algorithm 1 and derive the ratio of maximin headway under different minimum headway values. Figure 3.11 shows how the the ratio of maximin headway would change corresponding to different minimum headway values.

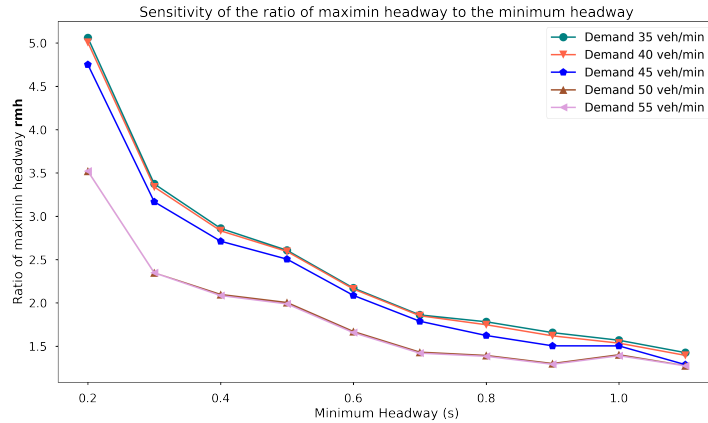


Figure 3.11: The ratio of maximin headway under different minimum headway settings.

One can see that in general, the ratio of maximin headway is relatively stabilized under different minimum headway, indicating that there is always room to enlarge the headway after achieving SO-DTA under different safety requirements (*i.e.*, minimum headway).

3.5.2 The Hong Kong network

In this section, we explore the effectiveness of the maximin headway control in real traffic conditions in the Hong Kong network.

3.5.2.1 Settings

We study the urban area of Kowloon Peninsula in Hong Kong to capture the real traffic conditions under headway control. The Hong Kong network consists of 55 nodes and 192 links, and the number of O-D pairs is 72, as referred to Figure 3.12. The total time horizon is 250 minutes and the length of time interval is set as 5 minutes. The minimum and maximum headway is set as a constant 0.5 seconds and 2.5 seconds, respectively. The link parameters and hour OD demand profile for the Hong Kong network are illustrated in Appendix A.10.

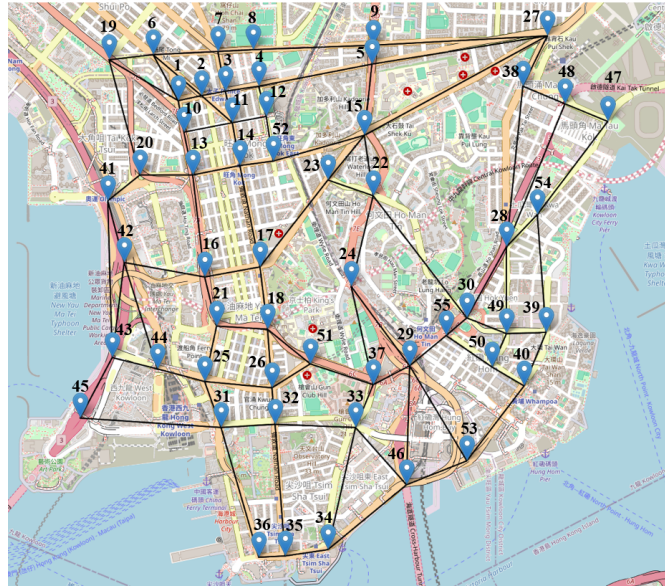


Figure 3.12: Overview of the Hong Kong network.

3.5.2.2 Experimental results

The results of **Min-HW** and **Maximin-HW** in Hong Kong network are illustrated as follows. Besides, we also present the high computation efficiency of Algorithm 1 in this large-scale network.

- **Min-HW**: we directly set the minimum headway for each link, and we have $TTT(\mathbf{h}_{min}) = 639,138$ minutes.
- **Maximin-HW**: The TTT is the same as that of the **Min-HW**, and the ratio of maximin headway $\mathbf{rmh}$ is 4.15, which indicates that the maximin headway control could significantly improve the minimum headway and also keeps SO-DTA.

In terms of the computation efficiency, the computation time of Algorithm 1 is about 2,802 seconds to solve the maximin headway. However, the sensitivity-based Algorithm 6 fails to solve the maximin headway as it requires impractical memory usage to compute the generalized inverse matrix $\mathbf{Q}^{-1}$ defined in Equation A.19. Therefore, our proposed framework of maximin control reveals its applicability in large-scale networks.

To this end, we plot the maximin headway against the link flow for all time intervals and links in Figure 3.13, where we set a homogeneous minimum headway as 0.5s and maximum headway and 2.5s for all time intervals and links as presented in Section 3.5.2. Due to different link parameters and queue capacity restrictions, the detailed free-flow and congestion states of each link vary.

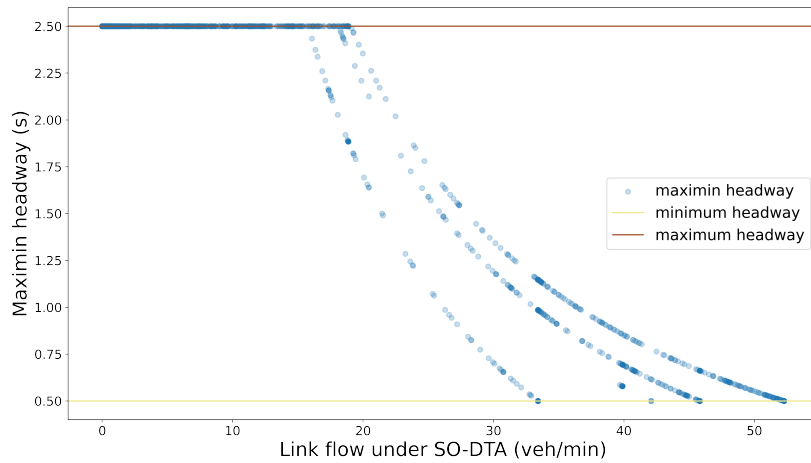


Figure 3.13: The distribution of maximin headway and link flow under SO-DTA.

The maximin headway control framework is more effective in the free-flow states when the link flow is relatively small, where the maximin headway could be approximating the maximum headway under SO-DTA. However, when the link flow increases, the maximin headway would decrease, but it is still larger than the minimum headway, representing the merits of maximin headway control in ensuring a higher safety level under SO-DTA. When the link flow is very high, the maximin headway would equal the minimum headway, indicating that there is no room to improve safety that still achieves SO-DTA. Overall, the maximin headway is highly dependent on the link flow under SO-DTA and, in general, maximin headway is higher under less congested links.

Chapter 4

The publicly-owned centralized platform for ride-hailing services

4.1 Introduction

The progression in mobile sensing technology and the widespread adoption of smartphones have significantly facilitated the extensive success of ride-hailing platforms. Enterprises like Uber and Didi Chuxing stand at the forefront of this massive transformation, fostering a robust digital hub that diligently bridges the gap between drivers and travelers. Uber's corporate trajectory exemplifies this revolutionary transformation, with expansive operations permeating 71 countries, and over 10,000 cities, and encompassing a staggering fleet of 3.5 million drivers. This expansive operational scope reflects a significant economic footprint, with the fiscal year 2022 witnessing projected aggregate values of reserved vehicular trips ranging between \$35 billion to \$51 billion (Dean, [2021](#)). The emergence of ride-hailing services has caused a major transformation in the transportation system. Data from October 2022 indicates that, on average, ride-hailing applications facilitated 622,700 vehicular trips each day, a figure that substantially surpasses the average daily count of 118,027 trips managed by traditional taxi services. This marked shift in consumer preference has paralleled a dramatic increase in the market share for ride-hailing services, which soared from a marginal 12% in January 2015 to a dominant 84% by October 2022 (Schneider, [2022](#)).

The rapid expansion of ride-hailing services has spurred research on the key stakeholders, including travelers, drivers, platform operators, and policymakers. These actors collectively in-

fluence the ecosystem, particularly the experiences of passengers and driver engagement, which affect supply and demand. Platform companies focus on profit maximization, while the rise of idle vehicles has worsened traffic congestion. In response, regulatory frameworks, such as subsidies (Mo et al., 2020; Zhao et al., 2023; Zhu, Xu, et al., 2021), congestion tolls (Li et al., 2021; Vignon et al., 2021), and service caps (Ke et al., 2021; Li et al., 2019; Liu et al., 2022), aim to mitigate the negative impacts. Understanding these measures is essential for policymakers to balance operational efficiency, improve system performance, and protect consumer interests, ultimately enhancing urban transportation sustainability.

Subsidies are widely used in the ride-hailing market to improve social welfare (Mo et al., 2020; Pantelidis et al., 2020; Zhao et al., 2023; Zhu, Xu, et al., 2021), increase platform profits (Guo and Huang, 2022; Tang et al., 2023; Zhu, Ke, and Wang, 2021), and ensure stable matching (Liu and Chow, 2023; Tafreshian and Masoud, 2020). However, research on the network-wide effects of spatial-temporal subsidies is limited. For example, previous studies focus on subsidies for local stability but overlook broader traffic impacts (Liu and Chow, 2023). Additionally, while Zhu, Ke, and Wang, 2021 model spatial-temporal subsidies using a mean-field Markov decision process, they don't fully capture large-scale traffic network dynamics. This study introduces a spatial-temporal subsidy model that integrates traffic assignment to better align with real-world conditions and explores different subsidy types for various ride-hailing behaviors.

The ownership structure of ride-hailing platforms significantly impacts the effectiveness of spatial-temporal subsidies aimed at improving services and promoting social welfare. Due to information asymmetry and competition between the government and platform operators, the government can only offer subsidies to platform operators, not directly to travelers. Previous studies, such as Zhu, Xu, et al., 2021, assume that government subsidies are limited to public transit users, excluding those using ride-hailing services on private platforms.

On the other hand, many researchers have proposed the concept of publicly-owned platforms to connect travelers and mobility operators, enabling the direct application of subsidies to travelers by the government (Liu and Chow, 2023; Tafreshian and Masoud, 2020). In practice, governments in numerous cities and countries are exploring the concept of operating ride-hailing platforms to achieve specific non-commercial objectives, distinguishing them from the popular platforms currently operated by third parties. For example, the Victorian government in Australia is actively considering the operation of ride-hailing apps to cater to the flexible

point-to-point transportation needs of individuals using wheelchairs and scooters (Dunn, 2016). Additionally, a recent state-supported endeavor in China encompassing a wide range of services, including ride-hailing, cargo trucking, road transport, railways, ferries, and flights referred to the website ¹. This platform aims to consolidate over 90% of the total capacity in the transportation market, with the overarching goal of providing user-friendly services while ensuring robust data protection and safeguarding personal privacy. This demonstrates the practical feasibility of designing publicly-owned platforms with network-wide spatial-temporal subsidies for travelers. Consequently, when aiming to implement spatial-temporal subsidies for travelers within a network, ride-hailing services administered by the government present a more viable option for the application of such subsidies compared to privately-owned ride-hailing platforms.

Dynamic traffic assignment (DTA) serves as a key mechanism for modeling network-wide traffic dynamics and the effects of applying spatial-temporal subsidy in publicly-owned ride-hailing platforms. This study utilizes a variant known as system optimal dynamic traffic assignment (SO-DTA) to discern optimal traffic states in a network. This approach allows us to draw significant insights into the optimal performance of a traffic network, primarily in terms of optimal objective value (i.e., the minimum social cost) and optimal solution value (i.e., flow dynamics and spatial-temporal subsidies) (Du and Ma, 2024). However, Shen and Zhang, 2014a states that the SO-DTA problem may have multiple optimal solutions and Tafreshian and Masoud, 2020 proposes the minimum subsidy problem that finds the minimum amount of subsidy required for stabilizing the matching in the platform. In this study, we formulate the publicly-owned ride-hailing platform with a spatial-temporal subsidy as a SO-DTA problem to maximize social welfare, which equals to minimizing the social cost. Considering multiple spatial-temporal subsidies may achieve minimum social cost, we develop the framework to obtain the optimal spatial-temporal subsidy representing the minimum amount of subsidy to achieve minimum social cost. Furthermore, we simultaneously account for the dynamics of both traveler and vehicle flows, thus enabling the modeling of shared automated vehicles (SAVs) with different occupancies and carpooling behaviors (Ma et al., 2020; Noruzoliaee and Zou, 2022).

Compared to human-driven vehicles (HVs), shared automated vehicles (SAVs) offer several advantages in SO-DTA: lower operational costs (Iacobucci et al., 2019; Tian et al., 2021), sim-

¹<https://www.reuters.com/business/autos-transportation/china-launch-state-backed-transport-platform-ride-hailing-trucking-2023-01-18/>

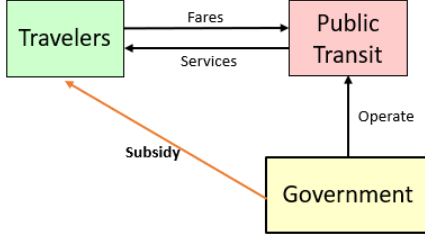
plified management (Guo et al., 2022; Narayanan et al., 2020), and enhanced safety (Fagnant and Kockelman, 2015; Teoh and Kidd, 2017). Additionally, SAVs can operate cooperatively by following route guidance provided by TNCs to maximize profit (Fan et al., 2023). Besides, SAVs have been shown and proven the possibility to replace personal vehicles representing the great potential for wide application in future mobility services including MaaS (Mobility as a Service) (Chen, Kockelman, and Hanna, 2016; Fagnant and Kockelman, 2014; Spieser et al., 2014). Most SAV studies are microscopic (Boesch et al., 2016; Hyland and Mahmassani, 2018; Loeb and Kockelman, 2019; Zhang et al., 2015), and there is little work based on dynamic traffic assignment (DTA). Levin, 2017 considers the congestion effects in formulating the SAVs routing problem as a SO-DTA problem. However, the operation and pricing perspectives are not considered.

In this study, we propose a novel insight for the regulation of government in dynamic traffic networks by designing the publicly-owned centralized platform (POCP) with a spatial-temporal subsidy, which receives order requests from travelers and sends the orders to the transportation network companies (TNCs). The key feature of the POCP is that it is owned by the government, and it aims to achieve the SO-DTA within a limited budget. This design allows the government to integrate management strategies into the POCP to enhance traffic efficiency. In particular, the POCP could adjust travelers' behaviors by subsidizing their trip fees with spatial-temporal discounts, thereby serving as a demand management tool for the dynamic networks and facilitating the achievement of SO-DTA. TNCs receive the orders from POCP and provide mobility services by routing SAVs. The summary of the structures of different platforms is illustrated in Figure 4.1.

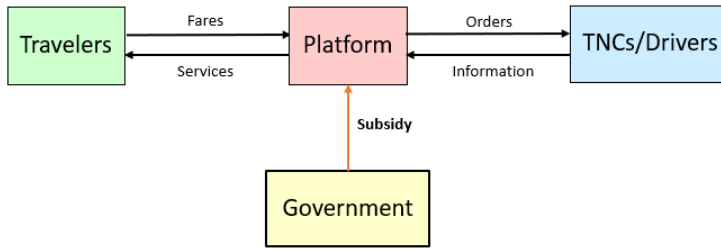
Furthermore, bi-level programming can be employed to formulate the POCP in this study, where the government operates the POCP to minimize social cost in the upper level, and TNCs are tasked with routing SAVs to maximize their profit in the lower level. The bi-level programming is effective in capturing the interactions of different stakeholders in the ride-hailing market. Yao and Zhang, 2024 designs an intermediary mobility-as-a-service (MaaS) platform formulated as bi-level programming, where the MaaS demand allocation is solved in the upper level and a multi-modal traffic equilibrium is determined in the lower level. Xi et al., 2022 proposes linear bi-level programming (LBP) where the agent in the upper level is the platform operator and the lower level is a two-sided market for travelers and transportation service providers

(TSPs). Algorithms for solving linear bi-level programming mainly include branch and bound algorithm (Fortuny-Amat and McCarl, 1981; Xi et al., 2022), extreme point-ranking algorithm (Papavassilopoulos, 1982) and complementary pivoting method (Candler and Townsley, 1982). However, when the lower problem is DTA problems, the LBP is generally complicated. There is a lack of effective solution algorithms with traceable analytical properties.

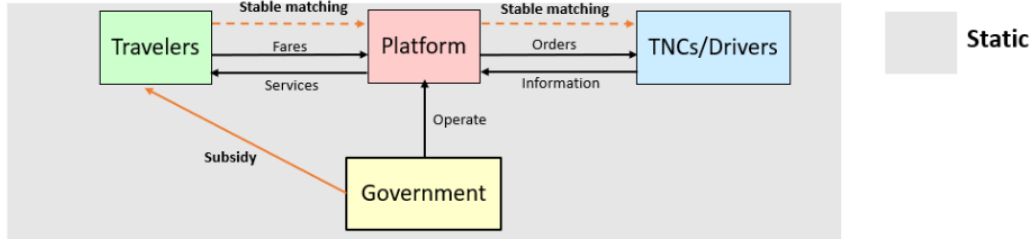
A. Public transit under government subsidy (Zhu et al.2020)



B. Privately-owned platform under government subsidy (Mo, Yu, and Chen 2020, Zhao et al 2023)



C. Publicly-owned platform under government subsidy (Tafreshian and Masoud 2020, Liu and Chow 2023)



D. Our work

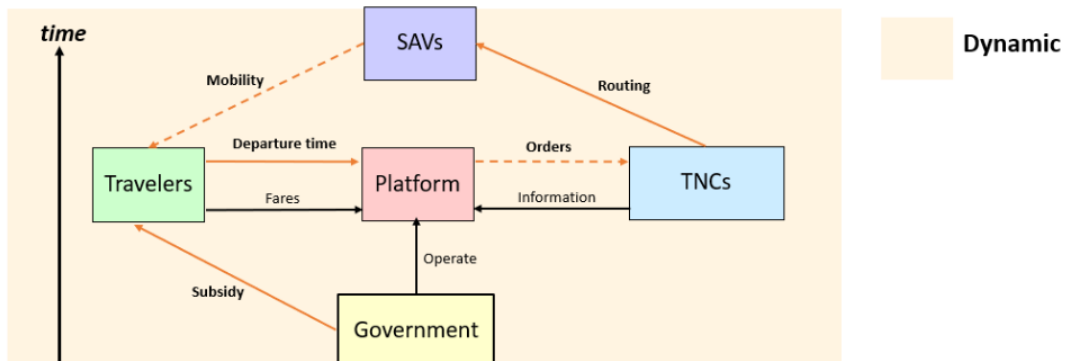


Figure 4.1: Comparisons of the structures of different platforms under government subsidy.

To summarize, this work proposes a novel insight to improve the current ride-hailing plat-

form and save social welfare in dynamic traffic networks from the government's perspective and designs a publicly-owned centralized platform (POCP). We formulate the operation of the POCP as a nonlinear bi-level programming where the government aims at minimizing the social cost by adjusting travelers' behaviors with the spatial-temporal discount in the upper level and TNC aims at maximizing its profit by providing mobility service of SAVs in the lower level. The bi-level problem can be decomposed by two sub-problems, where the first sub-problem could be solved by a series of LPs and the second sub-problem is a LP. Therefore, the solution process becomes tractable for large-scale networks and numerical experiments demonstrate the great potential of the design of POCP.

The contributions of this study are summarized as follows:

- This research proposes a design for a publicly-owned centralized platform (POCP) for shared automated vehicles (SAVs), which serves as an intermediary connecting travelers, and TNCs. The platform, operated by the government, aims to reduce social costs by subsidizing trips via spatial-temporal discounts. Conversely, TNCs focus on providing mobility services to travelers by optimally routing SAVs to maximize their profit.
- The study formulates a SO-DTA problem for the POCP as non-linear bi-level programming. This model incorporates both traveler and vehicle flows to capture network-wide traffic dynamics.
- We demonstrate that the proposed non-linear bi-level programming of POCP could be equivalently converted into two sub-problems with linear programming (LPs) to solve for the minimum social cost and minimum subsidy amount, respectively. The first LP could be obtained by the branch and bound algorithm. It provides us with a direct and efficient way to obtain the spatial-temporal subsidy to save the social cost using the minimum subsidy amount.

4.2 Problem settings

The POCP acts as an intermediary to connect travelers and TNCs. We would illustrate the roles of each stakeholder for POCP as follows. The government operates the platform to minimize the social cost and offers spatial-temporal subsidies to travelers by providing discounts on trip fees

in the POCP. These discounts are dependent on the destination, time of orders, and the type of SAVs.

Travelers, who are informed about the trip fees set by TNC and the discount set by the government in POCP, place trip orders on the platform. They pay the actual trip fare to the platform, which equals the trip fee multiplied by the discount. Additionally, travelers exhibit the following characteristics:

- Sensitivity to subsidies varies among travelers. Travel demand can be divided into regular demand and flexible demand. Regular demand pertains to travelers who would request an order at a specific time (denoted as r) regardless of the discounts. On the other hand, flexible demand is sensitive to discounts, and their requesting time may change due to varying travel costs at different times.
- Travelers have the option to utilize ride-sharing services and are divided into two groups: solo travelers and ride-sharing travelers (Li, Xu, et al., 2023; Ma et al., 2020, 2022). Solo travelers, who prefer not to use ride-sharing services, are the only occupants of a Shared Automated Vehicle (SAV). Ride-sharing travelers, on the other hand, opt for ride-sharing services. Accordingly, the type of SAV is defined as follows:
 - $x = 0$ represents an idle SAV.
 - $x = -1$ represents an SAV with a solo traveler (*e.g.*, UberX).
 - $x = 1, \dots, X$ represents an SAV with x ride-sharing travelers (*e.g.*, UberPool), where X is the maximum number of travelers in a SAV.

TNCs are alerted by the platform when travelers place orders, and they then provide mobility services by routing SAVs. The platform pays TNCs the total trip fees, which comprise the fares travelers pay and the subsidy offered by the government. Furthermore, the platform shares parking vacancy information with the TNCs for SAV parking selection, and TNCs remit parking fees to the platform.

4.3 Model

This section presents the formulation of the model. The flow dynamics for both SAVs and travelers are presented in Section 4.3.2, and the cost for different stakeholders is illustrated in Section

4.3.3. Lastly, we formulate the SO-DTA problem of POCP as a nonlinear bi-level programming in Section 4.3.4.

4.3.1 Notations

We represent the traffic network as a directed graph $G(\tilde{V}, \tilde{E})$, where $\tilde{V}$ is the set of all nodes including dummy nodes and $\tilde{E}$ is the set of all links including dummy links. We add the dummy nodes for origin nodes, destination nodes, and parking nodes. The dummy links are for connectors between actual networks and origins, destinations, and parking lots. We list the frequently used notations in Table 4.1, and all variables in Table 4.1 are non-negative.

Type	Notation	Description
Sets	V, E	Set of the nodes and links in traffic network, excluding the dummy nodes and dummy links.
	N, P	Set of the normal and parking nodes in traffic network
	O, S	Set of the origin and destination nodes in traffic network
	R, L	Set of the requesting order time and latest acceptable arrival time, where $R, L \subseteq 1, \dots, T$
Parameters	$\bar{C}_{i,j}^\mu, \bar{C}_{i,j}^\nu$	Inflow and outflow capacity of link (i, j)
	$\bar{Q}_{i,j}$	Upstream queue capacity of link (i, j)
	N_p	Number of SAVs at parking node p at initial
	$\bar{C}_p$	Parking capacity at parking node p
	X	Maximum number of travelers in a SAV
	T, Δ_t	Total time horizon and length of time interval
	$\tau_{i,j}^f, \tau_{i,j}^w$	Free-flow and shockwave travel time of link (i, j)
	λ	Sensitivity of flexible demand to the trip fee and discount
	η_x	Sensitivity of the trip fee and discount for taking a SAV of type x

	V_T, V_W, V_L	Value of travel, waiting and late time
	C_x	Trip fee set by TNC for taking type x of SAV
	C_S, C_{LP}	Travel cost of SAV; Late penalty
	C_p	Parking fee at parking node p
	$D_{o,s}^{\mathcal{E},S,l}(r)$	Regular demand of solo traveler with origin o , destination s and latest acceptable arrival time l at time r
	$D_{o,s}^{\mathcal{E},\mathcal{R},l}(r)$	Regular demand of ride-sharing traveler with origin o , destination s and latest acceptable arrival time l at time r
	$D_{o,s}^{\mathcal{F},S,l}[a, b]$	Total number of flexible demand of solo traveler with origin o , destination s and latest acceptable arrival time l in time interval $[a, b]$
	$D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b]$	Total number of flexible demand of ride-sharing traveler with origin o , destination s and latest acceptable arrival time l in time interval $[a, b]$
SAV-related Variables	$u_{i,j}(k), v_{i,j}(k)$	Inflow and outflow of SAVs of link (i, j) in the k th time interval
	$q_{i,j}^{\mathcal{U}}(k), q_{i,j}^{\mathcal{D}}(k)$	Upstream queue and downstream queue of SAVs of link (i, j) in the k th time interval
	$u_{i,j}^x(k), v_{i,j}^x(k)$	Inflow and outflow of SAVs with type x of link (i, j) in the k th time interval
	$q_{i,j}^{\mathcal{U},x}(k), q_{i,j}^{\mathcal{D},x}(k)$	Upstream queue and downstream queue of SAVs with type x of link (i, j) in the k th time interval
	$u_{p,p'}(k), v_{p,p'}(k)$	Inflow and outflow of SAVs of dummy parking link (p, p') in the k th time interval
	$n_p(k)$	Number of SAVs of parking node p in the k th time interval

Travelers ables	Vari-	$f_{i,j}^{x,s}(k), g_{i,j}^{x,s}(k)$	Inflow and outflow of travelers taking type x of SAVs with destination s of link (i, j) in the k th time interval
		$f_{i,j}^{x,s,r,l}(k), g_{i,j}^{x,s,r,l}(k)$	Inflow and outflow of travelers taking type x of SAVs and requesting order at time r with destination s with latest acceptable arrival time l of link (i, j) in the k th time interval
		$d_{o,s}^{S,l}(r), d_{o,s}^{R,l}(r)$	Demand of solo and ride-sharing travelers requesting order at time r with origin o , destination s and latest acceptable arrival time l
		$g_{o,o'}^{S,s,r,l}(k), g_{o,o'}^{R,s,r,l}(k)$	Outflow of solo and ride-sharing travelers requesting order at time r with origin o , destination s and latest acceptable arrival time l of dummy origin link (o, o') in the k th time interval
		$f_{s,s'}^{r,l}(k)$	Inflow of travelers requesting order at time r with latest acceptable arrival time l of dummy destination link (s, s') in the k th time interval
		$n_o^{S,s,r,l}(k), n_o^{R,s,r,l}(k)$	Number of solo and ride-sharing travelers requesting order at time r with destination s and latest acceptable arrival time l waiting at origin node o in the k th time interval
Discount ables	Vari-	$\alpha_s^x(r)$	Discount of taking SAVs with type x for orders with destination s and requesting order time r

Table 4.1: Notation list.

4.3.2 Dynamic network modeling for POCP

In this section, we model the traffic dynamics for both travelers and SAVs on general networks, and an overview of the link and node models is shown in Figures 4.2 and 4.3, respectively.

4.3.2.1 Traffic flow dynamics

We use the Double Queue (DQ) model to present the dynamic network loading process (Ma et al., 2014; Osorio et al., 2011). The dynamics of the DQ model are captured by the two dependent queues: the downstream queue and the upstream queue. The downstream and upstream queues of SAVs is presented in Equation 4.1 and 4.2 for $(i, j) \in E$, $x = -1, 0, 1, \dots, X$ and $k = 1, \dots, T$:

$$q_{i,j}^{\mathcal{D},x}(k) = \sum_{a=0}^{k-\tau_{i,j}^f} \Delta_t u_{i,j}^x(a) - \sum_{a=0}^k \Delta_t v_{i,j}^x(a), \quad (4.1)$$

$$q_{i,j}^{\mathcal{U},x}(k) = \sum_{a=0}^k \Delta_t u_{i,j}^x(a) - \sum_{a=0}^{k-\tau_{i,j}^w} \Delta_t v_{i,j}^x(a), \quad (4.2)$$

where $q_{i,j}^{\mathcal{D},x}(k)$ and $q_{i,j}^{\mathcal{U},x}(k)$ are the downstream queue and upstream queues of SAVs with type x of link (i, j) in the k th time interval, respectively; $u_{i,j}^x(k)$ and $v_{i,j}^x(k)$ are the inflow and outflow of SAVs with type x of link (i, j) in the k th time interval, respectively; $\tau_{i,j}^f$ and $\tau_{i,j}^w$ are the free-flow and shockwave travel time of link (i, j) , respectively.

Next, we would present double flow matching constraints to illustrate the quantitative relationship between travelers flow and SAVs flow in Equations 4.3 to 4.15. In a SAV with type $x > 0$, all ride-sharing travelers in the same SAV could have different origins and destinations.

The sum constraint of SAVs is presented in Equation 4.3 for $(i, j) \in E$ and $k = 1, \dots, T$:

$$u_{i,j}(k) = \sum_{x=-1}^X u_{i,j}^x(k); \quad v_{i,j}(k) = \sum_{x=-1}^X v_{i,j}^x(k); \quad q_{i,j}^{\mathcal{D}}(k) = \sum_{x=-1}^X q_{i,j}^{\mathcal{D},x}(k); \quad q_{i,j}^{\mathcal{U}}(k) = \sum_{x=-1}^X q_{i,j}^{\mathcal{U},x}(k), \quad (4.3)$$

where $u_{i,j}(k)$ and $v_{i,j}(k)$ are the inflow and outflow of SAVs of link (i, j) at the k th time interval, respectively; $q_{i,j}^{\mathcal{D}}(k)$ and $q_{i,j}^{\mathcal{U}}(k)$ are the downstream queue and upstream queue of link (i, j) at the k th time interval, respectively.

Equation 4.4 indicates the total inflow and outflow of travelers consists of the travelers with different destination in different types of SAVs when entering or leaving links, respectively, for $(i, j) \in E$, $x = -1, 1, \dots, X$, $s \in S$ and $k = 1, \dots, T$ as follows:

$$f_{i,j}^{x,s}(k) = \sum_{r \in R} \sum_{l \in L} f_{i,j}^{x,s,r,l}(k); \quad g_{i,j}^{x,s}(k) = \sum_{r \in R} \sum_{l \in L} g_{i,j}^{x,s,r,l}(k), \quad (4.4)$$

where $f_{i,j}^{x,s}(k)$ and $g_{i,j}^{x,s}(k)$ are the inflow and outflow of travelers taking type x of SAVs with destination s of link (i, j) at the k th time interval, respectively; $f_{i,j}^{x,s,r,l}(k)$ and $g_{i,j}^{x,s,r,l}(k)$ are the inflow and outflow of travelers taking type x of SAVs and requesting order at time r with latest acceptable arrival time l of link (i, j) at the k th time interval, respectively; Latest acceptable arrival time l refers to the case that if the actual arrival time k at destination is larger than l , then the TNC would pay the late cost for the time $k - l$ to travelers.

The quantitative relationship between travelers and SAVs is illustrated in Equation 4.5 for $(i, j) \in E$, $x = -1, 1, \dots, X$ and $k = 1, \dots, T$, as referred to Figure 4.2.

$$N(x)u_{i,j}^x(k) = \sum_{s \in S} f_{i,j}^{x,s}(k); \quad N(x)v_{i,j}^x(k) = \sum_{s \in S} g_{i,j}^{x,s}(k), \quad (4.5)$$

where $N(x) = x$ if $x = 1, \dots, X$ and $N(x) = 1$ if $x = -1$. For solo travelers, there would be only one traveler in a SAV with type $x = -1$. Conversely, for SAVs carrying ride-sharing travelers, SAV flow multiplied by the number of ride-sharing travelers in SAVs equals the traveler flow, where type x represents the number of ride-sharing travelers in a SAV when $x > 0$.

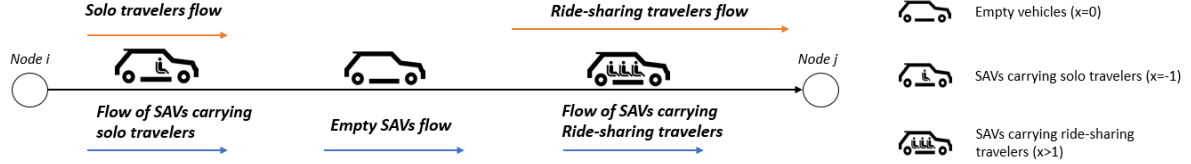


Figure 4.2: Overview of the double flow with POCP in link level.

Besides, the number of travelers of each link at any time should be greater than or equal to zero and this constraint is formulated in Equation 4.6 for $(i, j) \in E$, $s \in S$, $x = -1, 1, \dots, X$ and $k = 1, \dots, T$.

$$\sum_{a=1}^k \Delta_t f_{i,j}^{x,s}(a) - \sum_{a=1}^k \Delta_t g_{i,j}^{x,s}(a) \geq 0 \quad (4.6)$$

Travelers and SAVs demonstrate distinct flow conservation dynamics at various node types, including origin, destination, parking, and normal nodes, as illustrated in Figure 4.3. Below, we offer an in-depth analysis of these flow conservation constraints.

For travelers at node $j \in N \cup P$, where N and P are the sets of normal and parking nodes respectively, the flow conservation principle indicates that travelers will neither enter nor exit SAVs at node j in Equation 4.7 for $(i, j) \in E$, $s \in S$, $r \in R$, $l \in L$, $x = -1, 1, \dots, X$ and

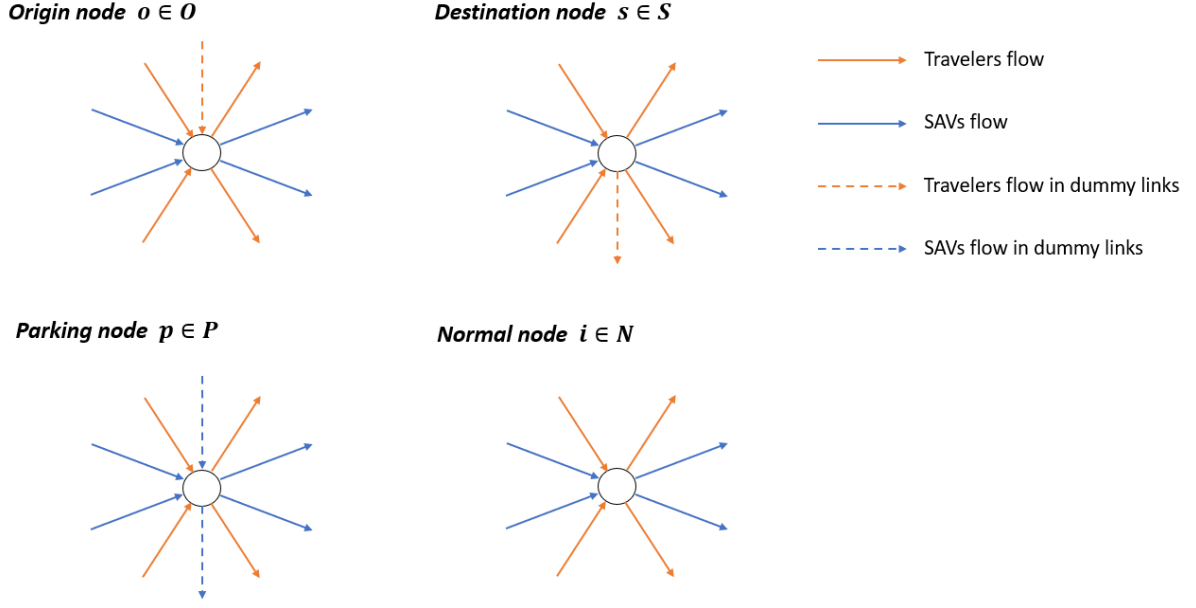


Figure 4.3: Overview of the double flow with POCP in node level.

$k = 1, \dots, T$.

$$\sum_i g_{i,j}^{x,s,r,l}(k) = \sum_m f_{j,m}^{x,s,r,l}(k) \quad (4.7)$$

At the origin node $o \in O$, the flow conservation constraints differ for solo travelers (ST) and ride-sharing travelers (RT). Solo travelers exclusively opt for Shared Autonomous Vehicles (SAVs) that do not carry other passengers (designated as SAVs of type $x = -1$), as delineated in Equation 4.8. Conversely, ride-sharing travelers are inclined to select SAVs that accommodate additional passengers (categorized as SAVs of type $x = 1, \dots, X$), as detailed in Equation 4.9 for $(i, j) \in E$, $s \in S$, $r \in R$, $l \in L$ and $k = 1, \dots, T$:

$$g_{o,o'}^{S,s,r,l}(k) + \sum_i g_{i,o}^{x=-1,s,r,l}(k) = \sum_j f_{o,j}^{x=-1,s,r,l}(k), \quad (4.8)$$

$$g_{o,o'}^{R,s,r,l}(k) + \sum_i \sum_{x=1}^X g_{i,o}^{x,s,r,l}(k) = \sum_j \sum_{x=1}^X f_{o,j}^{x,s,r,l}(k), \quad (4.9)$$

where $g_{o,o'}^{S,s,r,l}(k)$ and $g_{o,o'}^{R,s,r,l}(k)$ are outflow of solo and ride-sharing travelers requesting order at time r with origin o , destination s and latest acceptable arrival time l of dummy origin link (o, o') in the k th time interval, respectively.

At the destination node $s_0 \in S$, Equation 4.10 presents that travelers whose destination is not s_0 would not exit SAVs at node s_0 . Only those whose destination is s_0 would exit the SAVs at node s_0 , as shown in Equation 4.11 for $(i, j) \in E$, $s \in S$, $r \in R$, $l \in L$ and $k = 1, \dots, T$:

$$\sum_i \sum_{x \neq 0} g_{i,s}^{x,s,r,l}(k) = \sum_j \sum_{x \neq 0} f_{s,j}^{x,s,r,l}(k) \quad \text{for } s \neq s_0, \quad (4.10)$$

$$\sum_i \sum_{x \neq 0} g_{i,s}^{x,s,r,l}(k) = f_{s,s'}^{r,l}(k) \quad \text{for } s = s_0, \quad (4.11)$$

where $f_{s,s'}^{r,l}(k)$ is the inflow of travelers requesting order at time r with latest acceptable arrival time l of dummy destination link (s, s') in the k th time interval.

SAVs also have specific flow conservation constraints at different types of nodes. Particularly, at nodes $j \in O \cup D$, where O and D represent the set of origin and destination nodes respectively, Equation 4.12 implies that travelers choose to enter or exit a SAV at node j , potentially leading to a change in the type of SAV at node j .

$$\sum_i u_{i,j}(k) = \sum_m v_{j,m}(k). \quad (4.12)$$

At a normal node $j \in N$, the travelers would not enter or exit SAVs and SAVs would not also choose to park at node j . Therefore, the type of SAV does not change at node j , as represented by Equation 4.13 for $(i, j) \in E$ and $x = -1, \dots, X$.

$$\sum_i u_{i,j}^x(k) = \sum_m v_{j,m}^x(k) \quad (4.13)$$

At parking nodes $p \in P$, the flow conservation constraints for SAVs are represented by Equations 4.14 and 4.15. We assume that only idle SAVs ($x = 0$) will enter or leave the dummy parking link (p, p') for parking, as indicated in Equation 4.14. On the other hand, Equation 4.15 states that occupied SAVs ($x \neq 0$) would not choose to park at node p .

$$\sum_i v_{i,p}^{x=0}(k) + v_{p,p'}(k) = \sum_j u_{p,j}^{x=0}(k) + u_{p,p'}(k), \quad (4.14)$$

$$\sum_i v_{i,p}^x(k) = \sum_j u_{p,j}^x(k) \quad \text{for } x \neq 0, \quad (4.15)$$

where $u_{p,p'}(k)$ and $v_{p,p'}(k)$ are the inflow and outflow of SAVs of dummy parking link (p, p')

in the k th time interval, respectively.

The parking constraint is stated in Equation 4.16 to simulate the parking behavior for $p \in P$ and $k = 1, \dots, T$:

$$n_p(k) = N_p - \sum_{a=1}^k \Delta_t v_{p,p'}(a) + \sum_{a=1}^k \Delta_t u_{p,p'}(a), \quad (4.16)$$

where $n_p(k)$ represents the number of SAVs of parking node p in the k th time interval and N_p is the number of SAVs of parking node p at initial.

Equation 4.17 and 4.18 present the waiting constraints to represent the number of travelers waiting at the origin node as follows for $o \in O$, $s \in S$, $r \in R$, $l \in L$ and $k = 1, \dots, T$:

$$n_o^{S,s,r,l}(k) + \sum_{a=1}^k \Delta_t g_{o,o'}^{S,s,r,l}(a) = d_{o,s}^{S,l}(r), \quad (4.17)$$

$$n_o^{R,s,r,l}(k) + \sum_{a=1}^k \Delta_t g_{o,o'}^{R,s,r,l}(a) = d_{o,s}^{R,l}(r), \quad (4.18)$$

where $n_o^{S,s,r,l}(k)$ and $n_o^{R,s,r,l}(k)$ are the number of solo and ride-sharing travelers requesting order at time r with destination s and latest acceptable arrival time l waiting at origin o in the k th time interval, respectively; $d_{o,s}^{S,l}(r)$ and $d_{o,s}^{R,l}(r)$ are the demand of solo and ride-sharing travelers requesting order at time r with origin o , destination s and latest acceptable arrival time l , respectively.

The initial constraint is stated in Equation 4.19 representing the network is idle for SAVs and travelers at initial for $(i, j) \in E$.

$$q_{i,j}^U(0) = 0 \quad (4.19)$$

The end constraints are illustrated in Equations 4.20 to 4.22. Equation 4.20 and 4.21 present that no travelers are waiting at origin and the network is idle for travelers at the end. Besides, there are no SAVs in the network at the end, as illustrated in Equation 4.22 for $(i, j) \in E$, $s \in$

$S, r \in R$ and $l \in L$.

$$n_o^{S,s,r,l}(T) = 0; \quad n_o^{S,s,r,l}(T) = 0 \quad (4.20)$$

$$\sum_{k=1}^T \sum_{x \neq 0} \sum_{s \in S} f_{i,j}^{x,s}(k) - \sum_{k=1}^T \sum_{x \neq 0} \sum_{s \in S} g_{i,j}^{x,s}(k) = 0 \quad (4.21)$$

$$\sum_{k=1}^T u_{i,j}(k) - \sum_{k=1}^T v_{i,j}(k) = 0 \quad (4.22)$$

The capacity constraints are presented in Equation 4.23 for $(i, j) \in E$ and $k = 1, \dots, T$:

$$q_{i,j}^\mu(k) \leq \bar{Q}_{i,j}; \quad u_{i,j}(k) \leq \bar{C}_{i,j}^\mu; \quad v_{i,j}(k) \leq \bar{C}_{i,j}^\nu; \quad n_p(k) \leq \bar{C}_p, \quad (4.23)$$

where $\bar{Q}_{i,j}$ and $\bar{C}_p$ are the upstream queue capacity of link (i, j) and parking capacity at parking node p , respectively. $\bar{C}_{i,j}^\mu$ and $\bar{C}_{i,j}^\nu$ are the inflow and outflow capacity of link (i, j) , respectively.

4.3.2.2 Travel demand and government subsidy

As outlined in Section 4.2, it is important to recognize that not all travel demand choices are influenced by discounts on trip fees and overall travel demand is comprised of both regular and flexible demand. Both of them could get a discount given their destinations s , requesting order time r , and the type of SAVs x in the platform. The notations of demand-related parameters are shown in Table 4.1.

- Regular demand. $D_{o,s}^{\mathcal{E},S,l}(r)$ and $D_{o,s}^{\mathcal{E},\mathcal{R},l}(r)$ represent the regular demand of solo and ride-sharing travelers with origin o , destination s and latest acceptable arrival time l at time r , respectively. This category of travelers is characterized by their propensity to place orders at time r regardless of any discounts offered, and thus are considered as fixed parameters rather than variables in our analysis.
- Flexible demand. $d_{o,s}^{\mathcal{F},S,l}(r)$ and $d_{o,s}^{\mathcal{F},\mathcal{R},l}(r)$ are the flexible demand of solo and ride-sharing travelers with origin o , destination s and latest acceptable arrival time l at time r , respectively. This segment of travelers is more likely to be influenced by the level of discount available at the time of placing their order. A lower discount at request time r tends to make them more inclined to proceed with their order at that moment.

Equations 4.24 and 4.25 indicate that the demand of solo and ride-sharing travelers is the combination the regular demand and flexible demand for $(i, j) \in E$ and $r \in R$.

$$d_{o,s}^S(r) = \sum_{l \in L} D_{o,s}^{\mathcal{E},S,l}(r) + \sum_{l \in L} d_{o,s}^{\mathcal{F},S,l}(r) \quad (4.24)$$

$$d_{o,s}^R(r) = \sum_{l \in L} D_{o,s}^{\mathcal{E},R,l}(r) + \sum_{l \in L} d_{o,s}^{\mathcal{F},R,l}(r) \quad (4.25)$$

For flexible travelers, their requesting order time is usually not fixed, and they are willing to place an order within a certain time range. Their choices are influenced by the level of discount available at the time of placing their order. A lower discount at request time r tends to make them more inclined to proceed with their order at that moment. The Logit model is used to represent the effect of discounts on flexible demand in Equations 4.26 and 4.27 for $o \in O$, $s \in S$, $l \in L$ when requesting time $r \in R$ is in the time range $[a, b]$. And the Logit model also applies to other time ranges. λ represents the sensitivity of flexible demand to the trip fee and discount. A higher λ means that even a small discount can significantly increase the flexible demand.

The variable η_x on the other hand, represents the sensitivity of the trip fee and discount for taking a SAV of type x . In simpler terms, it signifies how much the flexible demand is influenced by the trip fee and discount when choosing an SAV of type x . A high value of η_x indicates that the flexible demand is highly sensitive to the cost and discount of taking an SAV of type x .

$$d_{o,s}^{\mathcal{F},S,l}(r) = D_{o,s}^{\mathcal{F},S,l}[a, b] \frac{e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}}{\sum_{i=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(i)}}, \quad (4.26)$$

$$d_{o,s}^{\mathcal{F},R,l}(r) = D_{o,s}^{\mathcal{F},R,l}[a, b] \frac{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}}{\sum_{i=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(i)}}, \quad (4.27)$$

where $\alpha_s^x(r)$ is the discount of taking SAVs with type x for orders with destination s and requesting order time r . $D_{o,s}^{\mathcal{F},S,l}[a, b]$ and $D_{o,s}^{\mathcal{F},R,l}[a, b]$ are the total number of flexible demand of solo and ride-sharing travelers with origin o , destination s and latest acceptable arrival time l in time interval $[a, b]$, respectively.

Besides, Equation 4.28 states that the discount variables $\alpha_s^x(r)$ offered by the government

should be limited to 0 and 1 for $s \in S$, $r \in R$ and $x = -1, 1, \dots, X$.

$$0 \leq \alpha_s^x(r) \leq 1 \quad (4.28)$$

4.3.3 Cost for different stakeholders

This section presents the costs for different stakeholders, including traveler cost, TNC profit, government cost, and social cost. The cost and demand-related parameters are shown in Table 4.1.

4.3.3.1 Travelers cost

Traveler costs encompass various components that contribute to the overall expenses incurred by individuals using ride-hailing services. These elements include the traveler travel time cost (TTTC), which reflects the cost of time spent traveling; the traveler trip fee cost (TTFC), representing the fare paid for the ride; the traveler waiting cost (TWC), which accounts for the time spent waiting for the ride to arrive; the traveler late cost (TLC), which is associated with the inconvenience and potential missed opportunities resulting from a delayed arrival; and the traveler late penalty (TLP), which is specifically related to the situation when a traveler arrives at their destination after the latest acceptable arrival time.

Travelers late penalty (TLP) becomes relevant when a traveler reaches their destination later than the time they deemed acceptable. In such cases, the Transportation Network Companies (TNCs) are responsible for compensating the traveler for the inconvenience caused by the delay. The compensation covers the time interval between the latest acceptable arrival time and the actual time the traveler arrives at their destination. This compensation is designed to alleviate the negative impact of the delay, ensuring that travelers are not financially burdened for situations beyond their control, such as delays caused by traffic, system errors, or other unforeseen circumstances. The late penalty system thus helps maintain fairness in the overall ride-hailing experience and ensures that travelers are treated equitably in instances of tardiness. The TLP amount is typically specified as a monetary rate per unit of delay, and is usually paid by the platform according to its compensation policy. Payment can take the form of an immediate refund, account credit, or direct reimbursement, and platforms commonly enforce caps or thresholds to limit financial exposure. Beyond compensation, TLP creates operational incentives for TNCs to

improve dispatching, routing, and reliability; platforms also maintain dispute-resolution procedures to handle contested claims about causation or acceptable arrival windows.

$$Travelers\ Cost = TTTC + TTFC + TWC + TLC - TLP \quad (4.29)$$

$$TTTC = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} V_T \Delta_t \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s}(a) - g_{i,j}^{x,s}(a)) \quad (4.30)$$

$$TTFC = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t \alpha_s^x(r) \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r,l}(a) - g_{i,j}^{x,s,r,l}(a)) \quad (4.31)$$

$$TWC = \sum_{k=1}^T \sum_{o \in O} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} V_W \Delta_t (n_o^{S,s,r,l}(k) + n_o^{R,s,r,l}(k)) \quad (4.32)$$

$$TLC = \sum_{l \in L} \sum_{k=l}^T \sum_{r \in R} \sum_{s \in S} V_L \Delta_t (k-l) \Delta_t f_{s,s'}^{r,l}(k) \quad (4.33)$$

$$TLP = \sum_{l \in L} \sum_{k=l}^T \sum_{r \in R} \sum_{s \in S} C_{LP} \Delta_t (k-l) \Delta_t f_{s,s'}^{r,l}(k) \quad (4.34)$$

4.3.3.2 TNC profit

TNC profit is determined by several factors including SAVs travel cost (STC), travelers late penalty (TLP), SAVs travel revenue (STR) and SAVs parking cost (SPC).

$$TNC\ Profit = STR - STC - TLP - SPC = -TNC\ Cost \quad (4.35)$$

$$STC = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x=-1}^X C_S \Delta_t \sum_{a=1}^k \Delta_t (u_{i,j}^x(a) - v_{i,j}^x(a)) \quad (4.36)$$

$$STR = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r,l}(a) - g_{i,j}^{x,s,r,l}(a)) \quad (4.37)$$

$$SPC = \sum_{p \in P} \sum_{k=1}^T C_p \Delta_t n_p(k) \quad (4.38)$$

4.3.3.3 Government cost

Government cost incorporates the subsidy amount (SA) and SAVs parking cost (SPC). In this research, we operate under the assumption that the government owns and oversees the parking spaces. However, this premise could be extended to include privately owned parking spaces,

assuming they are users of a centralized platform.

$$\text{Government Cost} = \text{subsidy amount} - \text{SPC} \quad (4.39)$$

$$\begin{aligned} \text{subsidy amount} &= \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t (1 - \alpha_s^x(r)) \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r}(a) - g_{i,j}^{x,s,r}(a)) \\ &= \text{STR} - \text{TTF C} \end{aligned} \quad (4.40)$$

4.3.3.4 Social cost

Social cost (SC) is the sum of travelers, TNC, and government cost and it only involves the cost of time (*TTTC*, *TWC* and *TLC*) and consumption (*STC*).

$$\text{Social cost} = \text{TTTC} + \text{TWC} + \text{TLC} + \text{STC} \quad (4.41)$$

4.3.4 Formulation

To summarize, the POCP design problem could be formulated as a nonlinear bi-level programming as follows:

- Upper level. This level involves the government's operation of the POCP with the goal of minimizing social costs, which equals maximizing social welfare. This is achieved by influencing travelers' behaviors via spatial-temporal discounts that subsidize their trip fees. The variables in this level, represented by $\mathbf{d}$, pertain to demand, while the spatial-temporal subsidy variables are represented by α .
- Lower Level. In this level, the TNCs handle the routing choices of SAVs and provides the mobility services upon receiving travel demand orders via a centralized platform. The aim is to maximize profit through the provision of mobility services. The variables in this level, denoted by $\mathbf{y}$, encapsulate the flow dynamics variables.

The formulation of the POCP design is presented as follows:

$$\begin{aligned}
& \min_{\mathbf{y}, \mathbf{d}, \alpha} \text{ Social Cost} \quad \text{Eqs. (4.41)} \\
& s.t. \quad \text{Demand Constraints} \quad \text{Eqs. (4.24) – (4.28)} \\
& \quad \quad \mathbf{y} \in \text{argmax}_{\mathbf{y}} \text{ TNC Profit} \quad \text{Eqs. (4.35)} \\
& \quad \quad s.t. \text{ Double Queue Constraints} \quad \text{Eqs. (4.1) – (4.2)} \\
& \quad \quad \text{Double Flow Matching Constraints} \quad \text{Eqs. (4.3) – (4.6)} \\
& \quad \quad \text{Node Conservation Constraints} \quad \text{Eqs. (4.7) – (4.15)} \\
& \quad \quad \text{Parking Constraints} \quad \text{Eqs. (4.16)} \\
& \quad \quad \text{Waiting Constraints} \quad \text{Eqs. (4.17) – (4.18)} \\
& \quad \quad \text{Other Constraints} \quad \text{Eqs. (4.19) – (4.23)}
\end{aligned} \tag{4.42}$$

Proposition 9 presents the feasibility of the proposed POCP design problem in Section 4.3.4. The parameters $W_{(o,s)_h}$, N_{o_h, s_h} , $\delta_{i,j}^{P_{o_h, s_h}}$, $\delta_{i,j}^{P_{s_h, o_h}}$, $\delta_{i,j}^{P_{s_h, o_{h+1}}}$, $\delta_{i,j}^{P_{p, o_1}}$ and $\delta_{i,j}^{P_{s_n, p}}$ are defined in Appendix B.1. The feasibility means the existence of at least one feasible solution of the non-linear bi-level optimization of POCP. It also represents that all travelers reach their destinations within the time horizon and all vehicles return to their assigned parking lots while satisfying all model constraints.

Proposition 9 (Feasibility). If we have

$$\begin{aligned}
T \geq \max_h W_{(o,s)_h} &+ \sum_{h=1}^n \sum_{(i,j) \in E} N_{o_h, s_h} \tau_{i,j}^f \delta_{i,j}^{P_{o_h, s_h}} + \sum_{h=1}^n \sum_{(i,j) \in E} (N_{o_h, s_h} - 1) \tau_{i,j}^f \delta_{i,j}^{P_{s_h, o_h}} \sum_{h=1}^{n-1} \sum_{(i,j) \in E} \\
&\tau_{i,j}^f \delta_{i,j}^{P_{s_h, o_{h+1}}} + \sum_{(i,j) \in E} \tau_{i,j}^f \delta_{i,j}^{P_{p, o_1}} + \sum_{(i,j) \in E} \tau_{i,j}^f \delta_{i,j}^{P_{s_n, p}} \text{ holds,}
\end{aligned}$$

then the bi-level programming in Equation 4.42 is always feasible.

Proof Sketch. The proof of feasibility is straightforward. Evidently, if we select a sufficiently large time horizon T , the POCP problem will always be feasible, implying that all travelers will reach their destinations and all SAVs will be parked at parking nodes by the time T . Therefore, the essence of proving feasibility in this context is to identify a large enough T , construct a feasible solution under this T , and thereby demonstrate the problem's feasibility. Detailed proof

of feasibility is provided in Appendix B.1.

4.4 Solution framework

The POCP problem, presented in Section 4.3.4, is characterized by its formulation as a nonlinear bi-level programming problem. Therefore, solving this problem and obtaining its analytical properties pose significant challenges. This section aims to introduce the method to obtain the solution of the nonlinear bi-level programming inherent in the POCP. This is achieved by equivalently converting it into two linear programming sub-problems (LPs) and subsequently determining the optimal subsidy strategy. This is particularly important given the non-uniqueness of optimal solutions of the System Optimal Dynamic Traffic Assignment (SO-DTA) problem. In Section 4.4.1, we demonstrate the analytical equivalence by reducing the proposed bi-level programming to a single-level optimization problem, showcasing the strong duality of the lower level. Moreover, in Section 4.4.2, we initially demonstrate that both the minimum social cost and the optimal flow dynamics can be determined through the first LP, which is formulated via the branch and bound algorithm. Subsequently, the second LP enables us to ascertain the optimal spatial-temporal subsidy alongside the minimal subsidy amount required.

4.4.1 Strong duality of lower level

In the formulation of POCP as a non-linear bi-level programming in Equation 4.42, the lower level consists of linear constraints and objectives for the lower-level variables $\mathbf{y}$, which represent the flow dynamics. Therefore, we utilize a generalized form of linear programming to depict the lower-level programming as follows:

$$\begin{aligned} & \underset{\mathbf{y}}{\text{minimize}} && TNC \text{ Cost} = \mathbf{P}^T \mathbf{y} \\ & \text{subject to} && \mathbf{A}\mathbf{y} = \mathbf{B} + \mathbf{C}\mathbf{d} \\ & && \mathbf{E}\mathbf{y} \leq \mathbf{F} \end{aligned}$$

where $\mathbf{d}$ represents the demand variables in the upper level and we treat it as fixed in the lower level.

The KKT (Karush-Kuhn-Tucker) conditions for the lower level of the optimization problem in Equation 4.42 are given as follows. These conditions are necessary for finding the optimal solution to a constrained optimization problem with inequality constraints. In this formulation, n_μ denotes both the number of inequality constraints and the dimension of the vector of decision variables μ in the lower level of the problem. The KKT conditions include the primal feasibility, dual feasibility, and complementary slackness conditions, ensuring that the solution satisfies both the constraints and the optimality conditions for the lower-level problem.

$$\begin{aligned}\mathbf{P} + \mathbf{A}^T \lambda + \mathbf{E}^T \mu &= \mathbf{0} \\ \mathbf{A} \mathbf{y} &= \mathbf{B} + \mathbf{C} \mathbf{d} \\ \mathbf{E} \mathbf{y} &\leq \mathbf{F} \\ \mu_i (\mathbf{E} \mathbf{y} - \mathbf{F})_i &= 0 \quad \text{for } i = 1, \dots, N \\ \mu &\geq 0\end{aligned}$$

We proceed to address the demand constraints described in Equations 4.24 to 4.28. Equations 4.24 and 4.25 linearly relate to the demand variables $\mathbf{d}$, which we represent utilizing the matrix $\mathbf{G}$ and $\mathbf{H}$. Additionally, the linear constraints related to the spatial-temporal discount variables α in Equation 4.28 are expressed through the matrix $\mathbf{J}$ and $\mathbf{K}$.

In contrast, Equations 4.26 and 4.27 exhibit the non-linearity with respect to the variables α , and these nonlinear relationships can be characterized using the nonlinear function f . In summary, the demand constraints outlined in Equations 4.24 to 4.28 can be generalized as follows:

$$\begin{aligned}\mathbf{G} \mathbf{d} &= \mathbf{H} \\ \mathbf{J} \alpha &\leq \mathbf{K} \\ f(\mathbf{d}, \alpha) &= 0\end{aligned}$$

where $f(\mathbf{d}, \alpha)$ is a nonlinear function for α . Mathematically, suppose $f : \mathbb{R}^{n_d} \times \mathbb{R}^{n_\alpha} \rightarrow \mathbb{R}^{n_f}$, then for $\forall \mathbf{M}_\alpha \in \mathbb{R}^{n_f \times n_\alpha}$, we have $\frac{\partial f}{\partial \alpha} \neq \mathbf{M}_\alpha$.

Therefore, the proposed nonlinear bi-level problem could be converted into single-level nonlinear programming, referred to as Programming (I), where n_μ is the dimension of variables μ .

Programming (I)

$$\underset{\mathbf{y}, \mathbf{d}, \alpha, \mu, \lambda}{\text{minimize}} \quad \text{Social Cost} = \mathbf{S}^T \mathbf{y} \quad (4.44a)$$

$$\text{subject to} \quad \mathbf{P} + \mathbf{A}^T \boldsymbol{\lambda} + \mathbf{E}^T \boldsymbol{\mu} = \mathbf{0} \quad (4.44b)$$

$$\mathbf{A} \mathbf{y} = \mathbf{B} + \mathbf{C} \mathbf{d} \quad (4.44c)$$

$$\mathbf{E} \mathbf{y} \leq \mathbf{F} \quad (4.44d)$$

$$\mu_i (\mathbf{E} \mathbf{y} - \mathbf{F})_i = 0 \quad \text{for } i = 1, \dots, n_\mu \quad (4.44e)$$

$$\boldsymbol{\mu} \geq \mathbf{0} \quad (4.44f)$$

$$\mathbf{G} \mathbf{d} = \mathbf{H} \quad (4.44g)$$

$$\mathbf{J} \boldsymbol{\alpha} \leq \mathbf{K} \quad (4.44h)$$

$$f(\mathbf{d}, \boldsymbol{\alpha}) = 0 \quad (4.44i)$$

where $f(\mathbf{d}, \boldsymbol{\alpha})$ is a nonlinear function for $\boldsymbol{\alpha}$.

Proposition 10 demonstrates that Programming (I) is equivalent to the proposed nonlinear bi-level programming delineated in Equation 4.42.

Proposition 10. Programming (I) is equivalent to the bi-level programming of POCP in Equation 4.42. Mathematically, $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*$ is the solution of bi-level programming in Equation 4.42 if and only if $\exists \boldsymbol{\mu}^*$ and $\boldsymbol{\lambda}^*$ such that $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ is the solution of Programming (I).

Proof Sketch. Proposition 10 could be directly obtained by the strong duality of the lower-level problem.

However, Programming (I) remains a nonlinear programming problem due to the presence of the demand constraints in Equation 4.44i and the complementary constraints in Equation 4.44e. As a result, the non-linearity complicates the solution process. Therefore, we aim to equivalently eliminate the non-linearity in both the demand constraints and the complementary constraints, as detailed in Section 4.4.2, to simplify the problem and make it more solvable.

4.4.2 Solution of bi-level programming

This section is devoted to obtaining the solution of the proposed bi-level optimization presented in Equations 4.42, which has been verified as equivalent to the nonlinear single-level Programming (I) in Section 4.4.1. However, it is important to note that solving the nonlinear Programming (I) directly is inherently challenging due to Equations 4.44e and 4.44i. Therefore, we introduce a methodology to obtain the solution for the bi-level programming by transforming it into two distinct linear programmings (LPs). The solution of Programming (I) consists of the optimal objective value SC^* and optimal solution $\mathbf{y}^*, \mathbf{d}^*, \alpha^*, \mu^*, \lambda^*$. We would first equivalently solve the minimum social cost SC^* and optimal flow dynamics by the branch and bound algorithm in Section 4.4.2.1, then considering the existence of multiple optimal solutions of SO-DTA problems, the optimal discount variables α^* is developed in Section 4.4.2.2, representing the minimum subsidy amount to achieve SO-DTA with the objective of saving financial resources.

4.4.2.1 Linearity in DTA

In our specific problem, Programming (I) demonstrates the non-linearity, which arises mainly from the presence of complementary constraints as described in Equation 4.44e, as well as the nonlinear function $f(\mathbf{d}, \alpha)$ outlined in Equation 4.44i. The non-linearity introduced by the complementary constraints in Equation 4.44e is a result of the interplay between the decision variables and the constraints, where the constraints must be satisfied in a complementary fashion. Similarly, the non-linearity in Equation 4.44i is caused by the demand constraints, which involve a nonlinear function of the decision variables $\mathbf{d}$ and α . To address this, we propose an approach where we eliminate the non-linearity in both the complementary constraints in Equation 4.44e and the demand constraints in Equation 4.44i. This can be achieved through specific reformulations or approximations that transform the problem into a more tractable linear or simplified form, thus making the optimization process more manageable.

Proposition 11 enlightens us on the existence of a nonlinear Programming (II) that is equivalent to Programming (I), where the complementary constraints in Equation 4.44e of Programming (I) are converted into two linear constraints as Equations 4.45e and 4.45f in Programming (II).

Proposition 11. Suppose n_μ is the dimension of variables μ , there always exists the set $\tilde{I} \subseteq \{1, \dots, n_\mu\}$ and $\tilde{J} \subseteq \{1, \dots, n_\mu\}$ such that Programming (II) is equivalent to Programming (I).

Programming (II)

$$\underset{\mathbf{y}, \mathbf{d}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\lambda}}{\text{minimize}} \quad \text{Social Cost} = \mathbf{S}^T \mathbf{y} \quad (4.45a)$$

$$\text{subject to} \quad \mathbf{P} + \mathbf{A}^T \boldsymbol{\lambda} + \mathbf{E}^T \boldsymbol{\mu} = \mathbf{0} \quad (4.45b)$$

$$\mathbf{A} \mathbf{y} = \mathbf{B} + \mathbf{C} \mathbf{d} \quad (4.45c)$$

$$\mathbf{E} \mathbf{y} \leq \mathbf{F} \quad (4.45d)$$

$$\boldsymbol{\mu}_i = \mathbf{0} \quad \text{for } i \in \tilde{I} \quad (4.45e)$$

$$(\mathbf{E} \mathbf{y} - \mathbf{F})_j = \mathbf{0} \quad \text{for } j \in \tilde{J} \quad (4.45f)$$

$$\boldsymbol{\mu} \geq \mathbf{0} \quad (4.45g)$$

$$\mathbf{G} \mathbf{d} = \mathbf{H} \quad (4.45h)$$

$$\mathbf{J} \boldsymbol{\alpha} \leq \mathbf{K} \quad (4.45i)$$

$$f(\mathbf{d}, \boldsymbol{\alpha}) = \mathbf{0} \quad (4.45j)$$

Proof Sketch. See Appendix B.2.

However, although the existence of Programming (II) has been theoretically demonstrated in Proposition 11, the precise method for deriving Programming (II) still remains unclear. The challenge lies in the fact that, despite its existence, Programming (II) is still a nonlinear optimization problem due to the nonlinear nature of the term in Equation 4.45j. As a result, solving Programming (II) directly is still difficult and computationally challenging. To address this issue, we propose to focus on deriving a linear-equivalent formulation for Programming (II), which could simplify the optimization process and make it more tractable. This approach would involve transforming the nonlinear constraints and objective functions into linear approximations or equivalents, thereby allowing us to apply linear programming techniques to find an optimal solution.

We first propose the transformation $\beta = \mathbf{g}(\mathbf{d}, \boldsymbol{\alpha})$ in Equations 4.46 and 4.47 for $o \in O$, $s \in$

$S, l \in L$ and $r \in R$.

$$\beta_{o,s}^{\mathcal{F},S,l}(r) = \frac{d_{o,s}^{\mathcal{F},S,l}(r)}{e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}} \quad (4.46)$$

$$\beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}} \quad (4.47)$$

The constraint $f(\mathbf{d}, \alpha) = \mathbf{0}$ could be converted into the Equations 4.48 to 4.55 for $r, r_1, r_2 \in \{a, a+1, \dots, b-1, b\}$, $l_1, l_2 \in L$ and $o_1, o_2 \in O$.

$$\sum_{r=a}^b d_{o,s}^{\mathcal{F},S,l}(r) = D_{o,s}^{\mathcal{F},S,l}[a, b] \quad (4.48)$$

$$\sum_{r=a}^b d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \quad (4.49)$$

$$D_{o_2,s}^{\mathcal{F},S,l_2}[a, b] d_{o_1,s}^{\mathcal{F},S,l_1}(r) = D_{o_1,s}^{\mathcal{F},S,l_1}[a, b] d_{o_2,s}^{\mathcal{F},S,l_2}(r) \quad (4.50)$$

$$D_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}[a, b] d_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}(r) = D_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}[a, b] d_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}(r) \quad (4.51)$$

$$D_{o_2,s}^{\mathcal{F},S,l_2}[a, b] \beta_{o_1,s}^{\mathcal{F},S,l_1}(r) = D_{o_1,s}^{\mathcal{F},S,l_1}[a, b] \beta_{o_2,s}^{\mathcal{F},S,l_2}(r) \quad (4.52)$$

$$D_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}[a, b] \beta_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}(r) = D_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}[a, b] \beta_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}(r) \quad (4.53)$$

$$\beta_{o,s}^{\mathcal{F},S,l}(r_1) = \beta_{o,s}^{\mathcal{F},S,l}(r_2) \quad (4.54)$$

$$\beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r_1) = \beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r_2) \quad (4.55)$$

Then the linear inequality constraints for variables α in Equation 4.44h are converted into the Equation 4.56 and 4.57 under the transformation $\beta = g(\mathbf{d}, \alpha)$ for $o \in O$, $s \in S$ and $l \in L$.

$$e^{-\lambda C_{-1}} \beta_{o,s}^{\mathcal{F},S,l}(r) \leq d_{o,s}^{\mathcal{F},S,l}(r) \leq \beta_{o,s}^{\mathcal{F},S,l}(r) \quad r \in \{a, a+1, \dots, b-1, b\} \quad (4.56)$$

$$e^{-\sum_{x=1}^X \lambda C_x \eta_x} \beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) \leq d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) \leq \beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) \quad r \in \{a, a+1, \dots, b-1, b\} \quad (4.57)$$

For Equations 4.48 to 4.57, we use the matrix to represent the generalized form of them in Equations 4.58 and 4.59.

$$\mathbf{Kd} + \mathbf{M}\beta = \mathbf{Q} \quad (4.58)$$

$$\mathbf{Rd} + \mathbf{V}\beta \leq \mathbf{0} \quad (4.59)$$

Therefore, Programming (III) is derived from Programming (II) by the proposed transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$ in Equations 4.46 and 4.47.

Programming (III)

$$\underset{\mathbf{y}, \mathbf{d}, \beta, \mu, \lambda}{\text{minimize}} \quad \text{Social Cost} = \mathbf{S}^T \mathbf{y} \quad (4.60a)$$

$$\text{subject to} \quad \mathbf{P} + \mathbf{A}^T \lambda + \mathbf{E}^T \mu = \mathbf{0} \quad (4.60b)$$

$$\mathbf{A} \mathbf{y} = \mathbf{B} + \mathbf{C} \mathbf{d} \quad (4.60c)$$

$$\mathbf{E} \mathbf{y} \leq \mathbf{F} \quad (4.60d)$$

$$\mu_i = 0 \quad \text{for } i \in \tilde{I} \quad (4.60e)$$

$$(\mathbf{E} \mathbf{y} - \mathbf{F})_j = 0 \quad \text{for } j \in \tilde{J} \quad (4.60f)$$

$$\mu \geq 0 \quad (4.60g)$$

$$\mathbf{G} \mathbf{d} = \mathbf{H} \quad (4.60h)$$

$$\mathbf{K} \mathbf{d} + \mathbf{M} \beta = \mathbf{Q} \quad (4.60i)$$

$$\mathbf{R} \mathbf{d} + \mathbf{V} \beta \leq \mathbf{0} \quad (4.60j)$$

To establish the equivalence of Programming (III) with Programming (II) under the transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$, we initially introduce Lemma 1.

Lemma 1. Suppose $\mathbf{y}, \mathbf{d}, \beta, \mu, \lambda$ are the feasible solution of Programming (III), then there always exists the spatial-temporal discount variables α such that α satisfies $\mathbf{J} \alpha \leq \mathbf{K}$ in Equation 4.44h and $\mathbf{f}(\mathbf{d}, \alpha) = \mathbf{0}$ in Equation 4.44i.

Proof Sketch. See Appendix B.3.

Then the Proposition 12 states that the Programming (II) and Programming (III) have same optimal objective SC^* and share optimal variables $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*$.

Proposition 12. Suppose $SC^*, \mathbf{y}^*, \mathbf{d}^*, \alpha^*, \mu^*, \lambda^*$ are the optimal objective and optimal variables of Programming (II), $\widetilde{SC}, \widetilde{\mathbf{y}}, \widetilde{\mathbf{d}}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\lambda}$ for Programming (III), then we have $SC^* = \widetilde{SC}$, $\mathbf{y}^* = \widetilde{\mathbf{y}}$, $\mathbf{d}^* = \widetilde{\mathbf{d}}$, $\mu^* = \widetilde{\mu}$ and $\lambda^* = \widetilde{\lambda}$.

Proof Sketch. See Appendix B.4.

Programming (III) is a linear programming and is proved as equivalent with Programming (I), as referred to Proposition 11 and 12. However, the set of $\tilde{I}$ and $\tilde{J}$ in Programming (III) are still unclear. Our next objective is to determine the set of $\tilde{I}$ and $\tilde{J}$ in Programming (III) to preserve the equivalency with Programming (I).

Assuming that $\mathbf{y}$, $\mathbf{d}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}$ are the optimal solutions of Programming (I), it follows that $\mathbf{y}$ and $\boldsymbol{\mu}$ satisfy the complementary constraints in Equation 4.61. Here, n_μ corresponds to the number of complementary constraints, which is also equivalent to the dimension of the variables $\boldsymbol{\alpha}$.

$$\mu_i = 0 \text{ or } (Ey - F)_i = 0 \quad \text{for } i = 1, \dots, n_\mu \quad (4.61)$$

To eliminate the non-linearity in the complementary constraints, we proceed as follows for each i from 1 to n_μ : we incorporate one of the constraints, either $\mu_i = 0$ or $(Ey - F)_i = 0$, into Programming (III) while simultaneously removing all complementary constraints. This transformation results in a total of 2^{n_μ} potential new programmings, each of them is a linear programming.

Subsequently, by selecting the new programming that yields the minimum optimal objective value from these 2^{n_μ} possibilities, we can transform Programming (I) into solving this selected linear Programming (III). The method mentioned above, while feasible, can be computationally expensive. To alleviate this, we propose a branch-and-bound-based algorithm (Algorithm 2) that effectively eliminates the non-linearity in the complementary constraints in Programming (I) and obtains the Programming (III).

Algorithm 2 Branch and Bound Based Minimum Social Cost Solving Algorithm

Input: parameters in Table 4.1.

Output: minimum social cost SC^* and optimal variables $\mathbf{y}^*$, $\mathbf{d}^*$, $\boldsymbol{\beta}^*$, $\boldsymbol{\mu}^*$, $\boldsymbol{\lambda}^*$; set $\tilde{I}$ and $\tilde{J}$.

Initialize $SC^* = +\infty$, $\tilde{I} = \emptyset$, $\tilde{J} = \emptyset$, $k = 0$ and $\delta = 0$.

- 1: **while** $\delta = 0$ **do**
- 2: Solve the Programming (IV) and derive the social cost SC , solution $\mathbf{y}$ and $\mathbf{d}$
- 3: Check all complementary constraints $\mu_i (Ey - F)_i = 0 \quad \text{for } i = 1, \dots, n_\mu$
- 4: **if** all complementary constraints are satisfied **then**
- 5: $\delta = 1$ and update $SC^* = SC$

```

6:  else
7:      Choose any infeasible complementary constraint indexed as  $l$ 
8:      Solve two Programming (IV) respectively:  $(\tilde{I} = \tilde{I} \cup \{l\}; \tilde{J})$  and  $(\tilde{I}; \tilde{J} = \tilde{J} \cup \{l\})$ 
9:      if Any of the two above programming is infeasible then
10:         Cut this branch
11:      end if
12:      Choose the smaller social cost among these two Programming (IV), update  $SC^*, \tilde{I}, \tilde{J}$ 
        and  $k = \text{len}(\tilde{I} \cup \tilde{J})$ 
13:      Conduct backward checking starting with the level  $k$ 
14:      for ( $p = k; p > 0; --p$ ) do
15:         Check the social cost for each leaf without any branch at level  $p$ .
16:         if the value of social cost at leaf  $l$  is lower than  $SC^*$  then
17:            Update  $SC^*, \tilde{I}$  and  $\tilde{J}$  at leaf  $l$ 
18:         end if
19:      end for
20:  end if
21: end while

```

Proposition 13 illustrates that Algorithm 2 outlines a method for the derivation of Programming (III) based on the outputs $\tilde{I}$ and $\tilde{J}$.

Proposition 13. Suppose $SC^*, \mathbf{y}^*, \mathbf{d}^*, \beta^*, \mu^*, \lambda^*$ are the output of Algorithm 2, then $\exists \alpha$ such that $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*$ and α constitute a optimal solution of Programming (I) and the optimal objective value of Programming (I) is SC^* .

Proof Sketch. See Appendix B.5.

Based on Propositions 10 and 13, we can determine the optimal objective SC^* , which represents the minimum social cost, along with the optimal flow dynamics variables $\mathbf{y}^*$ and optimal demand variables $\mathbf{d}^*$. However, it's important to note that the SO-DTA problems may have multiple optimal solutions, implying that achieving the minimum social cost could involve various spatial-temporal subsidy patterns. Considering the constraints posed by limited government resources, our goal is to minimize social costs using as little financial expenditure as possible.

Hence, in Section 4.4.2.2, we will establish a linear framework to determine the optimal spatial-temporal discount variables, ensuring the minimum subsidy required to attain the minimum social cost.

4.4.2.2 Linearity in subsidy

This section is dedicated to determining the optimal spatial-temporal discount variables α^* that minimize the subsidy amount SA^* under SO-DTA, which ultimately leads to the attainment of minimum social cost SC^* . Prior to delving into this analysis, we will initially consider the potential existence of multiple spatial-temporal discount variables α^* achieving SO-DTA.

Proposition 14 asserts that the minimum social cost can be attained under varying spatial-temporal discounts by attaining the same flow dynamics and demand variables, where $SC(\mathbf{y}, \mathbf{d}, \alpha, \mu, \lambda)$ represents the value of social cost under the feasible solution $(\mathbf{y}, \mathbf{d}, \alpha, \mu, \lambda)$ of Programming (I). Furthermore, SC^* denotes the optimal objective value of Programming (I), representing the minimum social cost.

Proposition 14. $\exists \alpha_1, \alpha_2$ with $\alpha_1 \neq \alpha_2$ and $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*$ such that $SC(\mathbf{y}^*, \mathbf{d}^*, \alpha_1, \mu^*, \lambda^*) = SC(\mathbf{y}^*, \mathbf{d}^*, \alpha_2, \mu^*, \lambda^*) = SC^*$.

Proof Sketch. See Appendix B.6.

Proposition 13 presents that the optimal flow dynamics variables $\mathbf{y}^*$ and optimal demand variables $\mathbf{d}^*$ could be solved by Algorithm 2. Then the subsidy amount as formulated in Equation 4.40 is only associated with the spatial-temporal discount variables α , as shown below:

$$\text{subsidy amount} = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t (1 - \alpha_s^x(r)) \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r,*}(a) - g_{i,j}^{x,s,r,*}(a)).$$

In scenarios where government budgets or financial resources are limited, identifying the spatial-temporal discount that achieves the desired outcomes with the minimum subsidy amount becomes crucial. Therefore, the goal of this section is to determine the optimal spatial-temporal discount variables α^* yielding the minimum subsidy amount given the optimal flow dynamics $\mathbf{y}^*$ and optimal demand variables $\mathbf{d}^*$ obtained by Algorithm 2.

To establish the linear framework to derive α^* , we need to find equivalent linear equations

for α^* to solve the equations $\beta^* = \mathbf{g}(\mathbf{d}^*, \alpha)$ in Equations 4.62 to 4.64 for $s \in S$ and $r \in R$.

$$-\lambda C_{-1} \alpha_s^{x=-1}(r) = \ln \frac{d_{o,s}^{\mathcal{F},S,l,*}(r)}{\beta_{o,s}^{\mathcal{F},S,l,*}(r)} \quad (4.62)$$

$$-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r) = \ln \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{\beta_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)} \quad (4.63)$$

$$0 \leq \alpha_s^x(r) \leq 1 \quad (4.64)$$

Therefore, we formulate a linear programming as Programming (IV) to solve the optimal spatial-temporal discount variables α^* and minimum subsidy amount SA^* given the optimal flow dynamics variables $\mathbf{y}^*$ and demand variables $\mathbf{d}^*$ under SO-DTA obtained by Algorithm 2.

Programming (IV)

$$\begin{aligned} \min_{\alpha} \quad & \text{Subsidy Amount} = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t (1 - \alpha_s^x(r)) \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r,*}(a) - g_{i,j}^{x,s,r,*}(a)) \\ \text{s.t.} \quad & -\lambda C_{-1} \alpha_s^{x=-1}(r) = \ln \frac{d_{o,s}^{\mathcal{F},S,l,*}(r)}{\beta_{o,s}^{\mathcal{F},S,l,*}(r)} \quad r \in \{a, a+1, \dots, b\} \\ & -\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r) = \ln \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{\beta_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)} \quad r \in \{a, a+1, \dots, b\} \\ & 0 \leq \alpha_s^x(r) \leq 1 \end{aligned}$$

4.4.3 Solution algorithm

Finally, we summarize the process of solving the proposed bi-level programming of POCP in Algorithm 3. The first step is to solve Programming (III) to obtain the minimum social cost SC^* , flow dynamics variables $\mathbf{y}^*$, and demand variables $\mathbf{d}^*$. Subsequently, we would use the values of $\mathbf{y}^*$ and $\mathbf{d}^*$ from the first step to solve Programming (IV), then we obtain the minimum subsidy amount SA^* and the optimal spatial-temporal discount variables α^* . Therefore, SC^* , $\mathbf{y}^*$, $\mathbf{d}^*$ and α^* together constitute a complete solution of the POCP problem.

Algorithm 3 Algorithm for Solving Minimum Social Cost and Minimum Subsidy Amount

Input: parameters in Table 4.1.

Output: minimum social cost SC^* ; optimal flow dynamics variables $\mathbf{y}^*$; optimal demand variables $\mathbf{d}^*$; minimum subsidy amount SA^* ; optimal spatial-temporal discount α^* .

- 1: Conduct Algorithm 2 to solve the linear Programming (III) and obtain the minimum social cost SC^* , optimal flow dynamics variables $\mathbf{y}^*$, optimal demand variables $\mathbf{d}^*$, and other optimal variables β^* , μ^* and λ^* .

Programming (III)

$$\begin{aligned}
& \underset{\mathbf{y}, \mathbf{d}, \beta, \mu, \lambda}{\text{minimize}} && \text{Social Cost} = \mathbf{S}^T \mathbf{y} \\
& \text{subject to} && \mathbf{P} + \mathbf{A}^T \lambda + \mathbf{E}^T \mu = \mathbf{0} \\
& && \mathbf{A} \mathbf{y} = \mathbf{B} + \mathbf{C} \mathbf{d} \\
& && \mathbf{E} \mathbf{y} \leq \mathbf{F} \\
& && \mu_i = 0 \quad \text{for } i \in \tilde{I}, \tilde{I} \subseteq \{1, \dots, n_\mu\} \\
& && (\mathbf{E} \mathbf{y} - \mathbf{F})_j = 0 \quad \text{for } j \in \tilde{J}, \tilde{J} \subseteq \{1, \dots, n_\mu\} \\
& && \mu \geq 0 \\
& && \mathbf{G} \mathbf{d} = \mathbf{H} \\
& && \mathbf{K} \mathbf{d} + \mathbf{M} \beta = \mathbf{Q} \\
& && \mathbf{R} \mathbf{d} + \mathbf{V} \beta \leq \mathbf{U}
\end{aligned}$$

- 2: Solve the linear Programming (IV) and obtain the minimum subsidy amount SA^* and optimal spatial-temporal discount α^* given the flow dynamics variables $\mathbf{y}^*$ and demand variables $\mathbf{d}^*$ under minimum social cost SC^* .

Programming (IV)

$$\begin{aligned}
\min_{\alpha} \quad & \text{Subsidy Amount} = \sum_{k=1}^T \sum_{(i,j) \in E} \sum_{x \neq 0} \sum_{s \in S} \sum_{r \in R} \sum_{l \in L} C_x \Delta_t (1 - \alpha_s^x(r)) \sum_{a=1}^k \Delta_t (f_{i,j}^{x,s,r,*}(a) - g_{i,j}^{x,s,r,*}(a)) \\
s.t. \quad & -\lambda C_{-1} \alpha_s^{x=-1}(r) = \ln \frac{d_{o,s}^{\mathcal{F},S,l,*}(r)}{\beta_{o,s}^{\mathcal{F},S,l,*}(r)} \quad r \in \{a, a+1, \dots, b\} \\
& -\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r) = \ln \frac{d_{o,s}^{\mathcal{F},R,l,*}(r)}{\beta_{o,s}^{\mathcal{F},R,l,*}(r)} \quad r \in \{a, a+1, \dots, b\} \\
& 0 \leq \alpha_s^x(r) \leq 1
\end{aligned}$$

Proposition 15 demonstrates the validity of Algorithm 3, which suggests that the nonlinear Programming (I) can be equivalently converted into two LPs.

Proposition 15. Suppose SC^* , $\mathbf{y}^*$, $\mathbf{d}^*$, $\boldsymbol{\mu}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\alpha}^*$ are the output of Algorithm 3 by solving linear Programming (III) and (IV), respectively. Then SC^* is the optimal objective value of Programming (I) and SC^* , $\mathbf{y}^*$, $\mathbf{d}^*$, $\boldsymbol{\mu}^*$, $\boldsymbol{\lambda}^*$, $\boldsymbol{\alpha}^*$ are a optimal solution of Programming (I).

Proof Sketch. It could be easily obtained by the proof of Lemma 1 and Proposition 13.

By Propositions 10 and 15, the proposed nonlinear bi-level optimization formulated in Equation 4.42 could be equivalently converted into two linear Programmings (III) and (IV) by solving the minimum social cost SC^* , optimal flow dynamics variables $\mathbf{y}^*$, minimum subsidy amount SA^* and optimal spatial-temporal discount variables $\boldsymbol{\alpha}^*$, respectively.

4.5 Numerical experiments

In this section, we would conduct the experiments are conducted on two networks. The synthetic small network, shown in Figure 4.4, is designed to validate the proposed model and compare the effects of POCP with other ride-hailing platforms and subsidy strategies in Section 4.5.1. The Hong Kong network contains the the Kowloon Peninsula, which is the core urban area of Hong Kong and we evaluate the effectiveness of the proposed POCP in real-world traffic conditions in Section 4.5.2. Besides, all the numerical experiments are conducted on a desktop with an AMD Ryzen 5 5600G CPU and 64 giga-bytes of memory. And the CPLEX 20.1.0.0 is applied as an LP solver for Algorithm 3.

4.5.1 Small Network

In this section, we would first examine the effectiveness and compare the computation performance of the proposed Algorithm 3 under different instances with different sizes of ride-hailing services by POCP in Section 4.5.1.1. Then the proposed POCP and subsidy scheme are examined and compared with other platforms with different subsidy strategies in Section 4.5.1.2. The synthetic small network is shown in Figure 4.4 and consists of 10 links and 5 nodes. In this small network, nodes 1 and 2 are origin nodes, nodes 4 and 5 are destination nodes, and node 3 is parking node.

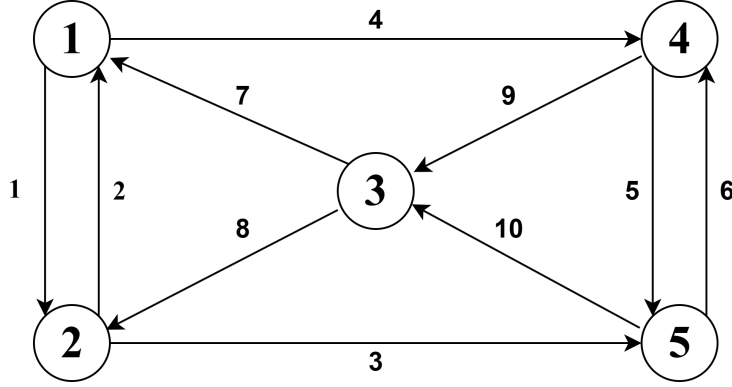


Figure 4.4: Structure of small network.

4.5.1.1 Computation performance

In this section, we would examine the computation performance of the proposed Algorithm 3, where the parameters and settings of this small network are summarized in Tables 4.2 and 4.3. The POCP problem is formulated as nonlinear bi-level programming and we have proved that it could be equivalently converted into two LPs by Algorithm 3. Therefore, we would compare the proposed Algorithm 3 with the heuristic method to solve the nonlinear bi-level programming in Equation 4.42, where we adopt the simulated annealing (SA) algorithm as the heuristic method. We choose the SA algorithm as heuristic method because SA algorithm is a widely used and well known heuristic algorithm. Besides, there are many previous work using SA algorithm in solving the bi-level optimization (Peng et al., 2008; Rui et al., 2010; Sahin and Ciric, 1998).

Link	Inflow Capacity	Outflow Capacity	Free flow time	Shockwave time	Upstream queue capacity
1	200 veh/min	200 veh/min	5 min	10 min	1050 veh
2	220 veh/min	220 veh/min	5 min	15 min	1200 veh
3	350 veh/min	350 veh/min	15 min	20 min	1800 veh
4	380 veh/min	380 veh/min	15 min	20 min	1950 veh
5	230 veh/min	230 veh/min	10 min	15 min	1250 veh
6	240 veh/min	240 veh/min	5 min	10 min	1350 veh
7	280 veh/min	280 veh/min	5 min	10 min	1450 veh
8	270 veh/min	270 veh/min	5 min	10 min	1500 veh
9	310 veh/min	310 veh/min	10 min	20 min	1750 veh
10	320 veh/min	320 veh/min	5 min	10 min	1850 veh

Table 4.2: The small network link properties.

Name	Notations	Value
Length of time interval	Δ_t	5 min
Total time horizon	T	125 min
Number of SAVs at parking node p at initial	N_p	2000 veh
Parking capacity at parking node p	$\bar{C}_p$	2100 veh
Maximum number of travelers in a SAV	X	3
Trip fare for solo traveler (ST)	$C_{x=-1}$	3.00 \$/min
Trip fare when only one ride-sharing traveler (RT) in an SAV	$C_{x=1}$	2.60 \$/min
Trip fare when only two ride-sharing travelers (RTs) in an SAV	$C_{x=2}$	2.15 \$/min
Trip fare when only three ride-sharing travelers (RTs) in an SAV	$C_{x=3}$	1.70 \$/min
Value of travel time	V_T	1.60 \$/min
Value of waiting time	V_W	1.60 \$/min
Value of late time	V_L	3.50 \$/min
Late penalty	C_{LP}	6.30 \$/min
Travel cost of SAVs	C_S	0.90 \$/min
Parking fee at parking node p	C_p	0.10 \$/min
Sensitivity of the flexible demand to the trip fee and discount	λ	0.25
Sensitivity of the trip fee and discount for taking an SAV of type 1	η_1	0.15
Sensitivity of the trip fee and discount for taking an SAV of type 2	η_2	0.30
Sensitivity of the trip fee and discount for taking an SAV of type 3	η_3	0.55

Table 4.3: Network parameters and model coefficients for a small network.

Table 4.4 illustrates that, when compared with the heuristic method used to solve the non-linear bi-level programming problem, the proposed Algorithm 3 results in both a lower social cost and reduced computation time. This outcome provides strong evidence supporting our theoretical work, demonstrating that nonlinear bi-level programming problems can be equivalently transformed into two LPs by Algorithm 3. Consequently, the social cost obtained using Algorithm 3 represents the actual optimal objective value of the bi-level programming problem presented in Equation 4.42, which is notably lower than the numerical solution for social cost derived using the heuristic method. The results correspond to the fact that the proposed algorithm guarantees to obtain the analytical solution of the proposed problem by Proposition 13, but the SA algorithm can only generate numerical approximations that approach the analytical solution.

Furthermore, our approach continues to maintain high computational efficiency, even as the problem scale increases with a larger number of travelers and SAVs, confirming its robustness and scalability.

Size		Heuristic Method		Algorithm 3	
Travelers	SAVs	Social cost (10^5)	Time (s)	Social cost	Time (s)
2,400	800	1.001	118	0.996	41
4,800	1,600	1.906	165	1.894	67
12,000	4,000	4.658	474	4.646	173
19,200	6,400	8.174	789	8.146	242
38,400	12,800	15.624	2229	15.599	636
48,000	16,000	19.737	4382	19.667	1052
72,000	24,000	31.986	7834	31.942	1795
120,000	40,000	51.607	17371	51.423	4641

Table 4.4: Computation performance of Algorithm 3 and heuristic method.

4.5.1.2 Comparison of different strategies

In this section, we conduct the Algorithm 3 in a small network and compare the effects of the proposed POCP with other ride-hailing platform strategies. Before this, we first introduce the link properties, network parameters, and flexible demand settings of small networks in Tables 4.2, 4.3 and 4.5. For regular demand, we set a constant regular demand as 25 and 55 for solo travelers (STs) and ride-sharing travelers (RTs), respectively.

Interval	l	$D_{1,4}^{\mathcal{F},S,l}[a,b]$	$D_{1,4}^{\mathcal{F},\mathcal{R},l}[a,b]$	$D_{1,5}^{\mathcal{F},S,l}[a,b]$	$D_{1,5}^{\mathcal{F},\mathcal{R},l}[a,b]$
[1, 3]	8	135	385	105	410
[4, 6]	12	115	425	160	320
[7, 9]	14	105	350	135	425
[10, 12]	17	105	280	105	410
Interval	l	$D_{2,4}^{\mathcal{F},S,l}[a,b]$	$D_{2,4}^{\mathcal{F},\mathcal{R},l}[a,b]$	$D_{2,5}^{\mathcal{F},S,l}[a,b]$	$D_{2,5}^{\mathcal{F},\mathcal{R},l}[a,b]$
[1, 3]	8	150	410	150	385
[4, 6]	12	175	425	150	400
[7, 9]	14	105	320	150	385
[10, 12]	17	150	280	105	280

Table 4.5: Flexible demand setting in the small network.

Then we compare the proposed POCP with other ride-hailing platform strategies as follows:

- **Private-owned platform.** The ride-hailing platform is owned and operated by third parties to maximize its profit. In this case, the government would not provide the subsidy as $\alpha_s^x(r) = 0$ and the optimization problem would be the single-level programming instead of bi-level programming.
- **Publicly-owned platform with spatial subsidy.** The ride-hailing platform is owned and operated by the government with the objective of minimizing the social cost. In this case, the government would provide the subsidy for the destination and requesting time but not specify the type of SAVs as $\alpha_s^{x_1}(r) = \alpha_s^{x_2}(r)$ for $x_1 \neq x_2$. The optimization problem would be bi-level programming as Equation 4.42 by adding $\alpha_s^{x_1}(r) = \alpha_s^{x_2}(r)$.
- **Publicly-owned platform with SAV type subsidy.** The ride-hailing platform is owned and operated by the government with the objective of minimizing the social cost. In this case, the government would provide the subsidy for SAV type and requesting time but not specify the destination as $\alpha_{s_1}^x(r) = \alpha_{s_2}^x(r)$ for $s_1 \neq s_2$. The optimization problem would be bi-level programming as Equation 4.42 by adding $\alpha_{s_1}^x(r) = \alpha_{s_2}^x(r)$.

The results of the proposed framework and other ride-hailing platform strategies are illustrated in Table 4.6. Compared with the private-owned ride-hailing platform, the proposed POCP framework would save 25,0765\$ of social cost by applying the 22,7179\$ of subsidy. For other publicly-owned ride-hailing platforms with different subsidy strategies, although the social cost would be almost minimized for all these strategies, the proposed POCP would use less subsidy to minimize social cost. Compared to privately-owned platforms, publicly-owned platforms reduce congestion by changing travelers' travel behavior by providing spatial-temporal subsidies, thereby reducing travelers' time costs. As a result, the travel time of SAVs owned by TNC under a public platform would decrease, and the parking costs would correspondingly increase. Furthermore, in this work, TNC charges passengers based on time. Due to the significant reduction in travel time, the profit for TNC in the public platform would decrease. For travelers, because the publicly-owned platform significantly alleviates congestion by providing spatial-temporal subsidies, their travel costs would decrease. In addition, the occurrence of SAVs arriving later than the latest acceptable arrival time will be less frequent, so travelers' late costs and late penalties would both decrease.

		No Subsidy	With Temporal Subsidy		
Stakeholder	Cost (HKD)(10 ⁴)	Privately-owned platform	Spatial subsidy	SAV type subsidy	Our subsidy
Traveler	Travel Time Cost	46.77	32.01	32.01	32.08
	Trip Fee Cost	58.92	16.11	15.72	17.93
	Waiting Cost	4.66	3.03	3.02	2.95
	Late Cost	9.28	4.06	4.05	3.75
	Late Penalty	16.71	7.31	7.29	6.76
	Total Cost	102.92	47.90	47.51	49.95
TNC	Travel Cost	18.84	15.68	15.68	15.69
	Travel Revenue	58.92	40.60	40.59	40.64
	Parking Cost	0.41	0.76	0.76	0.75
	Total Profit	22.96	16.85	16.86	17.44
Government	Subsidy Amount	0	24.49	24.87	22.72
	Social Cost	79.54	54.78	54.77	54.47

Table 4.6: Results of POCP on the small network.

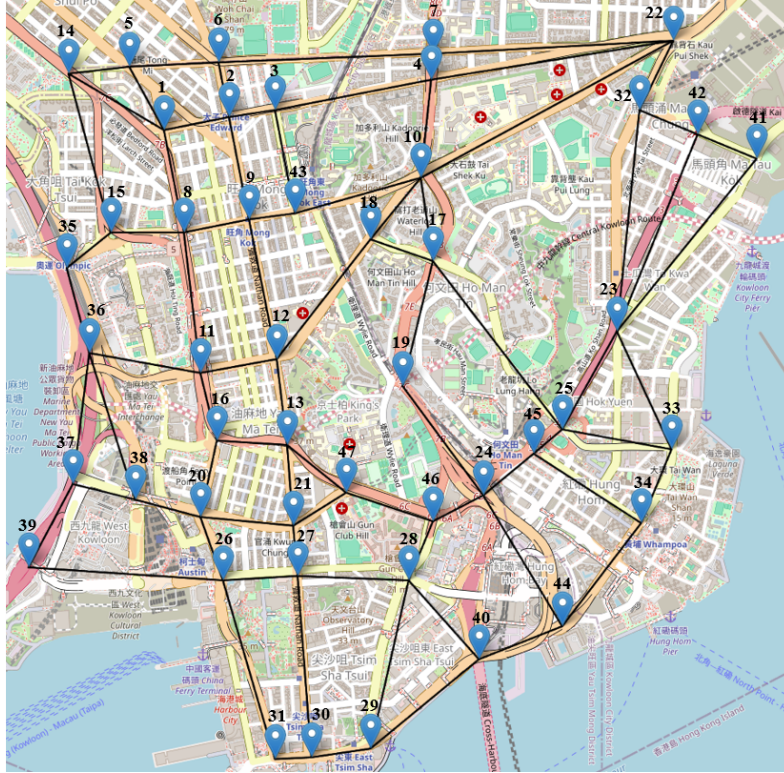
4.5.2 Hong Kong Network

To assess the scalability of the proposed POCP framework, we have implemented it in the urban area of the Kowloon Peninsula located in Hong Kong, as referred to in Figure 4.5. We would first introduce the experimental settings of POCP in the Hong Kong network. Then we evaluate the effects of applying subsidies in POCP to reduce social cost compared with the privately-owned platforms.

4.5.2.1 Experimental setting

The Hong Kong network is comprised of 47 nodes and 158 links, as illustrated in Figure 4.5. To simulate the real-world traffic conditions in Hong Kong, we first calibrate the parameters detailed in Table 4.1.

Ride-hailing services, such as Uber, are currently considered technically illegal in Hong Kong due to the existing regulatory framework (Centre, 2020). Consequently, the operations of traditional taxis in Hong Kong are utilized to calibrate the parameters presented in Table 4.1. Specifically, red taxis, which provide services across all areas of the New Territories, Kowloon, and Hong Kong Island, are examined. While previous studies have noted that the fare structure for red taxis is predominantly distance-based (Transport department, the government of Hong



Kong Special Administrative Region, 2023), in this analysis, we have derived an equivalent time-based fare model for red taxis in Hong Kong for our research.

On average, red taxi drivers in Hong Kong travel about 253.26 kilometers with passengers daily and complete around 36 trips per vehicle, as reported by (UITP, 2020). This information implies that the average distance per trip with passengers is roughly 7.035 kilometers. According to the standard taxi fare structure in Hong Kong (Transport department, the government of Hong Kong Special Administrative Region, 2023), the average fare per trip is calculated to be approximately 74.50 Hong Kong dollars (HKD). The proportion of time drivers spend with passengers is estimated at 0.63, based on data from (UITP, 2020). Consequently, the average duration per trip is estimated to be around 25.2 minutes. Using this data, we can deduce that the fare for solo travelers, represented as $C_{x=-1}$, is approximately 2.956 HKD per minute.

In this scenario, we assume that the maximum number of travelers who can use ride-sharing services in an SAV is 3. Furthermore, we establish a tiered discount system for ride-sharing services, offering discounts of 0.95, 0.70, and 0.45 based on $C_{x=-1}$ for groups with 1, 2, and 3 ride-sharing travelers (RTs), respectively, ($x = 1, 2, 3$). Furthermore, we employ Hong Kong's median hourly wage to quantify the value of travel (V_T) and waiting time (V_W) for travelers at

1.29 HKD per minute (Census and Statistics Department, 2023). Recognizing that late arrivals can result in additional losses, we establish the value of late time (V_L) as 1.8 times the value of travel time (V_T). The late penalty (C_{LP}) is set as 5.934 HKD per minute. Besides, considering the depreciation, maintenance, and electricity costs, we calculate the travel cost of SAVs (C_S) at 0.375 HKD per minute.

The parameters of the platform and Hong Kong network are summarized in Table 4.7. The Hong Kong network contains 52 O-D pairs. For this study, we simulate the operating period of POCP for 75 minutes. The total demand is estimated to be approximately 133,850, with the total number of SAVs being set at 20,250. Detailed information on demand and parking parameters can be found in Appendix B.7.

Name	Notations	Value
Length of time interval	Δ_t	3 min
Total time horizon	T	75 min
Maximum number of travelers in a SAV	X	3
Trip fare for solo traveler (ST)	$C_{x=-1}$	2.956 HKD/min
Trip fare when only one ride-sharing traveler (RT) in an SAV	$C_{x=1}$	2.808 HKD/min
Trip fare when only two ride-sharing travelers (RTs) in an SAV	$C_{x=2}$	2.069 HKD/min
Trip fare when only three ride-sharing travelers (RTs) in an SAV	$C_{x=3}$	1.330 HKD/min
Value of travel time	V_T	1.290 HKD/min
Value of waiting time	V_W	1.290 HKD/min
Value of late time	V_L	2.322 HKD/min
Late penalty	C_{LP}	5.934 HKD/min
Travel cost of SAVs	C_S	0.375 HKD/min
Sensitivity of the flexible demand to the trip fee and discount	λ	0.05
Sensitivity of the trip fee and discount for taking an SAV of type 1	η_1	0.15
Sensitivity of the trip fee and discount for taking an SAV of type 2	η_2	0.35
Sensitivity of the trip fee and discount for taking an SAV of type 3	η_3	0.50

Table 4.7: Network parameters and model coefficients for Hong Kong network.

Furthermore, to reduce the dimension of variables and enhance the computation efficiency, we aggregate the variables $f_{i,j}^{x,s,r,l}(k)$ and $g_{i,j}^{x,s,r,l}(k)$, where the node conservation for travelers flow are converted for $f_{i,j}^{x,s}(k)$ and $g_{i,j}^{x,s}(k)$. Subsequently, we use the variables $g_{o,o'}^{S,s,r,l}(k)$, $g_{o,o'}^{\mathcal{R},s,r,l}(k)$ and $f_{s,s'}^{r,l}(k)$ to infer the detailed traffic states of travelers in the network.

4.5.2.2 Results

Under the above settings in Section 4.5.2.1, we evaluate the effectiveness of the POCP framework and provide a comparison of two scenarios: the privately-owned platform with no subsidy, and POCP applied with the subsidy obtained by Algorithm 3. In the absence of a subsidy in the privately-owned platform, all discounts are set to 1, indicating that travelers would pay the full travel fee as set by the TNCS.

Table 4.8 presents the effect of applying optimal spatial-temporal discount by Algorithm 3 in the Hong Kong network. Compared with original traffic conditions that no subsidy would be applied, the proposed POCP with spatial-temporal subsidy saves the 1.108×10^6 social cost by 9.085×10^5 subsidy. Besides, Section 4.3.3.4 implies that social cost is sum of travelers, TNC, and government cost. And Equations 4.35 states that the TNC profit is the opposite of TNC cost. Therefore, it indicates that the decrease of travelers cost is the sum of the decrease of TNC profit, the increase of government cost and the decrease of social cost. Therefore, the reduction of travelers cost in POCP compared with the privately-owned platform is approximately 2.80×10^6 . This can be interpreted as the transfer of TNC profits amounting to 9.06×10^5 , government expenditures of 7.82×10^5 , and the decrease of social cost as 1.11×10^6 , which represents the reduction of congestion due to subsidy in POCP.

Stakeholder	Cost	Private-owned platform	POCP
Traveler	Traveler Travel Time Cost (10^6)	2.141	1.468
	Traveler Trip Fee Cost (10^6)	3.166	1.022
	Traveler Waiting Cost (10^5)	4.137	2.268
	Traveler Late Cost (10^5)	1.686	0.349
	Traveler Late Penalty (10^5)	4.309	0.891
	Traveler Cost (10^6)	5.459	2.663
TNC	SAV Travel Cost (10^5)	5.226	4.088
	SAV Travel Revenue (10^6)	3.166	1.931
	SAV Parking Cost (10^5)	0.444	1.708
	TNC Profit (10^6)	2.168	1.262
Government	Subsidy Amount (10^5)	0.000	9.085
	Social Cost (10^6)	3.246	2.138

Table 4.8: Results of POCP on Hong Kong network.

Then we calculate the average value of link flow under SO-DTA for all links and time intervals to explore the network-wide effects of applying subsidy in POCP in large-scale networks, as referred to in Figure 4.6.

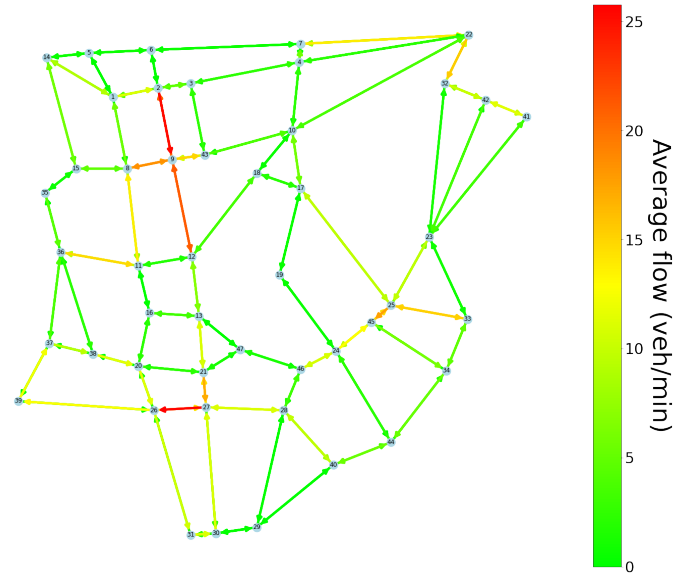


Figure 4.6: Overview of the link flow of POCP in Hong Kong network.

Finally, we presented the dynamics of the subsidy in POCP over time in Figure 4.7.

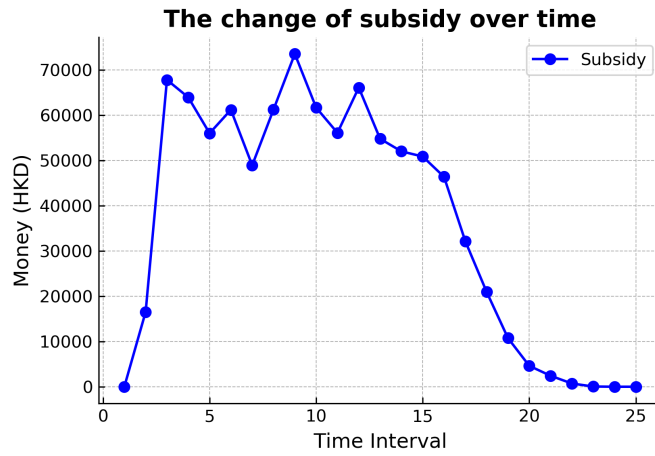


Figure 4.7: Dynamics of subsidy in POCP.

Chapter 5

Mixed sub-network equilibrium of automated vehicles in general traffic networks

5.1 Introduction

The rapid development of automated vehicle (AV) technology offers numerous advantages, particularly within large-scale transportation networks, offering significant potential to transform modern mobility (Hoogendoorn et al., 2014; Li et al., 2024; Luo et al., 2024). For individual users, in addition to the advantages of increased convenience and a personalized travel experience, users can choose routes that minimize their travel time, potentially avoiding congested areas and optimizing their journey based on real-time traffic information by controlling AVs (Qin et al., 2024; Sharon et al., 2017; Wadud and Chintakayala, 2021). From the perspective of the transportation system as a whole, the large-scale deployment and control of AVs promise to substantially reduce overall congestion through coordinated traffic management by the system (Barzegari et al., 2023; Du and Ma, 2024; Levin, 2017; Militão and Tirachini, 2021). However, while AVs can deliver substantial benefits, the challenge of effectively controlling AVs within large-scale road networks remains a critical issue (Li et al., 2017). In particular, determining how AVs should be controlled to balance individual and collective goals is an open question that requires careful consideration.

The User Equilibrium (UE) and System Optimal (SO) principles are widely applied in trans-

portation networks to represent the different behaviors of users and the system (Wardrop, 1952). Under UE, each vehicle aims to minimize its own travel without considering the impact on others, leading to an equilibrium where no vehicle can improve its travel time by changing routes (Sheffi, 1985). In contrast, SO seeks to minimize the total travel cost in the whole network, requiring a coordinated approach where vehicles may sacrifice individual travel efficiency for the benefit of the system as a whole. These principles have provided the foundation for research on traffic control and management (Boyles et al., 2020).

As AVs have the capacity to be centrally controlled, they provide an opportunity to implement the SO principle more effectively than human-driven vehicles (HVs). Previous studies have demonstrated that AVs can significantly improve network performance when controlled to follow SO principles, reducing total travel cost and mitigating congestion (Chen, Lin, et al., 2020). However, large-scale implementation of SO control across an entire traffic network can be impractical due to the complexity of managing every vehicle simultaneously (Shah et al., 2024). Therefore, the question arises: can AV control be applied more selectively within the network to achieve similar benefits without requiring full system-wide coordination?

This work addresses this question by introducing a novel concept of mixed sub-network equilibrium (MSNE) of AVs in general traffic networks. We propose dividing the entire traffic network into two types of sub-networks: the **user equilibrium sub-network (UESN)** and the **system optimal sub-network (SOSN)**. In the UESN, AVs are controlled by the users and follow the UE principle, with each vehicle minimizing its own travel cost. In contrast, in the SOSN, AVs are centrally controlled by the system to minimize the total travel cost, following the SO principle. This hybrid approach allows for localized system control in critical areas of the network while maintaining decentralized, user-driven control elsewhere.

In practice, the division of a traffic network into UESN and SOSN sub-networks has significant real-world relevance. For instance, in many urban areas, traffic congestion is often concentrated in specific regions, such as city centers or downtown areas, where traffic demand is consistently high, and congestion are common (Falcocchio and Levinson, 2015). For example, in February 2003, London introduced the London Congestion Charging Scheme (CCS), which imposed fees on vehicles entering the central London zone during weekday hours. This measure effectively reduced traffic volumes in the central area (Beevers and Carslaw, 2005). Additionally, Singapore implemented the Area License Scheme (ALS) in the previous century with the

goal of discouraging vehicles from entering the congested central region during morning peak hours (Dinh, 2019). To address these problems in real traffic conditions under fully automated environment, by applying the SOSN control of AVs only to these critical regions, traffic authorities can substantially reduce overall congestion and improve network efficiency without the need for full network-wide control. The remaining areas of the network, which may experience lower traffic demand or less severe congestion, can continue to operate under UESN, where AVs are controlled independently by users to minimize their own travel cost. Besides, a brief illustration of the UESN and SOSN in traffic network is shown in Figure 5.1. In UESN, AVs follow user-controlled routes (solid black lines), where navigation decisions are made by the user. However, if the AV enters an SOSN (blue-shaded area), a control transition is required at the boundary, facilitated by a Roadside Unit (RSU). This transition ensures a secure and seamless handover of control from the user to the system. Once inside the SOSN, AVs follow system-controlled routes (dashed black lines), where the system autonomously determines the optimal path based on traffic conditions and network efficiency. The system also decides the exit point for re-entering the UESN. At the boundary between SOSN and UESN, another RSU facilitates the transition of control back to the user, allowing the AVs to resume user-directed navigation. The deployment of RSUs at transition points is critical to ensuring a reliable, efficient, and secure exchange of control between users and the autonomous driving system. To conclude, this selective approach ensures that control resources are allocated efficiently to the areas where they are most needed, while still allowing a degree of autonomy in less critical areas.

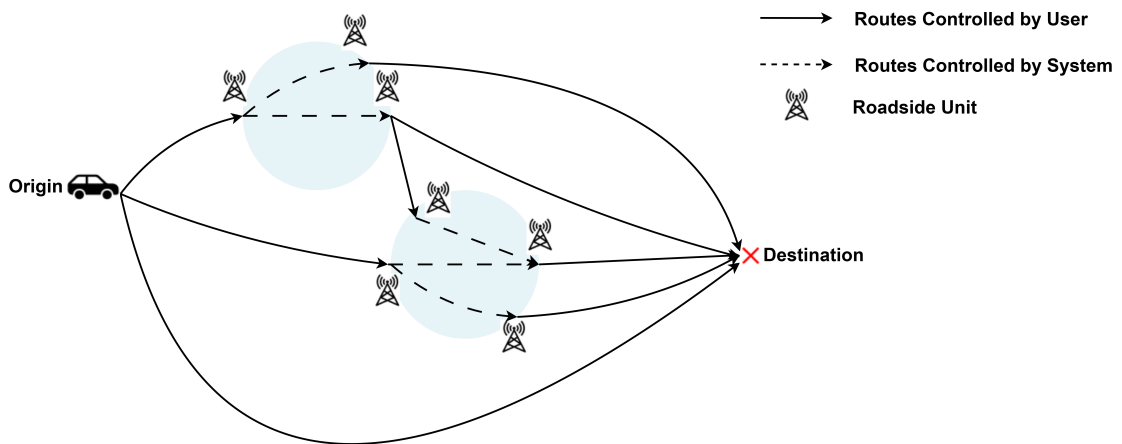


Figure 5.1: Illustration of SOSN and UESN.

Furthermore, excessive government or centralized control over AVs could face resistance

from users who may prefer to retain control over their vehicles or experience increased travel costs due to the complexity of system-wide management (Othman, 2021). The operational cost of implementing widespread system optimal control across an entire network can also be prohibitive, especially in large cities where infrastructure, data collection, and communication requirements are extensive (Yaqoob et al., 2019). By identifying the optimal size of the SOSN for effective centralized control, traffic managers can balance minimizing total travel cost with limiting the degree of system intervention.

To address this concern, we propose the concept of the **critical size of system optimal sub-network (CS-SOSN)**. Intuitively, as the size of SOSN increases, the total travel cost across the entire network decreases. When the size of SOSN is zero, the network operates under the UE principle, where the total travel cost would always be higher. Conversely, when the SOSN encompasses the entire network, it reaches the SO state, where the total travel cost is minimized. To represent the minimum government intervention required to achieve SO, CS-SOSN can be intuitively represented by a layman's version of the definition in Definition 5.1.1 and Figure 5.2. We will provide a rigorous definition of CS-SOSN in the later sections. To this end, we first propose an algorithm called sub-network-augmented method successive averages (SNA-MSA) to obtain the mixed sub-network equilibrium (MSNE) given the UESN and SOSN. Then we propose a novel adaptive variable neighborhoods search (AVNS) algorithm to solve the CS-SOSN.

Definition 5.1.1 (Critical Size of System Optimal Sub-Network - Layman's version). The smallest size of SOSN that still achieves the network-wide SO is referred to as the critical size of the system optimal sub-network (CS-SOSN).

In summary, this work presents a novel concept of MSNE of AVs in general traffic networks by dividing the network into two sub-networks: UESN and SOSN. While UESN allows AVs to be controlled by users following the UE principle, SOSN uses centralized control to follow the SO principle, minimizing the total travel cost. To determine the most efficient use of centralized control, we introduce the concept of CS-SOSN, which represents the minimum size of the SOSN required to achieve network-wide optimal performance. Finally, we validate the proposed CS-SOSN model on the real Hong Kong network. The results show that, in real-world traffic operations, only 44.17% of the network needs to be controlled to bring the whole network to a state of SO, without requiring full control of the whole network. This provides an effective

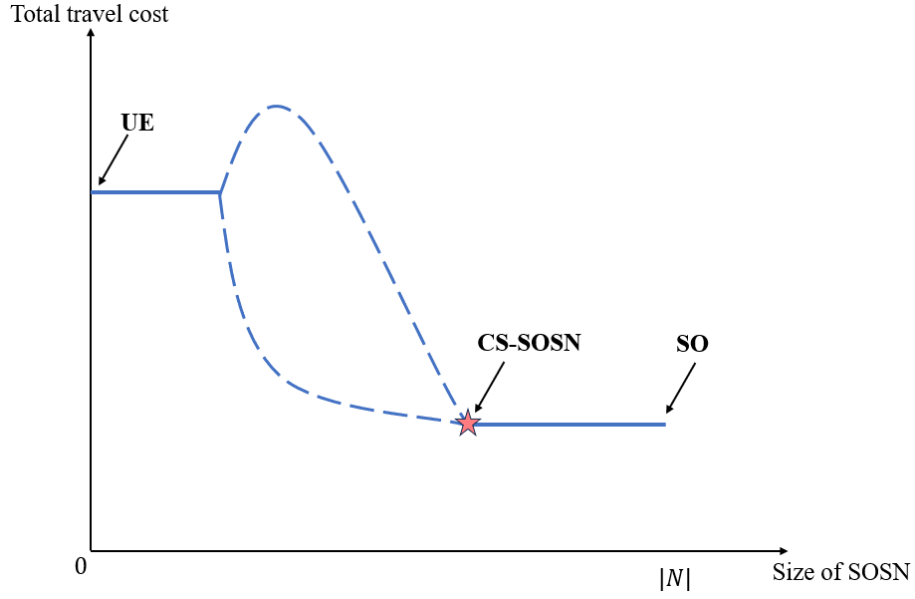


Figure 5.2: Critical size of system optimal sub-network.

method for managing the traffic network, reducing congestion, and improving performance in a fully AV environment.

The main contributions of this work are as follows:

- We develop a framework of mixed sub-network equilibrium (MSNE) for AVs by dividing the network into the user equilibrium sub-network (UESN) and the system optimal sub-network (SOSN).
- We analyze the properties of MSNE under the existence of UESN and SOSN, which is proved as equivalent to a variational inequality (VI) problem. A sub-network-augmented method successive averages (SNA-MSA) algorithm is proposed to obtain the MSNE given the UESN and SOSN.
- The concept of critical size of the system optimal sub-network (CS-SOSN) is proposed to quantify the minimum size of SOSN needed to achieve the optimal performance in a traffic network, ensuring maximum system performance with minimum government intervention. We propose a novel adaptive variable neighborhoods search (AVNS) algorithm to solve the CS-SOSN.

5.2 Problem Settings

The traffic network is represented as a directed graph $G(N, L)$, where N is the set of all nodes and L is the set of all links. The frequently used notations are summarized in Table 5.1. The whole network is divided into two types of sub-networks: user equilibrium sub-network (UESN) and system optimal sub-network (SOSN). In this study, we only consider the case of a single SOSN, but our proposed model can be easily extended to multiple SOSNs. We will introduce their basic concepts and assumptions as follows:

- In UESN, AVs are controlled by the users with the objective to minimize their own perceived travel cost following the UE principle, which equals finding the shortest path with minimum travel cost to the destination.
- The users could decide whether to enter the SOSN from the UESN and determine the entrance. AVs using the same entrance and exit of SOSN may experience different travel costs due to the SO principle in SOSN. We assume that AVs perceive their travel cost in SOSN to be the minimum travel cost among all minimum marginal cost paths from their entrances of SOSN to the exits of SOSN. It corresponds to the assumptions in (Chen et al., 2017b). This also applies to the case from the entrance of SOSN to the origin within SOSN.
- In SOSN, AVs are controlled by the system with the objective of minimizing the total travel cost in the whole network following the SO principle, which equals finding the shortest path with minimum marginal cost to the destination. Besides, the system has the right to decide whether to enter UESN and where to enter.
- The system could decide whether to exit the SOSN and enter the UESN, as well as from where to exit. AVs with the same exit of SOSN and origin in UESN may experience different marginal cost due to the UE principle in UESN. We assume that AVs perceive their marginal cost in UESN to be the minimum marginal cost among all minimum travel cost paths from their exit of SOSN to the origin in UESN, which also aligns with the assumptions presented in (Chen et al., 2017b).

Figure 5.3 serves as an example to illustrate SOSN and UESN. The area within the red dashed line represents the SOSN, while other areas represent the UESN. As shown in Figure 5.3, in a

Type	Notation	Description
Sets	N	Set of nodes
	L	Set of links
	Φ	Set of nodes in UESN
	Ψ	Set of nodes in SOSN
	Ω	Set of boundary nodes with upstream node in UESN and downstream node in SOSN. Mathematically, $\Omega = \{k \in N \mid \exists i \in \Phi, j \in \Psi, s.t. (i, k) \in L, (k, j) \in L\}$
	Θ	Set of boundary nodes with upstream node in SOSN and downstream node in UESN. Mathematically, $\Theta = \{k \in N \mid \exists i \in \Psi, j \in \Phi, s.t. (i, k) \in L, (k, j) \in L\}$
	$P_{m,n}$	Set of paths from node $m \in N$ to node $n \in N$
	O	Set of origins
	S	Set of destinations
	$D_{o,s}$	Travel demand from the origin $o \in O$ to destination $s \in S$
Parameters	$t_{i,j}^f$	Free-flow travel time of link $(i, j) \in L$
	$H_{i,j}$	Capacity of link $(i, j) \in L$
Variables	$f_{m,n}^r$	Path flow on path $r \in P_{m,n}$ from node $m \in N$ to node $n \in N$
	$x_{i,j}$	Link flow on link $(i, j) \in L$
	$c_{m,n}^r$	Perceived travel cost of path $r \in P_{m,n}$ from node $m \in N$ to node $n \in N$
	$h_{m,n}^r$	Perceived marginal cost of path $r \in P_{m,n}$ from node $m \in N$ to node $n \in N$
	$c_{i,j}$	Travel cost of link $(i, j) \in L$
	$h_{i,j}$	Marginal cost of link $(i, j) \in L$
	v_i	Binary variables indicating whether node i is in the SOSN. If $v_i = 1$, node $i \in \Psi$. Otherwise, node $i \in \Phi$.

Table 5.1: Notations table.

large network, when AVs enter the red area, they are controlled by the system, following the SO principle to reduce congestion across the entire network. In real-world traffic, SOSN often corresponds to highly congested areas in a city, such as downtown or CBD areas. The government typically only needs to control the AVs in these congested areas to alleviate congestion throughout the city and achieve the system optimal state.

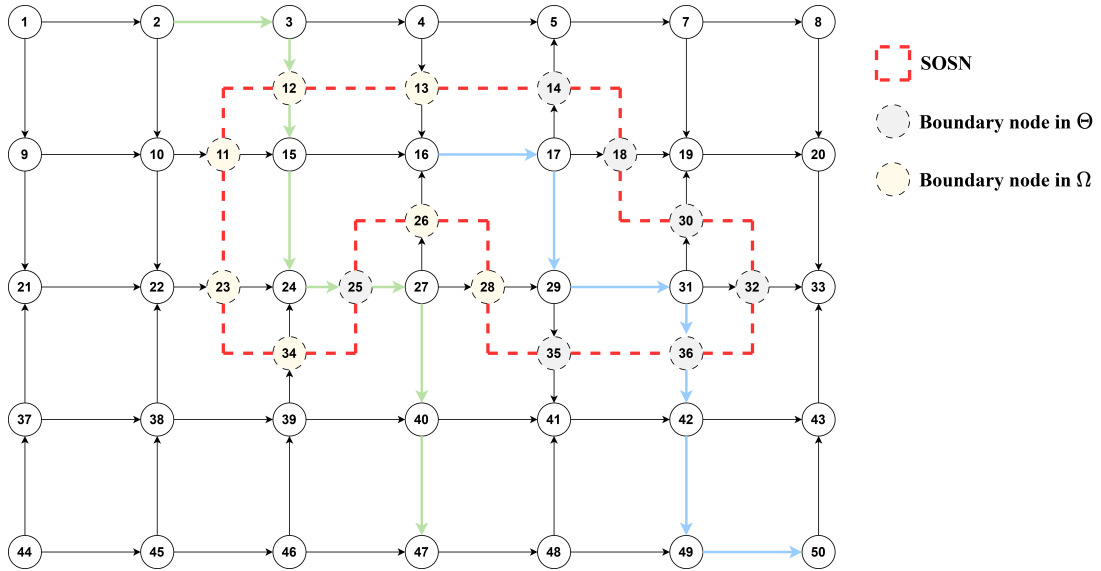


Figure 5.3: An example of SOSN and UESN.

Besides, to better represent the process of entering or exiting the UESN or SOSN, we have

added the following boundary nodes between the links connecting the UESN and SOSN based on the different link directions, as illustrated in Figure 5.3. This addition does not change the mixed sub-network equilibrium (MSNE) in the UESN and SOSN and we would illustrate them as follows:

- We add the boundary node $k \in \Omega$ at the middle of the link (i, j) connecting from UESN to SOSN, where $i \in \Phi$ and $j \in \Psi$. The boundary node $k \in \omega$ could represent the process by which the users decide to exit UESN and enter SOSN at boundary node k when $x_{i,k} > 0$. For any boundary node $k \in \Omega$, there uniquely exists a node i in UESN that is connected to k and serves as its upstream node, and there also uniquely exists a node j in SOSN that is connected to k and serves as its downstream node.
- We add the boundary node $k \in \Theta$ at the middle of the link (i, j) connecting from SOSN to UESN, where $i \in \Psi$ and $j \in \Phi$. The boundary node $k \in \Theta$ could represent the process that the system decides to exit SOSN and enter UESN at boundary node k when $x_{i,k} > 0$. For any boundary node $k \in \Theta$, there uniquely exists a node i in SOSN that is connected to k and serves as its upstream node, and there also uniquely exists a node j in UESN that is connected to k and serves as its downstream node.
- We set $t_{i,k}^f = t_{k,j}^f = 0.5t_{i,j}^f$ and $H_{i,k} = H_{j,k} = H_{i,j}$ for $\forall k \in \Omega \cup \Theta$. The setting of the boundary node indicates the flow conservation relationship as $x_{i,k} = x_{k,j} = x_{i,j}$. Therefore, the setting of the boundary node would not change the travel cost and marginal cost compared to the original network as follows, where we always set a same constant parameters α and β in the BPR function for all links.

$$\begin{aligned}
c_{i,k} + c_{k,j} &= t_{i,k}^f (1 + \alpha_{i,k} (\frac{x_{i,k}}{H_{i,k}})^{\beta_{i,k}}) + t_{k,j}^f (1 + \alpha_{k,j} (\frac{x_{k,j}}{H_{k,j}})^{\beta_{k,j}}) \\
&= t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}}{H_{i,j}})^{\beta_{i,j}}) \\
&= c_{i,j} \\
h_{i,k} + h_{k,j} &= c_{i,k} + x_{i,k} c'_{i,k} + c_{k,j} + x_{k,j} c'_{k,j} \\
&= c_{i,k} + c_{k,j} + t_{i,k}^f \alpha_{i,k} \beta_{i,k} (\frac{x_{i,k}}{H_{i,k}})^{\beta_{i,k}} + t_{k,j}^f \alpha_{k,j} \beta_{k,j} (\frac{x_{k,j}}{H_{k,j}})^{\beta_{k,j}} \\
&= c_{i,j} + x_{i,j} c'_{i,j} \\
&= h_{i,j}
\end{aligned}$$

5.3 Model

In this section, we first illustrate the formulation of MSNE in Section 5.3.1. Then we analyze the properties of MSNE in Section 5.3.2. Lastly, we propose the formulation of CS-SOSN in Section 5.3.3.

5.3.1 Formulation of mixed sub-network equilibrium

In this section, we propose the formulation of MSNE. Besides, AVs follow different routing principles depending on their locations. Specifically, in UESN, AVs follow the UE principle to choose the path that minimizes the travel cost. Conversely, in SOSN, they would follow the SO principle, choosing routes that minimize the marginal cost. In terms of node level, AVs follow the UE principle for origin node o in UESN and boundary $k \in \Theta$. In contrast, AVs follow the SO principle for origin node o in SOSN and boundary $k \in \Omega$.

We would first present an example to illustrate how AVs follow different routing principles depending on their locations, especially their nodes. As is shown in Figure 5.3, we choose two paths to illustrate it as follows:

- Path $2 \rightarrow 3 \rightarrow 12 \rightarrow 15 \rightarrow 24 \rightarrow 25 \rightarrow 27 \rightarrow 40 \rightarrow 47$: for the AVs choosing this path, the origin is node 2 in UESN, and the user chooses to leave UESN at boundary node $12 \in \Omega$ and enter SOSN, and the system would take over the control the AVs. After entering SOSN, the system chooses to leave SOSN at boundary node $25 \in \Theta$ and enter UESN, and the user controls the AVs to reach the destination node 47 in UESN.
 - Since the origin 2 is in UESN, so the path $2 \rightarrow \dots \rightarrow 47$ must satisfy the UE principle. This implies that the travel cost of this path would be minimized when the path flow is greater than zero.
 - For the boundary node $12 \in \Omega$, the path $12 \rightarrow \dots \rightarrow 47$ must satisfy the SO principle, meaning that the marginal cost of this path is minimized when the path flow is greater than zero.
 - For the boundary node $25 \in \Theta$, the path $25 \rightarrow \dots \rightarrow 47$ must follow the UE principle, ensuring that the travel cost of this path is minimized when the path flow is greater than zero.

- Path $16 \rightarrow 17 \rightarrow 29 \rightarrow 31 \rightarrow 36 \rightarrow 42 \rightarrow 49 \rightarrow 50$: for the AVs choosing this path, the origin is node 16 in SOSN, and the system chooses to leave SOSN at boundary node $35 \in \Theta$ and enter UESN, and the user would take over the control the AVs to reach the destination node 55 in UESN.
 - Since the origin 16 is in UESN, so the path $16 \rightarrow \dots \rightarrow 50$ must satisfy the SO principle. This implies that the marginal cost of this path would be minimized when the path flow is greater than zero.
 - For the boundary node $35 \in \Theta$, the path $35 \rightarrow \dots \rightarrow 50$ must follow the UE principle, ensuring that the travel cost of this path is minimized when the path flow is greater than zero.

Next, we will illustrate the distinct routing behaviors of AVs at four types of nodes: the origin node in UESN, the origin node in SOSN, the boundary node in Ω , and the boundary node in Θ , respectively.

When the origin is in UESN as $o \in \Phi \cap O$, AVs are controlled by the users by following the UE principle, which equals to find the path with minimum travel cost from the origin $o \in \Phi \cap O$ to the destination $s \in S$. Equations 5.1 to 5.3 are complementary slackness conditions to ensure that all used paths from $o \in \Phi \cap O$ to $s \in S$ have same travel cost, which is equal to or less than the travel cost of any unused path.

$$f_{o,s}^r(c_{o,s}^r - \pi_{o,s}) = 0 \quad \forall o \in O \cap \Phi, s \in S, r \in P_{o,s} \quad (5.1)$$

$$f_{o,s}^r \geq 0 \quad \forall o \in O \cap \Phi, s \in S, r \in P_{o,s} \quad (5.2)$$

$$c_{o,s}^r - \pi_{o,s} \geq 0 \quad \forall o \in O \cap \Phi, s \in S, r \in P_{o,s} \quad (5.3)$$

where $f_{o,s}^r$ is the path flow on path $r \in P_{o,s}$ from o to s . $c_{o,s}^r$ is the perceived travel cost of path $r \in P_{o,s}$ from o to s . $\pi_{o,s}$ is the Lagrangian multiplier representing the minimum travel cost of the path from o to s . Φ is the set of nodes in UESN. O and S are the set of origins and destinations, respectively. $P_{o,s}$ represents the set of paths from o to s .

Equation 5.4 indicates that the sum of all paths flow from $o \in \Phi$ to s equals to the demand of O-D pair (o, s) , where $D_{o,s}$ is the travel demand from the origin $o \in O$ to destination $s \in S$.

$$\sum_{r \in P_{o,s}} f_{o,s}^r = D_{o,s} \quad \forall o \in O \cap \Phi, s \in S \quad (5.4)$$

Equation 5.5 presents that the perceived path travel cost from the origin $o \in \Phi$ in UESN to the destination $s \in \Phi$ in UESN consists of the sum of travel cost of links in UESN and perceived travel cost in SOSN. The perceived travel cost in SOSN as $c_{k,l}$ is the minimum travel cost of used paths from the boundary node $k \in \Omega$ to $l \in \Theta$.

$$c_{o,s}^r = \sum_{i \in \Phi \cup \Theta} \sum_{j \in N} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r + \sum_{k \in \Omega} \sum_{l \in \Theta} c_{k,l} \delta_{k,l}^r \quad \forall o \in O \cap \Phi, s \in S \cap \Phi, r \in P_{o,s} \quad (5.5)$$

where $c_{i,j}$ is the travel cost of link (i, j) and $c_{k,l}$ represents the perceived travel cost from $k \in \Omega$ to $l \in \Theta$. Besides, $\delta_{i,j}^r$ is the indicator about whether link (i, j) is in the path r . $\delta_{k,l}^r$ is the indicator about whether both boundary node $k \in \Omega$ and $l \in \Theta$ are in the path r .

Equation 5.6 presents that the perceived path travel cost from the origin $o \in \Phi$ in UESN to the destination $s \in \Psi$ in SOSN consists of the sum of the travel cost of links in UESN and perceived travel cost in SOSN. The perceived travel cost in SOSN as $c_{k,s}$ is the minimum travel cost of used paths from the boundary node $k \in \Omega$ to $s \in \Psi$.

$$c_{o,s}^r = \sum_{i \in \Phi \cup \Theta} \sum_{j \in N} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r + \sum_{k \in \Omega} c_{k,s} \delta_{k,s}^r \quad \forall o \in O \cap \Phi, s \in S \cap \Psi, r \in P_{o,s} \quad (5.6)$$

where $\delta_{k,s}^r$ is the indicator about whether both boundary node $k \in \Omega$ and $s \in \Psi$ are in the path r .

When the origin is in SOSN as $o \in \Psi \cap O$, AVs are controlled by the system by following the SO principle, which equals to find the path with minimum marginal cost from the origin $o \in \Psi \cap O$ to the destination $s \in S$. Equations 5.7 to 5.9 are complementary slackness conditions to ensure that all used paths from $o \in \Psi \cap O$ to $s \in S$ have same marginal cost, which is equal to or less than the marginal cost of any unused path.

$$f_{o,s}^r(h_{o,s}^r - \eta_{o,s}) = 0 \quad \forall o \in O \cap \Psi, s \in S, r \in P_{o,s} \quad (5.7)$$

$$f_{o,s}^r \geq 0 \quad \forall o \in O \cap \Psi, s \in S, r \in P_{o,s} \quad (5.8)$$

$$h_{o,s}^r - \eta_{o,s} \geq 0 \quad \forall o \in O \cap \Psi, s \in S, r \in P_{o,s} \quad (5.9)$$

where $h_{o,s}^r$ is the perceived marginal cost of path $r \in P_{o,s}$ from o to s . $\eta_{o,s}$ is the Lagrangian multiplier representing the minimum marginal cost of path from o to s . Ψ is the set of nodes in SOSN.

Equation 5.10 indicates that the sum of all paths flow from $o \in \Psi$ to s equals to the demand of O-D pair (o, s) .

$$\sum_{r \in P_{o,s}} f_{o,s}^r = D_{o,s} \quad \forall o \in O \cap \Psi, s \in S \quad (5.10)$$

Equation 5.11 presents that the perceived marginal cost from the origin $o \in \Psi$ to the destination $s \in \Psi$ is the sum of marginal cost of links in the path, where $h_{i,j}$ is the marginal cost of link (i, j) .

$$h_{o,s}^r = \sum_{i \in N} \sum_{j \in N} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall o \in O \cap \Psi, s \in S \cap \Psi, r \in P_{o,s} \quad (5.11)$$

Equation 5.12 states that the perceived marginal cost from the origin $o \in \Psi$ to the destination $s \in \Phi$ consists of the sum of marginal cost of links in SOSN and perceived marginal cost in UESN, where $h_{l,s}$ is the perceived marginal cost from $l \in \Theta$ to $s \in \Phi$. The perceived marginal cost in UESN is the minimum marginal cost of the path from the boundary node $l \in \Theta$ to $s \in \Phi$ with minimum marginal cost.

$$h_{o,s}^r = \sum_{i \in \Phi} \sum_{j \in \Phi \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r + \sum_{l \in \Theta} h_{l,s} \delta_{l,s}^r \quad \forall o \in O \cap \Psi, s \in S \cap \Phi, r \in P_{o,s} \quad (5.12)$$

For the boundary node $k \in \Omega$, when users decide to leave UESN and enter SOSN at the boundary $k \in \Omega$, the AVs would be controlled by the system in SOSN following the SO principle, which implies that the routing choices behavior within SOSN and the choices of exit node from

the SOSN is made with the consideration of improving system performance. Equations 5.13 to 5.15 ensure that when AVs enter the SOSN from the UESN, they will follow the SO principle to select the path with the minimum marginal cost to their destinations.

$$f_{k,s}^r(h_{k,s}^r - \eta_{k,s}) = 0 \quad \forall k \in \Omega, s \in S, r \in P_{k,s} \quad (5.13)$$

$$f_{k,s}^r \geq 0 \quad \forall k \in \Omega, s \in S, r \in P_{k,s} \quad (5.14)$$

$$h_{k,s}^r - \eta_{k,s} \geq 0 \quad \forall k \in \Omega, s \in S, r \in P_{k,s} \quad (5.15)$$

The sum of path flow at the boundary $k \in \Omega$ depends on how many AVs decide to leave the UESN and enter the SOSN at boundary $k \in \Omega$. $\delta_{i,k}^L$ is the indicator about whether (i, k) is a link. As illustrated in Section 5.2, there only exists a node in UESN connected with the boundary $k \in \Omega$.

$$\sum_{r \in P_{k,s}} f_{k,s}^r = \sum_{i \in \Phi} \delta_{i,k}^L x_{i,k} \quad \forall k \in \Omega, s \in S, r \in P_{k,s} \quad (5.16)$$

Equation 5.17 states that the path marginal cost from the boundary $k \in \Omega$ to the destination $s \in S \cap \Psi$ in SOSN is the sum of the marginal cost of links in the path k to s .

$$h_{k,s}^r = \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, s \in S \cap \Psi, r \in P_{k,s} \quad (5.17)$$

Equation 5.18 presents that the path marginal cost from the boundary $k \in \Omega$ to the destination $s \in S \cap \Phi$ in UESN consists of the sum of marginal cost of links in SOSN and perceived marginal cost in UESN.

$$h_{k,s}^r = \sum_{i \in \Phi} \sum_{j \in \Phi \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r + \sum_{l \in \Theta} h_{l,s} \delta_{l,s}^r \quad \forall k \in \Omega \cap O, s \in S \cap \Phi, r \in P_{k,s} \quad (5.18)$$

Then we would calculate the perceived travel cost from the boundary $k \in \Omega$ to $l \in \Theta$ as follows. Equations 5.19 and 5.20 illustrate the actual travel cost and actual marginal cost for each path from $k \in \Omega$ to $l \in \Theta$, respectively. The minimum path marginal cost from $k \in \Omega$ to $l \in \Theta$ is presented as $h_{k,l}$ in Equation 5.21. Then the Equation 5.22 states that the perceived

travel cost from the boundary $k \in \Omega$ to $l \in \Theta$ is the minimum travel cost among all paths from $k \in \Omega$ to $l \in \Theta$ with minimum marginal cost $h_{k,l}$.

$$h_{k,l}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, l \in \Phi, r \in P_{k,l} \quad (5.19)$$

$$c_{k,l}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, l \in \Phi, r \in P_{k,l} \quad (5.20)$$

$$h_{k,l} = \min\{h_{k,l}^r \mid \forall r \in P_{k,l}\} \quad \forall k \in \Omega, l \in \Phi \quad (5.21)$$

$$c_{k,l} = \min\{c_{k,l}^r \mid h_{k,l}^r = h_{k,l}, \forall r \in P_{k,l}\} \quad \forall k \in \Omega, l \in \Phi \quad (5.22)$$

For the boundary node $k \in \Theta$, when the system decides to control the AVs to leave SOSN and enter UESN at the boundary node $k \in \Theta$, the AVs would be controlled by the users in UESN following the UE principle. Equations 5.23 to 5.25 ensure that when AVs enter the UESN from the SOSN, they will follow the UE principle to select the path with the minimum travel cost to their destinations.

$$f_{k,s}^r (c_{k,s}^r - \pi_{k,s}) = 0 \quad \forall k \in \Theta, s \in S, r \in P_{k,s} \quad (5.23)$$

$$f_{k,s}^r \geq 0 \quad \forall k \in \Theta, s \in S, r \in P_{k,s} \quad (5.24)$$

$$c_{k,s}^r - \pi_{k,s} \geq 0 \quad \forall k \in \Theta, s \in S, r \in P_{k,s} \quad (5.25)$$

The sum of path flow at the boundary $k \in \Theta$ depends on how many AVs decide to leave the SOSN and enter the UESN at boundary $k \in \Theta$. As illustrated in Section 5.2, there only exists an node in SOSN connected with the boundary $k \in \Theta$.

$$\sum_{r \in P_{k,s}^\Phi} f_{k,s}^r = \sum_{j \in \Psi} \delta_{i,k}^L x_{i,k} \quad \forall k \in \Theta, s \in S, r \in P_{k,s} \quad (5.26)$$

Equation 5.27 states that the path travel cost from the boundary $k \in \Theta$ to the destination $s \in S$ is the sum of travel cost of links in the path k to s .

$$c_{k,s}^r = \sum_{i \in \Theta \cup \Phi} \sum_{j \in \Phi} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r \quad \forall k \in \Theta, s \in S, r \in P_{k,s} \quad (5.27)$$

Then we would calculate the perceived marginal cost from the boundary $k \in \Theta$ to the destination $s \in S \cap \Phi$ in UESN as follows. Equation 5.28 illustrates the actual marginal cost for each path from the boundary $k \in \Theta$ to the destination $s \in S \cap \Phi$ in UESN, respectively. The minimum path travel cost from $k \in \Theta$ to $s \in S \cap \Phi$ is presented as $c_{k,s}$ in Equation 5.29. Then the Equation 5.22 states that the perceived travel cost from the boundary $k \in \Omega$ to $l \in \Theta$ is the minimum travel cost among all paths from $k \in \Omega$ to $l \in \Theta$ with minimum marginal cost $h_{k,l}$.

$$h_{k,s}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Theta, s \in S \cap \Phi, r \in P_{k,s} \quad (5.28)$$

$$c_{k,s} = \min\{c_{k,s}^r \mid \forall r \in P_{k,s}\} \quad \forall k \in \Theta, s \in S \cap \Phi \quad (5.29)$$

$$h_{k,s} = \min\{h_{k,s}^r \mid c_{k,s}^r = c_{k,s}, \forall r \in P_{k,s}\} \quad \forall k \in \Theta, s \in S \cap \Phi \quad (5.30)$$

There also exist other constraints for the MSNE problem given the existence of UESN and SOSN. Equation 5.31 illustrates the relationship between the link flow and path flow.

$$x_{i,j} = \sum_{o \in O} \sum_{s \in S} \sum_{r \in P_{o,s}} f_{o,s}^r \delta_{i,j}^r \quad \forall (i,j) \in L \quad (5.31)$$

Equations 5.32 and 5.33 present the formulation of travel cost and marginal cost in each link, respectively.

$$c_{i,j} = t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}}{H_{i,j}})^{\beta_{i,j}}) \quad \forall (i,j) \in L \quad (5.32)$$

$$h_{i,j} = c_{i,j} + x_{i,j} c'_{i,j} \quad \forall (i,j) \in L \quad (5.33)$$

To conclude, we propose the formulation of mixed sub-network equilibrium (MSNE) under the determined UESN and SOSN as follows.

MSNE Find the path flow variables $p_{o,s}^r, p_{k,s}^r$, link flow variables $x_{i,j}$, and cost variables $c_{k,s}^r, h_{k,s}^r$ for $o \in O, s \in S, (i,j) \in L, k \in \Omega \cup \Theta$ such that Equations (5.1)–(5.33) hold under the given Ψ, Φ, Ω and Θ .

Specifically, Equations (5.1) to (5.6) ensure that the users would select the path with the

minimum travel cost from their origins within the UESN. This includes decisions regarding whether and where to enter the SOSN from the UESN. Equations (5.7) to (5.12) guarantee that the system would choose the path with the minimum marginal cost from the origin within the SOSN, encompassing choices about whether and where to enter the UESN from the SOSN. Equations (5.13) to (5.22) ensure that when AVs enter the SOSN, their routing is controlled by the system in accordance with the SO principle. This takes into account the impact of routing choices within the SOSN and decisions regarding exits from the SOSN. Conversely, Equations (5.23) to (5.30) ensure that when AVs transition from the SOSN to the UESN, their routing is controlled by users following the UE principle.

5.3.2 Properties of mixed sub-network equilibrium

In this section, we present the properties of the MSNE and analyze the conditions under which users would choose to control AVs to enter the SOSN from the UESN, as well as the conditions where the system would control AVs to return to the UESN from the SOSN.

Proposition 16 presents that every used path connecting an origin in UESN and destination has equal and minimal travel cost. For users in UESN, the only factor they consider is their own travel cost. If the total travel cost of the route arranged by the system after entering SOSN is equal to less than the travel cost of any path that stays within UESN without passing through SOSN, the user will consider leaving UESN and entering SOSN. Otherwise, the user will not enter SOSN.

Proposition 16. In MSNE, for $o \in \mathcal{O} \cap \Phi$, $s \in \mathcal{S}$, $\forall r_1, r_2 \in P_{o,s}$ with $f_{o,s}^{r_1}, f_{o,s}^{r_2} > 0$, we have $c_{o,s}^{r_1} = c_{o,s}^{r_2} = \pi_{o,s}$. Besides, $\forall r \in P_{o,s}$ with $f_{o,s}^r = 0$, we have $c_{o,s}^r \geq \pi_{o,s}$.

Proof Sketch. Obviously obtained by Equations 5.1 to 5.3.

Proposition 17 illustrates that every used path connecting an origin in SOSN and destination has equal and minimal marginal cost. For AVs in SOSN, the system only considers the marginal cost. If the total marginal cost of the route arranged by the user after entering UESN is equal to less than the marginal cost of any path that stays within SOSN without passing through UESN, the user will consider leaving SOSN and entering UESN. Otherwise, the system will not enter UESN.

Proposition 17. In MSNE, for $o \in O \cap \Psi$, $s \in S$, $\forall r_1, r_2 \in P_{o,s}$ with $f_{o,s}^{r_1}, f_{o,s}^{r_2} > 0$, we have $h_{o,s}^{r_1} = h_{o,s}^{r_2} = \eta_{o,s}$. Besides, $\forall r \in P_{o,s}$ with $f_{o,s}^r = 0$, we have $h_{o,s}^r \geq \eta_{o,s}$.

Proof Sketch. Obviously obtained by Equations 5.7 to 5.9.

Proposition 18 illustrates that when users decide to enter the SOSN from the origin $o \in \Phi$ in UESN at the boundary node $k \in \Omega$, they will choose the path with the minimum travel cost. This travel cost is equal to or less than the cost of any unused path from $o \in \Phi$ to $k \in \Omega$.

Proposition 18. In MSNE, for $o \in O \cap \Phi$, $k \in \Omega$, $\forall r_1, r_2 \in P_{o,k}$ with $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$, we have $c_{o,k}^{r_1} = c_{o,k}^{r_2} = \pi_{o,k}$. Besides, $\forall r \in P_{o,k}$ with $f_{o,k}^r = 0$, we have $c_{o,k}^r \geq \pi_{o,k}$.

Proof Sketch. See Appendix C.1.

Proposition 19 illustrates that when the system decide to enter the UESN from the origin $o \in \Psi$ in SOSN at the boundary node $k \in \Theta$, they will choose the path with the minimum marginal cost. This marginal cost is equal to or less than the marginal cost of any unused path from $o \in \Psi$ to $k \in \Theta$.

Proposition 19. In MSNE, for $o \in O \cap \Psi$, $k \in \Theta$, $\forall r_1, r_2 \in P_{o,k}$ with $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$, we have $h_{o,k}^{r_1} = h_{o,k}^{r_2} = \eta_{o,k}$. Besides, $\forall r \in P_{o,k}$ with $f_{o,k}^r = 0$, we have $h_{o,k}^r \geq \eta_{o,k}$.

Proof Sketch. See Appendix C.2.

Proposition 20 states that when the users with destination $s \in S \cap \Phi$ enter SOSN at the boundary node $k \in \Omega$, the system would follow the SO principle and find the path with minimum marginal cost. Besides, among all paths with minimum marginal cost, the system would choose the path with minimum travel cost among them. In other words, only when the system promise to choose the path with the minimum travel cost among all paths from k to s that have the smallest marginal cost, then the users with destination s would consider entering SOSN at the boundary node $k \in \Omega$. It can be considered a necessary condition for the user to be willing to enter SOSN at boundary node $k \in \Omega$.

Proposition 20. In MSNE, for $s \in S \cap \Phi$, $k \in \Omega$, $\forall r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, we have $h_{k,s}^{r_1} = h_{k,s}^{r_2} = \eta_{k,s}$ and $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$. Besides, $\forall r \in P_{k,s}$ with $f_{k,s}^r = 0$, we have $h_{k,s}^r \geq \eta_{k,s}$.

Proof Sketch. See Appendix C.3.

Proposition 21 illustrates that when AVs enter the UESN from the SOSN at the boundary node $k \in \Theta$, they will choose the path with the minimum travel cost. This travel cost is equal to or less than the cost of any unused path from $k \in \Omega$ to destination $s \in S \in S \cap \Phi$.

Proposition 21. In MSNE, for $s \in S$, $k \in \Theta$, $\forall r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, we have $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \pi_{k,s}$. Besides, $\forall r \in P_{k,s}$ with $f_{k,s}^r = 0$, we have $c_{k,s}^r \geq \pi_{o,k}$.

Proof Sketch. Obviously obtained by Equations 5.7 to 5.9.

Next, to formulate the MSNE problem into an equivalent variational inequality (VI) formulation, we first define the feasible set $\Lambda = (\mathbf{f}, \mathbf{x}, \mathbf{c}, \mathbf{h})$. For $\forall (\mathbf{f}, \mathbf{x}, \mathbf{c}, \mathbf{h}) \in \Lambda$, it should satisfy the following conditions:

$$(5.2), (5.4) - (5.6), (5.8), (5.10) - (5.12), (5.14), (5.16) - (5.30), (5.24), (5.26) - (5.33) \quad (5.34)$$

Proposition 22 states that the MSNE problem formulated as Equations 5.1-5.33 is equivalent to finding the $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ solving the variational inequality in Equation 5.35.

Proposition 22 (VI formulation). $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ solves the variational inequality (5.35) if and only if $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*)$ satisfies the MSNE (5.1)-(5.33).

$$\begin{aligned} & \sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} (f_{o,s}^{r,*} - f_{o,s}^r) + \sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} (f_{o,s}^{r,*} - f_{o,s}^r) \\ & + \sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} (f_{k,s}^{r,*} - f_{k,s}^r) + \sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} (f_{k,s}^{r,*} - f_{k,s}^r) \leq 0 \quad \forall (\mathbf{f}, \mathbf{x}, \mathbf{c}, \mathbf{h}) \in \Lambda \end{aligned} \quad (5.35)$$

Proof Sketch. See Appendix C.4.

Proposition 23 illustrates the existence of the solution of the proposed MSNE problem.

Proposition 23 (Existence of solution). The variational inequality (5.35) has least one feasible solution.

Proof Sketch. See Appendix C.5.

5.3.3 Critical size of SOSN

In Section 5.3.2, the size and location of UESN and SOSN are assumed as fixed and we obtain the properties of MSNE under the fixed UESN and SOSN. However, as illustrated in Figure 5.2, the total travel cost (TTC) of the entire network is directly influenced by the size of the SOSN. Typically, when the size of the SOSN increases, the TTC tends to decrease correspondingly. This relationship suggests that expanding the SOSN helps to optimize the network, lowering overall costs. On the other hand, when the size of the SOSN is zero, the network operates purely under the UE principle, resulting in higher TTC due to the lack of optimization from the SOSN. In a contrasting scenario, when the SOSN extends to cover the entire network, the system reaches SO, where the TTC is minimized, representing the ideal situation in terms of cost-efficiency. This observation motivates us to further explore how dynamically adjusting the size and location of the SOSN influences both the MSNE and the TTC, as such changes could potentially lead to more efficient network configurations. The next steps will involve examining the impact of these dynamic adjustments, considering how variations in SOSN size and placement affect the equilibrium strategies and overall network performance.

We first introduce the binary variables v_i to represent whether the node i is in the SOSN or not for $i \in \bar{N}$, where $\bar{N}$ is the set of all nodes in the original network, excluding boundary nodes. If $v_i = 1$, node i is supposed to be in the SOSN. Otherwise, $v_i = 0$ indicates that node i is in the UESN. Then we add the node in the middle of each link in the original network, denoted as the set Q . However, for each node $k \in Q$, k may not belong to Ω or Θ , depending on the values of v_i and v_j , where node i and j are the upstream and downstream nodes connected with k , respectively.

The location and size of SOSN is related with the vector $\mathbf{v} = [v_i]_{i \in \bar{N}}$. When $\mathbf{v}$ is fixed, the set of Ω and Θ are also determined, allowing us to derive the TTC under the MSNE, as shown in Equations 5.1 to 5.33, denoted as $TTC(\mathbf{x})$.

Definition 5.3.1. Suppose TTC^* is the minimum total travel cost (TTC) and $\Gamma = \{\mathbf{v} \mid TTC(\mathbf{v}) = TTC^*\}$, then $|\mathbf{v}^*|$ is defined as the critical size of SOSN (CS-SOSN) if $\mathbf{v}^*$ satisfies the following two conditions:

- System Optimal: $TTC(\mathbf{v}^*) = TTC^*$. It implies that the system optimal could also be achieved under $\mathbf{v}^*$, representing that the TTC is minimized.

- Minimum Size: $|\mathbf{v}^*| \leq |\mathbf{v}|, \forall \mathbf{v} \in \Gamma$. It indicates that the minimized size of SOSN to achieve system optimal.

Although we have proposed the definition of CS-SOSN in Definition 5.3.1, due to the non-linearity of mixed equilibrium problem formulated in Equations (5.1)–(5.33), it is still difficult to obtain the SO, not to mention the minimize size of SOSN to achieve SO. Therefore, next, we will demonstrate how to obtain CS-SOSN.

Equation 5.36 ensures that each node in the SOSN is connected to at least one other node within the SOSN, thereby preventing the occurrence of an isolated node within the SOSN.

$$v_i \leq \sum_{l \in \tilde{N}} \sum_{(l,i) \in L} v_l + \sum_{j \in \tilde{N}} \sum_{(i,j) \in L} v_l \quad (5.36)$$

Equations 5.37 to 5.41 consider the scenario where the origin is within the UESN, under the condition that the UESN and SOSN are uncertain. When $v_o = 0$, it represents that the origin o is in UESN, then the Equations 5.37 to 5.39 and 5.68 are equivalent with the Equations 5.1 to 5.3, where M is large positive number. Conversely, when $v_o = 1$, variables $f_{o,s}^r, c_{o,s}^r$ and $\pi_{o,s}$ are actually not constrained by Equations 5.37 to 5.39.

$$f_{o,s}^r(c_{o,s}^r - \pi_{o,s}) \leq M * v_o \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.37)$$

$$f_{o,s}^r(c_{o,s}^r - \pi_{o,s}) \geq -M * v_o \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.38)$$

$$c_{o,s}^r - \pi_{o,s} \geq -M * v_o \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.39)$$

$$c_{o,s}^r = \sum_{i \in \Phi \cup \Theta} \sum_{j \in N} \sum_{(i,j)} c_{i,j} \delta_{i,j}^r + \sum_{k \in \Omega} \sum_{l \in \Theta} c_{k,l} \delta_{k,l}^r \quad \forall o \in O \cap \Phi, s \in S \cap \Phi \quad (5.40)$$

$$c_{o,s}^r = \sum_{i \in \Phi \cup \Theta} \sum_{j \in N} \sum_{(i,j)} c_{i,j} \delta_{i,j}^r + \sum_{k \in \Omega} c_{k,s} \delta_{k,s}^r \quad \forall o \in O \cap \Phi, s \in S \cap \Psi \quad (5.41)$$

Equations 5.42 to 5.46 consider the scenario where the origin is within the SOSN, under the condition that the UESN and SOSN are uncertain. When $v_o = 1$, it represents that the origin o is in SOSN, then the Equations 5.42 to 5.44 and 5.68 are equivalent with the Equations 5.7 to 5.9. Conversely, when $v_o = 0$, variables $f_{o,s}^r, h_{o,s}^r$ and $\eta_{o,s}$ are actually not constrained by Equations 5.42 to 5.44.

$$f_{o,s}^r(h_{o,s}^r - \eta_{o,s}) \leq M * (1 - v_o) \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.42)$$

$$f_{o,s}^r(h_{o,s}^r - \eta_{o,s}) \geq -M * (1 - v_o) \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.43)$$

$$h_{o,s}^r - \eta_{o,s} \geq -M * (1 - v_o) \quad \forall o \in O, s \in S, r \in P_{o,s} \quad (5.44)$$

$$h_{o,s}^r = \sum_{i \in N} \sum_{j \in N} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall o \in O \cap \Psi, s \in S \cap \Psi, r \in P_{o,s} \quad (5.45)$$

$$h_{o,s}^r = \sum_{i \in \Phi} \sum_{j \in \Phi \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r + \sum_{l \in \Theta} h_{l,s} \delta_{l,s}^r \quad \forall o \in O \cap \Psi, s \in S \cap \Phi \quad (5.46)$$

Equations 5.47 to 5.55 consider the scenario where the boundary node k is in the set Ω , under the condition that the UESN and SOSN are uncertain. When $v_i = 0$ and $v_j = 1$, it represents that the boundary node k is in Ω whose upstream and downstream node is in UESN and SOSN, respectively. Then the Equations 5.47 to 5.49 and 5.68 are equivalent with the Equations 5.13 to 5.15. Conversely, for others cases of v_i and v_j , variables $f_{k,s}^r, h_{k,s}^r$ and $\eta_{k,s}$ are actually not constrained by Equations 5.47 to 5.49.

$$f_{k,s}^r(h_{k,s}^r - \eta_{k,s}) \leq M * (1 + v_i - v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.47)$$

$$f_{k,s}^r(h_{k,s}^r - \eta_{k,s}) \geq -M * (1 + v_i - v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.48)$$

$$h_{k,s}^r - \eta_{k,s} \geq -M * (1 + v_i - v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.49)$$

$$h_{k,s}^r = \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, s \in S \cap \Psi, r \in P_{k,s} \quad (5.50)$$

$$h_{k,s}^r = \sum_{i \in \Phi} \sum_{j \in \Phi \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r + \sum_{l \in \Theta} h_{l,s} \delta_{l,s}^r \quad \forall k \in O \cap \Omega, s \in S \cap \Phi \quad (5.51)$$

$$h_{k,l}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, l \in \Phi, r \in P_{k,l} \quad (5.52)$$

$$c_{k,l}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r \quad \forall k \in \Omega, l \in \Phi, r \in P_{k,l} \quad (5.53)$$

$$h_{k,l} = \min\{h_{k,l}^r \mid \forall r \in P_{k,l}\} \quad \forall k \in \Omega, l \in \Phi \quad (5.54)$$

$$c_{k,l} = \min\{c_{k,l}^r \mid h_{k,l}^r = h_{k,l}, \forall r \in P_{k,l}\} \quad \forall k \in \Omega, l \in \Phi \quad (5.55)$$

Equations 5.56 to 5.62 consider the scenario where the boundary node k is in the set Θ , under the condition that the UESN and SOSN are uncertain. When $v_i = 1$ and $v_j = 0$, it represents that

the boundary node k is in Ω whose upstream and downstream nodes are in SOSN and UESN, respectively. Then the Equations 5.56 to 5.58 and 5.68 are equivalent with the Equations 5.23 to 5.25. Conversely, for others cases of v_i and v_j , variables $f_{k,s}^r, c_{k,s}^r$ and $\pi_{k,s}$ are actually not constrained by Equations 5.56 to 5.58.

$$f_{k,s}^r(c_{k,s}^r - \pi_{k,s}) \leq M * (1 - v_i + v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.56)$$

$$f_{k,s}^r(c_{k,s}^r - \pi_{k,s}) \geq -M * (1 - v_i + v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.57)$$

$$c_{k,s}^r - \pi_{k,s} \geq -M * (1 - v_i + v_j) \quad \forall k \in Q, s \in S, i, j \in \bar{N}, (i, k), (k, j) \in L \quad (5.58)$$

$$c_{k,s}^r = \sum_{i \in \Theta \cup \Phi} \sum_{j \in \Phi} \sum_{(i,j) \in L} c_{i,j} \delta_{i,j}^r \quad \forall k \in Q, s \in S, r \in P_{k,s} \quad (5.59)$$

$$h_{k,s}^r = \sum_{i \in \Omega} \sum_{j \in \Omega \cup \Theta} \sum_{(i,j) \in L} h_{i,j} \delta_{i,j}^r \quad \forall k \in \Theta, s \in S \cap \Phi, r \in P_{k,s} \quad (5.60)$$

$$c_{k,s} = \min\{c_{k,s}^r \mid \forall r \in P_{k,s}\} \quad \forall k \in \Theta, s \in S \cap \Phi \quad (5.61)$$

$$h_{k,s} = \min\{h_{k,s}^r \mid c_{k,s}^r = c_{k,s}, \forall r \in P_{k,s}\} \quad \forall k \in \Theta, s \in S \cap \Phi \quad (5.62)$$

There also exist other constraints for the critical size of SOSN (CS-SOSN) problem in Equations 5.63 to 5.68.

$$\sum_{r \in P_{o,s}} f_{o,s}^r = D_{o,s} \quad \forall o \in O, s \in S \quad (5.63)$$

$$\sum_{r \in P_{k,s}} f_{k,s}^r = \sum_{i \in \Phi} \delta_{i,k}^L x_{i,k} \quad \forall k \in Q, s \in S, r \in P_{k,s} \quad (5.64)$$

$$x_{i,j} = \sum_{o \in O} \sum_{s \in S} \sum_{r \in P_{o,s}} f_{o,s}^r \delta_{i,j}^r \quad \forall (i, j) \in L \quad (5.65)$$

$$c_{i,j} = t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}}{H_{i,j}})^{\beta_{i,j}}) \quad \forall (i, j) \in L \quad (5.66)$$

$$h_{i,j} = c_{i,j} + x_{i,j} c'_{i,j} \quad \forall (i, j) \in L \quad (5.67)$$

$$f_{o,s}^r, f_{k,s}^r \geq 0 \quad \forall o \in O, k \in Q, s \in S \quad (5.68)$$

To conclude, we propose the formulation of CS-SOSN problem in Equation 5.69, where TTC^* represents the minimum total travel cost (TTC) in the SO state. Equations 5.69 indicates that CS-SOSN represents the minimum size of SOSN to achieve the system optimal state.

$$\begin{aligned}
& \min_{\mathbf{f}, \mathbf{x}, \mathbf{v}} \sum_{i \in \bar{N}} v_i \\
& s.t. \text{ (5.36) - (5.68)} \\
& \sum_{(i,j) \in L} c_{i,j} x_{i,j} = TTC^*
\end{aligned} \tag{5.69}$$

Next, we construct a way to obtain the CS-SOSN by the following two steps:

Step 1: Solve the following system optimal (SO) problem, where all vehicles follow the SO principle in the whole network (Wardrop, 1952). We obtain the minimum total travel cost TTC^* .

$$\begin{aligned}
& \min_{\mathbf{f}, \mathbf{x}} TTC = \sum_{(i,j) \in L} c_{i,j} x_{i,j} \\
& s.t. \sum_{r \in P_{o,s}} f_{o,s}^r = D_{o,s} \quad \forall o \in O, s \in S \\
& x_{i,j} = \sum_{o \in O} \sum_{s \in S} \sum_{r \in P_{o,s}} f_{o,s}^r \delta_{i,j}^r \quad \forall (i,j) \in L \\
& c_{i,j} = t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}}{H_{i,j}})^{\beta_{i,j}}) \quad \forall (i,j) \in L \\
& f_{o,s} \geq 0
\end{aligned} \tag{5.70}$$

Step 2: Solve the proposed optimization in Equation 5.69 to obtain the CS-SOSN as $\sum_{i \in \bar{N}} v_i^*$, representing the minimize size of SOSN to achieve the SO.

5.4 Solution Method

In this section, we introduce the algorithms tailored to address the MSNE and CS-SOSN problems proposed in this study. In Section 5.4.1, we detail the steps of the SNA-MSA algorithm, which is used to solve the MSNE under the given SOSN and UESN. In Section 5.4.2, we propose a novel algorithm that extends the SNA-MSA algorithm and incorporates the AVNS algorithm to identify the CS-SOSN in large-scale networks.

5.4.1 Sub-network-augmented method of successive averages (SNA-MSA) for MSNE

We propose a sub-network-augmented method of successive averages (SNA-MSA) algorithm to obtain the MSNE under the given existence of SOSN and UESN, as summarized in Algorithm 4. Different from the traditional method of successive averages (MSA) algorithm (Sheffi, 1985), the SNA-MSA would begin with an empty network that only contains the nodes of the original network, without any links. Then, we update and select paths for the sub-network sequentially according to different principles in the proposed model, and simplify the corresponding paths into links to add to the empty network. After completing the traversal of all sub-networks, we find the shortest path with minimum travel cost or minimum marginal cost for the origin of the revised empty network based on its location and update the link flow of the original network accordingly. The output of the SNA-MSA is the link flow under the current SOSN and UESN, which helps us to find the MSNE and analyze it.

5.4.2 Adaptive variable neighborhoods search algorithm (AVNS) algorithm for CS-SOSN

In the previous section, we used the proposed SNA-MSA algorithm to solve the MSNE problem under the given UESN and SOSN conditions. In this section, we extend the SNA-MSA algorithm to address the CS-SOSN problem in large-scale networks. The CS-SOSN problem is a large and non-convex optimization whose objective is expensive to evaluate. AVNS is particularly well suited to this setting: it systematically explores a hierarchy of neighborhoods to escape local optima, adapts neighborhood sizes and shaking intensity based on search feedback to balance exploration and intensification, and easily incorporates problem-specific move operators (e.g., add/remove/swap links or nodes). The proposed AVNS algorithm consists of outer and inner loops. The number of outer loops is equal to the number of nodes in the input network, and the algorithm begins with the SOSN initialized to the total number of nodes minus one.

In each inner loop, the parallel method of successive average (PMSA) searches for the shortest paths between OD pairs to calculate the equilibrium flows of the traffic network under SO and UE conditions. Subsequently, these two equilibrium flows are used as inputs to the AVNS algorithm to identify the best SOSN for the current SOSN size, where the improvement of solu-

Algorithm 4 Sub-network-augmented method of successive averages (SNA-MSA)

1. **Input:** Traffic Network $G(N, L)$ with determined SOSN and UESN; Step size α_n ; OD demand $D_{o,s}$.
2. **Initialization:** Find a feasible link flow $\mathbf{x}^{(0)}$ in an empty network.
3. **Repeat** until convergence:

(a) Generate an empty network $\tilde{G}(N, \tilde{L})$ with no links, where $\tilde{L} = \emptyset$.

(b) For boundary node $k \in \Omega$ and destination $s \in S \cap \Psi$, choose all paths r as follows. Then we add link $l_{k,s}^r$ for each path r to $\tilde{G}$ as $\tilde{L} = \tilde{L} \cup \{l_{k,s}^r\}$.

$$h_{k,s}^r = \eta_{k,s}, \quad c_{k,s}^r = \min_{r \in P_{k,s}, h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$$

(c) For boundary node $k \in \Theta$ and destination $s \in S \cap \Phi$, choose all paths r as follows. Then we add link $l_{k,s}^r$ for each path r to $\tilde{G}$ as $\tilde{L} = \tilde{L} \cup \{l_{k,s}^r\}$.

$$c_{k,s}^r = \pi_{k,s}, \quad h_{k,s}^r = \min_{r \in P_{k,s}, c_{k,s}^r = \pi_{k,s}} \{h_{k,s}^r\}$$

(d) For boundary nodes $k_1 \in \Omega$ and $k_2 \in \Theta$, choose all paths r as follows. Then we add link l_{k_1,k_2}^r for each path r to $\tilde{G}$ as $\tilde{L} = \tilde{L} \cup \{l_{k_1,k_2}^r\}$.

$$h_{k_1,k_2}^r = \eta_{k_1,k_2}, \quad c_{k_1,k_2}^r = \min_{r \in P_{k_1,k_2}, h_{k_1,k_2}^r = \eta_{k_1,k_2}} \{c_{k_1,k_2}^r\}$$

(e) For $(i, j) \in L$ where $i, j \in \Psi \cup \Phi$, add link (i, j) with travel cost $c_{i,j}$ and marginal cost $h_{i,j}$ to $\tilde{G}$ as:

$$\tilde{L} = \tilde{L} \cup \{(i, j)\}$$

(f) In the revised network $\tilde{G}$, find the path flow as follows and obtain the link flows $\mathbf{z}^{(n)}$.

- Find the shortest path with minimum cost from origin $o \in O \cap \Phi$ to destination $s \in S$.
- Find the shortest path with minimum marginal cost from the origin $o \in O \cap \Psi$ to the destination $s \in S$.

(g) Map the link flows $\mathbf{z}^{(n)}$ from $\tilde{G}$ to the original network G to obtain $\mathbf{y}^{(n)}$.

(h) Update the link flow in network G :

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \alpha_n(\mathbf{y}^{(n)} - \mathbf{x}^{(n)})$$

(i) Update the link travel cost and marginal cost in G using $\mathbf{x}^{(n+1)}$.

(j) Increment $n = n + 1$.

4. **End Repeat**

5. **Output:** Link flow $\mathbf{x}$ under MSNE given the UESN and SOSN.

tions during the local search is evaluated using the SNA-MSA algorithm. The AVNS algorithm gradually reduces the SOSN size during the external iterations and stops when it can no longer maintain the SO, outputting the final CS-SOSN. Figure 5.4 illustrates the workflow of the proposed AVNS algorithm.

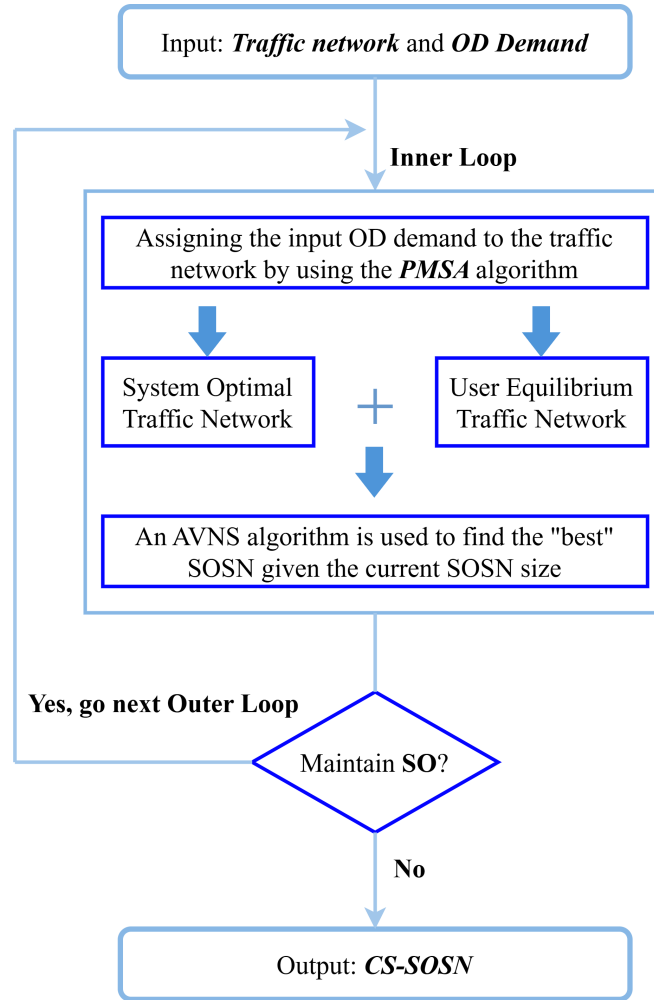


Figure 5.4: Workflow of AVNS algorithm.

The PMSA algorithm enhances the traditional MSA by integrating parallelized shortest path calculations for all OD pairs, significantly reducing computational time on large-scale networks. By leveraging parallelized Dijkstra's algorithm, PMSA efficiently computes shortest paths and updates link flows iteratively, improving scalability and efficiency while ensuring convergence. This makes PMSA particularly effective for analyzing large-scale traffic networks.

After obtaining the equilibrium UE and SO traffic network flows using the PMSA algorithm, we apply the AVNS algorithm to identify the CS-SOSN. The AVNS algorithm consists of the

AVNS algorithm and the SNA-MSA algorithm. The AVNS algorithm performs neighborhood search operations on the solution, while the SNA-MSA algorithm computes the traffic flow under MSNE for the new solution obtained through the neighborhood search. The AVNS algorithm begins by setting the maximum outer and inner loop iteration numbers as i_{outer} and i_{inner} , respectively. Each neighborhood N_i for $i = 1, \dots, 4$ is initialized with the same weight $t_i = 1$. An initial solution x is generated by using the SO traffic network, and this network is set as the initial solution. This solution is then set as the current best solution x^{best} . At the beginning of each outer iteration, the current solution x undergoes random perturbation using a shaking operator to escape local optima. Within the inner iterations, the perturbed solution x' is subjected to local search operations within neighborhoods N_i , selected based on their probabilities p_i . These probabilities are adaptively updated using past performance, calculated as $p_i = t_i / \sum t_i$ at the start of each outer iteration. The improvement of solutions by the local search is evaluated using the SNA-MSA algorithm.

If the local search improves the solution, the updated solution is retained, and the corresponding neighborhood weight t_i is increased to reflect its success. Otherwise, the search continues within the current neighborhood until i_{inner} is reached. In the context of CS-SOSN, AVNS leverages its flexibility to navigate the solution space, searching for an optimal or near-optimal candidate sub-network. Algorithm 5 explains the details of the AVNS algorithm proposed in this work. 4 tailored neighborhoods are introduced to improve the solution quality.

Neighborhoods

- **1-1 Swap:** Swap 1 randomly selected link from the CS-SOSN that is directly connected to the UESN with 1 randomly selected link from the UESN that is directly connected to the CS-SOSN.
- **2-2 Swap:** Swap 2 randomly selected links from the CS-SOSN that are directly connected to the UESN with 2 randomly selected links from the UESN that are directly connected to the CS-SOSN.
- **Random swap:** Swap random numbers randomly selected link from the CS-SOSN that is directly connected to the UESN with the same numbers randomly selected link from the UESN that is directly connected to the CS-SOSN.

- **Best swap:** Swap the lowest travel cost link from the CS-SOSN that is directly connected to the UESN with 1 randomly selected link from the UESN that is directly connected to the CS-SOSN.

Algorithm 5 Adaptive Variable Neighborhoods Search algorithm (AVNS)

```

1: Input: System Optimal Traffic Network and User Equilibrium Traffic Network
2:   Initialize weight  $t_i$  of neighborhoods  $N_i$  for  $i = 1, \dots, 4$ ,  $i_{outer}$  and  $i_{inner}$ 
3:   Generate initial solution  $x$  and set it as the current best solution  $x^{best}$ 
4:   Repeat until the  $i_{outer}$  is reached
5:     Generate a solution  $x'$  from current solution  $x$  by random choosing a shaking
6:     operator
7:     Repeat until the  $i_{inner}$  is reached
8:       Generate a new solution  $x''$  from  $x'$  by neighborhoods based on probability
9:        $p_i$ 
10:      If  $x''$  is better than  $x'$ 
11:         $x' = x''$ , update  $t_i = t_i + t_i/i_{outer}$ 
12:        If  $x''$  is better than  $x^{best}$ 
13:           $x^{best} = x''$ 
14:        End If
15:      Select next neighborhood based on probability  $p_i$ 
16:    Else
17:      Repeat in current neighborhood and update  $t_i = t_i - t_i/i_{outer}$ 
18:    End If
19:    Update selection probability  $p_i$  of each neighborhoods by  $p_i = t_i / \sum t_i$ 
20:  End Repeat
21:  If  $x'$  is better than  $x$ 
22:     $x = x'$ 
23:  End If
24: Output: Best solution

```

5.5 Numerical Results

The proposed model and propositions are validated three networks in this section, containing a small network, a large network and a real Hong Kong network. We explore and analyze the MSNE of the proposed model in a small network in Section 5.5.1. Then we solve the CS-SOSN of the large-scale network in Section 5.5.2. Lastly, we validate the effectiveness of our proposed model in real traffic conditions of Hong Kong network in Section 5.5.3.

5.5.1 Small network

The small network consists of 10 nodes and 15 links, as referred to 5.5. The origin node is 1 and the destination node is 10. The demand from nodes 1 to 10 is 30. The SOSN consists of three nodes: node 3, 4, and 5. The UESN consists of four nodes: node 1, 8, 9, and 10. The boundary nodes include that node 3 $\in \Omega$ and nodes 4, 5 $\in \Theta$. The parameters of the small network are illustrated in Figure 5.2, where the link performance function is assumed to be $t_{i,j}(x_{i,j}) = a_{i,j} + b_{i,j}x_{i,j}$.

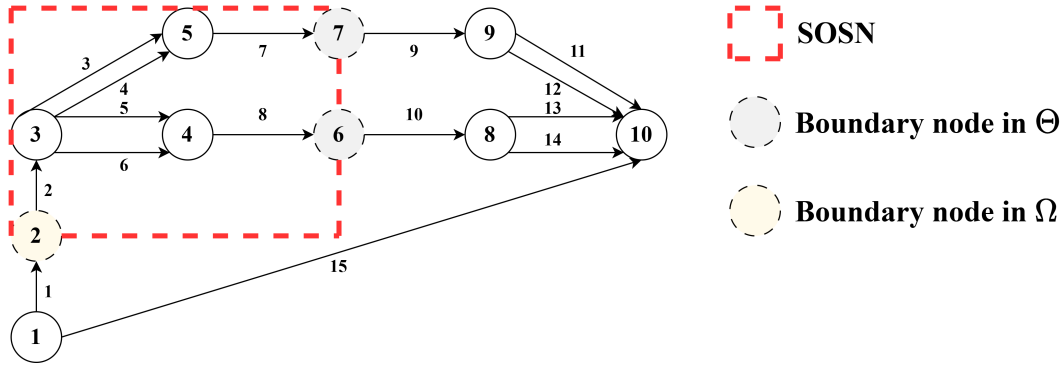


Figure 5.5: Illustration of small network.

Link	a_i (min)	b_i	Link	a_i (min)	b_i	Link	a_i (min)	b_i
1	0.50	0.75	6	10.00	9.50	11	6.00	4.00
2	0.50	0.75	7	1.00	0.30	12	8.50	3.00
3	4.50	3.00	8	1.00	0.84	13	10.00	9.00
4	3.50	4.00	9	1.00	0.30	14	10.00	10.50
5	2.00	8.00	10	1.00	0.84	15	27.30	14.50

Table 5.2: Small network parameters.

Given the small network parameters in Table 5.2 and the existence of SOSN in Figure 5.5, we obtain the MSNE of the proposed model by conducting the Algorithm 4. The link flow, perceived path travel cost, of and perceived marginal cost are presented in Tables 5.3 and 5.4, respectively. For the O-D pair (1,10), all three used paths have an equal perceived travel cost as 46.323 minutes, which corresponds to Proposition 16. Then for the paths from the boundary node in Θ to the destination in UESN, all used have equal perceived travel cost as indicated by Proposition 21. For example, all used paths have equal perceived travel cost of 46.323 for the

pair (6,10) and 43.503 for the pair (7,10). Besides, Proposition 20 states that all used paths from the boundary node $k \in \Omega$ to the destination $s \in \Phi$ in UESN have the same minimum perceived marginal cost and perceived travel cost, where perceived travel cost refers to the minimum path travel cost among all paths with marginal cost from k to s . In the small network as shown in Figure 5.5, we have node $2 \in \Omega$ and destination $s \in S \cap \Phi$. Table 5.4 presents that both paths from node 2 to 10 have equal perceived travel cost and marginal cost. Besides, for path $2 \rightarrow 3 \rightarrow 4 \rightarrow 6 \rightarrow 8 \rightarrow 10$ which leaves the SOSN and enter UESN at the boundary node $6 \in \Theta$, the perceived travel cost is the sum of the perceived travel cost from node 2 to 6 as 55.243 minutes and perceived travel cost from node 6 to 10 as 46.323 minutes. It also corresponds to the assumption that AVs perceive their travel cost in SOSN to be the minimum travel cost of used paths from their entrances of SOSN to the exits of SOSN.

Link	Flow	Travel Cost (min)	Marginal Cost (min)	Link	Flow	Travel Cost (min)	Marginal Cost (min)
1	23.622	18.217	35.933	9	17.410	6.223	11.446
2	23.622	18.217	35.933	10	6.212	6.218	11.436
3	9.877	34.131	63.673	11	7.820	37.280	68.560
4	7.533	33.632	63.673	12	9.590	37.280	66.060
5	3.601	30.808	59.616	13	3.345	40.105	70.210
6	2.611	34.805	59.616	14	2.867	40.105	70.210
7	17.410	6.223	11.446	15	6.378	119.781	212.262
8	6.212	6.218	11.436				

Table 5.3: Link flow under MSNE in small network.

Pair	Path	Perceived Travel Cost (min)	Perceived Marginal Cost (min)
6 – 10	6 → 8 → 10 (link: 10 → 13)	46.323	81.646
	6 → 8 → 10 (link: 10 → 14)	46.323	81.646
7 – 10	7 → 9 → 10 (link: 9 → 11)	43.503	80.006
	7 → 9 → 10 (link: 9 → 12)	43.503	77.506
2 – 6	2 → 3 → 4 → 6 (link: 2 → 5 → 8)	55.243	106.985
	2 → 3 → 4 → 6 (link: 2 → 6 → 8)	59.240	106.985
2 – 7	2 → 3 → 5 → 7 (link: 2 → 3 → 7)	58.571	111.052
	2 → 3 → 5 → 7 (link: 2 → 4 → 7)	58.072	111.052
2 – 10	2 → 3 → 4 → 6 → 8 → 10	101.566	188.631
	2 → 3 → 5 → 7 → 9 → 10	101.575	188.558
1 – 10	1 → 2 → 3 → 4 → 6 → 8 → 10	119.783	224.564
	1 → 2 → 3 → 5 → 7 → 9 → 10	119.792	224.491
	1 → 10	119.781	212.262

Table 5.4: MSNE of path in small network.

5.5.2 Large network

In this section, we would solve the CS-SOSN in a regular large-scale network, which consists of 72 nodes and 254 links, as referred to Figure 5.6. There are a total of 40 O-D pairs. The detailed O-D pairs and their demand are illustrated in Table 5.5. We would conduct the proposed Algorithm 5 to obtain the CS-SOSN. The output of the Algorithm 5 contains the size of CS-SOSN and the location of CS-SOSN, which is a continuous region.

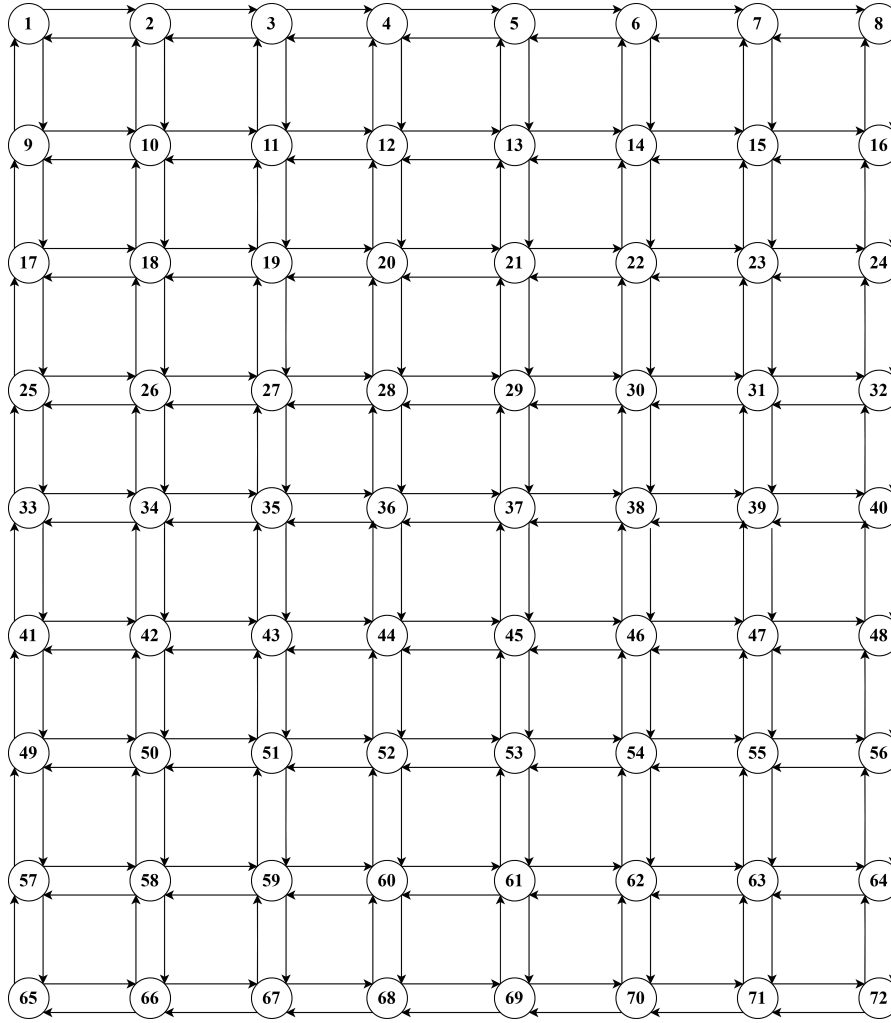


Figure 5.6: Illustration of large network.

Figure 5.7 shows the variation in the system's total travel cost as the size of SOSN increases in a large-scale traffic network. It shows that as the size of the SOSN increases, the system's total travel cost generally decreases. When the size of SOSN reaches 29, the system's total travel cost has been minimized. Therefore, we conclude that in this large-scale traffic network, the size of CS-SOSN is 29. Figure 5.8 reveals the specific location of this CS-SOSN. When the

O-D	Demand	O-D	Demand	O-D	Demand	O-D	Demand
1-29	40	1-35	45	1-47	45	1-52	50
1-72	55	5-29	60	5-35	60	5-47	45
5-52	40	5-72	55	8-29	45	8-35	45
8-47	60	8-52	50	8-72	40	15-29	55
15-35	40	15-47	60	15-52	55	15-72	45
19-29	55	19-35	60	19-47	55	19-52	55
19-72	55	55-29	55	55-35	55	55-47	60
55-52	40	55-72	50	61-29	50	61-35	40
61-47	60	61-52	50	61-72	40	66-29	45
66-35	45	66-47	50	66-52	60	66-72	45

Table 5.5: O-D demand.

system controls all AVs in the area within the red dashed box as SOSN, the system will reach the SO state, meaning the system's total travel cost is minimized. This is also the smallest area of the SOSN required to control the system to achieve the SO state. Figure 5.8 indicates the effectiveness of our proposed CS-SOSN framework. Compared with centrally controlling the whole network with 72 nodes to achieve minimum total travel cost, the system only needs to control an area with 29 nodes to follow the SO principle such that the total travel cost would be minimized. Equivalently, the result of CS-SOSN in a large-scale network indicates that we only need to control the AVs in about 40% of the whole large-scale network to make the whole network to the SO state. Therefore, the proposed framework of the CS-SOSN is an effective method to optimally control AVs in a region on a large-scale network to achieve the SO state.

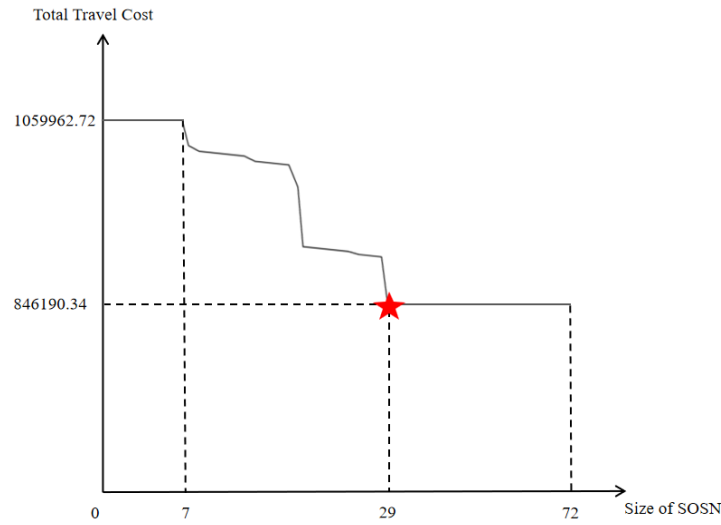


Figure 5.7: The variation of the total travel cost for different size of SOSN in large network.

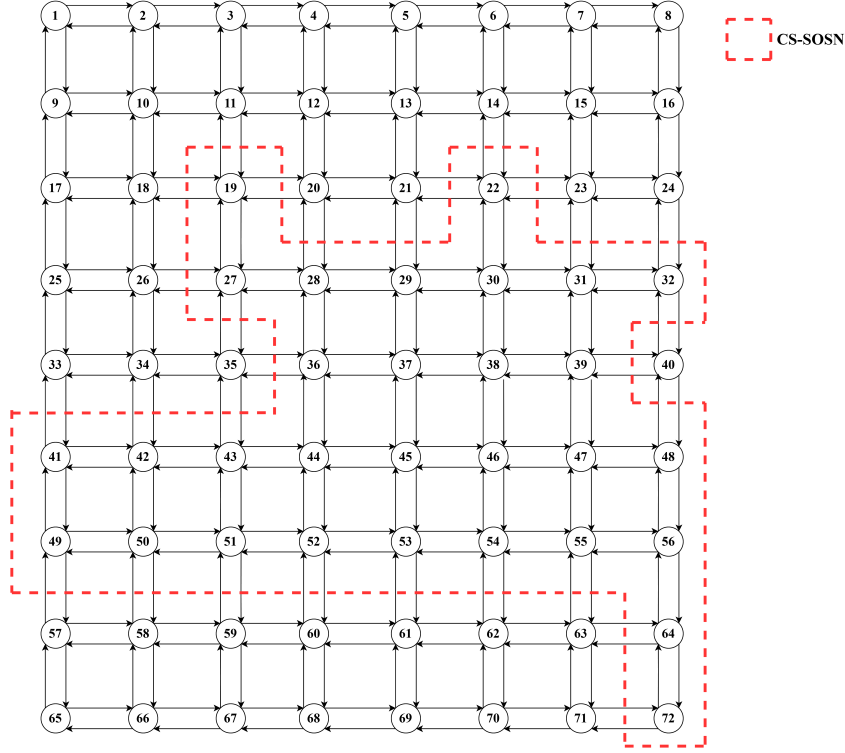


Figure 5.8: The location of CS-SOSN in large network.

5.5.3 Hong Kong network

In this section, we would apply the proposed framework of CS-SOSN to the real Hong Kong traffic network to test its effectiveness in real traffic conditions. The Hong Kong network is comprised of 78 nodes and 262 links, as illustrated in Figure 5.9. Besides, the links in Figure 5.9 are all bidirectional. The Hong Kong network contains 144 O-D pairs and the total demand is estimated to be about 100,800.

Figure 5.10 further demonstrates the relationship between the total travel cost and the size of the SOSN in the real Hong Kong network. As the size of the SOSN increases, the total travel cost decreases. When the SOSN size reaches 26, the total travel cost reaches its minimum value of 70,6741. This suggests that achieving the system's optimal state does not require controlling all regions in the Hong Kong network; instead, controlling a sub-network with just 26 nodes is sufficient for optimization. This result highlights the effectiveness of the proposed CS-SOSN framework in real traffic networks, which provides a valuable insight for government to optimally controlling AVs in real traffic conditions to achieve the SO state.

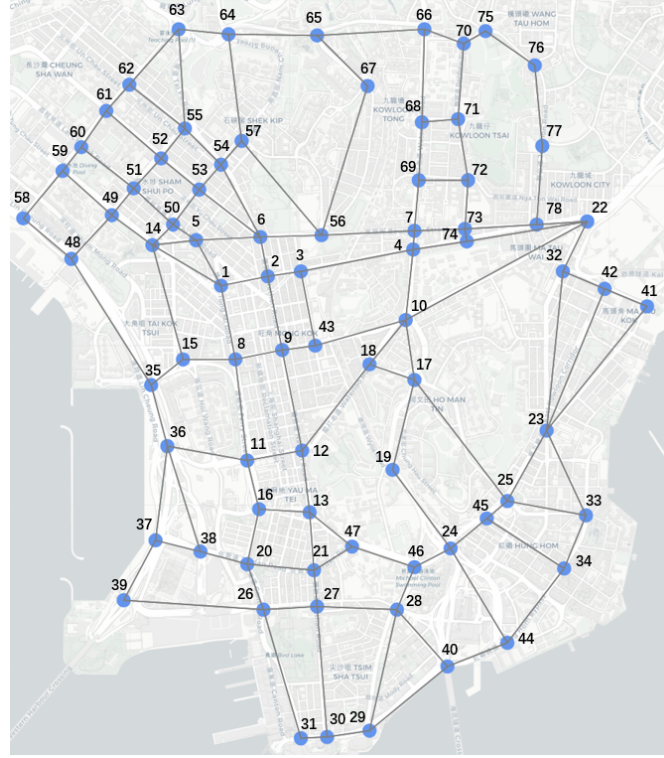


Figure 5.9: The Hong Kong network.

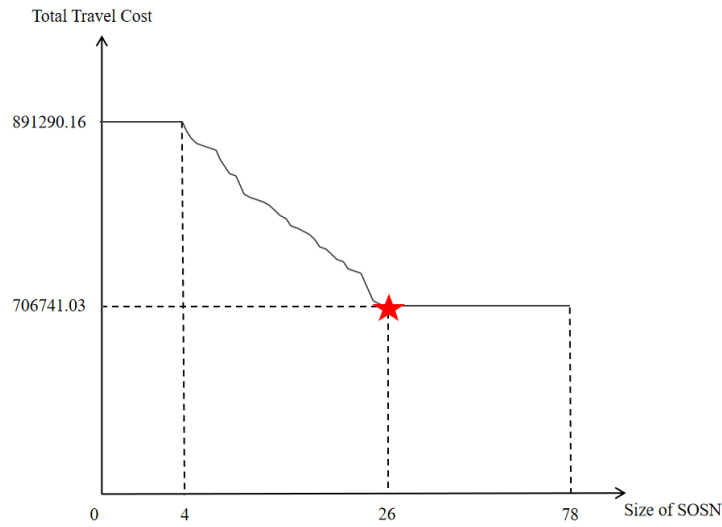


Figure 5.10: The variation of the total travel cost for different size of SOSN in the Hong Kong network.

Figure 5.11 illustrates the specific location of the CS-SOSN in the Hong Kong network. When the system controls the area within the red box as the SOSN, the system reaches the SO state, meaning the total travel cost is minimized. This also represents the smallest SOSN area required to achieve the SO state. The total area of the HK network in Figure 5.11 is about 14.15

square kilometers, while the controlled CS-SOSN area is about 6.23 square kilometers. This means that, for the government's traffic management department, only 44.17% of the Hong Kong network needs to be controlled in actual traffic operations to bring the whole network to a state of SO, without requiring full control of the whole network. This provides an effective method for managing the traffic network and reducing congestion in a pure AV environment, improving overall performance. The CS-SOSN in the Figure 5.11 mainly includes the areas of Tsim Sha Tsui, Mong Kok, and Hung Hom, which are considered the heart of Hong Kong, serving as vibrant hubs that draw both locals and tourists. These areas are bustling with activity, featuring a mix of commercial, cultural, and residential spaces. As a result, they play a critical role in the city's infrastructure. The vast majority of transportation demand in Hong Kong are connected to these districts, as they are central to economic activities, shopping, dining, and entertainment. Reducing congestion in these core regions would not only improve commuter experiences but also facilitate smoother transit across the city, underpinning the vitality and economic dynamism of Hong Kong as a whole.

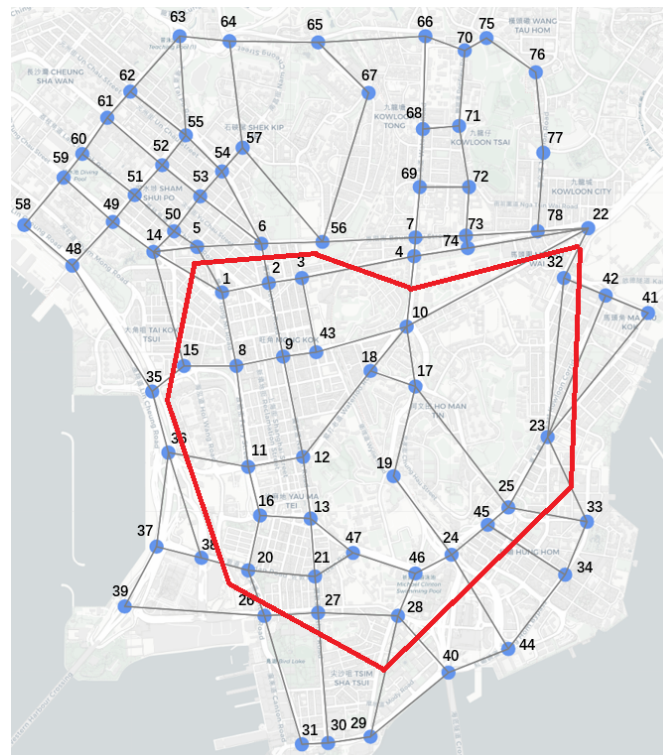


Figure 5.11: The location of CS-SOSN in the Hong Kong network.

Chapter 6

Summary and future directions

This dissertation has explored the roles of AVs in SO for traffic assignment problems and developed several strategies for AVs to achieve SO in traffic assignment problems. The developed strategies can be categorized into three aspects, (1) controlling the headway of AVs to achieve SO-DTA, (2) designing the publicly-owned centralized platform by providing spatial-temporal subsidies for travelers taking SAVs to achieve SO-DTA, and (3) analyzing the mixed sub-network equilibrium for AVs, where AVs follows different routing principles in UESN and SOSN.

6.1 Research findings and contributions

The primary findings and contributions of this dissertation can be categorized into three main parts:

- **Maximin headway control.**
 - We develop the headway-dependent fundamental diagram (HFD) and headway-dependent double queue model (HDQ) to capture the effects of dynamic AV headway control. Particularly, the HDQ is proved to be a generalized version of the double queue (DQ) model. Besides, it is rigorously proved that the minimum headway could achieve SO-DTA.
 - Considering the non-uniqueness of SO-DTA and safety concerns, the maximin headway of AVs is innovatively defined as the “largest” headway setting that still achieves

SO-DTA. We further propose an efficient solution algorithm to derive the maximin headway.

- **Publicly-owned centralized platform.**

- We propose the publicly-owned centralized platform (POCP) as a novel concept for the regulation of government to achieve the SO-DTA. Operated by the government, POCP receives trip order requests from travelers and sends the orders to the TNCs. The government could integrate management strategies into the POCP to improve network-wide traffic efficiency, such as offering a spatial-temporal discount.
- We formulate POCP as a non-linear bi-level problem, which could be equivalently decomposed into two sub-problems to solve minimum social cost and minimum subsidy amount, by incorporating a transmission of demand and spatial-temporal discount variables, respectively. The solution process of POCP becomes tractable for large-scale networks because the first sub-problem can be solved by a series of LPs and the second sub-problem is also a LP.

- **Mixed sub-network equilibrium.**

- We propose a novel concept of mixed sub-network equilibrium (MSNE) of automated vehicles (AVs) in general traffic networks. By segmenting the traffic network into two distinct sub-networks, namely the user equilibrium sub-network (UESN) and the system optimal sub-network (SOSN), we offer an effective approach to traffic management that combines both decentralized user control and centralized system optimization, where AVs are controlled by the users to follow the UE principle in UESN and controlled by the system to follow the SO principle in SOSN.
- We analyze the properties of MSNE given the existence of UESN and SOSN and prove that it is equivalent to a variational inequality (VI) problem. Then we propose a novel concept of the critical size of the system optimal sub-network (CS-SOSN), which provides a practical guide for policymakers and traffic managers aiming to maximize the system performance with minimum regulatory intervention.

6.2 Future direction

We first acknowledge that this study has a narrow scope because it is based on an ideal scenario and ignores some realistic factors, such as user acceptance of AVs and the cost of AVs themselves. Besides, this study relies on some assumptions. For example, AVs will be completely controlled, whether by the user or the system, and they must follow the assigned routes and headway. Besides, it is assumed that traffic safety and headway have a monotonically increasing relationship. It is believed that traffic safety and headway have a monotonically increasing relationship. The future direction mainly includes five aspects to address these problems: mixed traffic conditions, heterogeneous travel behaviors, stochastic traffic demand, interaction of transportation facilities, and realistic traffic environment.

Mixed traffic conditions: While the current study focuses on the assumption of an environment with fully AVs, future research can expand by exploring mixed traffic conditions, where both AVs and human-driven vehicles (HVs) coexist. In such scenarios, the interactions between AVs and HVs present unique challenges that require the development of new routing strategies for AVs. These strategies should take into account the behavior and unpredictability of HVs, ensuring safe and efficient integration of both vehicle types on the road. Additionally, future work could investigate the impact of mixed traffic on traffic flow, safety, and vehicle-to-vehicle communication, with an emphasis on optimizing the performance of AVs in environments with varying levels of human intervention. Besides, we would also incorporate the cost of AVs, which would influence the travelers' choices from HVs to AVs.

Heterogeneous travel behaviors: The current research assumes that users of AVs fully comply with the routing strategies in place, but this assumption may not hold true in real-world scenarios. Future work can explore the impact of heterogeneous travel behaviors, where AV users may not always follow the assigned routes. By considering the variability in human behavior and decision-making, research could focus on developing more adaptive control systems for AVs that can effectively manage and respond to non-compliant user actions.

Stochastic traffic demand: The current study does not account for the variability and uncertainty in traffic demand, which is often stochastic in nature. Future research could explore the

integration of stochastic traffic demand models to better reflect the real-world conditions where traffic volume fluctuates due to various factors. By incorporating probabilistic demand models, researchers can develop more robust traffic management strategies that adapt to dynamic demand patterns.

Interaction of transportation facilities: As the deployment of AVs becomes more widespread, addressing the operational challenges related to charging infrastructure and parking strategies will be crucial for optimizing the performance and efficiency of AVs. One promising area for future research is the integration of charging station networks into the routing models of AVs. This involves not only determining the optimal paths for AVs to follow but also ensuring that charging stations are effectively incorporated into these paths to minimize energy consumption. Furthermore, parking strategies will play an important role in enhancing the operational efficiency of AVs. Future research will investigate optimal parking locations and strategies for AVs, aiming to minimize dead-head trips and ensure that vehicles are parked in areas that maximize accessibility, safety, and readiness for the next trip.

Realistic traffic environment: In reality, traffic safety would also be dependent on other factors, such as relative speed, lane-change rates, and road conditions. Future work would extend the proposed maximin headway framework into a more complicated environment, incorporating the above factors to capture the actual conditions of traffic safety. Besides, the weather condition is also an important realistic factor. The headway would have different effects on the traffic states under different weather conditions. Future research will improve our proposed model to make it remain effective in realistic traffic environments.

Bibliography

- Ayres, T., Li, L., Schleuning, D., & Young, D. (2001). Preferred time-headway of highway drivers. *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585)*, 826–829.
- Bagloee, S., Sarvi, M., Patriksson, M., & Rajabifard, A. (2017). A mixed user-equilibrium and system-optimal traffic flow for connected vehicles stated as a complementarity problem: Mixed user-equilibrium and system-optimal traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, 32. <https://doi.org/10.1111/mice.12261>
- Ban, X. J., Pang, J.-S., Liu, H. X., & Ma, R. (2012). Continuous-time point-queue models in dynamic network loading. *Transportation Research Part B: Methodological*, 46(3), 360–380.
- Bao, Y., Zang, G., Yang, H., Gao, Z., & Long, J. (2023). Mathematical modeling of the platform assignment problem in a ride-sourcing market with a third-party integrator. *Transportation Research Part B: Methodological*, 178, 102833.
- Barzegari, V., Edrisi, A., & Nourinejad, M. (2023). Fleet cost and capacity effects of automated vehicles in mixed traffic networks: A system optimal assignment problem. *Transportation Research Part C: Emerging Technologies*, 148, 104020.
- Battifarano, M., & Qian, S. (2023). The impact of optimized fleets in transportation networks. *Transportation Science*.
- Becker, S., Brandenburg, S., & Thüring, M. (2022). Driver-initiated take-overs during critical braking maneuvers in automated driving—the role of time headway, traction usage, and trust in automation. *Accident Analysis & Prevention*, 174, 106725.
- Beckmann, M., McGuire, C. B., & Winsten, C. B. (1956). *Studies in the economics of transportation* (tech. rep.).

- Beevers, S. D., & Carslaw, D. C. (2005). The impact of congestion charging on vehicle emissions in london. *Atmospheric environment*, 39(1), 1–5.
- Bellei, G., Gentile, G., & Papola, N. (2005). A within-day dynamic traffic assignment model for urban road networks. *Transportation Research Part B: Methodological*, 39(1), 1–29.
- Ben-Akiva, M. E., Gao, S., Wei, Z., & Wen, Y. (2012). A dynamic traffic assignment model for highly congested urban networks. *Transportation research part C: emerging technologies*, 24, 62–82.
- Beojone, C. V., & Geroliminis, N. (2021). On the inefficiency of ride-sourcing services towards urban congestion. *Transportation research part C: emerging technologies*, 124, 102890.
- Biswas, R. K. (2022). *How do drivers avoid crashes: The role of driving headway* [Doctoral dissertation, UNSW Sydney].
- Boesch, P. M., Ciari, F., & Axhausen, K. W. (2016). Autonomous vehicle fleet sizes required to serve different levels of demand. *Transportation Research Record*, 2542(1), 111–119.
- Boyles, S. D., Lownes, N. E., & Unnikrishnan, A. (2020). Transportation network analysis. *Vol. I: Static and Dynamic Traffic Assignment*.
- Candler, W., & Townsley, R. (1982). A linear two-level programming problem. *Computers & Operations Research*, 9(1), 59–76.
- Census & Statistics Department, t. g. o. H. K. S. A. R. (2023). Wages and labour earnings.
- Centre, C. L. I. (2020). Is uber legal?
- Chan, C., Kuncheria, A., Zhao, B., Cabannes, T., Keimer, A., Wang, B., Bayen, A., & Macfarlane, J. (2021). Quasi-dynamic traffic assignment using high performance computing. *arXiv preprint arXiv:2104.12911*.
- Chen, J., Liang, H., Li, J., & Lv, Z. (2020). Connected automated vehicle platoon control with input saturation and variable time headway strategy. *IEEE Transactions on Intelligent Transportation Systems*, 22(8), 4929–4940.
- Chen, J., Liu, X., Liu, W., & Jiang, H. (2023). Two-sided instant incentive optimization under a shared budget in ride-hailing services. *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 3481–3493.
- Chen, T. D., Kockelman, K. M., & Hanna, J. P. (2016). Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice*, 94, 243–254.

- Chen, Z., He, F., Yin, Y., & Du, Y. (2017a). Optimal design of autonomous vehicle zones in transportation networks. *Transportation Research Part B: Methodological*, 99, 44–61. <https://doi.org/https://doi.org/10.1016/j.trb.2016.12.021>
- Chen, Z., He, F., Yin, Y., & Du, Y. (2017b). Optimal design of autonomous vehicle zones in transportation networks. *Transportation Research Part B: Methodological*, 99, 44–61.
- Chen, Z., He, F., Zhang, L., & Yin, Y. (2016a). Optimal deployment of autonomous vehicle lanes with endogenous market penetration. *Transportation Research Part C: Emerging Technologies*, 72, 143–156. <https://doi.org/https://doi.org/10.1016/j.trc.2016.09.013>
- Chen, Z., He, F., Zhang, L., & Yin, Y. (2016b). Optimal deployment of autonomous vehicle lanes with endogenous market penetration. *Transportation research part C: emerging technologies*, 72, 143–156.
- Chen, Z., Lin, X., Yin, Y., & Li, M. (2020). Path controlling of automated vehicles for system optimum on transportation networks with heterogeneous traffic stream. *Transportation Research Part C: Emerging Technologies*, 110, 312–329.
- Chiu, H. Y., Stupp Jr, G., & Brown Jr, S. (1977). Vehicle-follower control with variable-gains for short headway automated guideway transit systems.
- Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, S., & Hicks, J. (2011). Dynamic traffic assignment: A primer (transportation research circular e-c153).
- Chiu, Y.-C., Zheng, H., Villalobos, J., & Gautam, B. (2007). Modeling no-notice mass evacuation using a dynamic traffic flow optimization model. *Iie Transactions*, 39(1), 83–94.
- Chow, A. H. (2009). Properties of system optimal traffic assignment with departure time choice and its solution method. *Transportation Research Part B: Methodological*, 43(3), 325–344.
- Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation research part B: methodological*, 28(4), 269–287.
- Daganzo, C. F. (1995). The cell transmission model, part ii: Network traffic. *Transportation Research Part B: Methodological*, 29(2), 79–93. [https://doi.org/https://doi.org/10.1016/0191-2615\(94\)00022-R](https://doi.org/https://doi.org/10.1016/0191-2615(94)00022-R)
- de La Fortelle, A., & Qian, X. (2015). Autonomous driving at intersections: Combining theoretical analysis with practical considerations. *Its world congress 2015*.

- Dean. (2021). Uber statistics 2022: How many people ride with uber?
- Diao, M., Kong, H., & Zhao, J. (2021). Impacts of transportation network companies on urban mobility. *Nature Sustainability*, 4(6), 494–500.
- Dinh, T. T. (2019). Managing traffic congestion in a city: A study of singapore’s experiences. *Research gate*, 1.
- Du, J., & Ma, W. (2024). Maximin headway control of automated vehicles for system optimal dynamic traffic assignment in general networks. *Transportation Research Part E: Logistics and Transportation Review*, 188, 103628.
- Dunn. (2016). Ridesharing (uber) bill 2016.
- Elmorshedy, L., Smirnov, I., & Abdulhai, B. (2023). Freeway congestion management with reinforcement learning headway control of connected and autonomous vehicles. *Transportation Research Record*, 03611981231152459.
- Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science advances*, 5(5), eaau2670.
- Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77, 167–181.
- Fagnant, D. J., & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, 40, 1–13.
- Falocchio, J. C., & Levinson, H. S. (2015). *Road traffic congestion: A concise guide* (Vol. 7). Springer.
- Fan, Q., van Essen, J. T., & Correia, G. H. (2023). Optimising fleet sizing and management of shared automated vehicle (sav) services: A mixed-integer programming approach integrating endogenous demand, congestion effects, and accept/reject mechanism impacts. *Transportation Research Part C: Emerging Technologies*, 157, 104398.
- Fenton, R. E. (1979). A headway safety policy for automated highway operations. *IEEE Transactions on Vehicular Technology*, 28(1), 22–28.

- Fortuny-Amat, J., & McCarl, B. (1981). A representation and economic interpretation of a two-level programming problem. *Journal of the operational Research Society*, 32(9), 783–792.
- Fraedrich, E., Heinrichs, D., Bahamonde-Birke, F. J., & Cyganski, R. (2019). Autonomous driving, the built environment and policy implications. *Transportation research part A: policy and practice*, 122, 162–172.
- Gawron, C. (1998). An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C*, 9(03), 393–407.
- Gong, S., Shen, J., & Du, L. (2016a). Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. *Transportation Research Part B: Methodological*, 94, 314–334.
- Gong, S., Shen, J., & Du, L. (2016b). Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. *Transportation Research Part B: Methodological*, 94, 314–334. [https://doi.org/https://doi.org/10.1016/j.trb.2016.09.016](https://doi.org/10.1016/j.trb.2016.09.016)
- Guo, G., & Yue, W. (2012). Autonomous platoon control allowing range-limited sensors. *IEEE Transactions on vehicular technology*, 61(7), 2901–2912.
- Guo, H., Chen, Y., & Liu, Y. (2022). Shared autonomous vehicle management considering competition with human-driven private vehicles. *Transportation research part C: emerging technologies*, 136, 103547.
- Guo, Q., & Ban, X. (2020). Macroscopic fundamental diagram based perimeter control considering dynamic user equilibrium. *Transportation Research Part B: Methodological*, 136, 87–109. [https://doi.org/https://doi.org/10.1016/j.trb.2020.03.004](https://doi.org/10.1016/j.trb.2020.03.004)
- Guo, R.-Y., & Huang, H.-J. (2022). Day-to-day dynamics in a duopoly ride-sourcing market. *Transportation Research Part C: Emerging Technologies*, 135, 103528.
- Han, K., Liu, H., Gayah, V. V., Friesz, T. L., & Yao, T. (2016). A robust optimization approach for dynamic traffic signal control with emission considerations. *Transportation Research Part C: Emerging Technologies*, 70, 3–26.
- Han, K., Piccoli, B., & Szeto, W. (2016). Continuous-time link-based kinematic wave model: Formulation, solution existence, and well-posedness. *Transportmetrica B: Transport Dynamics*, 4(3), 187–222.

- Harker, P. T., & Pang, J.-S. (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical programming*, 48(1), 161–220.
- Hatipoglu, C., Ozguner, U., & Sommerville, M. (1996). Longitudinal headway control of autonomous vehicles. *Proceeding of the 1996 IEEE International Conference on Control Applications IEEE International Conference on Control Applications held together with IEEE International Symposium on Intelligent Control*, 721–726.
- Hoogendoorn, R., van Arerm, B., & Hoogendoorn, S. (2014). Automated driving, traffic flow efficiency, and human factors: Literature review. *Transportation Research Record*, 2422(1), 113–120.
- Hu, B., Hu, M., & Zhu, H. (2022). Surge pricing and two-sided temporal responses in ride hailing. *Manufacturing & Service Operations Management*, 24(1), 91–109.
- Hyland, M., & Mahmassani, H. S. (2018). Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign avs to immediate traveler demand requests. *Transportation Research Part C: Emerging Technologies*, 92, 278–297.
- Iacobucci, R., McLellan, B., & Tezuka, T. (2019). Optimization of shared autonomous electric vehicles operations with charge scheduling and vehicle-to-grid. *Transportation Research Part C: Emerging Technologies*, 100, 34–52.
- Jabari, S. E. (2016). Node modeling for congested urban road networks. *Transportation Research Part B: Methodological*, 91, 229–249.
- Janson, B. N. (1991). Dynamic traffic assignment for urban road networks. *Transportation Research Part B: Methodological*, 25(2-3), 143–161.
- Jiang, N. (2017). Optimal signal design for mixed equilibrium networks with autonomous and regular vehicles. *Journal of Advanced Transportation*, 2017(1), 5649823.
- Jiao, G., & Ramezani, M. (2022). Incentivizing shared rides in e-hailing markets: Dynamic discounting. *Transportation Research Part C: Emerging Technologies*, 144, 103879.
- Jin, W.-L. (2015). Continuous formulations and analytical properties of the link transmission model. *Transportation Research Part B: Methodological*, 74, 88–103.
- Kashmiri, F. A., Lo, H. K., Huai, Y., Liang, Z., & Li, L. (2025). Routing of autonomous vehicles for optimal total system cost under equilibrium. *Transportation Research Part C: Emerging Technologies*, 171, 105004.

- Ke, J., Li, X., Yang, H., & Yin, Y. (2021). Pareto-efficient solutions and regulations of congested ride-sourcing markets with heterogeneous demand and supply. *Transportation Research Part E: Logistics and Transportation Review*, 154, 102483.
- Ke, J., Yang, H., Li, X., Wang, H., & Ye, J. (2020). Pricing and equilibrium in on-demand ride-pooling markets. *Transportation Research Part B: Methodological*, 139, 411–431.
- Lavieri, P. S., & Bhat, C. R. (2019). Modeling individuals' willingness to share trips with strangers in an autonomous vehicle future. *Transportation research part A: policy and practice*, 124, 242–261.
- Lebacque, J.-P. (2005). First-order macroscopic traffic flow models: Intersection modeling, network modeling. *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory* University of Maryland, College Park.
- Lebacque, J. (1996). The godunov scheme and what it means for first order traffic flow models. *transportation and traffic theory: The proceedings of the 13th international symposium on transportation and traffic theory (isttt)*, jb lesort, ed.
- Levin, M. W. (2017). Congestion-aware system optimal route choice for shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 82, 229–247.
- Levin, M. W. (2019). Linear program for system optimal parking reservation assignment. *Journal of Transportation Engineering, Part A: Systems*, 145(12), 04019049.
- Levin, M. W., & Boyles, S. D. (2016a). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, 62, 103–116.
- Levin, M. W., & Boyles, S. D. (2016b). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, 62, 103–116. [https://doi.org/https://doi.org/10.1016/j.trc.2015.10.005](https://doi.org/10.1016/j.trc.2015.10.005)
- Li, G., Jiao, Y., Calvert, S. C., & van Lint, J. H. (2024). Lateral conflict resolution data derived from argoverse-2: Analysing safety and efficiency impacts of autonomous vehicles at intersections. *Transportation Research Part C: Emerging Technologies*, 167, 104802.
- Li, L., & Chen, X. M. (2017). Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: A survey. *Transportation Research Part C: Emerging Technologies*, 76, 170–188.

- Li, S., Tavafoghi, H., Poolla, K., & Varaiya, P. (2019). Regulating tncs: Should uber and lyft set their own rules? *Transportation Research Part B: Methodological*, 129, 193–225.
- Li, S., Yang, H., Poolla, K., & Varaiya, P. (2021). Spatial pricing in ride-sourcing markets under a congestion charge. *Transportation Research Part B: Methodological*, 152, 18–45.
- Li, S. E., Zheng, Y., Li, K., Wu, Y., Hedrick, J. K., Gao, F., & Zhang, H. (2017). Dynamical modeling and distributed control of connected and automated vehicles: Challenges and opportunities. *IEEE Intelligent Transportation Systems Magazine*, 9(3), 46–58.
- Li, T., Cao, Y., Xu, M., & Sun, H. (2023). Optimal intersection design and signal setting in a transportation network with mixed hvs and cavs. *Transportation Research Part E: Logistics and Transportation Review*, 175, 103173.
- Li, T., Xu, M., Sun, H., Xiong, J., & Dou, X. (2023). Stochastic ridesharing equilibrium problem with compensation optimization. *Transportation Research Part E: Logistics and Transportation Review*, 170, 102999.
- Li, X. (2022). Trade-off between safety, mobility and stability in automated vehicle following control: An analytical method. *Transportation research part B: methodological*, 166, 1–18.
- Li, Y., & Liu, Y. (2021). Optimizing flexible one-to-two matching in ride-hailing systems with boundedly rational users. *Transportation Research Part E: Logistics and Transportation Review*, 150, 102329.
- Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the royal society of london. series a. mathematical and physical sciences*, 229(1178), 317–345.
- Litman, T. (2017). Autonomous vehicle implementation predictions.
- Liu, B., & Chow, J. Y. (2023). On-demand mobility-as-a-service platform assignment games with guaranteed stable outcomes. *arXiv preprint arXiv:2305.00818*.
- Liu, J., Mirchandani, P., & Zhou, X. (2020). Integrated vehicle assignment and routing for system-optimal shared mobility planning with endogenous road congestion. *Transportation Research Part C: Emerging Technologies*, 117, 102675.
- Liu, Y., Lai, X., & Chang, G.-L. (2006). Cell-based network optimization model for staged evacuation planning under emergencies. *Transportation Research Record*, 1964(1), 127–135.

- Liu, Z., & Song, Z. (2019). Strategic planning of dedicated autonomous vehicle lanes and autonomous vehicle/toll lanes in transportation networks. *Transportation Research Part C: Emerging Technologies*, 106, 381–403.
- Liu, Z., Chen, Z., Yin, Y., & Xu, Z. (2022). Regulatory policies to electrify ridesourcing systems. *Transportation Research Part C: Emerging Technologies*, 141, 103743.
- Lo, H. K. (2001). A cell-based traffic control formulation: Strategies and benefits of dynamic timing plans. *Transportation Science*, 35(2), 148–164.
- Loeb, B., & Kockelman, K. M. (2019). Fleet performance and cost evaluation of a shared autonomous electric vehicle (saev) fleet: A case study for austin, texas. *Transportation Research Part A: Policy and Practice*, 121, 374–385.
- Long, J., Chen, J., Szeto, W., & Shi, Q. (2018). Link-based system optimum dynamic traffic assignment problems with environmental objectives. *Transportation Research Part D: Transport and Environment*, 60, 56–75.
- Long, J., & Szeto, W. Y. (2019). Link-based system optimum dynamic traffic assignment problems in general networks. *Operations Research*, 67(1), 167–182.
- Lu, C.-C., Liu, J., Qu, Y., Peeta, S., Roupail, N. M., & Zhou, X. (2016). Eco-system optimal time-dependent flow assignment in a congested network. *Transportation Research Part B: Methodological*, 94, 217–239.
- Luo, L., Portugal-Pereira, J., Takami, K., & Parady, G. (2024). Unintended environmental impacts of private automated vehicles: Insights from gunma prefecture, japan. *Transportation Research Part D: Transport and Environment*, 134, 104298.
- Lyu, G., Cheung, W. C., Teo, C.-P., & Wang, H. (2023). Multiobjective stochastic optimization: A case of real-time matching in ride-sourcing markets. *Manufacturing & Service Operations Management*.
- Ma, J., Meng, Q., Cheng, L., & Liu, Z. (2022). General stochastic ridesharing user equilibrium problem with elastic demand. *Transportation Research Part B: Methodological*, 162, 162–194.
- Ma, J., Xu, M., Meng, Q., & Cheng, L. (2020). Ridesharing user equilibrium problem under od-based surge pricing strategy. *Transportation Research Part B: Methodological*, 134, 1–24.

- Ma, R., Ban, X. (, & Pang, J.-S. (2014). Continuous-time dynamic system optimum for single-destination traffic networks with queue spillbacks. *Transportation Research Part B: Methodological*, 68, 98–122. <https://doi.org/https://doi.org/10.1016/j.trb.2014.06.003>
- Ma, R., Ban, X. J., & Szeto, W. (2017). Emission modeling and pricing on single-destination dynamic traffic networks. *Transportation Research Part B: Methodological*, 100, 255–283.
- Ma, W., & Qian, Z. S. (2018). Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data. *Transportation Research Part C: Emerging Technologies*, 96, 96–121.
- Madadi, B., van Nes, R., Snelder, M., & Van Arem, B. (2020). A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm. *Computer-Aided Civil and Infrastructure Engineering*, 35(1), 80–96.
- Madadi, B., van Nes, R., Snelder, M., & van Arem, B. (2021). Multi-stage optimal design of road networks for automated vehicles with elastic multi-class demand. *Computers & Operations Research*, 136, 105483.
- Middleton, R. H., & Braslavsky, J. H. (2010). String instability in classes of linear time invariant formation control with limited communication range. *IEEE Transactions on Automatic Control*, 55(7), 1519–1530.
- Militão, A. M., & Tirachini, A. (2021). Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. *Transportation research part A: policy and practice*, 151, 52–80.
- Mo, D., Yu, J., & Chen, X. M. (2020). Modeling and managing heterogeneous ride-sourcing platforms with government subsidies on electric vehicles. *Transportation Research Part B: Methodological*, 139, 447–472.
- Narayanan, S., Chaniotakis, E., & Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111, 255–293.
- Ngoduy, D., Hoang, N., Vu, H. L., & Watling, D. (2016). Optimal queue placement in dynamic system optimum solutions for single origin-destination traffic networks. *Transportation Research Part B: Methodological*, 92, 148–169.

- Ngoduy, D., Hoang, N., Vu, H., & Watling, D. (2021). Multiclass dynamic system optimum solution for mixed traffic of human-driven and automated vehicles considering physical queues. *Transportation Research Part B: Methodological*, 145, 56–79. <https://doi.org/10.1016/j.trb.2020.12.008>
- Nguyen, C. H., Hoang, N. H., Lee, S., & Vu, H. L. (2021). A system optimal speed advisory framework for a network of connected and autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 5727–5739.
- Nie, X., & Zhang, H. M. (2005). A comparative study of some macroscopic link models used in dynamic traffic assignment. *Networks and Spatial Economics*, 5, 89–115.
- Noruzoliaee, M., & Zou, B. (2022). One-to-many matching and section-based formulation of autonomous ridesharing equilibrium. *Transportation Research Part B: Methodological*, 155, 72–100.
- Osorio, C., Flötteröd, G., & Bierlaire, M. (2011). Dynamic network loading: A stochastic differentiable model that derives link state distributions. *Procedia-Social and Behavioral Sciences*, 17, 364–381.
- Othman, K. (2021). Public acceptance and perception of autonomous vehicles: A comprehensive review. *AI and Ethics*, 1(3), 355–387.
- Pantelidis, T. P., Chow, J. Y., & Rasulkhani, S. (2020). A many-to-many assignment game and stable outcome algorithm to evaluate collaborative mobility-as-a-service platforms. *Transportation Research Part B: Methodological*, 140, 79–100.
- Papavassilopoulos, G. P. (1982). Algorithms for static stackelberg games with linear costs and polyhedra constraints. *1982 21st IEEE Conference on Decision and Control*, 647–652.
- Parkopedia. (2023). Parking locations in mong kok, kowloon.
- Patriksson, M. (2015). *The traffic assignment problem: Models and methods*. Courier Dover Publications.
- Peeta, S., & Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and spatial economics*, 1, 233–265.
- Peng, R., Rui-hua, X., & Jin, Q. (2008). Bi-level simulated annealing algorithm for facility location problem. *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*, 3, 17–22.

- Penrose, R. (1955). A generalized inverse for matrices. *Mathematical proceedings of the Cambridge philosophical society*, 51(3), 406–413.
- Qian, Z. S., & Rajagopal, R. (2014). Optimal dynamic parking pricing for morning commute considering expected cruising time. *Transportation Research Part C: Emerging Technologies*, 48, 468–490.
- Qian, Z. S., Shen, W., & Zhang, H. (2012). System-optimal dynamic traffic assignment with and without queue spillback: Its path-based formulation and solution via approximate path marginal cost. *Transportation research part B: methodological*, 46(7), 874–893.
- Qin, H., Yu, B., & Zhang, Y. (2024). Exploring commuters' mode preference to autonomous vehicles based on a personalized travel experience survey. *Transportation*, 1–34.
- Richards, P. I. (1956). Shock waves on the highway. *Operations research*, 4(1), 42–51.
- Rui, X., Gongyuan, L., & Dong, H. (2010). Simulated annealing algorithm for solving a bi-level optimization model on high-speed railway station location. *2010 Third International Conference on Information and Computing*, 2, 159–162.
- Sahin, K. H., & Ciric, A. R. (1998). A dual temperature simulated annealing approach for solving bilevel programming problems. *Computers & chemical engineering*, 23(1), 11–25.
- Samaranayake, S., Krichene, W., Reilly, J., Monache, M. L. D., Goatin, P., & Bayen, A. (2018). Discrete-time system optimal dynamic traffic assignment (so-dta) with partial control for physical queuing networks. *Transportation Science*, 52(4), 982–1001.
- Schaller, B. (2017). Empty seats, full streets: Fixing manhattan's traffic problem. *Schaller Consulting*, 1(3), 1–27.
- Schneider. (2022). Taxi and ridehailing usage in new york city.
- Seo, T., & Asakura, Y. (2022). Multi-objective linear optimization problem for strategic planning of shared autonomous vehicle operation and infrastructure design. *IEEE Transactions on Intelligent Transportation Systems*, 23(4), 3816–3828. <https://doi.org/10.1109/TITS.2021.3071512>
- Shah, G., Zaman, M., Saifuddin, M., Toghi, B., & Fallah, Y. (2024). Scalable cellular v2x solutions: Large-scale deployment challenges of connected vehicle safety networks. *Automotive Innovation*, 1–10.

- Sharon, G., Albert, M., Rambha, T., Boyles, S., & Stone, P. (2018). Traffic optimization for a mixture of self-interested and compliant agents. *Proceedings of the AAAI conference on artificial intelligence*, 32(1).
- Sharon, G., Levin, M. W., Hanna, J. P., Rambha, T., Boyles, S. D., & Stone, P. (2017). Network-wide adaptive tolling for connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 84, 142–157.
- Sheffi, Y. (1985). *Urban transportation networks* (Vol. 6). Prentice-Hall, Englewood Cliffs, NJ.
- Shen, W., & Zhang, H. (2014a). System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network. *Transportation Research Part B: Methodological*, 65, 1–17.
- Shen, W., & Zhang, H. (2014b). System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network. *Transportation Research Part B: Methodological*, 65, 1–17. <https://doi.org/https://doi.org/10.1016/j.trb.2014.02.002>
- Shi, X., & Li, X. (2021). Constructing a fundamental diagram for traffic flow with automated vehicles: Methodology and demonstration. *Transportation Research Part B: Methodological*, 150, 279–292. <https://doi.org/https://doi.org/10.1016/j.trb.2021.06.011>
- Shladover, S. E., Desoer, C. A., Hedrick, J. K., Tomizuka, M., Walrand, J., Zhang, W.-B., McMahon, D. H., Peng, H., Sheikholeslam, S., & McKeown, N. (1991). Automated vehicle control developments in the path program. *IEEE Transactions on vehicular technology*, 40(1), 114–130.
- Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D., & Pavone, M. (2014). Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: A case study in singapore. *Road vehicle automation*, 229–245.
- Sumalee, A., Zhong, R., Pan, T., & Szeto, W. (2011). Stochastic cell transmission model (sctm): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B: Methodological*, 45(3), 507–533.
- Sun, H., Wang, H., & Wan, Z. (2019). Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity. *Transportation Research Part B: Methodological*, 125, 76–93.
- Szeto, W., & Lo, H. K. (2006). Dynamic traffic assignment: Properties and extensions. *Transportmetrica*, 2(1), 31–52.

- Tafreshian, A., & Masoud, N. (2020). Using subsidies to stabilize peer-to-peer ridesharing markets with role assignment. *Transportation Research Part C: Emerging Technologies*, 120, 102770.
- Tampère, C. M., Corthout, R., Cattrysse, D., & Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, 45(1), 289–309.
- Tan, Y., Ma, R., Sun, Z., & Zhang, P. (2021). Emission exposure optimum for a single-destination dynamic traffic network. *Transportation Research Part D: Transport and Environment*, 94, 102817.
- Tang, W., Xie, N., Mo, D., Cai, Z., Lee, D.-H., & Chen, X. M. (2023). Optimizing subsidy strategies of the ride-sourcing platform under government regulation. *Transportation Research Part E: Logistics and Transportation Review*, 173, 103112.
- Tarduno, M. (2021). The congestion costs of uber and lyft. *Journal of Urban Economics*, 122, 103318.
- Teoh, E. R., & Kidd, D. G. (2017). Rage against the machine? google's self-driving cars versus human drivers. *Journal of safety research*, 63, 57–60.
- Tian, Z., Feng, T., Timmermans, H. J., & Yao, B. (2021). Using autonomous vehicles or shared cars? results of a stated choice experiment. *Transportation Research Part C: Emerging Technologies*, 128, 103117.
- Transport department, the government of Hong Kong Special Administrative Region. (2023). Taxi fare of hong kong.
- UITP. (2020). Global taxi benchmarking study 2019.
- van Essen, M., Thomas, T., van Berkum, E., & Chorus, C. (2016). From user equilibrium to system optimum: A literature review on the role of travel information, bounded rationality and non-selfish behaviour at the network and individual levels. *Transport reviews*, 36(4), 527–548.
- Vignon, D. A., Yin, Y., & Ke, J. (2021). Regulating ridesourcing services with product differentiation and congestion externality. *Transportation Research Part C: Emerging Technologies*, 127, 103088.
- Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident analysis & prevention*, 35(3), 427–433.

- Wadud, Z., & Chintakayala, P. K. (2021). To own or not to own—that is the question: The value of owning a (fully automated) vehicle. *Transportation Research Part C: Emerging Technologies*, 123, 102978.
- Wadud, Z., MacKenzie, D., & Leiby, P. (2016). Help or hindrance? the travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, 86, 1–18.
- Waller, S. T., Mouskos, K. C., Kamaryiannis, D., & Ziliaskopoulos, A. K. (2006). A linear model for the continuous network design problem. *Computer-Aided Civil and Infrastructure Engineering*, 21(5), 334–345.
- Waller, S. T., & Ziliaskopoulos, A. K. (2001). Stochastic dynamic network design problem. *Transportation Research Record*, 1771(1), 106–113.
- Wang, G., Qi, H., Xu, H., & Ryu, S. (2020). A mixed behaviour equilibrium model with mode choice and its application to the endogenous demand of automated vehicles [Special Issue on Traffic and Transportation in the Era of Big Data and Shared Economy]. *Journal of Management Science and Engineering*, 5(4), 227–248. [https://doi.org/https://doi.org/10.1016/j.jmse.2020.05.003](https://doi.org/10.1016/j.jmse.2020.05.003)
- Wang, H., & Yang, H. (2019). Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological*, 129, 122–155.
- Wang, J., Lu, L., Peeta, S., & He, Z. (2021). Optimal toll design problems under mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 125, 102952.
- Wang, J., Peeta, S., & He, X. (2019a). Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. *Transportation Research Part B: Methodological*, 126, 139–168. [https://doi.org/https://doi.org/10.1016/j.trb.2019.05.022](https://doi.org/10.1016/j.trb.2019.05.022)
- Wang, J., Peeta, S., & He, X. (2019b). Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. *Transportation Research Part B: Methodological*, 126, 139–168.
- Wang, X., Liu, W., Yang, H., Wang, D., & Ye, J. (2019). Customer behavioural modelling of order cancellation in coupled ride-sourcing and taxi markets. *Transportation Research Procedia*, 38, 853–873.

- Wang, Y., Szeto, W. Y., Han, K., & Friesz, T. L. (2018). Dynamic traffic assignment: A review of the methodological advances for environmentally sustainable road transportation applications. *Transportation Research Part B: Methodological*, 111, 370–394.
- Wardrop, J. G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3), 325–362.
- Xi, H., Aussel, D., Liu, W., Waller, S. T., & Rey, D. (2022). Single-leader multi-follower games for the regulation of two-sided mobility-as-a-service markets. *European Journal of Operational Research*.
- Xiao, L., & Gao, F. (2011). Practical string stability of platoon of adaptive cruise control vehicles. *IEEE Transactions on intelligent transportation systems*, 12(4), 1184–1194.
- Xiao, X., Zhang, Y., Wang, X. B., & Guo, X. (2023). Adaptive headway control algorithm for mixed-traffic stabilization and optimization with automated cars and trucks. *Transportation Research Record*, 03611981231156587.
- Yao, R., & Zhang, K. (2024). Design an intermediary mobility-as-a-service (maas) platform using many-to-many stable matching framework. *Transportation Research Part B: Methodological*, 102991.
- Yaqoob, I., Khan, L. U., Kazmi, S. A., Imran, M., Guizani, N., & Hong, C. S. (2019). Autonomous driving cars in smart cities: Recent advances, requirements, and challenges. *IEEE Network*, 34(1), 174–181.
- Yperman, I. (2007). The link transmission model for dynamic network loading.
- Yu, H., Jiang, R., He, Z., Zheng, Z., Li, L., Liu, R., & Chen, X. (2021). Automated vehicle-involved traffic flow studies: A survey of assumptions, models, speculations, and perspectives. *Transportation Research Part C: Emerging Technologies*, 127, 103101. <https://doi.org/https://doi.org/10.1016/j.trc.2021.103101>
- Yu, H., Ma, R., & Zhang, H. M. (2018). Optimal traffic signal control under dynamic user equilibrium and link constraints in a general network. *Transportation research part B: methodological*, 110, 302–325.
- Zhang, F., Lu, J., & Hu, X. (2020). Traffic equilibrium for mixed traffic flows of human-driven vehicles and connected and autonomous vehicles in transportation networks under tradable credit scheme. *Journal of advanced transportation*, 2020(1), 8498561.

- Zhang, F., Lu, J., & Hu, X. (2022). Integrated path controlling and subsidy scheme for mobility and environmental management in automated transportation networks. *Transportation Research Part E: Logistics and Transportation Review*, 167, 102906.
- Zhang, F., Lu, J., Hu, X., & Meng, Q. (2023). Integrated deployment of dedicated lane and roadside unit considering uncertain road capacity under the mixed-autonomy traffic environment. *Transportation research part B: methodological*, 174, 102784.
- Zhang, K., & Nie, Y. M. (2018a). Mitigating the impact of selfish routing: An optimal-ratio control scheme (orcs) inspired by autonomous driving. *Transportation Research Part C: Emerging Technologies*, 87, 75–90.
- Zhang, K., & Nie, Y. (2018b). Mitigating the impact of selfish routing: An optimal-ratio control scheme (orcs) inspired by autonomous driving. *Transportation Research Part C: Emerging Technologies*, 87, 75–90. <https://doi.org/https://doi.org/10.1016/j.trc.2017.12.011>
- Zhang, K., & Nie, Y. M. (2022). Mitigating traffic congestion induced by transportation network companies: A policy analysis. *Transportation Research Part A: Policy and Practice*, 159, 96–118.
- Zhang, P., & Qian, S. (2020). Path-based system optimal dynamic traffic assignment: A subgradient approach. *Transportation Research Part B: Methodological*, 134, 41–63.
- Zhang, Q., Liu, Y., & Fan, Z.-P. (2023). Short-term subsidy strategy for new users of ride-hailing platform with user base. *Computers & Industrial Engineering*, 179, 109177.
- Zhang, W., Guhathakurta, S., Fang, J., & Zhang, G. (2015). Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable Cities and Society*, 19, 34–45.
- Zhao, M., Li, B., Ren, J., & Hao, Z. (2023). Competition equilibrium of ride-sourcing platforms and optimal government subsidies considering customers' green preference under peak carbon dioxide emissions. *International Journal of Production Economics*, 255, 108679.
- Zhong, Y., Yang, T., Cao, B., & Cheng, T. (2022). On-demand ride-hailing platforms in competition with the taxi industry: Pricing strategies and government supervision. *International Journal of Production Economics*, 243, 108301.
- Zhou, C., Li, X., & Chen, L. (2024). Impact of negotiation on subsidy scheme to incentivise the integrated use of metro and ride-hailing services. *International Journal of Logistics Research and Applications*, 27(10), 1870–1890.

- Zhou, J., & Zhu, F. (2020a). Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. *Transportation research part C: emerging technologies*, 115, 102614.
- Zhou, J., & Zhu, F. (2020b). Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. *Transportation Research Part C: Emerging Technologies*, 115, 102614. [https://doi.org/https://doi.org/10.1016/j.trc.2020.102614](https://doi.org/10.1016/j.trc.2020.102614)
- Zhu, F., & Ukkusuri, S. V. (2015). A linear programming formulation for autonomous intersection control within a dynamic traffic assignment and connected vehicle environment. *Transportation Research Part C: Emerging Technologies*, 55, 363–378.
- Zhu, Y., & Zhu, F. (2018). Distributed adaptive longitudinal control for uncertain third-order vehicle platoon in a networked environment. *IEEE Transactions on Vehicular Technology*, 67(10), 9183–9197.
- Zhu, Z., Ke, J., & Wang, H. (2021). A mean-field markov decision process model for spatial-temporal subsidies in ride-sourcing markets. *Transportation Research Part B: Methodological*, 150, 540–565.
- Zhu, Z., Xu, A., He, Q.-C., & Yang, H. (2021). Competition between the transportation network company and the government with subsidies to public transit riders. *Transportation Research Part E: Logistics and Transportation Review*, 152, 102426.

Appendix A

Appendix in maximin headway control

A.1 Notations

Type	Notation	Description
Sets	V	Set of nodes
	R	Set of origins
	$\bar{R}$	Set of dummy origins
	S	Set of destinations
	$\bar{S}$	Set of dummy destinations
	E	Set of links
	L_R	Set of origin connectors
	L_S	Set of destination connectors
	$v_{i,j}^f$	Free-flow speed of link (i, j)
	$L_{i,j}$	Length of link (i, j)
Parameters	L	Length of an AV
	$\bar{Q}_{i,j}^U$	Upstream queue capacity of link (i, j)
	$\bar{Q}_{i,j}^D$	Downstream queue capacity of link (i, j)
	$\bar{C}_{i,j}^\mu$	Inflow capacity of link (i, j)
	$\bar{C}_{i,j}^\nu$	Outflow capacity of link (i, j)
	T_l	Time horizon of demand
	T	Total time horizon
	$h_{i,j;t}^{min}$	Lower bound of headway in link (i, j) at time t
	$h_{i,j;t}^{max}$	Upper bound of headway in link (i, j) at time t
Variables	$u_{i,j}(t)$	Inflow of the flow area of link (i, j) at time t
	$u_{i,j}^{s'}(t)$	Inflow of the flow area of link (i, j) with destination s' at time t
	$v_{i,j}(t)$	Outflow of the buffer area of link (i, j) at time t
	$v_{i,j}^{s'}(t)$	Outflow of the buffer area of link (i, j) with destination s' at time t
	$f_{i,j}(t)$	Inflow at the boundary of the flow area and buffer area of link (i, j) at time t
	$f_{i,j}^{s'}(t)$	Inflow at the boundary of the flow area and buffer area of link (i, j) with destination s' at time t
	$\rho_{i,j}(t)$	Density of the flow area of link (i, j) at time t
	$\rho_{i,j}^{s'}(t)$	Density of the flow area of link (i, j) with destination s' at time t
	$q_{i,j}^D(t)$	Downstream queue of link (i, j) at time t
	$q_{i,j}^{D,s'}(t)$	Downstream queue of link (i, j) with destination s' at time t
	$q_{i,j}^U(t)$	Upstream queue of link (i, j) at time t
	$q_{i,j}^{U,s'}(t)$	Upstream queue of link (i, j) with destination s' at time t
	$h_{i,j}(t)$	Headway of automated vehicles in link (i, j) at time t
	$\tau_{i,j}^w(t)$	Wave travel time in congested states on link (i, j) at time t
	$d_{r',s'}(t)$	Travel demand of O-D pair (r', s') at time t
	$w_{r',r}^{s'}(t)$	Number of AVs waiting at origins in the origin connectors (r', r) with destination s' at time t
	$U_{i,j}^{s'}(t)$	Cumulative inflow of the flow area of link (i, j) with destination s' at time t
	$F_{i,j}^{s'}(t)$	Cumulative inflow at the boundary of the flow area and buffer area of link (i, j) with destination s' at time t
	$V_{i,j}^{s'}(t)$	Cumulative outflow of the buffer area of link (i, j) with destination s' at time t

Table A.1: Notation table.

A.2 Proof of Proposition 1

Proposition 1 indicates that the headway-dependent double queue (HDQ) model is a generalized form of the double queue (DQ) model. The formulation of the downstream queue $q_{i,j}^{d,s'}(t)$ and upstream queue $q_{i,j}^{u,s'}(t)$ in DQ for $(i, j) \in E$, $s' \in \widetilde{S}$ and $t \in [0, T]$ are shown as follows:

$$\begin{aligned} q_{i,j}^{d,s'}(t) &= \int_0^{t-\tau_{i,j}^f} u_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^t v_{i,j}^{s'}(\hat{t}) d\hat{t} + q_{i,j}^{d,s'}(0) = U_{i,j}^{s'}(t - \tau_{i,j}^f) - V_{i,j}^{s'}(t), \\ q_{i,j}^{u,s'}(t) &= \int_0^t u_{i,j}^{s'}(\hat{t}) d\hat{t} - \int_0^{t-\tau_{i,j}^w} v_{i,j}^{s'}(\hat{t}) d\hat{t} + q_{i,j}^{u,s'}(0) = U_{i,j}^{s'}(t) - V_{i,j}^{s'}(t - \tau_{i,j}^w), \end{aligned}$$

where $q_{i,j}^{d,s'}(t)$ and $q_{i,j}^{u,s'}(t)$ represent the downstream and upstream queue in link (i, j) with destination s' at time t in DQ, respectively, and the two queues are interpreted as follows:

- Downstream queue $q_{i,j}^{d,s'}(t)$ measures the number of vehicles with destination s' at the end of link (i, j) at time t . DQ assumes the inflow $u_{i,j}^{s'}$ would travel from the entrance to the exit of link (i, j) at the free-flow speed, and the outflow of link (i, j) at time t is determined by the outflow capacity link (i, j) and upstream queue in next link.
- Upstream queue $q_{i,j}^{u,s'}(t)$ measures the number of vehicles at the entrance of link (i, j) at time t . DQ assumes the space of outflow $v_{i,j}^{s'}$ would travel from the exit to the entrance of link (i, j) at the wave travel speed in congested states.

However, the assumptions of DQ model may have the following limitations:

- Congestion may occur in the flow area of a link so that AVs may not travel from the entrance to the exit of a link at the free-flow speed. The influence of headway on the flow of a link is also not considered in DQ.
- Space may not propagate to the entrance of a link in time and retain in the downstream. Take an extreme example, if the front AV leaves the link but other AVs behind are still, the space left by the front AV would still retain in the downstream and it is not accurate to use the outflow to calculate the propagation of space in wave travel speed. Besides, the influence of headway on the wave travel speed on a link is also not considered in DQ.

The HDQ addresses the above limitations by exploring the effects of headway. Equations 3.3 and 3.4 formulate the downstream and upstream queue in HDQ. The relationship between the HDQ and DQ is presented in Equation A.1 and A.2.

$$\begin{aligned}
q_{i,j}^{d,s'}(t) &= U_{i,j}^{s'}(t - \tau_{i,j}^f) - V_{i,j}^{s'}(t) \\
&= U_{i,j}^{s'}(t - \tau_{i,j}^f) - F_{i,j}^{s'}(t) + F_{i,j}^{s'}(t) - V_{i,j}^{s'}(t) \\
&= q_{i,j}^{\mathcal{D}}(t) + U_{i,j}^{s'}(t - \tau_{i,j}^f) - F_{i,j}^{s'}(t)
\end{aligned} \tag{A.1}$$

$$\begin{aligned}
q_{i,j}^{u,s'}(t) &= U_{i,j}^{s'}(t) - V_{i,j}^{s'}(t - \tau_{i,j}^w) \\
&= U_{i,j}^{s'}(t) - F_{i,j}^{s'}(t - \tau_{i,j}^w) + F_{i,j}^{s'}(t - \tau_{i,j}^w) - V_{i,j}^{s'}(t - \tau_{i,j}^w) \\
&= q_{i,j}^{u,s'}(t) + q_{i,j}^{\mathcal{D},s'}(t - \tau_{i,j}^w)
\end{aligned} \tag{A.2}$$

In Equation A.1, $U_{i,j}^{s'}(t - \tau_{i,j}^f) - F_{i,j}^{s'}(t)$ represents the congestion in the flow area in HDQ. However, under the assumption of DQ, we have $u_{i,j}^{s'}(t - \tau_{i,j}^f) = f_{i,j}^{s'}(t)$, meaning that vehicles would travel from the entrance of link to the boundary of the flow area and buffer area at free-flow speed. Therefore, the downstream queue of HDQ is equivalent to the DQ as follows:

$$q_{i,j}^{d,s'}(t) = q_{i,j}^{\mathcal{D},s'}(t) + U_{i,j}^{s'}(t - \tau_{i,j}^f) - F_{i,j}^{s'}(t) = q_{i,j}^{\mathcal{D},s'}(t).$$

In Equation A.2, the upstream queue in DQ actually consists of the queue in both the flow area and buffer area in HDQ. The upstream queue in HDQ actually considers the flow entering the buffer area to fill the leaving space. If we ignore it, we would derive the upstream queue by outflow v as DQ.

Different from DQ, the HDQ specifies the traffic state in a link and incorporates the effect of headway. To calculate the number of AVs at the entrance and end of a link, we just focus on the flow area and buffer area, respectively. Compared with the DQ, the HDQ provides a more precise way to calculate the downstream and upstream queue if we are clear about the state of the boundary of the flow area and buffer area compared with the DQ. If under the assumption of DQ, the effect of headway is ignored and HDQ would be reduced to DQ.

A.3 Proof of Proposition 2

This appendix proves the properties of the proposed headway-dependent double queue (HDQ) model, and we mainly focus on exploring the constraints on the flow rates of the inflow and outflow in the HDQ model.

With Eqs.(4.29) in Yperman, 2007 and the formulation of the downstream queue in the HDQ model, we extend it to the HDQ model and Equation A.3 presents the constraint on the sending flow $S_{i,j}(t)$ of HDQ.

$$S_{i,j}(t) \leq V_{i,j}(t + \Delta_t) - V_{i,j}(t) \leq F_{i,j}(t + \Delta_t) - V_{i,j}(t) \quad (\text{A.3})$$

where $V_{i,j}(t)$ and $F_{i,j}(t)$ are the cumulative outflow of the buffer area and cumulative inflow at the boundary of the flow area and buffer area of link (i, j) at time t , respectively.

With Eqs.(4.33) in Yperman, 2007 and the formulation of the upstream queue in HDQ, the constraint on the receiving flow $R_{i,j}(t)$ of the HDQ is stated in Equation A.4.

$$R_{i,j}(t) \leq U_{i,j}(t + \Delta_t) - U_{i,j}(t) \leq \bar{Q}_{i,j}^u + F_{i,j}(t + \Delta_t - \tau_{i,j}^w(t)) - U_{i,j}(t) \quad (\text{A.4})$$

where $U_{i,j}(t)$ is the cumulative inflow of the flow area of link (i, j) at time t .

When the time step Δ_t approaches zero, the constraint of outflow v is shown in A.5 and the constraint of inflow u is shown in A.6.

$$\begin{aligned}
v_{i,j}(t) &\leq \lim_{\Delta_t \rightarrow 0} \frac{S_{i,j}(t)}{\Delta_t} \leq \lim_{\Delta_t \rightarrow 0} \min \left\{ \frac{F_{i,j}(t + \Delta_t) - V_{i,j}(t)}{\Delta_t}, \bar{C}_{i,j}^\nu \right\} \\
&= \lim_{\Delta_t \rightarrow 0} \min \left\{ \frac{F_{i,j}(t + \Delta_t) - F_{i,j}(t)}{\Delta_t} + \frac{F_{i,j}(t) - V_{i,j}(t)}{\Delta_t}, \bar{C}_{i,j}^\nu \right\} \\
&= \lim_{\Delta_t \rightarrow 0} \min \left\{ \frac{F_{i,j}(t + \Delta_t) - F_{i,j}(t)}{\Delta_t} + \frac{q_{i,j}^{\mathcal{D}}(t)}{\Delta_t}, \bar{C}_{i,j}^\nu \right\} \\
&= \begin{cases} \bar{C}_{i,j}^\nu, & q_{i,j}^{\mathcal{D}}(t) > 0 \\ \min \{f_{i,j}(t), \bar{C}_{i,j}^\nu\}, & q_{i,j}^{\mathcal{D}}(t) = 0 \end{cases} \tag{A.5}
\end{aligned}$$

$$\begin{aligned}
u_{i,j}(t) &\leq \lim_{\Delta_t \rightarrow 0} \frac{R_{i,j}(t)}{\Delta_t} \leq \lim_{\Delta_t \rightarrow 0} \min \left\{ \frac{\bar{Q}_{i,j}^u + F_{i,j}(t + \Delta_t - \tau_{i,j}^w(t)) - V_{i,j}(t)}{\Delta_t}, \bar{C}_{i,j}^\mu \right\} \\
&= \lim_{\Delta_t \rightarrow 0} \min \left\{ \frac{\bar{Q}_{i,j}^u - q_{i,j}^u(t)}{\Delta_t} + \frac{F_{i,j}(t + \Delta_t - \tau_{i,j}^w(t)) - F_{i,j}(t - \tau_{i,j}^w(t))}{\Delta_t}, \bar{C}_{i,j}^\mu \right\} \\
&= \begin{cases} \bar{C}_{i,j}^\mu, & q_{i,j}^u(t) < \bar{Q}_{i,j}^u \\ \min \{f_{i,j}(t - \tau_{i,j}^w(t)), \bar{C}_{i,j}^\mu\}, & q_{i,j}^u(t) = \bar{Q}_{i,j}^u \end{cases} \tag{A.6}
\end{aligned}$$

A.4 Proof of Proposition 3

In general, we could construct a feasible solution to Equation 3.18 using the following three steps:

- Let all the demands of each O-D pair arrive at the downstream queue in the origin connectors while not releasing them. We define the time period for such a process as N_I . We don't consider the capacity of the number of AVs waiting at origins here.
- For each O-D pair (r', s') , we divide the total demand $D^{r',s'}$ into many equal batches $B^{r',s'}$, and $D^{r',s'} = n^{r',s'} B^{r',s'}$. We only use one path of each O-D pair (r', s') , which is denoted as $P^{r',s'}$. We only release one batch of an O-D pair only once. After all the AVs in this batch arrive at the destination, we will release another batch of this O-D pair. Repeat this process until finishing all batches of an O-D pair and continue another O-D pair.
- In each batch, only when all the AVs in this batch arrive at the downstream queue in each link, we will release them into the next link in the selected path. The time of passing link (i, j) for batch $B^{r',s'}$ is denoted by $N_{i,j}^{r',s'}$.

Therefore, the total travel time consumed by the above three steps is $N_I + \sum_{(r',s')} n^{r',s'} \sum_{(i,j) \in Pr^{r',s'}} N_{i,j}^{r',s'}$.

If we choose the time horizon N that is larger than $N_I + \sum_{(r',s')} n^{r',s'} \sum_{(i,j) \in Pr^{r',s'}} N_{i,j}^{r',s'}$, we could construct a solution so that the problem is feasible. To this end, we present the details of how to construct the flow as follows.

In order to satisfy the queue capacity constraint in Equation 3.29, the size of the batch should be constrained as follows:

$$B^{r',s'} < \min_{(i,j) \in Pr^{r',s'}} \left\{ \bar{Q}_{i,j}^{\mathcal{U}}, \bar{Q}_{i,j}^{\mathcal{D}} \right\}.$$

To keep the AVs in each batch following the free-flow region of HFD in the flow area in each link, the size of the batch should satisfy the following equation.

$$B^{r',s'} < \min_{(i,j) \in Pr^{r',s'}, k \in [1, N]} \left\{ \frac{L_{i,j}}{h_{i,j;k}(t)v_{i,j}^f + L} \right\}$$

Therefore, the size of batch $B^{r',s'}$ are restricted as follows.

$$\sum_{k=1}^{N_I} d_{r',s'}(k) = D^{r',s'} = n^{r',s'} B^{r',s'}$$

$$B^{r',s'} < \min_{(i,j) \in Pr^{r',s'}, k \in [1, N]} \left\{ \bar{Q}_{i,j}^{\mathcal{U}}, \bar{Q}_{i,j}^{\mathcal{D}}, \frac{L_{i,j}}{h_{i,j;k}(t)v_{i,j}^f + L} \right\}$$

After determining the size of each batch, the next step is to determine the flow rate of each batch. Suppose we release the batch at a constant rate, so $B^{r',s'} = N_B^{r',s'} F_B^{r',s'}$, where $N_B^{r',s'}$ is the time of releasing all AVs in this batch and $F_B^{r',s'}$ is the flow rate of releasing AVs. In order to satisfy the inflow and outflow capacity constraints in Equation 3.30, $F_B^{r',s'}$ should satisfy the following equation.

$$F_B^{r',s'} < \min_{(i,j) \in Pr^{r',s'}} \left\{ \bar{C}_{i,j}^{\mu}, \bar{C}_{i,j}^{\nu} \right\}.$$

Then we calculate $N_{i,j}^{r',s'}$, which represents the time of passing link (i, j) for batch $B^{r',s'}$ and $N_{i,j}^{r',s'}$ is comprised of two time periods.

The first time period refers to the case that we release the batch from the prior link to link

(i, j) . In this time period, we have $u_{i,j}(k) = F_B^{r',s'}$ for $k \in [N_{B,i,j}^{r',s'}, N_{B,i,j}^{r',s'} + N_B^{r',s'}]$, where $N_{B,i,j}^{r',s'}$ is the time ready to release the batch B with O-D (r', s') to link (i, j) . In this time period, the flow in the flow area follows the free-flow region in HFD, so we have the following equations.

$$\begin{aligned} f_{i,j}(k) &= v_{i,j}^f \rho_{i,j}(k) \\ \rho_{i,j}^{s'}(k) &= \rho_{i,j}^{s'}(k-1) + \Delta_t \frac{F_B^{r',s'} - f_{i,j}^{s'}(k)}{L_{i,j}} \end{aligned}$$

Combining the above two equations, we have

$$\begin{aligned} \rho_{i,j}(k) &= \frac{L_{i,j} \rho_{i,j}(k-1) + \Delta_t F_B^{r',s'}}{L_{i,j} + \Delta_t v_{i,j}^f} \\ &= a_{i,j} \rho_{i,j}(k-1) + b_{i,j}^{B,r',s'} \end{aligned}$$

where $a_{i,j} = \frac{L_{i,j}}{L_{i,j} + \Delta_t v_{i,j}^f}$ and $b_{i,j}^{B,r',s'} = \frac{\Delta_t F_B^{r',s'}}{L_{i,j} + \Delta_t v_{i,j}^f}$.

Therefore, we have the formulation of density at time $N_{B,i,j}^{r',s'} + N_B^{r',s'}$ as follows.

$$\rho_{i,j}(N_{B,i,j}^{r',s'} + N_B^{r',s'}) = b_{i,j}^{B,r',s'} (a_{i,j}^{(N_{B,i,j}^{r',s'})} + a_{i,j}^{(N_{B,i,j}^{r',s'}-1)} + \dots + 1) = c_{i,j}^{B,r',s'}$$

In the second period, all AVs in this batch have already been released and we calculate the time when all AVs in this batch enter the buffer area to satisfy the end constraint. So we have $u_{i,j}(k) = 0$ for $k \in [N_{B,i,j}^{r',s'} + N_B^{r',s'}, N_{B,i,j}^{r',s'} + N_{i,j}^{r',s'}]$. Then we have following equations in this time period.

$$\begin{aligned} f_{i,j}(k) &= v_{i,j}^f \rho_{i,j}(k) \\ \rho_{i,j}^{s'}(k) &= \rho_{i,j}^{s'}(k-1) - \Delta_t \frac{f_{i,j}^{s'}(k)}{L_{i,j}} \end{aligned}$$

Combining the above two equations, we have

$$\begin{aligned}
\rho_{i,j}(k) &= \frac{L_{i,j} \rho_{i,j}(k-1)}{L_{i,j} + \Delta_t v_{i,j}^f} \\
&= a_{i,j} \rho_{i,j}(k-1)
\end{aligned}$$

To satisfy the end constraint in Equation 3.34, the density at time $N_{B,i,j}^{r',s'} + N_{i,j}^{r',s'}$ should satisfy the following equation.

$$\rho_{i,j}(N_{B,i,j}^{r',s'} + N_{i,j}^{r',s'}) = c_{i,j} a_{i,j}^{(N_{i,j}^{r',s'} - N_{B,i,j}^{r',s'})} < \frac{1}{L_{i,j}}$$

So we have

$$N_{i,j}^{r',s'} > N_{B,i,j}^{r',s'} - \frac{\ln c_{i,j} L_{i,j}}{\ln a_{i,j}}$$

Then we set $N_{i,j}^{r',s'} = \max\{N_{B,i,j}^{r',s'}, \lceil N_{B,i,j}^{r',s'} - \frac{\ln c_{i,j} L_{i,j}}{\ln a_{i,j}} \rceil\}$. It is obvious that other constraints are satisfied by constructing the solution based on the above procedures. Therefore, if we choose the time horizon N so that N is larger than $N_I + \sum_{(r',s')} n^{r',s'} \sum_{(i,j)} N_{i,j}^{r',s'} \delta_{i,j}^{P^{r',s'}}$, where $\delta_{i,j}^{P^{r',s'}}$ indicates whether link (i, j) is in the path $P^{r',s'}$, then we could always find a feasible solution of proposed problem.

A.5 Sensitivity analysis of the headway control for SO-DTA

This section proposes a sensitivity-based algorithm to solve the SO-DTA problem. We first transform the discretized version of the headway-dependent SO-DTA formulation into linear programming (LP) given an exogenous headway variable $\mathbf{h}$, as shown in A.5.1, then the sensitivity-based algorithm is presented.

A.5.1 LP formulation

This section transforms the discretized version of SO-DTA problem formulation from Equations 3.18 to 3.19 and Equations 3.22 to 3.36 into LP.

The objective of section 3.3 is to find the optimal headway $\{h_{i,j}(k)\}$. In the proposed pro-

gramming, the involved variables include $\{\rho_{i,j}(k)\}$, $\{\rho_{i,j}^{s'}(k)\}$, $\{q_{i,j}^{\mathcal{D}}(k)\}$, $\{q_{i,j}^{\mathcal{D},s'}(k)\}$, $\{q_{i,j}^{\mathcal{U}}(k)\}$, $\{q_{i,j}^{\mathcal{U},s'}(k)\}$, $\{u_{i,j}(k)\}$, $\{u_{i,j}^{s'}(k)\}$, $\{f_{i,j}(k)\}$, $\{f_{i,j}^{s'}(k)\}$, $\{v_{i,j}(k)\}$, $\{v_{i,j}^{s'}(k)\}$, $\{n_{i,j}(k)\}$, $\{h_{i,j}(k)\}$ and $\{w_{i,j}(k)\}$. However, the proposed problem is hard to solve due to the nonlinearity in Equations 3.19 and Equations 3.22 to 3.23.

We note that if the headway $\{h_{i,j}(k)\}$ is fixed, then the variables are $\{\rho_{i,j}(k)\}$, $\{\rho_{i,j}^{s'}(k)\}$, $\{q_{i,j}^{\mathcal{D}}(k)\}$, $\{q_{i,j}^{\mathcal{D},s'}(k)\}$, $\{q_{i,j}^{\mathcal{U}}(k)\}$, $\{q_{i,j}^{\mathcal{U},s'}(k)\}$, $\{u_{i,j}(k)\}$, $\{u_{i,j}^{s'}(k)\}$, $\{f_{i,j}(k)\}$, $\{f_{i,j}^{s'}(k)\}$, $\{v_{i,j}(k)\}$, $\{v_{i,j}^{s'}(k)\}$, $\{n_{i,j}(k)\}$ and $\{w_{i,j}(k)\}$. Hence the proposed formulation becomes a linear programming program (LP).

For simplicity in notations, we use a generalized form of mixed integer linear programming to represent the formulation, as shown in Equation A.7.

$$\begin{aligned} \min_{\mathbf{x}} \quad & TTT = \mathbf{P}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}(\mathbf{h})\mathbf{x} = \mathbf{B}(\mathbf{h}) \\ & \mathbf{C}(\mathbf{h})\mathbf{x} \leq \mathbf{D}(\mathbf{h}) \end{aligned} \tag{A.7}$$

where we have:

- $\mathbf{x}$ is the flow dynamics variables and the dimension of $\mathbf{x}$ is n .
- $\mathbf{h} = \{h_{i,j}(k) \mid (i,j) \in E \setminus (L_R \cup L_S); 1 \leq k \leq N\} \in \mathfrak{R}^m$, and the dimension of parameters $\mathbf{h}$ is m .

$$\bullet \mathbf{A}(\mathbf{h}) = \begin{bmatrix} a_1(\mathbf{h}) \\ \vdots \\ a_{l_1}(\mathbf{h}) \end{bmatrix} = \begin{bmatrix} a_{11}(\mathbf{h}) & \cdots & a_{1n}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ a_{l_1 1}(\mathbf{h}) & \cdots & a_{l_1 n}(\mathbf{h}) \end{bmatrix} \in \mathfrak{R}^{l_1 \times n}; \quad \mathbf{B}(\mathbf{h}) = \begin{bmatrix} b_1(\mathbf{h}) \\ \vdots \\ b_{l_1}(\mathbf{h}) \end{bmatrix} \in \mathfrak{R}^{l_1}.$$

l_1 represents the number of equality constraints.

$$\bullet \mathbf{C}(\mathbf{h}) = \begin{bmatrix} c_1(\mathbf{h}) \\ \vdots \\ c_{l_2}(\mathbf{h}) \end{bmatrix} = \begin{bmatrix} c_{11}(\mathbf{h}) & \cdots & c_{1n}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ c_{l_2 1}(\mathbf{h}) & \cdots & c_{l_2 n}(\mathbf{h}) \end{bmatrix} \in \mathfrak{R}^{l_2 \times n}; \quad \mathbf{D}(\mathbf{h}) = \begin{bmatrix} d_1(\mathbf{h}) \\ \vdots \\ d_{l_2}(\mathbf{h}) \end{bmatrix} \in \mathfrak{R}^{l_2}.$$

l_2 represents the number of inequality constraints.

A.5.2 Sensitivity-based optimal headway control solution algorithm

We then present the sensitivity analysis for the developed LP, and the focus is to derive the gradient of TTT with respect to the headway. The Lagrange multiplier of Equation A.7 is shown in Equation A.8 and the corresponding Karush–Kuhn–Tucker (KKT) conditions are presented from Equations A.9 to A.13.

$$\min_{\mathbf{x}} \quad L = \mathbf{P}^T \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{A}(\mathbf{h})\mathbf{x} - \mathbf{B}(\mathbf{h})) + \boldsymbol{\mu}^T (\mathbf{C}(\mathbf{h})\mathbf{x} - \mathbf{D}(\mathbf{h})) \quad (\text{A.8})$$

$$\mathbf{P}^T + \boldsymbol{\lambda}^{*T} \mathbf{A}(\mathbf{h}) + \boldsymbol{\mu}^{*T} \mathbf{C}(\mathbf{h}) = 0 \quad (\text{A.9})$$

$$\mathbf{A}(\mathbf{h})\mathbf{x}^* = \mathbf{B}(\mathbf{h}) \quad (\text{A.10})$$

$$\mathbf{C}(\mathbf{h})\mathbf{x}^* \leq \mathbf{D}(\mathbf{h}) \quad (\text{A.11})$$

$$\boldsymbol{\mu}_i^* (\mathbf{C}(\mathbf{h})\mathbf{x}^* - \mathbf{D}(\mathbf{h}))_i = 0 \quad (\text{A.12})$$

$$\boldsymbol{\mu}^* \geq 0 \quad (\text{A.13})$$

where,

- $\boldsymbol{\lambda}^* \in \Re^{l_1}$ and $\boldsymbol{\mu}^* \in \Re^{l_2}$.
- $K = \{i \mid \mu_i^* > 0; 1 \leq i \leq l_2\}$ is the set of index of all positive μ_i^* . And the size of K is k . If $i \in K$, then $\mu_i^* > 0$.

Then we do the differential to the Equation A.7 as $\mathbf{z} = \mathbf{P}^T \mathbf{X}$ and Equations A.9 to A.12.

$$\mathbf{P}^T d\mathbf{x} - d\mathbf{z} = 0 \quad (\text{A.14})$$

$$\left(\sum_{i=1}^{l_1} \lambda_i^* \nabla_{\mathbf{h}} a_i(\mathbf{h}) + \sum_{j=1}^{l_2} \mu_j^* \nabla_{\mathbf{h}} c_j(\mathbf{h}) \right) d\mathbf{h} + (\mathbf{A}(\mathbf{h}))^T d\lambda + (\mathbf{C}(\mathbf{h}))^T d\mu = 0 \quad (\text{A.15})$$

$$\mathbf{A}(\mathbf{h}) d\mathbf{x} + [\nabla_{\mathbf{h}} (\mathbf{A}(\mathbf{h}) \mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{B}(\mathbf{h})] d\mathbf{h} = 0 \quad (\text{A.16})$$

$$\mathbf{C}_{\mathbf{k}}(\mathbf{h}) d\mathbf{x} + [\nabla_{\mathbf{h}} (\mathbf{C}_{\mathbf{k}}(\mathbf{h}) \mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{D}_{\mathbf{k}}(\mathbf{h})] d\mathbf{h} = 0 \quad (\text{A.17})$$

where,

$$\bullet \nabla_{\mathbf{h}} a_i(\mathbf{h}) = \nabla_{\mathbf{h}} \begin{bmatrix} a_{i1}(\mathbf{h}) & \cdots & a_{in}(\mathbf{h}) \end{bmatrix} = \begin{bmatrix} \frac{\partial a_{i1}(\mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial a_{i1}(\mathbf{h})}{\partial \mathbf{h}_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{in}(\mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial a_{in}(\mathbf{h})}{\partial \mathbf{h}_m} \end{bmatrix} \in \mathfrak{R}^{n \times m}$$

$$\bullet \nabla_{\mathbf{h}} c_j(\mathbf{h}) = \nabla_{\mathbf{h}} \begin{bmatrix} c_{j1}(\mathbf{h}) & \cdots & c_{jn}(\mathbf{h}) \end{bmatrix} = \begin{bmatrix} \frac{\partial c_{j1}(\mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial c_{j1}(\mathbf{h})}{\partial \mathbf{h}_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial c_{jn}(\mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial c_{jn}(\mathbf{h})}{\partial \mathbf{h}_m} \end{bmatrix} \in \mathfrak{R}^{n \times m}$$

$$\bullet \mathbf{A}^x = \mathbf{A}(\mathbf{h}) \mathbf{x}^* = \begin{bmatrix} a_1^x(x^*, \mathbf{h}) \\ \vdots \\ a_{l_1}^x(x^*, \mathbf{h}) \end{bmatrix} \in \mathfrak{R}^{l_1}; \nabla_{\mathbf{h}} \mathbf{A}^x = \begin{bmatrix} \frac{\partial a_1^x(x^*, \mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial a_1^x(x^*, \mathbf{h})}{\partial \mathbf{h}_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{l_1}^x(x^*, \mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial a_{l_1}^x(x^*, \mathbf{h})}{\partial \mathbf{h}_m} \end{bmatrix} \in \mathfrak{R}^{l_1 \times m}$$

$$\bullet \mathbf{C}_{\mathbf{k}}(\mathbf{h}) = \begin{bmatrix} \vdots \\ c_j(\mathbf{h}) \\ \vdots \end{bmatrix}_{j \in K} \in \mathfrak{R}^{k \times n}; \quad \mathbf{C}_{\mathbf{k}}^x = \mathbf{C}_{\mathbf{k}}(\mathbf{h}) \mathbf{x}^* = \begin{bmatrix} \vdots \\ c_j^x(x^*, \mathbf{h}) \\ \vdots \end{bmatrix}_{j \in K} \in \mathfrak{R}^k; \quad \mathbf{D}_{\mathbf{k}}(\mathbf{h}) =$$

$$\begin{bmatrix} \vdots \\ d_j(\mathbf{h}) \\ \vdots \end{bmatrix}_{j \in K} \in \mathfrak{R}^k$$

$$\bullet \nabla_{\mathbf{h}} \mathbf{C}_{\mathbf{k}}^x = \begin{bmatrix} \vdots & \cdots & \vdots \\ \frac{\partial c_j^x(x^*, \mathbf{h})}{\partial \mathbf{h}_1} & \cdots & \frac{\partial c_j^x(x^*, \mathbf{h})}{\partial \mathbf{h}_m} \\ \vdots & \cdots & \vdots \end{bmatrix}_{j \in K} \in \mathfrak{R}^{k \times m}$$

We cast Equations A.14 to A.17 into the form of block matrices.

$$\begin{bmatrix} \mathbf{P}^T & 0 & 0 & -1 \\ 0 & (\mathbf{A}(\mathbf{h}))^T & (\mathbf{C}(\mathbf{h}))^T & 0 \\ \mathbf{A}(\mathbf{h}) & 0 & 0 & 0 \\ \mathbf{C}_k(\mathbf{h}) & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} d\mathbf{x} \\ d\lambda \\ d\mu \\ d\mathbf{z} \end{bmatrix} = - \begin{bmatrix} 0 \\ \sum_{i=1}^{l_1} \lambda_i^* \nabla_{\mathbf{h}} a_i(\mathbf{h}) + \sum_{j=1}^{l_2} \mu_j^* \nabla_{\mathbf{h}} c_j(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{A}(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{B}(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{C}_k(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{D}_k(\mathbf{h}) \end{bmatrix} d\mathbf{h} \quad (\text{A.18})$$

We further define

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P}^T & 0 & 0 & -1 \\ 0 & (\mathbf{A}(\mathbf{h}))^T & (\mathbf{C}(\mathbf{h}))^T & 0 \\ \mathbf{A}(\mathbf{h}) & 0 & 0 & 0 \\ \mathbf{C}_k(\mathbf{h}) & 0 & 0 & 0 \end{bmatrix}. \quad (\text{A.19})$$

The dimension of matrix $\mathbf{Q}$ is $(n + l_1 + k + 1) \times (n + l_1 + l_2 + 1)$. If there exists $\mu_i^* = 0$, then the matrix $\mathbf{Q}$ is not a square matrix and hence not invertible. So we calculate the generalized inverse matrix of $\mathbf{Q}$ and denote it as $\mathbf{Q}^{-1}$ (Penrose, 1955).

$$\begin{bmatrix} d\mathbf{x} \\ d\lambda \\ d\mu \\ d\mathbf{z} \end{bmatrix} = -\mathbf{Q}^{-1} \begin{bmatrix} 0 \\ \sum_{i=1}^{l_1} \lambda_i^* \nabla_{\mathbf{h}} a_i(\mathbf{h}) + \sum_{j=1}^{l_2} \mu_j^* \nabla_{\mathbf{h}} c_j(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{A}(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{B}(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{C}_k(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{D}_k(\mathbf{h}) \end{bmatrix} d\mathbf{h} \quad (\text{A.20})$$

Then we replace $d\mathbf{h}$ with the identity matrix I_m and derive the gradient $\frac{\partial \mathbf{z}}{\partial \mathbf{h}}$.

$$\begin{bmatrix} \frac{\partial \mathbf{x}}{\partial \mathbf{h}} \\ \frac{\partial \lambda}{\partial \mathbf{h}} \\ \frac{\partial \mu}{\partial \mathbf{h}} \\ \frac{\partial \mathbf{z}}{\partial \mathbf{h}} \end{bmatrix} = -\mathbf{Q}^{-1} \begin{bmatrix} 0 \\ \sum_{i=1}^{l_1} \lambda_i^* \nabla_{\mathbf{h}} a_i(\mathbf{h}) + \sum_{j=1}^{l_2} \mu_j^* \nabla_{\mathbf{h}} c_j(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{A}(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{B}(\mathbf{h}) \\ \nabla_{\mathbf{h}} (\mathbf{C}_k(\mathbf{h})\mathbf{x}^*) - \nabla_{\mathbf{h}} \mathbf{D}_k(\mathbf{h}) \end{bmatrix} \quad (\text{A.21})$$

Therefore, a sensitivity-based optimal headway control algorithm is proposed based on exploring the gradient descent of TTT, as shown in Algorithm 6.

Algorithm 6 Sensitivity-based Optimal Headway Solving Algorithm

Input: learning rate η .

Output: time-dependent and link-specific headway $\{h_{i,j}(k)\}$.

Initialize the headway $\mathbf{h}$.

- 1: **for** ($i = 0; i < I; ++i$) **do**
- 2: Solve the Formulation A.7 given the current headway setting $\mathbf{h}$ and derive $\mathbf{z}^*, \lambda^*, \mu^*$ and $\mathbf{x}^*$.
- 3: Conduct the sensitivity analysis and derive $\frac{\partial \mathbf{z}^*}{\partial \mathbf{h}}$.
- 4: Update the headway by $\mathbf{h} = \mathbf{h} - \eta \frac{\partial \mathbf{z}^*}{\partial \mathbf{h}}$.
- 5: **end for**

Return: optimal headway $\mathbf{h}^*$.

A.6 Proof of Proposition 4

We first present Lemma 2 to show that a smaller headway $h_{i,j}(k)$ could generate a larger flow $f_{i,j}(k)$.

Lemma 2. Given $u_{i,j}(k)$ and $\rho_{i,j}(k-1)$, if $h_{i,j}^{(1)}(k) \leq h_{i,j}^{(2)}(k)$, then $f_{i,j}^{(1)}(k) \geq f_{i,j}^{(2)}(k)$.

Proof. We solve $f_{i,j}(k)$ and $\rho_{i,j}(k)$ under given $u_{i,j}(k)$ and $\rho_{i,j}(k-1)$ in both free-flow and congested states:

- Free-flow state:

$$\begin{aligned} f_{i,j}(k) &= v_{i,j}^f \rho_{i,j}(k) \\ \rho_{i,j}(k) &= \rho_{i,j}(k-1) + \Delta_t \frac{u_{i,j}(k) - f_{i,j}(k)}{L_{i,j}} \end{aligned}$$

Then we obtain the following equations:

$$\begin{aligned} f_{i,j}(k) &= \frac{L_{i,j} v_{i,j}^f \rho_{i,j}(k-1) + \Delta_t v_{i,j}^f u_{i,j}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \\ \rho_{i,j}(k) &= \frac{L_{i,j} \rho_{i,j}(k-1) + \Delta_t u_{i,j}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \end{aligned}$$

- Congested state:

$$f_{i,j}(k) = \frac{1 - \rho_{i,j}(k) L}{h_{i,j}(k)}$$

$$\rho_{i,j}(k) = \rho_{i,j}(k-1) + \Delta_t \frac{u_{i,j}(k) - f_{i,j}(k)}{L_{i,j}}$$

Then the two quantities can be derived as:

$$f_{i,j}(k) = \frac{L_{i,j} - L L_{i,j} \rho_{i,j}(k-1) - L \Delta_t u_{i,j}(k)}{L_{i,j} h_{i,j}(k) - L \Delta_t}$$

$$\rho_{i,j}(k) = \frac{h_{i,j}(k) L_{i,j} \rho_{i,j}(k-1) + h_{i,j}(k) \Delta_t u_{i,j}(k) - \Delta_t}{L_{i,j} h_{i,j}(k) - L \Delta_t}$$

Next, We will discuss different states of $\rho_{i,j}^{(2)}(k)$:

- If $\rho_{i,j}^{(2)}(k)$ is under free-flow state: Because $h_{i,j}^{(1)}(k) \leq h_{i,j}^{(2)}(k)$, as is shown in Figure 3.4, it is obvious that $f_{i,j}^{(1)}(k) = f_{i,j}^{(2)}(k)$.
- If $\rho_{i,j}^{(2)}(k)$ is under congested state:

- The density in the free-flow state should be infeasible:

$$\rho_{i,j}^{(2)}(k) = \frac{L_{i,j} \rho_{i,j}^{(2)}(k-1) + \Delta_t u_{i,j}^{(2)}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \geq \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} \quad (\text{A.22})$$

- The density in the congested state should be feasible:

$$\rho_{i,j}^{(2)}(k) = \frac{h_{i,j}^{(2)}(k) L_{i,j} \rho_{i,j}^{(2)}(k-1) + h_{i,j}^{(2)}(k) \Delta_t u_{i,j}^{(2)}(k) - \Delta_t}{L_{i,j} h_{i,j}^{(2)}(k) - L \Delta_t} \geq \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} \quad (\text{A.23})$$

For the free-flow state, if $L_{i,j} h_{i,j}^{(2)}(k) - L \Delta_t < 0$, we have

$$\begin{aligned}\rho_{i,j}^{(2)}(k) &= \frac{h_{i,j}^{(2)}(k) L_{i,j} \rho_{i,j}^{(2)}(k-1) + h_{i,j}^{(2)}(k) \Delta_t u_{i,j}^{(2)}(k) - \Delta_t}{L_{i,j} h_{i,j}^{(2)}(k) - L \Delta_t} \\ &\leq \frac{\frac{L_{i,j} + \Delta_t v_{i,j}^f}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} h_{i,j}^{(2)}(k) - \Delta_t}{L_{i,j} h_{i,j}^{(2)}(k) - L \Delta_t} = \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L}\end{aligned}$$

It is contradicted with $\rho_{i,j}^{(2)}(k) \geq \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L}$. Therefore, $L_{i,j} h_{i,j}^{(2)}(k) - L \Delta_t > 0$. It is also similar to prove that if $\rho_{i,j}(k)$ is congested under the headway $h_{i,j}(k)$, then $L_{i,j} h_{i,j}(k) - L \Delta_t > 0$.

Next, we discuss the states of $\rho_{i,j}^{(1)}(k)$.

* $\rho_{i,j}^{(1)}(k)$ is in congested state, then we have

$$\rho_{i,j}^{(1)}(k) = \frac{L_{i,j} \rho_{i,j}^{(2)}(k-1) + \Delta_t u_{i,j}^{(2)}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \geq \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} \geq \frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L}$$

The congested state of $\rho_{i,j}^{(1)}(k)$ can be calculated as follows:

$$\begin{aligned}\rho_{i,j}^{(1)}(k) &= \frac{h_{i,j}^{(1)}(k) L_{i,j} \rho_{i,j}^{(2)}(k-1) + h_{i,j}^{(2)}(k) \Delta_t u_{i,j}^{(2)}(k) - \Delta_t}{L_{i,j} h_{i,j}^{(1)}(k) - L \Delta_t} \\ &\geq \frac{\frac{L_{i,j} + \Delta_t v_{i,j}^f}{h_{i,j}^{(1)}(k) v_{i,j}^f + L} h_{i,j}^{(1)}(k) - \Delta_t}{L_{i,j} h_{i,j}^{(1)}(k) - L \Delta_t} = \frac{1}{h_{i,j}^{(1)}(k) v_{i,j}^f + L}\end{aligned}$$

The flow $f_{i,j}(k) = \frac{L_{i,j} - L_{i,j} \rho_{i,j}^{(2)}(k-1) - L \Delta_t u_{i,j}^{(2)}(k)}{L_{i,j} h_{i,j}^{(1)}(k) - L \Delta_t}$ and we have proved that both numerator and denominator are positive, so if $h_{i,j}^{(1)}(k) \leq h_{i,j}^{(2)}(k)$, we have $f_{i,j}^{(1)}(k) \geq f_{i,j}^{(2)}(k)$.

* $\rho_{i,j}^{(1)}(k)$ is in the free-flow state, then the free-flow state of $\rho_{i,j}^{(1)}(k)$ should be feasible, as shown in the following equation.

$$\frac{1}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} \leq \rho_{i,j}^{(1)}(k) = \frac{L_{i,j} \rho_{i,j}^{(2)}(k-1) + \Delta_t u_{i,j}^{(2)}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \leq \frac{1}{h_{i,j}^{(1)}(k) v_{i,j}^f + L}$$

Therefore, for the flow $f_{i,j}^{(1)}(k)$, we have:

$$f_{i,j}^{(1)}(k) = \frac{L_{i,j} v_{i,j}^f \rho_{i,j}^{(2)}(k-1) + \Delta_t v_{i,j}^f u_{i,j}^{(2)}(k)}{L_{i,j} + \Delta_t v_{i,j}^f} \geq \frac{v_{i,j}^f}{h_{i,j}^{(2)}(k) v_{i,j}^f + L} \geq f_{i,j}^{(2)}(k)$$

Combining the above proofs for the two states, we have $f_{i,j}^{(1)}(k) \geq f_{i,j}^{(2)}(k)$ hold. $\square$

Then we develop Lemma 3 to prove that the assumption of Lemma 2 is reachable, which means that when $h_{i,j}^1(k) \leq h_{i,j}^2(k)$, we could always find the feasible $u_{i,j}(k)$ and $\rho_{i,j}(k-1)$ for both $h_{i,j}^1(k)$ and $h_{i,j}^2(k)$.

Lemma 3. For $\mathbf{h}_1 = \{h_{i,j}^{(1)}(k)\}$ and $\mathbf{h}_2 = \{h_{i,j}^{(2)}(k)\}$, if there exist only one (i, j) and k such that $h_{i,j}^{(1)}(k) \leq h_{i,j}^{(2)}(k)$ and $h_{m,n}^{(1)}(l) = h_{m,n}^{(2)}(l)$ for other links $(m, n) \neq (i, j)$ or time intervals $l \neq k$, suppose $u_{i,j}^{s',(2)}(k) \in \mathbf{x}_2, \forall s'$, where $\mathbf{x}_2 \in \Omega_{\mathbf{h}_2}$, then there always exists $\mathbf{x}_1 \in \Omega_{\mathbf{h}_1}$, such that $u_{i,j}^{s',(1)}(k) \in \mathbf{x}_1$ and $u_{i,j}^{s',(1)}(k) = u_{i,j}^{s',(2)}(k)$.

Proof. In addition to the flow $f_{i,j}(k)$, headway affects the wave travel time in congested states $n_{i,j}^w(k)$ as follows:

$$\begin{aligned} n_{i,j}^w(k) \Delta_t \frac{L}{h_{i,j}(k)} &\leq L_{i,j}, \\ (n_{i,j}^w(k) + 1) \Delta_t \frac{L}{h_{i,j}(k)} &\geq L_{i,j}. \end{aligned}$$

To show the existence of a feasible $\mathbf{x}_1$ such that $u_{i,j}^{s',(1)}(k) = u_{i,j}^{s',(2)}(k)$, we need to show that $u_{i,j}^{s',(2)}(k)$ satisfies the upstream queue capacity under $h_{i,j}^{(1)}(k)$ and other related variables in $\mathbf{x}_2$ if we set $u_{i,j}^{s',(1)}(l) = u_{i,j}^{s',(2)}(l), \forall 1 \leq l \leq k-1$. Therefore, we could also derive $f_{i,j}^{s',(1)}(l) = f_{i,j}^{s',(2)}(l), \forall 1 \leq l \leq k-1$ by both the proof of Lemma 2 and the assumption that $h_{i,j}^{(1)}(l) = h_{i,j}^{(2)}(l), \forall 1 \leq l \leq k-1$.

If $h_{i,j}^{(1)}(k) \leq h_{i,j}^{(2)}(k)$, then $n_{i,j}^{w,(1)}(k) \leq n_{i,j}^{w,(2)}(k)$, which implies that if we use a smaller the headway in link (i, j) at the k_{th} time interval, the wave travel time in congested states would be smaller. By Equation 3.25, we have $q_{i,j}^{u,s',(1)}(k) \leq q_{i,j}^{u,s',(2)}(k)$. In order to keep $u_{i,j}^{s',(2)}(k)$ unchanged under $h_{i,j}^{(1)}(k)$, we next prove $q_{i,j}^{u,s',(1)}(k) \geq 0$.

$$\begin{aligned}
q_{i,j}^{u,s',(1)}(k) &= \sum_{l=0}^k \Delta_t u_{i,j}^{s',(1)}(l) - \sum_{l=0}^{k-n_{i,j}^{w,(1)}(k)} \Delta_t f_{i,j}^{s',(1)}(l) \\
&= \sum_{l=0}^k \Delta_t u_{i,j}^{s',(2)}(l) - \sum_{l=0}^{k-n_{i,j}^{w,(1)}(k)} \Delta_t f_{i,j}^{s',(1)}(l) \\
&\geq \sum_{l=0}^k \Delta_t u_{i,j}^{s',(2)}(l) - \sum_{l=0}^{k-1} \Delta_t f_{i,j}^{s',(1)}(l) \\
&= \sum_{l=0}^k \Delta_t u_{i,j}^{s',(2)}(l) - \sum_{l=0}^{k-1} \Delta_t f_{i,j}^{s',(2)}(l) \\
&= L_{i,j} \rho_{i,j}^{s',(2)}(k-1) + \Delta_t u_{i,j}^{s',(2)}(k) \geq 0
\end{aligned}$$

□

By Lemma 2 and 3, for link (j, s) , where $s \in S$, if $h_{j,s}^{(1)}(k) \leq h_{j,s}^{(2)}(k)$, $h_{j,s}^{(1)}(l) = h_{j,s}^{(2)}(l)$, $\forall l \neq k$ and we set $u_{j,s}^{(1)}(l) = u_{j,s}^{(2)}(l)$ for all $1 \leq l \leq N$, we have $f_{j,s}^{(1)}(k) \geq f_{j,s}^{(2)}(k)$. For density, we have the following equation hold:

$$\rho_{j,s}(k) = \rho_{j,s}(k-1) + \Delta_t \frac{u_{j,s}(k) - f_{j,s}(k)}{L_{j,s}}. \quad (\text{A.24})$$

If $\rho_{j,s}^{(1)}(k-1) = \rho_{j,s}^{(2)}(k-1)$, then we have $\rho_{i,j}^{(1)}(k) \leq \rho_{i,j}^{(2)}(k)$. If $\rho_{i,j}^{(2)}(k+1)$ is in the free-flow state, the following equations hold in the proof of Lemma 2 for $\rho_{j,s}^{(2)}(k+1)$ and $f_{j,s}^{(2)}(k+1)$:

$$\begin{aligned}
f_{j,s}(k) &= \frac{L_{j,s} v_{j,s}^f \rho_{j,s}(k-1) + \Delta_t v_{j,s}^f u_{j,s}(k)}{L_{j,s} + \Delta_t v_{j,s}^f} \\
\rho_{j,s}(k) &= \frac{L_{j,s} \rho_{j,s}(k-1) + \Delta_t u_{j,s}(k)}{L_{j,s} + \Delta_t v_{j,s}^f}
\end{aligned}$$

So we have $f_{i,j}^{(1)}(k+1) \leq f_{i,j}^{(2)}(k+1)$ and $\rho_{i,j}^{(1)}(k+1) \leq \rho_{i,j}^{(2)}(k+1)$ if $\rho_{i,j}^{(2)}(k+1)$ hold in the free-flow state. Then if $\rho_{i,j}^{(2)}(k+1)$ is in the congested state, we have the following equations in the proof of Lemma 2 hold for $\rho_{j,s}^{(2)}(k+1)$ and $f_{j,s}^{(2)}(k+1)$:

$$\begin{aligned}
f_{j,s}(k) &= \frac{L_{j,s} - L L_{j,s} \rho_{j,s}(k-1) - L \Delta_t u_{j,s}(k)}{L_{j,s} h_{j,s}(k) - L \Delta_t} \\
\rho_{j,s}(k) &= \frac{h_{j,s}(k) L_{j,s} \rho_{j,s}(k-1) + h_{j,s}(k) \Delta_t u_{j,s}(k) - \Delta_t}{L_{j,s} h_{j,s}(k) - L \Delta_t}
\end{aligned}$$

Similarly, $f_{i,j}^{(1)}(k+1) \geq f_{i,j}^{(2)}(k+1)$ and $\rho_{i,j}^{(1)}(k+1) \leq \rho_{i,j}^{(2)}(k+1)$ hold for $\rho_{i,j}^{(2)}(k+1)$ in the congested state.

Therefore, we have $\rho_{i,j}^{(1)}(k+1) \leq \rho_{i,j}^{(2)}(k+1)$. We conduct the same procedure for $k+2, \dots, N$ and $\rho_{i,j}^{(1)}(l) \leq \rho_{i,j}^{(2)}(l)$ is derived for $l \geq k$ and $\rho_{i,j}^{(1)}(l) = \rho_{i,j}^{(2)}(l)$ for $l < k$. Besides, by $F_{j,s}(k) = L_{j,s} \rho_{j,s}(k) + U_{j,s}(k)$, so we have $F_{j,s}^{(1)}(l) = F_{j,s}^{(2)}(l)$ for $l < k$ and $F_{j,s}^{(1)}(l) \geq F_{j,s}^{(2)}(l)$ for $l \geq k$.

Similarly, we could also get the $\rho_{j,s}^{s'_0, (1)}(l) \leq \rho_{j,s}^{s'_0, (2)}(l)$ for $l \geq k$ and $\rho_{i,j}^{s'_0, (1)}(l) = \rho_{i,j}^{s'_0, (2)}(l)$ for $l < k$ by the same procedure for any destination s_0 . Besides, for any destination s_0 , we also have $F_{j,s}^{s'_0, (1)}(k) = F_{j,s}^{s'_0, (2)}(k)$ for $l < k$ and $F_{j,s}^{s'_0, (1)}(k) \geq F_{j,s}^{s'_0, (2)}(k)$ for $l \geq k$.

Therefore, for destination $s_0 \neq s$, we could set $v_{j,s}^{s'_0, (1)}(l) = v_{j,s}^{s'_0, (2)}(l), \forall 1 \leq l \leq N$. For destination s , we have $V_{j,s}^{s', (1)}(k) = V_{j,s}^{s', (2)}(k)$ for $l < k$ and $V_{j,s}^{s', (1)}(k) \geq V_{j,s}^{s', (2)}(k)$ for $l \geq k$, where $V_{j,s}^{s'}(k) = \sum_{l=1}^k \Delta_t v_{j,s}^{s'}(l)$ represents the cumulative outflow with destination s' of link (j, s) in the k th time interval. Then the total travel time in link (j, s) can be written as follows:

$$\begin{aligned}
TTT_{j,s} &= \sum_{k=1}^N \Delta_t \sum_{l=1}^k \Delta_t u_{j,s}(l) - v_{j,s}(l) \\
&= \sum_{k=1}^N \sum_{s_0} \Delta_t \Delta_t k u_{j,s}^{s'_0}(k) - \Delta_t \Delta_t k v_{j,s}^{s'_0}(k) \\
&= \sum_{k=1}^N \sum_{s_0 \neq s} \left(\Delta_t \Delta_t k u_{j,s}^{s'_0}(k) - \Delta_t \Delta_t k v_{j,s}^{s'_0}(k) \right) + \sum_{k=1}^N \left(\Delta_t \Delta_t k u_{j,s}^{s'}(k) - \Delta_t \Delta_t k v_{j,s}^{s'}(k) \right) \\
&= \sum_{k=1}^N \sum_{s_0 \neq s} \left(\Delta_t \Delta_t k u_{j,s}^{s'_0}(k) - \Delta_t \Delta_t k v_{j,s}^{s'_0}(k) \right) + \sum_{k=1}^N \Delta_t \Delta_t k u_{j,s}^{s'}(k) - \sum_{k=1}^N \Delta_t V_{j,s}^{s'}(k)
\end{aligned}$$

Therefore, we have proved that $u_{j,s}^{s'_0, (1)}(l) = u_{j,s}^{s'_0, (2)}(l), v_{j,s}^{s'_0, (1)}(l) = v_{j,s}^{s'_0, (2)}(l)$ for $s_0 \neq s$ and $1 \leq l \leq N$. For destination s , we have $u_{j,s}^{s', (1)}(l) = u_{j,s}^{s', (2)}(k)$ and $V_{j,s}^{s', (1)}(l) \geq V_{j,s}^{s', (2)}(l)$ for $1 \leq l \leq N$. Then we have

$$TTT_{j,s}^{(1)} \leq TTT_{j,s}^{(2)}.$$

The total travel time consists of the total travel time in link (j, s) , total travel time in other links and waiting time at origins. The inflow and outflow variables in other links and origins are not affected by the change of headway from $h_{i,j}^{(1)}(k)$ to $h_{i,j}^{(2)}(k)$, so $TTT_{\mathbf{h}_1} \leq TTT_{\mathbf{h}_2}$ if $h_{j,s}^{(1)}(k) \leq h_{j,s}^{(2)}(k)$. For link (i, j) , the node j acts like the destination node, and we want to push more flow with destination s to leave link (i, j) and enter link (j, s) . The process is similar to what we have done for link (j, s) by only changing one headway in link (i, j) and keeping flow variables in other links unchanged. By repeating the above procedure for all the links and time intervals, we can show that TTT could be minimized under the minimum headway.

A.7 Proof of Proposition 5

Suppose $\mathbf{x}^*$ and $\mathbf{h}^*$ are the optimal solution of the proposed SO-DTA problem. If $\exists (i, j) \in E \setminus (L_R \cup L_S)$ and $1 \leq k \leq N$, s.t. $v_{i,j}^f \rho_{i,j}^*(k) < \frac{1-\rho_{i,j}^*(k)L}{h_{i,j}^*(k)}$, then $\mathbf{h}^*$ and $\mathbf{x}^*$ should satisfy the Equations 3.22, 3.23 and A.25 as follows.

$$\rho_{i,j}^*(k) \leq \frac{1}{h_{i,j}^*(k)v_{i,j}^f + L} \quad (\text{A.25})$$

If $h_{i,j;k}^{\min} \leq h_{i,j}^*(k) < h_{i,j;k}^{\max}$, we could always find an extreme small constant c such that $ch_{i,j}^*(k)$ also satisfy the Equations 3.22, 3.23 and A.25 such that we find another optimal headway. If $h_{i,j}^*(k) = h_{i,j;k}^{\max}$, by Proposition 4, minimum headway setting achieves SO-DTA, so we show that SO-DTA can have at least two optimal headway that achieves the same TTT.

A.8 Proof of Proposition 6

Suppose $\mathbf{x}^*$ with $TTT(\mathbf{x}^*) = TTT^*$ is unique, if $\hat{\mathbf{h}}$ and $\bar{\mathbf{h}}$ are different optimal solutions of Formulation 3.37, then there must exist $(i_1, j_1), (i_2, j_2) \in E$ and $1 \leq k_1, k_2 \leq N$ such that $\hat{h}_{i_1,j_1}(k_1) < \bar{h}_{i_1,j_1}(k_1)$ and $\hat{h}_{i_2,j_2}(k_2) > \bar{h}_{i_2,j_2}(k_2)$. By the uniqueness of $\mathbf{x}^*$, we have $v_{i_1,j_1}^f \rho_{i_1,j_1}^*(k_1) < \frac{1-\rho_{i_1,j_1}^*(k_1)L}{h_{i_1,j_1}^*(k_1)}$ and $v_{i_2,j_2}^f \rho_{i_2,j_2}^*(k_2) < \frac{1-\rho_{i_2,j_2}^*(k_2)L}{h_{i_2,j_2}^*(k_2)}$.

Therefore, if we construct a new headway $\mathbf{h}$ as follows:

$$h_{i,j}(k) = \begin{cases} \hat{h}_{i,j}(k), & (i,j) \neq (i_1, j_1) \text{ or } k \neq k_1 \\ \bar{h}_{i,j}(k), & (i,j) = (i_1, j_1) \text{ and } k = k_1 \end{cases} \quad (\text{A.26})$$

We note that $\mathbf{h}$ is the optimal headway of the discretized SO-DTA formulation (Equations 3.18 to 3.36) as $\mathbf{x}^* \in \Omega_{\mathbf{h}}$. However, we also have $\|\mathbf{h}\|_1 > \|\hat{\mathbf{h}}\|_1$, and this contradicts to the assumption that $\hat{\mathbf{h}}$ is the solution of Formulation 3.37, which completes the proof.

A.9 Settings of the small network

For simplicity, the inflow capacity is set equal to the outflow capacity and the downstream queue capacity is set equal to the upstream queue capacity in each link. The detailed configurations of each link are presented in Table A.2.

Link	Flow capacity (vehicles/min)	Queue capacity (vehicles)	Free flow speed (km/min)	Link length (km)
1 → 3	50	600	0.9	1.6
1 → 4	45	600	1.0	1.2
2 → 3	45	600	1.2	3.6
2 → 4	50	600	1.1	3.3
3 → 5	60	600	1.0	4.0
4 → 5	60	600	1.0	3.0

Table A.2: Parameters settings in the small network.

The settings of minimum headway and maximum headway for each link are illustrated in Table A.3 and Table A.4, respectively.

A.10 Settings of the Hong Kong network

We introduce the geographical positioning of each origin-destination (O-D) pair and detail their mean demand profiles within the first hour, as showcased in Figure A.1. The color saturation reflects the magnitude of the O-D demand, with a darker hue denoting a greater number of trips between the corresponding O-D pairs. Details of the link parameters are presented in Table A.5.

Time Interval \ Link	Link					
	1 $\rightarrow$ 3	1 $\rightarrow$ 4	2 $\rightarrow$ 3	2 $\rightarrow$ 4	3 $\rightarrow$ 5	4 $\rightarrow$ 5
1	0.20s	0.20s	0.20s	0.20s	0.20s	0.20s
2	0.20s	0.20s	0.20s	0.20s	0.20s	0.20s
3	0.20s	0.20s	0.20s	0.20s	0.20s	0.20s
4	0.75s	0.60s	0.45s	0.55s	0.50s	0.50s
5	0.75s	0.60s	0.45s	0.55s	0.50s	0.50s
6	0.75s	0.60s	0.45s	0.55s	0.50s	0.50s
7	0.75s	0.60s	0.45s	0.55s	0.50s	0.50s
8	1.05s	1.15s	0.85s	0.75s	0.80s	0.90s
9	1.05s	1.15s	0.85s	0.75s	0.80s	0.90s
10	1.05s	1.15s	0.85s	0.75s	0.80s	0.90s
11	1.05s	1.15s	0.85s	0.75s	0.80s	0.90s
12	1.25s	1.35s	1.10s	0.95s	1.00s	1.20s
13	1.25s	1.35s	1.10s	0.95s	1.00s	1.20s
14	1.25s	1.35s	1.10s	0.95s	1.00s	1.20s
15	1.25s	1.35s	1.10s	0.95s	1.00s	1.20s
16	1.55s	1.65s	1.30s	1.25s	1.25s	1.30s
17	1.55s	1.65s	1.30s	1.25s	1.25s	1.30s
18	1.55s	1.65s	1.30s	1.25s	1.25s	1.30s

Table A.3: Settings of minimum headway.

Time Interval \ Link	Link					
	1 $\rightarrow$ 3	1 $\rightarrow$ 4	2 $\rightarrow$ 3	2 $\rightarrow$ 4	3 $\rightarrow$ 5	4 $\rightarrow$ 5
1	1.80s	1.80s	1.80s	1.80s	1.80s	1.80s
2	1.80s	1.80s	1.80s	1.80s	1.80s	1.80s
3	1.80s	1.80s	1.80s	1.80s	1.80s	1.80s
4	2.20s	2.10s	2.45s	2.20s	2.15s	2.20s
5	2.20s	2.10s	2.45s	2.20s	2.15s	2.20s
6	2.20s	2.10s	2.45s	2.20s	2.15s	2.20s
7	2.20s	2.10s	2.45s	2.20s	2.15s	2.20s
8	2.55s	2.45s	2.90s	2.75s	2.40s	2.65s
9	2.55s	2.45s	2.90s	2.75s	2.40s	2.65s
10	2.55s	2.45s	2.90s	2.75s	2.40s	2.65s
11	2.55s	2.45s	2.90s	2.75s	2.40s	2.65s
12	3.05s	2.95s	3.50s	3.15s	2.90s	3.05s
13	3.05s	2.95s	3.50s	3.15s	2.90s	3.05s
14	3.05s	2.95s	3.50s	3.15s	2.90s	3.05s
15	3.05s	2.95s	3.50s	3.15s	2.90s	3.05s
16	3.55s	3.30s	3.85s	3.65s	3.25s	3.35s
17	3.55s	3.30s	3.85s	3.65s	3.25s	3.35s
18	3.55s	3.30s	3.85s	3.65s	3.25s	3.35s

Table A.4: Results of the maximum headway.

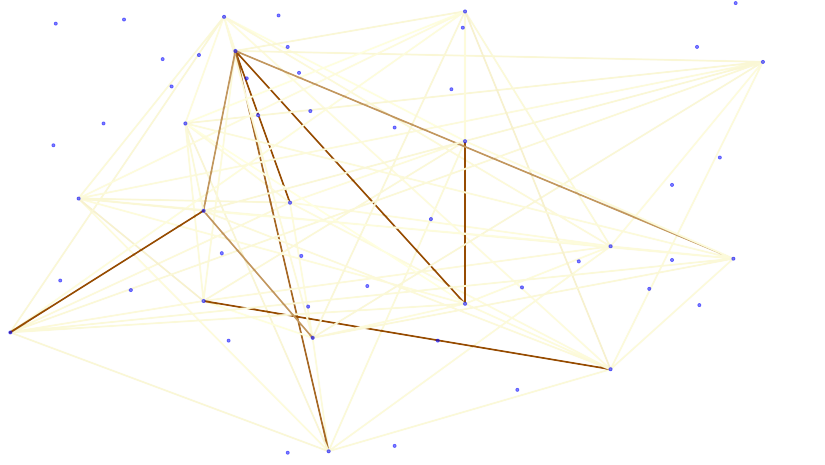


Figure A.1: The hour O-D demand profile in the Hong Kong network. A darker color indicates more trips between the corresponding O-D pair.

Link		19 → 6	6 → 19	6 → 7	7 → 6	7 → 8	8 → 7	8 → 9	9 → 8	9 → 5	5 → 9	5 → 27
Parameters												
Flow	capacity	285	465	465	450	505	190	465	475	195	465	505
	(veh/min)											
Queue capacity (veh)		4200.0	1400.0	1300.0	2800.0	1583.0	1300.0	3600.0	1300.0	2600.0	3900.0	5200.0
Free	flow											
	speed	0.232	0.232	0.37	0.463	0.231	0.232	0.463	0.37	0.463	0.232	0.232
	(km/min)											
Link length (km)		0.311	0.311	0.454	0.454	0.248	0.248	0.845	0.845	0.133	0.133	1.252
Link		9 → 27	27 → 9	1 → 19	19 → 1	1 → 2	2 → 1	1 → 6	6 → 1	2 → 6	6 → 2	2 → 3
Parameters												
Flow	capacity	285	195	475	475	505	505	190	190	505	195	195
	(veh/min)											
Queue capacity (veh)		3900.0	2600.0	1400.0	5600.0	4200.0	3166.0	1583.0	3166.0	1583.0	2800.0	2800.0
Free	flow											
	speed	0.232	0.231	0.370	0.370	0.232	0.463	0.231	0.370	0.232	0.463	0.232
	(km/min)											
Link length (km)		1.228	1.228	0.563	0.563	0.168	0.168	0.366	0.366	0.446	0.446	0.168
Link		3 → 4	4 → 3	3 → 7	7 → 3	4 → 5	5 → 4	1 → 10	10 → 1	10 → 13	13 → 10	13 → 20
Parameters												
Flow	capacity	195	465	195	465	475	380	465	380	380	190	380
	(veh/min)											
Queue capacity (veh)		2800.0	1300.0	4200.0	3166.0	1583.0	1583.0	4200.0	2800.0	3166.0	2600.0	2800.0
Free	flow											
	speed	0.232	0.232	0.463	0.463	0.463	0.370	0.463	0.370	0.232	0.185	0.463
	(km/min)											
Link length (km)		0.239	0.239	0.282	0.282	0.808	0.808	0.225	0.225	0.305	0.305	0.371

Parameters		Link 19 → 20 → 10 → 11 → 11 → 12 → 2 → 11 11 → 2 3 → 11 11 → 3 4 → 12										
		20	19	11	10	12	11					
Flow capacity (veh/min)		285	475	505	450	380	285	450	450	190	380	475
Queue capacity (veh)		3166.0	4200.0	2600.0	2250.0	3166.0	1583.0	1583.0	1583.0	1583.0	3166.0	1583.0
Free flow speed (km/min)		0.232	0.370	0.370	0.185	0.463	0.463	0.231	0.232	0.463	0.185	0.370
Link length (km)		0.837	0.837	0.346	0.346	0.241	0.241	0.287	0.287	0.227	0.227	0.217
Parameters		Link 12 → 52 → 14 → 52 → 11 → 14 → 22 → 23 → 23 → 15 → 5 → 15										
		52	12	52	14	14	11	23	22	15	23	
Flow capacity (veh/min)		195	465	505	195	450	465	380	285	380	380	380
Queue capacity (veh)		1583.0	4200.0	1583.0	2800.0	4200.0	1400.0	2800.0	4200.0	1583.0	1300.0	3166.0
Free flow speed (km/min)		0.463	0.232	0.463	0.463	0.463	0.463	0.324	0.232	0.232	0.370	0.463
Link length (km)		0.314	0.314	0.239	0.239	0.303	0.303	0.338	0.338	0.403	0.403	0.501
Parameters		Link 18 → 21 → 16 → 17 → 16 → 42 → 31 → 32 → 31 → 25 → 20 →										
		21	18	17	16	42	16	32	31	25	31	41
Flow capacity (veh/min)		190	380	190	475	505	475	505	190	450	450	475
Queue capacity (veh)		4200.0	3166.0	1583.0	1300.0	1583.0	6332.0	6332.0	1400.0	3166.0	1583.0	1300.0
Free flow speed (km/min)		0.231	0.370	0.370	0.232	0.185	0.463	0.463	0.232	0.463	0.463	0.232
Link length (km)		0.361	0.361	0.397	0.397	0.575	0.575	0.382	0.382	0.341	0.341	0.288
Parameters		Link 41 → 42 → 42 → 43 → 43 → 45 → 43 → 44 → 43 → 41 → 31 →										
		42	41	43	42	45	43	44	43	41	43	36
Flow capacity (veh/min)		475	285	195	190	505	190	380	190	285	380	190
Queue capacity (veh)		2800.0	2800.0	1400.0	1583.0	2800.0	2600.0	1583.0	4200.0	2800.0	3166.0	1300.0
Free flow speed (km/min)		0.463	0.370	0.463	0.463	0.324	0.370	0.463	0.232	0.370	0.231	0.231
Link length (km)		0.446	0.446	0.670	0.670	0.478	0.478	0.329	0.329	1.097	1.097	0.947
Parameters		Link 35 → 36 → 32 → 35 → 32 → 26 → 25 → 26 → 38 → 48 → 48 →										
		36	35	35	32	26	32	26	25	48	38	47
Flow capacity (veh/min)		475	380	190	285	505	380	195	190	505	380	475
Queue capacity (veh)		1583.0	1400.0	4749.0	1583.0	3166.0	1300.0	2800.0	1583.0	3166.0	1300.0	3600.0
Free flow speed (km/min)		0.370	0.463	0.232	0.463	0.463	0.463	0.232	0.370	0.463	0.232	0.463
Link length (km)		0.186	0.186	0.922	0.922	0.256	0.256	0.476	0.476	0.323	0.323	0.323
Parameters		Link 27 → 38 → 15 → 27 → 13 → 14 → 13 → 16 → 39 → 40 → 14 →										
		38	27	27	15	14	13	16	13	40	39	17

Flow capacity (veh/min)	450	190	450	195	465	195	190	285	195	465	190
Queue capacity (veh)	4200.0	2600.0	6332.0	4749.0	2800.0	2800.0	5600.0	2600.0	1583.0	2800.0	4200.0
Free flow speed (km/min)	0.231	0.232	0.231	0.463	0.324	0.463	0.232	0.232	0.463	0.463	0.232
Link length (km)	0.395	0.395	1.465	1.465	0.336	0.336	0.713	0.713	0.407	0.407	0.723
Parameters	Link 17 → 23	23 → 17	15 → 52	52 → 15	22 → 24	24 → 22	15 → 22	22 → 15	22 → 30	30 → 22	29 → 37
Flow capacity (veh/min)	285	450	190	285	195	190	465	190	190	465	190
Queue capacity (veh)	2800.0	5200.0	2800.0	4200.0	2800.0	3166.0	1300.0	2400.0	4749.0	2800.0	1300.0
Free flow speed (km/min)	0.463	0.463	0.463	0.463	0.463	0.463	0.232	0.463	0.370	0.463	0.232
Link length (km)	0.772	0.772	0.663	0.663	0.650	0.650	0.425	0.425	1.078	1.078	0.290
Parameters	Link 24 → 29	29 → 24	18 → 51	51 → 18	37 → 51	51 → 37	34 → 46	46 → 34	34 → 35	35 → 34	33 → 34
Flow capacity (veh/min)	380	195	450	195	450	475	380	380	380	465	505
Queue capacity (veh)	4200.0	3166.0	2800.0	4200.0	1583.0	4200.0	3900.0	2600.0	4749.0	2800.0	1400.0
Free flow speed (km/min)	0.463	0.370	0.463	0.370	0.232	0.232	0.463	0.463	0.185	0.324	0.463
Link length (km)	0.690	0.690	0.386	0.386	0.466	0.466	0.718	0.718	0.302	0.302	0.875
Parameters	Link 32 → 33	33 → 32	40 → 53	53 → 40	30 → 49	49 → 30	38 → 28	28 → 38	28 → 30	30 → 28	28 → 54
Flow capacity (veh/min)	285	285	285	285	190	380	450	195	190	475	505
Queue capacity (veh)	2600.0	3166.0	2800.0	4200.0	4200.0	3600.0	2800.0	2800.0	4749.0	1583.0	1300.0
Free flow speed (km/min)	0.232	0.463	0.463	0.232	0.232	0.232	0.463	0.463	0.463	0.463	0.463
Link length (km)	0.567	0.567	0.658	0.658	0.299	0.299	1.124	1.124	0.571	0.571	0.310
Parameters	Link 54 → 47	47 → 54	28 → 48	48 → 28	39 → 54	54 → 39	46 → 53	53 → 46	46 → 29	29 → 46	29 → 53
Flow capacity (veh/min)	285	195	505	285	380	475	195	380	465	465	380
Queue capacity (veh)	1583.0	1300.0	1583.0	1583.0	3166.0	1583.0	5600.0	2800.0	1300.0	2600.0	2800.0
Free flow speed (km/min)	0.324	0.463	0.463	0.185	0.232	0.370	0.232	0.463	0.463	0.370	0.232
Link length (km)	0.819	0.819	1.078	1.078	0.822	0.822	0.454	0.454	0.831	0.831	0.776
Parameters	Link 29 → 55	55 → 29	30 → 55	55 → 30	50 → 55	55 → 50	40 → 50	50 → 40	33 → 37	37 → 33	33 → 46
Flow capacity (veh/min)	190	195	450	190	195	380	380	450	380	465	380
Queue capacity (veh)	1300.0	2600.0	1300.0	5600.0	2600.0	4200.0	1300.0	4749.0	4200.0	1300.0	1400.0

Free flow speed	0.370	0.231	0.231	0.463	0.463	0.463	0.231	0.231	0.463	0.370	0.324
(km/min)											
Link length (km)	0.333	0.333	0.189	0.189	0.389	0.389	0.263	0.263	0.324	0.324	0.538
Parameters	Link 21 → 25	25 → 21	18 → 26	26 → 18	25 → 44	44 → 25	42 → 44	44 → 42	31 → 45	45 → 31	49 → 50
Flow capacity	190	450	475	475	285	380	195	465	475	190	285
(veh/min)											
Queue capacity (veh)	4200.0	1583.0	1583.0	2800.0	1583.0	5600.0	3166.0	1583.0	2800.0	5200.0	1583.0
Free flow speed	0.232	0.370	0.463	0.324	0.463	0.324	0.370	0.370	0.463	0.463	0.232
(km/min)											
Link length (km)	0.396	0.396	0.411	0.411	0.341	0.341	0.779	0.779	0.991	0.991	0.254
Parameters	Link 26 → 51	51 → 26	17 → 18	18 → 17	16 → 21	21 → 16	24 → 37	37 → 24	39 → 49	49 → 39	49 → 28
Flow capacity	505	285	285	450	505	475	190	380	505	380	505
(veh/min)											
Queue capacity (veh)	1300.0	4200.0	5200.0	4200.0	2600.0	4200.0	1583.0	1583.0	2800.0	2800.0	1300.0
Free flow speed	0.463	0.231	0.463	0.370	0.370	0.370	0.370	0.370	0.370	0.231	0.232
(km/min)											
Link length (km)	0.315	0.315	0.435	0.435	0.353	0.353	0.704	0.704	0.278	0.278	0.609
Parameters	Link 27 → 5	3 → 2	20 → 13	12 → 4	15 → 5	41 → 20	36 → 31	47 → 48	17 → 14	37 → 29	34 → 33
Flow capacity	475	195	285	380	450	285	475	380	475	475	285
(veh/min)											
Queue capacity (veh)	1583.0	3166.0	2600.0	1583.0	2600.0	3166.0	1583.0	3166.0	2600.0	1583.0	2800.0
Free flow speed	0.370	0.370	0.232	0.370	0.232	0.370	0.232	0.463	0.232	0.370	0.463
(km/min)											
Link length (km)	1.252	0.168	0.371	0.217	0.501	0.288	0.947	0.323	0.723	0.290	0.875
Parameters	Link 54 → 28	53 → 29	46 → 33	50 → 49	28 → 49						
Flow capacity	505	380	195	505	195						
(veh/min)											
Queue capacity (veh)	2800.0	1400.0	2800.0	5600.0	1583.0						
Free flow speed	0.232	0.232	0.370	0.463	0.185						
(km/min)											
Link length (km)	0.310	0.776	0.538	0.254	0.609						

Table A.5: Link parameters in the Hong Kong network.

Appendix B

Appendix in publicly-owned centralized platform

B.1 Proof of Feasibility of POCP Problem

To demonstrate the feasibility of the proposed POCP design problem, it is necessary to find feasible demand variables $\mathbf{d}$ and spatial-temporal discount variables α in upper level, and feasible flow dynamics variables $\mathbf{y}$ in lower level. This implies that all travelers will reach their destinations and all SAVs will park at designated parking nodes by a large time T under the given demand $\mathbf{d}$. The construction of a feasible solution is illustrated by the following four steps:

- All origin-destination (O-D) pairs are ordered as $(o, s)_1, (o, s)_2, \dots, (o, s)_n$, where o_i denotes the origin node of the i_{th} O-D pair, s_i denotes the destination node of the i_{th} O-D pair, and n is the total number of O-D pairs.
- A suitable number of SAVs are selected to transport travelers, while the remaining SAVs are parked at the parking node.
- The travel demand for each O-D pair is split into several batches. Each batch of travel demand is then sequentially addressed, which means that only when all travelers in the j_{th} batch of O-D pair $(o, s)_i$ arrive the destination s_i , all selected SAVs would return from s_i to o_i , then pick up travelers in the $j+1_{th}$ batch of O-D pair $(o, s)_i$ to s_i .
- We would pick up each origin-destination travel demand in an increasing order. It means

that only when all travelers in O-D pair $(o, s)_i$ arrive at the destination s_i , all SAVs would pick up travelers in O-D pair $(o, s)_{i+1}$ from s_i to o_{i+1} .

For O-D pair $(o, s)_i$, a path P_{o_i, d_i} is chosen from origin o_i to destination s_i , and a path P_{s_i, o_i} is selected from destination s_i to origin o_i . $P_{s_i, o_{i+1}}$ denotes the path from destination s_i of O-D pair $(o, s)_i$ to the origin o_{i+1} of O-D pair $(o, s)_{i+1}$.

For simplicity, we assume there is only one parking node. For multiple parking nodes, the proof is similar; it is merely necessary to specify the number of SAVs at each parking node to satisfy the parking capacity constraint 4.23.

In addition, P_{p, o_1} denotes the path from the parking node p to the origin o_1 of O-D pair $(o, s)_1$, and $P_{s_n, p}$ denotes the path from destination s_n of O-D pair $(o, s)_n$ to the parking node p .

Next, we would consider the demand $\mathbf{d}$ and discount α in the upper level. The travel demand is shown as follows:

$$\sum_{r=a}^b d_{o,s}^{\mathcal{F}, \mathcal{S}, l}(r) = D_{o,s}^{\mathcal{F}, \mathcal{S}, l}[a, b] ; \quad \sum_{r=a}^b d_{o,s}^{\mathcal{F}, \mathcal{R}, l}(r) = D_{o,s}^{\mathcal{F}, \mathcal{R}, l}[a, b]$$

$$d_{o,s}^{\mathcal{S}}(r) = \sum_{l \in L} D_{o,s}^{\mathcal{E}, \mathcal{S}, l}(r) + \sum_{l \in L} d_{o,s}^{\mathcal{F}, \mathcal{S}, l}(r) ; \quad d_{o,s}^{\mathcal{R}}(r) = \sum_{l \in L} D_{o,s}^{\mathcal{E}, \mathcal{R}, l}(r) + \sum_{l \in L} d_{o,s}^{\mathcal{F}, \mathcal{R}, l}(r)$$

For any feasible discount α , we assume all travel demand of each O-D pair waiting at origin node o_i and total travel demand of $(o, s)_i$ is stated following. And we suppose all ride-sharing travelers would only take the SAVs with type $x = 1$.

$$D_{o_i, s_i} = d_{o_i, s_i}^{\mathcal{S}}(r) + d_{o_i, s_i}^{\mathcal{R}}(r)$$

$$u_{i,j}^x(k), u_{i,j}^x(k) = 0 \quad \text{for } x = 2, \dots, X$$

$$u_{i,j}(k) = \sum_{x=-1}^1 u_{i,j}^x(k); \quad v_{i,j}(k) = \sum_{x=-1}^1 v_{i,j}^x(k)$$

The travel demand of each O-D pair is divided into many batches as follows, where N_{o_i, s_i} is the number of batches of O-D pair $(o, s)_i$.

$$D_{o_i, s_i} = N_{o_i, s_i} B_{o_i, s_i}$$

Next, we would determine the size of inflow and outflow of SAVs. For the link in path

$P_{o_i, s_i} \cup P_{s_i, o_i} \cup P_{s_i, o_{i+1}} \cup P_{p, o_1} \cup P_{s_n, p}$, we suppose all SAVs would travel link in a free-flow travel time and each batch could be carried by SAVs in a time interval.

$$\begin{aligned} v_{i,j}(k + \tau_{i,j}^f) &= u_{i,j}(k) \quad \text{for } (i, j) \in P_{o_h, s_h} \cup P_{s_h, o_h} \cup P_{s_h, o_{h+1}} \cup P_{p, o_1} \cup P_{s_n, p} \\ \Delta_t u_{i,j}(k) &\geq B_{o_h, s_h} \quad \text{for } (i, j) \in P_{o_h, d_h} \end{aligned}$$

Then we consider the dynamic traffic assignment in the lower level. The double queue constraints in equation 4.1 and 4.2 are converted as follows. Normally, from the fundamental diagram, shockwave travel speed is always slower than free-flow travel speed, so we have $\tau_{i,j}^w > \tau_{i,j}^f$.

$$\begin{aligned} q_{i,j}^{\mathcal{D}}(k) &= \sum_{a=0}^{k-\tau_{i,j}^f} \Delta_t u_{i,j}(a) - \sum_{a=0}^k \Delta_t v_{i,j}(a) \\ &= 0 \\ q_{i,j}^{\mathcal{U}}(k) &= \sum_{a=0}^k \Delta_t u_{i,j}(a) - \sum_{a=0}^{k-\tau_{i,j}^w} \Delta_t v_{i,j}(a) \\ &= \sum_{a=k-\tau_{i,j}^w-\tau_{i,j}^f}^k \Delta_t u_{i,j}(a) \end{aligned}$$

The inflow, outflow, and upstream queue of SAVs should satisfy the capacity constraints as follows:

$$\begin{aligned} u_{i,j}(k) &\leq \bar{C}_{i,j}^\mu; \quad v_{i,j}(k) \leq \bar{C}_{i,j}^\nu, \\ q_{i,j}^u(k) &= \sum_{a=k-\tau_{i,j}^w-\tau_{i,j}^f}^k \Delta_t u_{i,j}(a) \leq \bar{Q}_{i,j}. \end{aligned}$$

The total travel time of O-D pair $(o, s)_h$, the travel time from s_h to o_{h+1} , travel time from p

to o_1 and travel time from s_n to p are presented as following:

$$\begin{aligned}
T_{(o,s)_h} &= N_{o_h,s_h} T_{o_h,s_h} + (N_{o_h,s_h} - 1) T_{s_h,o_h}, \\
&= \sum_{(i,j) \in P_{o_h,s_h}} N_{o_h,s_h} \tau_{i,j}^f + \sum_{(i,j) \in P_{s_h,o_h}} (N_{o_h,s_h} - 1) \tau_{i,j}^f, \\
T_{s_h,o_{h+1}} &= \sum_{(i,j) \in P_{s_h,o_{h+1}}} \tau_{i,j}^f, \\
T_{p,o_1} &= \sum_{(i,j) \in P_{p,o_1}} \tau_{i,j}^f, \\
T_{s_n,p} &= \sum_{(i,j) \in P_{s_n,p}} \tau_{i,j}^f.
\end{aligned}$$

We use $W_{(o,s)_h}$ represents the total time of travelers requesting demand with O-D pair $(o, s)_h$, which means that $d_{o_h,s_h}^S(r), d_{o_h,s_h}^R(r) = 0$ if $r \geq W_{(o,s)_h}$. So the total time $\tilde{T}$ is shown as following:

$$\begin{aligned}
\tilde{T} &= \max_h W_{(o,s)_h} + \sum_{h=1}^n T_{(o,s)_h} + \sum_{h=1}^{n-1} T_{s_h,o_{h+1}} + T_{p,o_1} + T_{s_n,p}, \\
&= \max_h W_{(o,s)_h} + \sum_{h=1}^n \sum_{(i,j) \in P_{o_h,s_h}} N_{o_h,s_h} \tau_{i,j}^f + \sum_{(i,j) \in P_{s_h,o_h}} (N_{o_h,s_h} - 1) \tau_{i,j}^f + \sum_{h=1}^{n-1} \sum_{(i,j) \in P_{s_h,o_{h+1}}} \tau_{i,j}^f \\
&\quad + \sum_{(i,j) \in P_{p,o_1}} \tau_{i,j}^f + \sum_{(i,j) \in P_{s_n,p}} \tau_{i,j}^f, \\
&= \max_h W_{(o,s)_h} + \sum_{h=1}^n \sum_{(i,j) \in E} N_{o_h,s_h} \tau_{i,j}^f \delta_{i,j}^{P_{o_h,s_h}} + \sum_{h=1}^n \sum_{(i,j) \in E} (N_{o_h,s_h} - 1) \tau_{i,j}^f \delta_{i,j}^{P_{s_h,o_h}} \sum_{h=1}^{n-1} \sum_{(i,j) \in E} \tau_{i,j}^f \delta_{i,j}^{P_{s_h,o_{h+1}}} \\
&\quad + \sum_{(i,j) \in E} \tau_{i,j}^f \delta_{i,j}^{P_{p,o_1}} + \sum_{(i,j) \in E} \tau_{i,j}^f \delta_{i,j}^{P_{s_n,p}},
\end{aligned}$$

where $\delta_{i,j}^{P_{o_h,s_h}}, \delta_{i,j}^{P_{s_h,o_h}}, \delta_{i,j}^{P_{s_h,o_{h+1}}}, \delta_{i,j}^{P_{p,o_1}}$ and $\delta_{i,j}^{P_{s_n,p}}$ are the indicator about whether link (i, j) is in the path $P_{o_h,s_h}, P_{s_h,o_h}, P_{s_h,o_{h+1}}, P_{p,o_1}$ and $P_{s_n,p}$, respectively.

Therefore, if we set the time horizon T as larger than $\tilde{T}$, then we could construct a feasible solution to make all travelers arrive at their destinations and all SAVs park at parking nodes under a feasible demand and discount by a large time T .

B.2 Proof of Proposition 11

Suppose $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ is the optimal variables of Programming (I). Then we have

$$\mu_i^* = 0 \text{ or } (Ey^* - F)_i = 0 \text{ for } i = 1, \dots, n_\mu$$

Define $\tilde{I} = \{i | \mu_i^* = 0\}$ and $\tilde{J} = \{j | (Ey^* - F)_j = 0\}$. Then we would prove that Programming (I) would be equivalent with Programming (II) under the set $\tilde{I}$ and $\tilde{J}$. Suppose SC^* and $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ are the optimal objective and optimal solution for Programming (I). $\widetilde{SC}$ and $\widetilde{\mathbf{y}}, \widetilde{\mathbf{d}}, \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}}$ for Programming (II).

It is obvious that $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ also satisfies the Equations 4.45b to 4.45j. Therefore, $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*$ are the feasible solution of Programming (II). Therefore, we have $SC^* \geq \widetilde{SC}$.

As $\tilde{I} \cup \tilde{J} = \{1, \dots, n_\mu\}$, we have $\widetilde{\mu}_i (\mathbf{E}\widetilde{\mathbf{y}} - \mathbf{F})_i = 0$ for $i = 1, \dots, n_\mu$. Therefore, $\widetilde{\mathbf{y}}, \widetilde{\mathbf{d}}, \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}}$ satisfies the Equations 4.44b to 4.44i and they are the feasible solution of Programming (I). Therefore, we have $SC^* \leq \widetilde{SC}$.

To conclude, we have $SC^* = \widetilde{SC}$. We find the set $\tilde{I}$ and $\tilde{J}$ such that Programming (I) would be equivalent with Programming (II).

B.3 Proof of Lemma 1

Suppose $\mathbf{y}, \mathbf{d}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ are the feasible solution of Programming (III), then the objective is to find the spatial-temporal discount variables $\boldsymbol{\alpha}$ such that $\mathbf{J}\boldsymbol{\alpha} \leq \mathbf{K}$ and $f(\mathbf{d}, \boldsymbol{\alpha}) = 0$. Next, we would prove that the $\boldsymbol{\alpha} = \mathbf{g}^{-1}(\mathbf{d}, \boldsymbol{\beta})$ obtained by the transformation $\boldsymbol{\beta} = \mathbf{g}(\mathbf{d}, \boldsymbol{\alpha})$ in Equations B.1 and B.2 satisfy $\mathbf{J}\boldsymbol{\alpha} \leq \mathbf{K}$ and $f(\mathbf{d}, \boldsymbol{\alpha}) = 0$.

$$e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} = \frac{d_{o,s}^{\mathcal{F}, \mathcal{S}, l}(r)}{\beta_{o,s}^{\mathcal{F}, \mathcal{S}, l}(r)} \quad (\text{B.1})$$

$$e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} = \frac{d_{o,s}^{\mathcal{F}, \mathcal{R}, l}(r)}{\beta_{o,s}^{\mathcal{F}, \mathcal{R}, l}(r)} \quad (\text{B.2})$$

Since $\mathbf{y}, \mathbf{d}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ are the feasible solution of Programming (III), we have that $\mathbf{d}$ and $\boldsymbol{\beta}$ satisfy

the Equations 4.56 and 4.57, then we obtain the Equations B.3 and B.4.

$$e^{-\lambda C_{-1}} \leq \frac{d_{o,s}^{\mathcal{F},S,l}(r)}{\beta_{o,s}^{\mathcal{F},S,l}(r)} \leq 1 \quad r \in \{a, a+1, \dots, b-1, b\} \quad (\text{B.3})$$

$$e^{-\sum_{x=1}^X \lambda C_x \eta_x} \leq \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l}(r)}{\beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r)} \leq 1 \quad r \in \{a, a+1, \dots, b-1, b\} \quad (\text{B.4})$$

By Equations B.1 to B.4, we have Equations B.5 and B.6.

$$0 \leq \alpha_s^{x=-1}(r) \leq 1 \quad (\text{B.5})$$

$$0 \leq \sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r) \leq \sum_{x=1}^X \lambda C_x \eta_x \quad (\text{B.6})$$

The $\mathbf{J}\alpha \leq \mathbf{K}$ could be specified by Equation B.7. It is obvious that for α satisfying Equations B.5 and B.6, there always exists such α to make Equation B.7 feasible. Therefore, we have proved that for $\alpha = \mathbf{g}^{-1}(\mathbf{d}, \beta)$ obtained by the transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$ in Equations B.1 and B.2, we have $\mathbf{J}\alpha \leq \mathbf{K}$.

$$0 \leq \alpha_s^x(r) \leq 1 \quad (\text{B.7})$$

Then we would prove that the $\alpha = \mathbf{g}^{-1}(\mathbf{d}, \beta)$ obtained by the transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$ in Equations B.1 and B.2 satisfy and $f(\mathbf{d}, \alpha) = \mathbf{0}$. And the constraint $f(\mathbf{d}, \alpha) = \mathbf{0}$ are specified by following Equations B.8 and B.9.

$$d_{o,s}^{\mathcal{F},S,l}(r) = D_{o,s}^{\mathcal{F},S,l}[a, b] \frac{e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}}{\sum_{i=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(i)}} \quad (\text{B.8})$$

$$d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \frac{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}}{\sum_{i=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(i)}} \quad (\text{B.9})$$

For α obtained by Equations B.1 and B.2, we have Equations B.10 and B.11 for $o_1 \neq o_2$ and

$l_1 \neq l_2$.

$$e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} = \frac{d_{o_1,s}^{\mathcal{F},S,l_1}(r)}{\beta_{o_1,s}^{\mathcal{F},S,l_1}(r)} = \frac{d_{o_2,s}^{\mathcal{F},S,l_2}(r)}{\beta_{o_2,s}^{\mathcal{F},S,l_2}(r)} \quad (\text{B.10})$$

$$e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} = \frac{d_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}(r)}{\beta_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}(r)} = \frac{d_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}(r)}{\beta_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}(r)} \quad (\text{B.11})$$

Since $\mathbf{y}, \mathbf{d}, \beta, \mu, \lambda$ are the feasible solution of Programming (III), by Equations 4.54 and 4.55, we have Equations B.12 and B.13 for $\forall r \in \{a, a+1, \dots, b-1, b\}$, where we set $\beta_{o,s}^{\mathcal{F},S,l}[a, b]$ and $\beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b]$ as the common value in Equations 4.54 and 4.55, respectively.

$$\beta_{o,s}^{\mathcal{F},S,l}(r) = \beta_{o,s}^{\mathcal{F},S,l}[a, b] \quad (\text{B.12})$$

$$\beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = \beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \quad (\text{B.13})$$

By Equations 4.48, 4.49, B.12 and B.13, we have Equations B.14 to B.17.

$$\sum_{r=a}^b d_{o,s}^{\mathcal{F},S,l}(r) = \sum_{r=a}^b \beta_{o,s}^{\mathcal{F},S,l}(r) e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} \quad (\text{B.14})$$

$$= \beta_{o,s}^{\mathcal{F},S,l}[a, b] \sum_{r=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} = D_{o,s}^{\mathcal{F},S,l}[a, b] \quad (\text{B.15})$$

$$\sum_{r=a}^b d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = \sum_{r=a}^b \beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} \quad (\text{B.16})$$

$$= \beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \sum_{r=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} = D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \quad (\text{B.17})$$

Then we obtain the formulation of $\beta_{o,s}^{\mathcal{F},S,l}[a, b]$ and $\beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b]$ in Equations B.18 and B.19.

$$\beta_{o,s}^{\mathcal{F},S,l}[a, b] = \frac{D_{o,s}^{\mathcal{F},S,l}[a, b]}{\sum_{r=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}} \quad (\text{B.18})$$

$$\beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] = \frac{D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b]}{\sum_{r=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}} \quad (\text{B.19})$$

Therefore, we obtain the formulation of $\mathbf{d}$ and α in Equations B.20 and B.21, which indicates

that the $\mathbf{d}$ and α satisfy the $f(\mathbf{d}, \alpha) = 0$ in Equations B.8 and B.9.

$$\begin{aligned}
d_{o,s}^{\mathcal{F},\mathcal{S},l}(r) &= \beta_{o,s}^{\mathcal{F},\mathcal{S},l}(r) e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} \\
&= \beta_{o,s}^{\mathcal{F},\mathcal{S},l}[a, b] e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)} \\
&= D_{o,s}^{\mathcal{F},\mathcal{S},l}[a, b] \frac{e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}}{\sum_{i=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(i)}} \tag{B.20}
\end{aligned}$$

$$\begin{aligned}
d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) &= \beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} \\
&= \beta_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} \\
&= D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \frac{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}}{\sum_{i=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(i)}} \tag{B.21}
\end{aligned}$$

To conclude, we have proved that the $\alpha = \mathbf{g}^{-1}(\mathbf{d}, \beta)$ obtained by the transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$ in Equations B.1 and B.2 satisfy $\mathbf{J}\alpha \leq \mathbf{K}$ and $f(\mathbf{d}, \alpha) = 0$. Therefore, Lemma 1 holds.

B.4 Proof of Proposition 12

Since Programming (II) and Programming (III) have the same objective function, the idea to prove that Programming (II) and Programming (III) have the same optimal objective value and optimal variables is as follows: If any feasible solution of Programming (II) also satisfies all the constraints of Programming (III), then the optimal objective value of Programming (II) is greater than or equal to the optimal objective value of Programming (III). Conversely, if any feasible solution of Programming (III) also satisfies all the constraints of Programming (II), then the optimal objective value of Programming (II) is less than or equal to the optimal objective value of Programming (III).

Firstly, we add the variables β for Programming (I), but the optimal objective and optimal variables of Programming (I) are unchanged because β does not exist in both the objective and constraints of Programming (I). Similarly, we add the variables α for Programming (III).

Suppose $SC^*, \mathbf{y}^*, \mathbf{d}^*, \alpha^*, \mu^*, \lambda^*$ are the optimal objective and optimal variables of Program-

ming (I), then we have Equations B.22 and B.23 for $f(\mathbf{d}^*, \alpha^*) = 0$.

$$d_{o,s}^{\mathcal{F},S,l,*}(r) = D_{o,s}^{\mathcal{F},S,l}[a, b] \frac{e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(r)}}{\sum_{i=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(i)}} \quad (\text{B.22})$$

$$d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r) = D_{o,s}^{\mathcal{F},\mathcal{R},l}[a, b] \frac{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(r)}}{\sum_{i=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(i)}} \quad (\text{B.23})$$

The Equations B.22 and B.23 imply the relationship of $\mathbf{d}^*$ and α^* in following equations.

$$\frac{d_{o_1,s}^{\mathcal{F},S,l_1,*}(r)}{D_{o_1,s}^{\mathcal{F},S,l_1}[a, b]} = \frac{d_{o_2,s}^{\mathcal{F},S,l_2,*}(r)}{D_{o_2,s}^{\mathcal{F},S,l_2}[a, b]} \quad (\text{B.24})$$

$$\frac{d_{o_1,s}^{\mathcal{F},\mathcal{R},l_1,*}(r)}{D_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}[a, b]} = \frac{d_{o_2,s}^{\mathcal{F},\mathcal{R},l_2,*}(r)}{D_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}[a, b]} \quad (\text{B.25})$$

$$D_{o_2,s}^{\mathcal{F},S,l_2}[a, b] \frac{d_{o_1,s}^{\mathcal{F},S,l_1,*}(r)}{e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(r)}} = D_{o_1,s}^{\mathcal{F},S,l_1}[a, b] \frac{d_{o_2,s}^{\mathcal{F},S,l_2,*}(r)}{e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(r)}} \quad (\text{B.26})$$

$$D_{o_2,s}^{\mathcal{F},\mathcal{R},l_2}[a, b] \frac{d_{o_1,s}^{\mathcal{F},\mathcal{R},l_1,*}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(r)}} = D_{o_1,s}^{\mathcal{F},\mathcal{R},l_1}[a, b] \frac{d_{o_2,s}^{\mathcal{F},\mathcal{R},l_2,*}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(r)}} \quad (\text{B.27})$$

Besides, we have $\mathbf{J}\alpha^* \leq \mathbf{K}$ as $0 \leq \alpha_s^{x,*}(r) \leq 1$, then we have the Equation B.28 and B.29.

$$d_{o,s}^{\mathcal{F},S,l,*}(r) \leq \frac{d_{o,s}^{\mathcal{F},S,l,*}(r)}{e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(r)}} \leq \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{e^{-\lambda C_{-1}}} \quad (\text{B.28})$$

$$d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r) \leq \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(r)}} \leq \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x}} \quad (\text{B.29})$$

If we define the variables β as $\beta = \mathbf{g}(\mathbf{d}^*, \alpha^*)$ in Equations B.30 and B.31, the Equations B.24 to B.29 could be converted as Equations 4.48 to 4.55. Therefore, $SC^*, \mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*, \beta$ satisfy the Equations 4.60b to 4.60j, then they are the feasible solution of Programming (III). Therefore, we have $SC^* \geq \widetilde{SC}$.

$$\beta_{o,s}^{\mathcal{F},S,l}(r) = \frac{d_{o,s}^{\mathcal{F},S,l,*}(r)}{e^{-\lambda C_{-1} \alpha_s^{x=-1,*}(r)}} \quad (\text{B.30})$$

$$\beta_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = \frac{d_{o,s}^{\mathcal{F},\mathcal{R},l,*}(r)}{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^{x,*}(r)}} \quad (\text{B.31})$$

Besides, by Lemma 1, for optimal variables $\widetilde{\mathbf{y}}, \widetilde{\mathbf{d}}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\lambda}$ of Programming (III), there always exists the spatial-temporal discount variables α such that $\mathbf{J}\alpha \leq \mathbf{K}$ in Equation 4.44h and $f(\mathbf{d}, \alpha) =$

0 in Equation 4.44i.

Therefore, $\tilde{\mathbf{y}}, \tilde{\mathbf{d}}, \alpha, \tilde{\mu}, \tilde{\lambda}$ satisfy the Equations 4.44b to 4.44i, then they are the feasible solution of Programming (I). Therefore, we have $SC^* \leq \tilde{SC}$.

To conclude, we have proved that $SC^* = \tilde{SC}$. And $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*, \beta$ constitute the optimal solution of Programming (III), $\tilde{\mathbf{y}}, \tilde{\mathbf{d}}, \alpha, \tilde{\mu}, \tilde{\lambda}$ for Programming (I), so we also have that $\mathbf{d}^* = \tilde{\mathbf{d}}$, $\mu^* = \tilde{\mu}$ and $\lambda^* = \tilde{\lambda}$.

B.5 Proof of Proposition 13

Suppose SC is the optimal objective of Programming (I), $\mathbf{y}, \mathbf{d}, \mu, \lambda, \alpha$ is the optimal solution of Programming (I).

1) We first prove $SC^* \geq SC$.

By Algorithm 2, $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*, \beta^*$ is the optimal variables of Programming (III) under set $\tilde{I}$ and $\tilde{J}$ by Algorithm 2. Therefore, $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*$ satisfy the Equations 4.44b to 4.44d, 4.44f to 4.44g. By the process of Algorithm 2, $\mathbf{y}, \mathbf{d}, \mu, \lambda$ should satisfy all complementary constraints as Equation 4.44e in Programming (I). Besides, by Lemma 1, there always exists the spatial-temporal discount variables α such that α satisfies $\mathbf{J}\alpha \leq \mathbf{K}$ in Equation 4.44h and $\mathbf{f}(\mathbf{d}, \alpha) = \mathbf{0}$ in Equation 4.44i. Therefore, $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*, \alpha$ are the feasible solution Programming (I) by satisfying all constraints as Equations 4.44a to 4.44i, so we have $SC^* \geq SC$.

2) Then we prove $SC^* \leq SC$.

For $\mathbf{y}, \mathbf{d}, \mu, \lambda, \alpha, \mu$ and $\mathbf{y}$ should satisfy all complementary constraints as $\mu_i (\mathbf{E}\mathbf{y} - \mathbf{F})_i = 0$ for $i = 1, \dots, n_\mu$. Then we have the set of $\hat{I}$ and $\hat{J}$ as $\mu_i = 0$ for $i \in \hat{I}$ and $(\mathbf{E}\mathbf{y} - \mathbf{F})_j = 0$ for $j \in \hat{J}$, where $\hat{I} \subseteq \{1, \dots, n_\mu\}$ and $\hat{J} \subseteq \{1, \dots, n_{mu}\}$. The complementary constraints of Programming (I) in Equation 4.44e could be equivalently replaced by $\mu_i = 0$ for $i \in \hat{I}$ and $(\mathbf{E}\mathbf{y} - \mathbf{F})_j = 0$ for $j \in \hat{J}$. By transformation $\beta = \mathbf{g}(\mathbf{d}, \alpha)$ in Equations 4.46 and 4.47, we could find the variables β for $\mathbf{d}$ and α . In the process of Algorithm 2, we find the Programming (III) under $\tilde{I}$ and $\tilde{J}$ with the minimum social cost SC^* , which is smaller than the value of Programming (III) under $\hat{I}$ and $\hat{J}$. So we have $SC^* \leq SC$.

To conclude, we have proved that $SC^* \geq SC$ and $\mathbf{y}^*, \mathbf{d}^*, \mu^*, \lambda^*, \alpha$ are the optimal solution Programming (I).

B.6 Proof of Proposition 14

In the case of Programming (I), the existence of an optimal solution $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*$, and $\boldsymbol{\lambda}^*$ is guaranteed due to the proven feasibility of the proposed bi-level programming in B.1 and the equivalence to the bi-level programming as stated in Proposition 10.

For flexible demand, their requesting order time choices are related to the discount for both solo travelers and ride-sharing travelers.

$$\begin{aligned} d_{o,s}^{\mathcal{F},\mathcal{S},l}(r) &= D_{o,s}^{\mathcal{F},\mathcal{S},l}[a,b] \frac{e^{-\lambda C_{-1} \alpha_s^{x=-1}(r)}}{\sum_{i=a}^b e^{-\lambda C_{-1} \alpha_s^{x=-1}(i)}} \\ d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) &= D_{o,s}^{\mathcal{F},\mathcal{R},l}[a,b] \frac{e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}}{\sum_{i=a}^b e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(i)}} \end{aligned}$$

As we define $z(r) = e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}$, this function captures the effect of the discount at time r for ride-sharing travelers. Notably, the value of $z(r)$ lies within the range of $e^{-\sum_{x=1}^X \lambda C_x \eta_x}$ and 1. This means that the relationship between the demand $d_{o,s}^{\mathcal{F},\mathcal{R},l}(r)$ and the discount can be expressed as a system of linear equations. This linear representation facilitates the mathematical tractability and computational efficiency of the problem.

$$\mathbf{M}_{o,s}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}) \mathbf{Z}_{a,b} = \mathbf{0}$$

where,

$$\begin{aligned} \bullet \mathbf{M}_{o,s}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}) &= \begin{bmatrix} d_{o,s}^{\mathcal{F},\mathcal{R},l}(a) - D_{o,s}^{\mathcal{F},\mathcal{R},l}[a,b] & \cdots & d_{o,s}^{\mathcal{F},\mathcal{R},l}(a) \\ d_{o,s}^{\mathcal{F},\mathcal{R},l}(a+1) & \cdots & d_{o,s}^{\mathcal{F},\mathcal{R},l}(a+1) \\ \vdots & \cdots & \cdots \\ d_{o,s}^{\mathcal{F},\mathcal{R},l}(b) & \cdots & d_{o,s}^{\mathcal{F},\mathcal{R},l}(b) - D_{o,s}^{\mathcal{F},\mathcal{R},l}[a,b] \end{bmatrix} \\ \bullet \mathbf{Z}_{a,b} &= \begin{bmatrix} z(a) = e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(a)} \\ \vdots \\ z(r) = e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)} \\ \vdots \\ z(b) = e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(b)} \end{bmatrix} \end{aligned}$$

As we've established that $\mathbf{y}^*, \mathbf{d}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*$, and $\boldsymbol{\lambda}^*$ constitute an optimal solution for Programming (I), we can also deduce that $d_{o,s}^{\mathcal{F},\mathcal{R},l}(r)$ for $r \in a, a+1, \dots, b-1, b$ must satisfy the

following constraint: $\sum_{r=a}^b d_{o,s}^{\mathcal{F},\mathcal{R},l}(r) = D_{o,s}^{\mathcal{F},\mathcal{R},l}[a,b]$. Likewise, the optimal $z^*(r)$ for $r \in a, a+1, \dots, b-1, b$ must satisfy the inequality $e^{-\sum_{x=1}^X \lambda C_x \eta_x} \leq z^*(r) \leq 1$. Therefore, $\mathbf{Z}_{a,b}^*$ is a non-zero solution of the system of linear equations $\mathbf{M}_{o,s}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}) \mathbf{Z}_{a,b} = \mathbf{0}$. Hence, the system $\mathbf{M}_{o,s}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}^*) \mathbf{Z}_{a,b} = \mathbf{0}$ possesses an infinite number of non-zero solutions and may have other non-zero solutions that satisfy $e^{-\sum_{x=1}^X \lambda C_x \eta_x} \leq z(r) \leq 1$.

For clarification, consider the following set of equations, where $\mathbf{m}_{o,s,i}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d})$ is the i th column of $\mathbf{M}_{o,s}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}^*)$ and $\mathbf{z}_{a,b,i}$ is the i th row of $\mathbf{Z}_{a,b}^*$.

$$\begin{aligned} \sum_{i=1}^{b-a+1} \mathbf{m}_{o,s,i}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}^*) \mathbf{z}_{a,b,i}^* &= \mathbf{0} \\ e^{-\sum_{x=1}^X \lambda C_x \eta_x} \leq z(i)^* \leq 1 \quad i &\in \{a, a+1, \dots, b-1, b\} \end{aligned}$$

If the condition $\delta(z(i)^* = e^{-\sum_{x=1}^X \lambda C_x \eta_x}) \cdot \delta(z(j)^* = 1) = 0$ holds for all $i, j \in a, a+1, \dots, b-1, b$, it implies that there exists a scalar value θ such that the new matrix $\mathbf{Z}_{a,b}$ can be obtained by scaling the original solution $\mathbf{Z}_{a,b}^*$ by θ . This scaled solution still satisfies the necessary conditions as follows.

$$\begin{aligned} \sum_{i=1}^{b-a+1} \mathbf{m}_{o,s,i}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}^*) \mathbf{z}_{a,b,i} &= \sum_{i=1}^{b-a+1} \mathbf{m}_{o,s,i}^{\mathcal{F},\mathcal{R},l,a,b}(\mathbf{d}^*) \theta \mathbf{z}_{a,b,i}^* = \mathbf{0} \\ e^{-\sum_{x=1}^X \lambda C_x \eta_x} \leq z(i) = \theta z(i)^* \leq 1 \quad i &\in \{a, a+1, \dots, b-1, b\} \end{aligned}$$

We can then derive a new set of spatial-temporal discount variables α by solving the equation $z(r) = e^{-\sum_{x=1}^X \lambda C_x \eta_x \alpha_s^x(r)}$ for the newly scaled $\mathbf{Z}_{a,b}$.

Therefore, the new set of variables $\mathbf{y}^*, \mathbf{d}^*, \alpha, \mu^*$, and λ^* also constitutes an optimal solution for the Programming (I), demonstrating that there are indeed multiple optimal solutions.

B.7 Parameters of POCP

We first illustrate the parameters of parking fee, number of SAVs at the parking node at initial, and the parking capacity of the Kowloon Peninsula in Hong Kong (Parkopedia, 2023), as referred to Table B.1.

Parking fee (HKD/hour)		Number of SAVs (veh)		Parking capacity (veh)	
Notations	Values	Notations	Values	Notations	Values
C_{p39}	26	N_{p39}	1,950	$\tilde{C}_{p39}$	2,200
C_{p16}	27	N_{p16}	2,000	$\tilde{C}_{p16}$	2,150
C_{p3}	27	N_{p3}	2,050	$\tilde{C}_{p3}$	2,600
C_{p1}	24	N_{p1}	2,100	$\tilde{C}_{p1}$	2,700
C_{p10}	26	N_{p10}	2,150	$\tilde{C}_{p10}$	2,950
C_{p28}	36	N_{p28}	1,950	$\tilde{C}_{p28}$	2,550
C_{p32}	18	N_{p32}	2,050	$\tilde{C}_{p32}$	2,550
C_{p34}	16	N_{p34}	1,950	$\tilde{C}_{p34}$	2,750
C_{p18}	26	N_{p18}	2,050	$\tilde{C}_{p18}$	2,400
C_{p35}	20	N_{p35}	2,000	$\tilde{C}_{p35}$	2,600

Table B.1: Parking parameters for Hong Kong network.

The study simulates the demand over the first 36 minutes, divided into four intervals: 0-9 minutes, 9-18 minutes, 18-27 minutes, and 27-36 minutes. The flexible demand of solo travelers (STs) for each origin-destination (O-D) pair in each interval ranges from 95 to 210, while for ride-sharing travelers (RTs) it ranges from 325 to 510. A constant regular demand is set at 10 for STs and 25 for RTs. The latest acceptable arrival time for each demand interval is set as 27, 33, 42, and 51 minutes, respectively.

Appendix C

Appendix in mixed sub-network equilibrium

C.1 Proof of Proposition 18

First, we define the $\subseteq$ operator for path in Definition C.1.1. If $\tilde{r} \subseteq \bar{r}$, it means that the path $\tilde{r}$ is part of path $\bar{r}$.

Definition C.1.1. Suppose $\tilde{r} = (n_1, n_2, \dots, n_w) \in P_{n_1, n_w}$ and $\bar{r} = (m_1, m_2, \dots, m_u) \in P_{m_1, m_u}$, where $n_1, \dots, n_w, m_1, \dots, m_u \in N$, then we define $\tilde{r} \subseteq \bar{r}$ if $\exists i \in [1, u - w + 1]$ such that $m_i = n_1, m_{i+1} = n_2, \dots, m_{i+w-1} = n_w$

Then we define the $\cup$ operator for path in Definition C.1.2, which implies that a path can be viewed as the concatenation of two sub-paths. And it would be easily extended to the case for multiple sub-paths.

Definition C.1.2. Suppose $r = (n_1, n_2, \dots, n_w) \in P_{n_1, n_w}$, where $n_1, \dots, n_w \in N$, then we define $r = \tilde{r} \cup \bar{r}$ for $\forall i \in [1, w]$ with $\tilde{r} = (n_1, \dots, n_i) \in P_{n_1, n_i}$ and $\bar{r} = (n_i, \dots, n_w) \in P_{n_i, n_w}$

Assume $\exists o \in O \cap \Phi, k \in \Omega, r_1, r_2 \in P_{o, k}$ with $f_{o, k}^{r_1}, f_{o, k}^{r_2} > 0$ and $r_1 \neq r_2$ such that $c_{o, k}^{r_1} > c_{o, k}^{r_2}$. Since $f_{o, k}^{r_1}, f_{o, k}^{r_2} > 0$, there must exist $s \in S$ and $r_3, r_4 \in P_{o, s}$ such that $r_1 \subseteq r_3$ and $r_2 \subseteq r_4$. When the destination s is in the UESN as $s \in S \cap \Phi$, then there must exist $r_5, r_6 \in P_{k, l}$ with $l \in \Theta$ and $r_7, r_8 \in P_{l, s}$ such that $r_3 = r_1 \cup r_5 \cup r_7$ and $r_4 = r_2 \cup r_6 \cup r_8$. Then we have the following equations.

$$c_{o,s}^{r_3} - \pi_{o,s} = 0 \quad \text{for } s \in S \cap \Phi \quad (\text{C.1})$$

$$c_{o,s}^{r_4} - \pi_{o,s} = 0 \quad \text{for } s \in S \cap \Phi \quad (\text{C.2})$$

$$f_{o,s}^{r_3} = f_{o,s}^{r_1} = f_{o,s}^{r_5} = f_{o,s}^{r_7} > 0 \quad \text{for } s \in S \cap \Phi \quad (\text{C.3})$$

$$f_{o,s}^{r_4} = f_{o,s}^{r_2} = f_{o,s}^{r_6} = f_{o,s}^{r_8} > 0 \quad \text{for } s \in S \cap \Phi \quad (\text{C.4})$$

$$c_{o,s}^{r_3} = c_{o,k}^{r_1} + c_{k,l} + c_{k,s}^{r_7} \quad \text{for } s \in S \cap \Phi \quad (\text{C.5})$$

$$c_{o,s}^{r_4} = c_{o,k}^{r_2} + c_{k,l} + c_{k,s}^{r_8} \quad \text{for } s \in S \cap \Phi \quad (\text{C.6})$$

Therefore, we have $c_{l,s}^{r_7} < c_{l,s}^{r_8}$. Then for the path $r_9 \in P_{o,s}$ with $r_9 = r_2 \cup r_4 \cup r_7$, the travel cost of path r_9 is presented as follows:

$$c_{o,s}^{r_9} - \pi_{o,s} = c_{o,k}^{r_2} + c_{k,l} + c_{l,s}^{r_7} - (c_{o,k}^{r_1} + c_{k,l} + c_{l,s}^{r_7}) \quad (\text{C.7})$$

$$= c_{o,k}^{r_2} - c_{o,k}^{r_1} < 0 \quad (\text{C.8})$$

Equation C.8 contradicts with Equation 5.3. Then we would discuss the case when the destination s is in the SOSN as $s \in S \cap \Psi$, then there must exist $r_9, r_{10} \in P_{k,s}$ such that $r_3 = r_1 \cup r_9$ and $r_4 = r_2 \cup r_{10}$. In this case, it is obvious that we have $c_{o,s}^{r_3} > c_{o,s}^{r_5}$, which contradicts with the Equation 5.1. Therefore, for $o \in O \cap \Phi$, $k \in \Omega$, $\forall r_1, r_2 \in P_{o,k}$ with $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$, we have $c_{o,k}^{r_1} = c_{o,k}^{r_2} = \pi_{o,k}$.

Besides, for $\forall r \in P_{o,k}$ with $f_{o,k}^r = 0$, Equations 5.1 and 5.3 imply that $c_{o,s}^r - \pi_{o,s} \geq 0$.

C.2 Proof of Proposition 19

Assume $\exists o \in O \cap \Psi, k \in \Theta, r_1, r_2 \in P_{o,k}$ with $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$ and $r_1 \neq r_2$ such that $h_{o,k}^{r_1} > h_{o,k}^{r_2}$. Since $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$, there must exist $s \in S \cap \Phi$ and $r_3, r_4 \in P_{o,s}$ such that $r_1 \subseteq r_3$ and $r_2 \subseteq r_4$. Besides, there also exists $r_5, r_6 \in P_{k,s}$ such that $r_3 = r_1 \cup r_5$ and $r_4 = r_2 \cup r_6$. Therefore, we have the following equations based on Equation 5.7 to 5.9.

$$h_{o,s}^{r_3} - \eta_{o,s} = 0 \quad (\text{C.9})$$

$$h_{o,s}^{r_4} - \eta_{o,s} = 0 \quad (\text{C.10})$$

$$f_{o,s}^{r_3} = f_{o,k}^{r_1} = f_{k,s}^{r_5} > 0 \quad (\text{C.11})$$

$$f_{o,s}^{r_4} = f_{o,k}^{r_2} = f_{k,s}^{r_6} > 0 \quad (\text{C.12})$$

$$h_{o,s}^{r_3} = h_{o,k}^{r_1} + h_{k,s} \quad (\text{C.13})$$

$$h_{o,s}^{r_4} = h_{o,k}^{r_2} + h_{k,s} \quad (\text{C.14})$$

It is obvious that we $h_{o,s}^{r_3} > h_{o,s}^{r_4}$ by Equations C.13 and C.14, which contradicts with Equations C.9 and C.10. Therefore, for $o \in O \cap \Psi$, $k \in \Theta$, $\forall r_1, r_2 \in P_{o,k}$ with $f_{o,k}^{r_1}, f_{o,k}^{r_2} > 0$, we have $h_{o,k}^{r_1} = h_{o,k}^{r_2} = \eta_{o,k}$.

Besides, for $\forall r \in P_{o,k}$ with $f_{o,k}^r = 0$, Equations 5.7 and 5.9 imply that $h_{o,s}^r - \eta_{o,s} \geq 0$.

C.3 Proof of Proposition 20

By Equations 5.13 to 5.15, it is obvious that for $s \in S \in \Phi$, $k \in \Omega$, $\forall r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, we have $h_{k,s}^{r_1} = h_{k,s}^{r_2} = \eta_{k,s}$ and $h_{k,s}^r \geq \eta_{k,s}$ for $\forall r \in P_{k,s}$ with $f_{k,s}^r = 0$. Next, we would prove that for $s \in S \in \Phi$, $k \in \Omega$, if $\sum_{i \in \Phi} \delta_{i,k}^L x_{i,k} > 0$, $\forall r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, we have $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$.

Assume $\exists s \in S \in \Phi$, $k \in \Omega$, $r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$ and $r_1 \neq r_2$ such that $c_{k,s}^{r_1} > c_{k,s}^{r_2}$. Since $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, there must exist $o \in O \cap \Phi$ and $r_3, r_4 \in P_{o,s}$ such that $r_1 \subseteq r_3$ and $r_2 \subseteq r_4$. Besides, there also exists $r_5, r_6 \in P_{o,k}$ such that $r_3 = r_1 \cup r_5$ and $r_4 = r_2 \cup r_6$.

When $o \in \Phi$, Proposition 18 indicates that $c_{o,k}^{r_5} = c_{o,k}^{r_6}$. Besides, we have the following equations:

$$c_{o,s}^{r_3} - \pi_{o,s} = 0 \quad (\text{C.15})$$

$$c_{o,s}^{r_4} - \pi_{o,s} = 0 \quad (\text{C.16})$$

$$c_{o,s}^{r_3} = c_{o,k}^{r_1} + c_{k,s}^{r_5} \quad (\text{C.17})$$

$$c_{o,s}^{r_4} = c_{o,k}^{r_2} + c_{k,s}^{r_6} \quad (\text{C.18})$$

Therefore, we have $c_{k,s}^{r_1} = c_{k,s}^{r_2}$, which contradicts with $c_{k,s}^{r_1} > c_{k,s}^{r_2}$. Next, we would prove that $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$. Suppose $\exists r_7 \in P_{k,s}$ with $f_{k,s}^{r_7} > 0$, $h_{k,s}^{r_7} = \eta_{k,s}$ and $c_{k,s}^{r_7} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$, if $c_{r_1} \neq c_{r_7}$, then we construct a new path from $k \in \Omega$ to the destination $s \in S \cap \Phi$ as $r_8 = r_7 \cup r_5$. Since $f_{k,s}^{r_7} > 0$ and $f_{k,s}^{r_5} = f_{k,s}^{r_1} > 0$, we obtain that $f_{o,s}^{r_8}$ is greater than 0. Then we have the following Equations:

$$c_{o,s}^{r_8} - \eta_{o,s} = c_{o,k}^{r_5} + c_{o,k}^{r_7} - \eta_{o,s} \quad (\text{C.19})$$

$$< c_{o,k}^{r_5} + c_{o,k}^{r_1} - \eta_{o,s} \quad (\text{C.20})$$

$$= c_{o,s}^{r_3} - \eta_{o,s} \quad (\text{C.21})$$

$$= 0 \quad (\text{C.22})$$

So we have $c_{o,s}^{r_8} < \eta_{o,s}$, which contradicts with the Equation 5.1 given that $f_{o,s}^{r_8} > 0$. Therefore, we prove that $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$.

To conclude, we obtain that for $s \in S$, $k \in \Omega$, $\forall r_1, r_2 \in P_{k,s}$ with $f_{k,s}^{r_1}, f_{k,s}^{r_2} > 0$, we have $c_{k,s}^{r_1} = c_{k,s}^{r_2} = \min_{r \in P_{k,s}; h_{k,s}^r = \eta_{k,s}} \{c_{k,s}^r\}$.

C.4 Proof of Proposition 22

Equation 5.35 could be equivalently written as “vector $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ does not solve the variational inequality 5.35 $\iff$ it does not satisfy the MSNE in Equations 5.1 to 5.33”.

(1) We first prove that $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ does not solve the variational inequality (5.35) $\implies$ it does not satisfy the MSNE in Equations 5.1 to 5.33”.

Assume $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*)$ does not solve the variational inequality (5.35). Then there would

exist $(\mathbf{f}, \mathbf{x}, \mathbf{c}, \mathbf{h}) \in \Lambda$ such that the left side of the variational inequality (5.35) has at least one term greater than 0.

When $\sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} (f_{o,s}^{r,*} - f_{o,s}^r) > 0$, it is equivalently represented as Equation C.23 as follows:

$$\sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} f_{o,s}^{r,*} > \sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} f_{o,s}^r \quad (\text{C.23})$$

where $\sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} f_{o,s}^{r,*}$ represents the total travel cost for the vehicles whose origin is in UESN when the path flows are $f_{o,s}^{r,*}$ and $\sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} f_{o,s}^r$ represents the total travel cost for the vehicles whose origin is in UESN if the travel cost is fixed at $c_{o,s}^{r,*}$ while the flows changes to $f_{o,s}^r$.

For the latter to be strictly less than the former, at least one vehicle's travel cost must have decreased, even though the path travel cost does not change. This can only happen if that vehicle is not initially on a minimum travel cost path, meaning that does not satisfy the MSNE when origin is in UESN.

When $\sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} (f_{o,s}^{r,*} - f_{o,s}^r) > 0$, it is equivalently represented as Equation C.24 as follows:

$$\sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} f_{o,s}^{r,*} > \sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} f_{o,s}^r \quad (\text{C.24})$$

where $\sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} f_{o,s}^{r,*}$ represents the total marginal cost for the vehicles whose origin is in SOSN when the path flows are $f_{o,s}^{r,*}$ and $\sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} f_{o,s}^r$ represents the total marginal cost for the vehicles whose origin is in UESN if the marinal cost is fixed at $h_{o,s}^{r,*}$ while the flows changes to $f_{o,s}^r$.

For the latter to be strictly less than the former, at least one vehicle's marginal cost must have decreased, even though the path marginal cost does not change. This can only happen if that vehicle is not initially on a minimum marginal cost path, meaning that does not satisfy the MSNE when origin is in SOSN.

When $\sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} (f_{k,s}^{r,*} - f_{k,s}^r) > 0$, it is equivalently represented as Equation C.25 as follows:

$$\sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} f_{k,s}^{r,*} > \sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} h_{o,s}^{r,*} f_{k,s}^r \quad (\text{C.25})$$

where $\sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} f_{k,s}^{r,*}$ represents the total marginal cost for the vehicles entering SOSN from UESN when the path flows are $f_{k,s}^{r,*}$ and $\sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} h_{o,s}^{r,*} f_{k,s}^r$ represents the total marginal cost for the vehicles entering SOSN from UESN if the marginal cost is fixed at $h_{k,s}^{r,*}$ while the flows changes to $f_{o,s}^r$.

For the latter to be strictly less than the former, at least one vehicle's marginal cost must have decreased, even though the path marginal cost does not change. This can only happen if that vehicle is not initially on a minimum marginal cost path, meaning that does not satisfy the MSNE when vehicles enter SOSN from UESN.

When $\sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} (f_{k,s}^{r,*} - f_{k,s}^r) > 0$, it is equivalently represented as Equation C.26 as follows:

$$\sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} f_{k,s}^{r,*} > \sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} h_{o,s}^{r,*} f_{k,s}^r \quad (\text{C.26})$$

where $\sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} f_{k,s}^{r,*}$ represents the total travel cost for the vehicles entering UESN from SOSN when the path flows are $f_{k,s}^{r,*}$ and $\sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} f_{k,s}^r$ represents the total travel cost for the vehicles entering UESN from SOSN if the travel cost is fixed at $c_{k,s}^{r,*}$ while the flows changes to $f_{k,s}^r$.

For the latter to be strictly less than the former, at least one vehicle's travel cost must have decreased, even though the path travel cost does not change. This can only happen if that vehicle is not initially on a minimum travel cost path, meaning that does not satisfy the MSNE when vehicles enter UESN from SOSN.

(2) Then we would prove that $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ does not solve the variational inequality (5.35) $\iff$ it does not satisfy the MSNE in Equations 5.1 to 5.33".

If $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*) \in \Lambda$ does not satisfy the MSNE in Equations 5.1 to 5.33, then we have the following equations:

$$\begin{aligned}
& \sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} \delta(f_{o,s}^{r,*} > 0) \delta(c_{o,s}^{r,*} > \pi_{o,s}) + \sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} \delta(f_{o,s}^{r,*} > 0) \delta(h_{o,s}^{r,*} > \eta_{o,s}) \\
& + \sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} \delta(f_{k,s}^{r,*} > 0) \delta(h_{k,s}^{r,*} > \pi_{k,s}) + \sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} \delta(f_{k,s}^{r,*} > 0) \delta(c_{k,s}^{r,*} > \pi_{k,s}) > 0
\end{aligned}$$

where $\delta(\cdot)$ is an indicator function. For example, if $f_{o,s}^{r,*} > 0$, then $\delta(f_{o,s}^{r,*} > 0) = 1$. Otherwise, we have $\delta(f_{o,s}^{r,*} > 0) = 0$. Besides, $\pi_{o,s}$ and $\eta_{o,s}$ represents the minimum travel cost and marginal cost from o to s , respectively.

The above equation indicates that at least one component on the left side is greater than 0. Since the proof process is similar, we assume that $\sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} \delta(f_{o,s}^{r,*} > 0) \delta(c_{o,s}^{r,*} > \pi_{o,s}) > 0$.

Therefore, $\exists o \in O \cap \Phi, s \in S, \hat{r} \in P_{o,s}$ such that $f_{o,s}^{\hat{r},*} > 0$ and $c_{o,s}^{\hat{r},*} > \pi_{o,s}$. We construct a new path flow vector $\mathbf{g}$ that is the same as f^* , except except that $g_{o,s}^{\hat{r}}$ is reduced by a small positive amount ϵ and $g_{o,s}^{\bar{r}}$ is increased by ϵ , where $\bar{r} \in P_{o,s}$, $f_{o,s}^{\bar{r},*} > 0$ and $\bar{r} \neq \hat{r}$. For convenience, we assume that both $\hat{r}$ and $\bar{r}$ do not pass through SOSN. If $\hat{r}$ and $\bar{r}$ pass through SOSN, the proof process is similar. Then we have the Equation C.27.

$$\begin{aligned}
& \sum_{o \in O \cap \Phi} \sum_{s \in S} \sum_{r \in P_{o,s}} c_{o,s}^{r,*} (f_{o,s}^{r,*} - g_{o,s}^r) + \sum_{o \in O \cap \Psi} \sum_{s \in S} \sum_{r \in P_{o,s}} h_{o,s}^{r,*} (f_{o,s}^{r,*} - g_{o,s}^r) + \\
& \sum_{k \in \Omega} \sum_{s \in S} \sum_{r \in P_{k,s}} h_{k,s}^{r,*} (f_{k,s}^{r,*} - g_{k,s}^r) + \sum_{k \in \Theta} \sum_{s \in S} \sum_{r \in P_{k,s}} c_{k,s}^{r,*} (f_{k,s}^{r,*} - g_{k,s}^r) \\
& = c_{o,s}^{\hat{r},*} (f_{o,s}^{\hat{r},*} - g_{o,s}^{\hat{r}}) + c_{o,s}^{\bar{r},*} (f_{o,s}^{\bar{r},*} - g_{o,s}^{\bar{r}}) \\
& = \epsilon (c_{o,s}^{\hat{r},*} - c_{o,s}^{\bar{r},*}) > 0
\end{aligned} \tag{C.27}$$

Therefore, $(\mathbf{f}^*, \mathbf{x}^*, \mathbf{c}^*, \mathbf{h}^*)$ does not solve the variational inequality (5.35).

C.5 Proof of Proposition 23

$c_{o,s}^{r,*}$, $h_{o,s}^{r,*}$, $c_{k,s}^{r,*}$ and $h_{k,s}^{r,*}$ could be represented as the function of the path vector $\mathbf{f}^*$ and link vector $\mathbf{x}^*$.

$$\begin{aligned}
c_{o,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*) &= c_{o,s}^{r,*} = \sum_{(i,j) \in L} c_{i,j}^* \delta_{i,j}^r \quad \text{for } o \in O \cap \Phi, s \in S, r \in P_{o,s} \\
&= \sum_{(i,j) \in L} \delta_{i,j}^r t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}})
\end{aligned} \tag{C.28}$$

$$\begin{aligned}
h_{o,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*) &= h_{o,s}^{r,*} = \sum_{(i,j) \in L} h_{i,j}^* \delta_{i,j}^r \quad \text{for } o \in O \cap \Psi, s \in S, r \in P_{o,s} \\
&= \sum_{(i,j) \in L} \delta_{i,j}^r t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}}) + t_{i,j}^f \alpha_{i,j} \beta_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}}
\end{aligned} \tag{C.29}$$

$$\begin{aligned}
c_{k,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*) &= c_{k,s}^{r,*} = \sum_{(i,j) \in L} c_{i,j}^* \delta_{i,j}^r \quad \text{for } k \in \Omega, s \in S, r \in P_{o,s} \\
&= \sum_{(i,j) \in L} \delta_{i,j}^r t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}})
\end{aligned} \tag{C.30}$$

$$\begin{aligned}
h_{k,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*) &= h_{k,s}^{r,*} = \sum_{(i,j) \in L} h_{i,j}^* \delta_{i,j}^r \quad \text{for } k \in \Theta, s \in S, r \in P_{o,s} \\
&= \sum_{(i,j) \in L} \delta_{i,j}^r t_{i,j}^f (1 + \alpha_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}}) + t_{i,j}^f \alpha_{i,j} \beta_{i,j} (\frac{x_{i,j}^*}{H_{i,j}})^{\beta_{i,j}}
\end{aligned} \tag{C.31}$$

Therefore, $c_{o,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*)$, $h_{o,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*)$, $c_{k,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*)$ and $h_{k,s}^{r,*}(\mathbf{f}^*, \mathbf{x}^*)$ are the continuous for the path vector $\mathbf{f}^*$ and link vector $\mathbf{x}^*$.

For the feasible set Λ , (5.2), (5.4), (5.8), (5.10), (5.14), (5.16), (5.24), (5.26), (5.31) are linear for the path vector $\mathbf{f}^*$ and link vector $\mathbf{x}^*$. Besides, $\mathbf{f}^*$ and $\mathbf{x}^*$ are bounded by the demand D . Therefore, Λ is a compact and convex set for $\mathbf{f}^*$ and $\mathbf{x}^*$. The MSNE problem has least one solution (Harker and Pang, 1990).