SYSTEM MODELING FOR CONNECTED AND AUTONOMOUS VEHICLES

by

Jian Wang

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Lyles School of Civil Engineering

West Lafayette, Indiana

December, 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Srinivas Peeta, Chair
Lyles School of Civil Engineering
Dr. Konstantina “Nadia” Gkritza
Lyles School of Civil Engineering
Agricultural and Biological Engineering
Dr. Shaoshuai Mou
School of Aeronautics and Astronautics
Dr. Vijay Raghunathan
Electrical and Computer Engineering

Approved by:
Dr. Dulcy Abraham
Head of the Graduate Program

Dedicated to my beloved parents and fiancée

ACKNOWLEDGMENTS

I would like to convey my sincere gratitude to my advisor and mentor Prof. Srinivas Peeta, for his expert supervision and guidance throughout my Ph.D. study. I am deeply grateful to him for his support, patience, and suggestions during all phases of the research. His devotion and enthusiasm for education and research have been and will always be an inspiration in my life.

I would like to express my appreciation to my Ph.D. advisory committee: Prof. Konstantina "Nadia" Gkritza, Prof. Vijay Raghunathan, Prof. Shaoshuai Mou. I would also like to thank Dr. Xiaozheng He in Rensselaer Polytechnic Institute for his help and collaborations during my Ph.D. study.

I would like to thank my labmates – Dr. Yonghoon Kim, Mr. Yuntao Guo, Mr. Choungryeol Lee, Mr. Shubham Agrawal, Ms. Irina Benedyk and other members of the NEXTRANS Center for their support and company. They have made my time at Purdue University memorable. Moreover, I thank my fiancée, Lili Lu; she has always encouraged me during difficult times and provided me useful guidance. Her kind love and support have made this accomplishment possible for me.

I would like to thank my beloved and respected parents. They have sacrificed a lot to build a better future for me. I am blessed to have such kind, understanding, and supportive people in my life.

This dissertation is partially supported by the NSF grant "Coordinated real-time traffic management based on dynamic information propagation and aggregation under connected vehicle systems", the NEXTRANS Center (the former USDOT Region 5 University Transportation Center), and the Center for Connected and Automated Transportation (CCAT), the current Region 5 University Transportation Center, Award #69A3551747105. However, the findings reflect the opinions of the authors alone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Wang, Jian, PhD.
Institution: Purdue University
Degree Received: December 2018
Title: System Modeling for Connected and Autonomous Vehicles
Committee Chair: Srinivas Peeta

Connected and autonomous vehicle (CAV) technologies provide disruptive and transformational opportunities for innovations toward intelligent transportation systems. Compared with human driven vehicles (HDVs), the CAVs can reduce reaction time and human errors, increase traffic mobility and will be more knowledgeable due to vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. CAVs' potential to reduce traffic accidents, improve vehicular mobility and promote eco-driving is immense. However, the new characteristics and capabilities of CAVs will significantly transform the future of transportation, including the dissemination of traffic information, traffic flow dynamics and network equilibrium flow. This dissertation seeks to realize and enhance the application of CAVs by specifically advancing the research in three connected topics: (1) modeling and controlling information flow propagation within a V2V communication environment, (2) designing a real-time deployable cooperative control mechanism for CAV platoons, and (3) modeling network equilibrium flow with a mix of CAVs and HDVs.

Vehicular traffic congestion in a V2V communication environment can lead to congestion effects for information flow propagation due to full occupation of the communication channel. Such congestion effects can impact not only whether a specific information packet of interest is able to reach a desired location, but also the timeliness needed to influence traffic system performance. This dissertation begins with exploring spatiotemporal information flow propagation under information congestion effects, by introducing a two-layer macroscopic model and an information packet relay control strategy. The upper layer models the information dissemination in the information flow regime, and the lower layer model captures the impacts of traffic flow dynamics on information propagation. Analytical and numerical solutions of the information flow propagation wave (IFPW) speed are provided, and the density of informed vehicles is derived under different traffic conditions. Hence, the proposed model can be

leveraged to develop a new generation of information dissemination strategies focused on enabling specific V2V information to reach specific locations at specific points in time.

In a V2V-based system, multiclass information (e.g., safety information, routing information, work zone information) needs to be disseminated simultaneously. The application needs of different classes of information related to vehicular reception ratio, the time delay and spatial coverage (i.e., distance it can be propagated) are different. To meet the application needs of multiclass information under different traffic and communication environments, a queuing strategy is proposed for each equipped vehicle to disseminate the received information. It enables control of multiclass information flow propagation through two parameters: 1) the number of communication servers and 2) the communication service rate. A two-layer model is derived to characterize the IFPW under the designed queuing strategy. Analytical and numerical solutions are derived to investigate the effects of the two control parameters on information propagation performance in different information classes.

Third, this dissertation also develops a real-time implementable cooperative control mechanism for CAV platoons. Recently, model predictive control (MPC)-based platooning strategies have been developed for CAVs to enhance traffic performance by enabling cooperation among vehicles in the platoon. However, they are not deployable in practice as they require anembedded optimal control problem to be solved instantaneously, with platoon size and prediction horizon duration compounding the intractability. Ignoring the computational requirements leads to control delays that can deteriorate platoon performance and cause collisions between vehicles. To address this critical gap, this dissertation first proposes an idealized MPC-based cooperative control strategy for CAV platooning based on the strong assumption that the problem can be solved instantaneously. It then develops a deployable model predictive control with first-order approximation (DMPC-FOA) that can accurately estimate the optimal control decisions of the idealized MPC strategy without entailing control delay. Application of the DMPC-FOA approach for a CAV platoon using real-world leading vehicle trajectory data shows that it can dampen the traffic oscillation effectively, and can lead to smooth deceleration and acceleration behavior of all following vehicles.

Finally, this dissertation also develops a multiclass traffic assignment model for mixed traffic flow of CAVs and HDVs. Due to the advantages of CAVs over HDVs, such as reduced value of time, enhanced quality of travel experience, and seamless situational awareness and connectivity,

CAV users can differ in their route choice behavior compared to HDV users, leading to mixed traffic flows that can significantly deviate from the single-class HDV traffic pattern. However, due to a lack of quantitative models, there is limited knowledge on the evolution of mixed traffic flows in a traffic network. To partly bridge this gap, this dissertation proposes a multiclass traffic assignment model. The multiclass model captures the effect of knowledge level of traffic conditions on route choice of both CAVs and HDVs. In addition, it captures the characteristics of mixed traffic flow such as the difference in value of time between HDVs and CAVs and the asymmetry in their driving interactions, thereby enhancing behavioral realism in the modeling. New solution algorithms will be developed to solve the multiclass traffic assignment model. The study results can assist transportation decision-makers to design effective planning and operational strategies to leverage the advantages of CAVs and manage traffic congestion under mixed traffic flows.

This dissertation deepens our understanding of the characteristics and phenomena in domains of traffic information dissemination, traffic flow dynamics and network equilibrium flow in the age of connected and autonomous transportation. The findings of this dissertation can assist transportation managers in designing effective traffic operation and planning strategies to fully exploit the potential of CAVs to improve system performance related to traffic safety, mobility and energy consumption.

CHAPTER 1. INTRODUCTION

1.1 Background

In the past century, the automobile significantly enhanced the mobility of goods and people and expedited economic development. However, the increasing ownership of automobiles in the past few decades has brought three major challenges to the transportation system: traffic accidents, mobility issues, and environmental pollution. In 2016, one and one-quarter million people died worldwide due to traffic accidents, while in the US, traffic accidents took almost 40,000 peoples' lives and injured over four million people. Traffic congestion caused almost seven billion hours of travel delay in the US and an equivalent $160 billion lost in terms of productivity; additionally, it caused almost three billion gallons of additional fuel consumption and 56 billion pounds of additional carbon dioxide (Lapuerta et al., 2017)—all of which, to a certain extent, contribute to global warming, as is becoming evident today. Unfortunately, most of these issues are due to human failings; for example, over 90% percent of traffic accidents are due to human error. Heterogeneous driving behaviors and human reaction time dramatically reduce the mobility of traffic flow. There is little if any cooperation between drivers concerning routes and timing, which increases the possibility for traffic congestion. Further, drivers generally do not drive at an environmentally-friendly speed; nor do they choose 'eco-routes,' often—all adding to air pollution.

The emerging technologies of connected and autonomous vehicles have great potential to address the challenges of traffic accidents, mobility issues, and environmental pollution. A connected vehicle can exchange information and data with other connected vehicles and infrastructures through vehicle-to-vehicle (V2V) communication and vehicle-to-infrasturcture (V2I) communication, respectively. In the V2V-based system, vehicles themselves generate information and relay it to other vehicles through a 'multi-hop' process, enabling drivers to make informed decisions by providing them with the real-time information needed to understand the evolving traffic network conditions better. Autonomous vehicles can independently detect the surrounding environment using a variety of advanced sensors for self-driving—all without human control, enhancing the mobility of the less-able population (such as the elderly) and improving traffic safety by reducing human error. Compared to regular human-driven vehicles

(HDVs), a connected and autonomous vehicle (CAV) (a vehicle that is equipped with both autonomous and connected technologies) will significantly alleviate the three major issues in transportation system due to the following characteristics:

(1). Fewer errors. The CAV can drive itself with fewer human errors due to advances in computing and sensing technologies, thereby dramatically reducing traffic accidents.

(2). Greater knowledge of traffic conditions. CAVs can generate and exchange information with each other through V2V communication in real time. Additionally, they can access the information from the control center through V2V communication. Hence, CAVs may be better informed about traffic conditions compared to HDVs.

(3). Greater mobility. CAVs can form a 'platoon' to drive cooperatively with each other, significantly increasing link capacity and reducing traffic congestion.

(4). Less reaction time. CAVs can process information faster than human beings, especially from multiple sources, leading to smaller reaction times.

(5). Easier to control traffic. CAVs can choose routes cooperatively to avoid possible traffic congestion.

The novel characteristics and capabilities of CAVs will significantly transform the future of transportation, including the dissemination of traffic information, traffic flow dynamics, and network equilibrium flow. Further, many technical barriers and research questions need to be addressed to fully exploit the potential of CAVs. In this dissertation, we seek to realize and enhance the application of CAVs by specifically advancing knowledge through the following three interrelated topics:

(1). Modeling and controlling information flow propagation in V2V communication environments. Understanding the characteristics of spatiotemporal information flow propagation in a V2V-based traffic system is important, as most applications require timely and reliable information delivery. However, modeling information flow propagation in space and time is challenging. Factors in both traffic and communication regimes significantly affect the reliability of V2V communication and information spread. The traffic flow dynamics contribute to the occurrence of V2V communications. The communication constraints such as communication frequency, channel capacity, and communication range etc. significantly contribute to the reliability of V2V communications. A better understanding of the effects of factors in both traffic flow and information flow

regimes on spatiotemporal propagation of information flow will provide insights to aid in the design of traffic management strategies built upon V2V communications.

(2). Designing cooperative control strategies for 'platoons' of CAVs. CAV platoons can benefit transportation systems in many ways, including the need to increase road capacity, reduce energy consumption and tailpipe emissions, and facilitate V2V-based applications (e.g., involving data sharing and dissemination) due to the relatively fixed position of the vehicles within a platoon. The control model for a platoon of CAVs would include adaptive cruise control (ACC) and cooperative adaptive cruise control (CACC), designed to optimize an individual vehicle's performance. The behavior of the vehicles in the platoon controlled by ACC and CACC models are non-cooperative. In comparison, the model predictive control (MPC)-based cooperative control strategies can coordinate the behaviors (accelerations or decelerations) of all of the following vehicles in a CAV platoon to maximize the platoon performance, including efficiency, safety, and comfort, based on information collected from these vehicles through V2V communications, enabling CAVs in a platoon to collaborate with each other and operate under a common goal. Prior studies suggest that these strategies can lead to smoother deceleration behavior and more responsive and agile acceleration behavior compared with non-cooperative controllers.

(3). Developing a multiclass traffic assignment model for mixed traffic flow with CAVs and HDVs. CAVs offer users the potential to save valuable time, enhance the quality of the travel experience, and create seamless situational awareness and connectivity. CAV users might differ in their route choice behavior compared to HDV users, leading to mixed traffic flows that could significantly deviate from the single-class HDV traffic pattern. The multiclass traffic assignment model characterizes the interactions of route choices of CAVs and HDVs, providing a modeling framework to estimate the network equilibrium flow in the transition period when both CAVs and HDVs exist. Thereby, it assists transportation decision-makers to design effective planning strategies to reduce traffic congestion under mixed traffic flows.

As shown in Figure 1-1, the relationships of the three topics can be summarized in three ways. First, information flow propagation can significantly impact the dynamics of CAV platoons through providing CAVs with information, such as adjacent vehicles' kinematic states. The platoon of CAVs will also impact V2V communications as the traffic flow dynamics

significantly contribute to the occurrence and reliability of V2V communications. Second, a platoon of CAVs can enhance mobility, increase link capacity, and reduce travel time and energy consumption. These advantages will impact the route choices of CAVs. Additionally, the multiclass traffic assignment model can estimate the distribution of both CAV and HDV flows in the network. It will be useful to deploy AV-dedicated lanes to improve the system performance; therein, providing CAVs the useful information related to the locations where platooning is possible. Third, information flow propagation can provide drivers real-time traffic information, which can impact their route choice behavior, significantly changing the network flows.

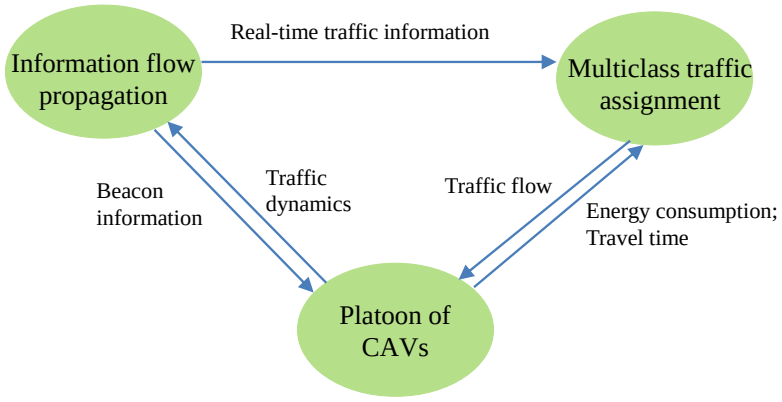


Figure 1.1 The framework of the three research topics

Although recent literature has sought to address the modeling needs of the three research topics (see e.g., Kim et al., 2017a; Wang et al., 2014; Gong and Du, 2018; Chen et al., 2016; Chen et al., 2017; Levin and Boyles, 2015), the characteristics and emerging phenomena arising from CAVs are not fully captured by these models. Thereby, gaps exist between the capabilities of the existing models and the application needs of CAVs in different traffic and communication environments.

(1). A unified theory to model information flow propagation under congested V2V communication environments is lacking. Due to limited channel capacity, a V2V communication channel can be fully occupied in certain traffic and communication environments, preventing vehicles from sending and receiving information. Effective strategies are needed to prevent congestion; therefore, a modeling framework is needed to characterize the information flow propagation in space and time under congested V2V communication environments.

(2). There are no effective mechanisms to control multiclass information propagation under V2V communications. Most of the existing studies in the domain of information flow propagation are only descriptive: they lack control of information flow propagation. Thereby, propagation performance related to information spread (defined as the proportion of vehicles informed with a specific information packet), time delay bounds, and spatial coverage (i.e., distance it can be propagated) may not satisfy the application needs of the information under certain traffic and communication environments. Additionally, V2V-based systems may need to simultaneously propagate multiclass information (e.g., safety information, routing information, and work zone information); therefore, the application needs related to propagation performance can be different. For example, urgent traffic accident information (e.g., road is blocked by an accident) needs to be delivered to all vehicles in the impacted area with low latency. By comparison, routing information needs to reach only a certain proportion of vehicles to avoid possible congestion arising from the provision of information on the suggested route. Also, work zone information or sudden hard brake information may need to be propagated in a small area in the vicinity of where they are generated. Research efforts are needed to design effective mechanisms to control multiclass information propagation under V2V communications.

(3). Currently, there is no real-time deployable cooperative control mechanism for a platoon of CAVs. The existing MPC-based cooperative control mechanisms are not deployable in practice as they require the embedded optimal control problem to be solved instantaneously, with platoon size and prediction horizon duration compounding the intractability. Ignoring the computational requirements leads to control delays that can deteriorate platoon performance, and may cause collisions between vehicles. Therefore, a new cooperative control mechanism is needed to enable effective and efficient control of CAV platoons in real time.

(4). Also, there is no traffic assignment model currently that captures the characteristics of CAVs and the interactions between the route choices of CAVs and HDVs. Compared to HDVs, CAVs can reduce the value of time, increase link capacity and reduce energy consumption by forming a platoon. In addition, due to V2V and V2I communications, CAVs can understand traffic conditions better than HDVs. The effects of these characteristics of CAVs on route choice of both CAVs and HDVs have not been fully captured by existing models.

1.2 Objectives of the dissertation

The objectives of this dissertation are threefold: (1) development of an integrated framework for modeling and controlling information flow propagation under V2V communication environments; (2) design of a cooperative control strategy for CAV platoons, and (3) modeling network equilibrium under the mixed traffic flow of CAVs and HDVs. This dissertation seeks to provide traffic managers efficient and effective traffic operational and planning strategies in the age of connected and autonomous transportation to systematically reduce travel costs and improve traffic mobility and safety. These specific objectives will be achieved through the following tasks:

(1). Develop a macroscopic model to characterize the spatiotemporal propagation of information flow in a congested V2V communication environment. To address congestion effects associated with information flow propagation induced by communication channel capacity being fully occupied, an information flow relay control strategy will be proposed. This strategy seeks to exclude information that is dated in the communication buffer under a first-in, first-out queue discipline, from being relayed if the information flow regime is congested. It trades off the need to enable the dissemination of every information packet as far as possible, against the congestion effects that accrue because of the presence of multiple information packets. A macroscopic model will be developed to characterize the information flow propagation wave (IFPW) for an information packet in a congested V2V communication environment under the designed information relay control strategy. The model will help to analytically determine solutions for asymptotic IFPW speed and density of vehicles informed with the specific information of interest, which can aid in the design of traffic management strategies built upon the timely propagation of information through V2V communications.

(2). Develop a queuing-based modeling approach to control the propagation of multiclass information under V2V communication environments. A queuing strategy will be developed for each V2V equipped vehicle to propagate the received information packets in multiple classes. This strategy will enable the control of propagation performance of multiclass information related to information spread, time delay bounds, and spatial coverage to meet the heterogeneous application needs of this information. The analytical model will be used

to generate insights on the effects of the control parameters in the queuing strategy on information flow propagation speed and density of informed vehicles.

(3). Develop a real-time deployable MPC-based cooperative platooning strategy for platoons of CAVs to maximize performance with regard to safety, mobility and comfort. An idealized MPC-based cooperative control strategy is proposed for CAV platooning based on the strong assumption that the imbedded optimal control problem can be solved instantaneously. The idealized MPC strategy can coordinate the behaviors of all following CAVs in the platoon efficiently and maneuver them safely. To address the issue of control delay of the idealized MPC strategy induced by the computational time for solving the embedded optimal control problem, a deployable model predictive control (DMPC) with first-order approximation (DMPC-FOA) approach will be developed. It not only addresses the issues of control delay of the idealized MPC strategy efficiently, but can also accurately characterize the optimal control decisions of the idealized MPC strategy. Thereby, it can be applied for real-time cooperative control of a CAV platoon. The conditions for stability of the idealized MPC strategy will also be discussed to better dampen traffic oscillations in the platoon.

(4). Develop a multiclass traffic assignment model for mixed traffic flow of CAVs and HDVs. This model considers the effects of both CAVs and HDVs on travel costs—including the heterogeneous value of time, the asymmetric travel costs of HDVs and CAVs, and the energy saved by CAVs through platooning on AV-dedicated lanes. Additionally, it also characterizes the interactions of route choices between CAVs and HDVs due to different knowledge levels of traffic conditions. The proposed multiclass traffic assignment model can provide better behavioral realism in charactering the mixed traffic flow, and can assist transportation decision-makers to design effective planning and operational strategies to leverage the advantages of CAVs to manage traffic congestion under mixed traffic flows.

1.3 Organization of the dissertation

This dissertation consists of six chapters. Figure 1.2 provides an overview and illustrates the relationship between the four main chapters. Chapter 2 proposes an information relay control strategy to address the information congestion effects that accrue because of the presence of multiple information packets. A two-layer framework is developed to model the information

flow propagation under congested V2V communication environments. It consists of integro-differential equations in the upper layer to model the information flow dissemination due to V2V communications and partial differential equations in the lower layer to characterize traffic flow dynamics. Based on the two-layer model, the analytical solutions of IFPW speed, density of informed vehicles and necessary conditions for the existence of IFPW are derived under homogeneous traffic conditions (i.e., unidirectional traffic flow with uniform density). A numerical method is also developed to solve the two-layer model to estimate the spatiotemporal dissemination of information under heterogeneous traffic conditions.

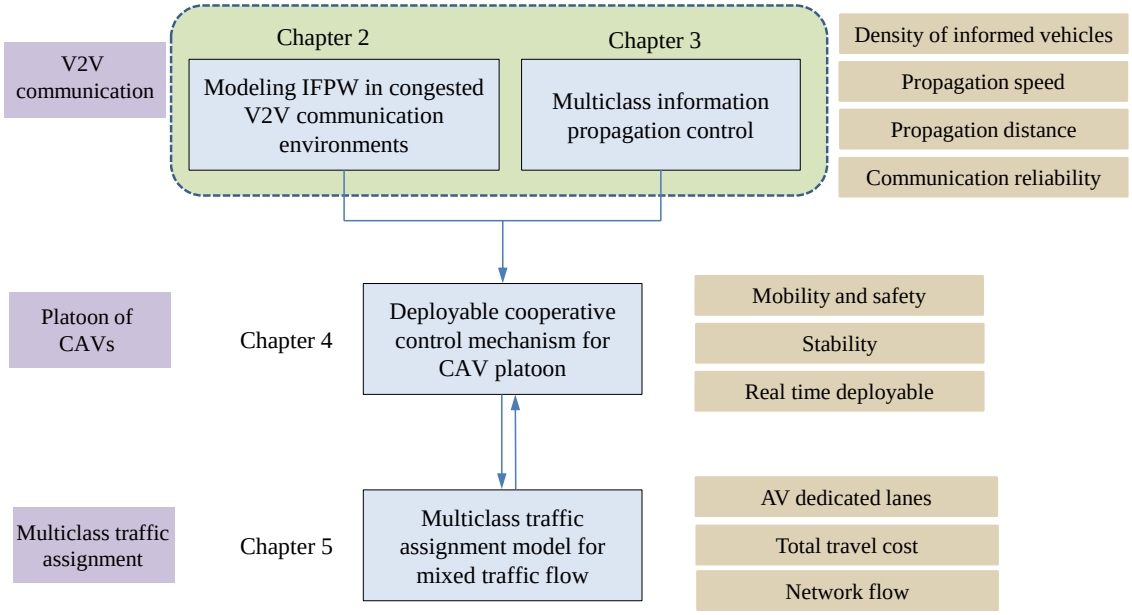


Figure 1.2 Organization of the dissertation

Chapter 3 proposes a control method for the propagation of multiclass information to meet the application needs of information of each information class related to information spread, time delay bounds, and spatial coverage under different traffic and communication environments. The control method is constructed on a queuing strategy developed for equipped vehicles to disseminate the received information packets, enabling control of the propagation of multiclass information through two control parameters: number of communication servers and communication service rates. A two-layer model is derived to characterize the information flow propagation in space and time under the designed queuing strategy. Analytical and numerical solutions will be developed to solve the two-layer model under different traffic flow conditions,

which provide insights on the effects of the two control parameters on information flow propagation performance.

Chapter 4 proposes an idealized MPC-based cooperative control strategy for CAV platooning based on the strong assumption that the embedded optimal control problem can be solved instantaneously. It also proposes a solution algorithm for the embedded optimal control problem to maximize platoon performance. It then develops two approaches to deploy the idealized strategy: the deployable MPC (DMPC) and the DMPC with first-order approximation (DMPC-FOA), that estimate the optimal control decisions of the idealized MPC strategy without entailing control delay. However, under the DMPC approach, the optimal control decisions of the following vehicles in the platoon may deviate significantly from those of the idealized MPC strategy, while the DMPC-FOA approach can accurately characterize them. An analytical method is derived for the sensitivity analysis of the optimal control decisions. Further, stability analysis is performed for the idealized MPC strategy, and a sufficient condition is derived to ensure its asymptotic stability under certain conditions.

Chapter 5 develops a multiclass traffic assignment model for mixed traffic flow with CAVs and HDVs. The route choices of HDV and CAV users are characterized by the cross-nested logit (CNL) model and user equilibrium (UE) model, respectively. The CNL model captures HDV users' uncertainty associated with limited knowledge of traffic conditions while overcoming the route overlap issue of logit-based stochastic user equilibrium. The UE model characterizes CAV's capability for acquiring accurate information on traffic conditions. Additionally, the multiclass model can capture the characteristics of mixed traffic flow, such as the difference in value-of-time between HDVs and CAVs, and the asymmetry in their driving interactions—thereby enhancing behavioral realism in the modeling. A route-swapping-based solution algorithm embedded with a self-regulated step size choice technique is proposed to solve the proposed model efficiently. Sensitivity analysis of the proposed model is performed to gain insights into the effects of perturbations on the mixed traffic equilibrium, which facilitates the estimation of traffic state and identification of critical elements under expected or unexpected events.

Chapter 6 concludes this dissertation with a summary of the insights. Novelties and significant contributions are identified. Finally, potential directions for future research are discussed.

CHAPTER 2. ANALYTICAL MODEL FOR INFORMATION FLOW PROPAGATION WAVE UNDER AN INFORMATION RELAY CONTROL STRATEGY IN A CONGESTED VEHICLE-TO-VEHICLE COMMUNICATION ENVIRONMENT

2.1 Introduction

Over the past few decades, dynamic traffic assignment models have sought to address how information affects traffic flow (e.g., Mahmassani and Peeta, 1995; Peeta and Yu, 2002; Paz and Peeta, 2009). In the context of vehicle-to-vehicle (V2V) communication, we seek to address how traffic flow impacts information propagation so as to develop a new generation of information dissemination strategies that can ensure that information reaches a desired location at specific time in order to achieve systemwide or individual level objectives for the V2V-based traffic system.

Connectivity in a V2V-based traffic system can be enabled by two types of communication: periodic and event-driven. Information broadcast by either type of communication is sealed into an information packet. Periodic communication using an information packet, often labeled beacon message, is used to proactively broadcast a vehicle's position, speed, heading, brake status, and other data to all neighboring vehicles. Such information can be of critical importance to the "receiver" vehicles; for example, it can enable the detection of an unsafe road condition. A beacon message is characterized by its short lifetime that spans only one hop or a few hops of broadcasting communication, with a high frequency of up to 10 times per second. Hence, beacon messages are appropriate for communicating with the local neighbors of the "sender" vehicle, and can be leveraged to alert drivers of potential collisions and hazards by providing up-to-date status information (Yang et al., 2004; Yeo et al, 2010; Talebpour et al., 2014).

Event-driven communication is triggered by a specific event, such as accident, sudden brake, or congestion, etc. It is useful for warning vehicles to approach the affected area with caution or adopt an alternate route to their destinations (Ding et al., 2010). It requires the multi-hop dissemination of an information packet that contains information related to the event (such as congestion or route guidance). The information packet released by the sender can be relayed by other vehicles inside and outside the initial sender's communication range, depending on the traffic characteristics and dynamics. Therefore, vehicles store the received information packet in

their communication buffers and retransmit it to other vehicles. Since traffic safety and efficiency related applications have requirements related to information coverage and latency, it is critical to understand how information propagates in space and time in a V2V-based traffic system. This study considers the event-driven communication context, where multiple information packets, triggered by different events, are propagated simultaneously through multi-hop dissemination.

Modeling the information flow propagation through a multi-hop dissemination mechanism is challenging for a V2V-based traffic system, because the following characteristics need to be modeled appropriately. First, factors in both the information flow regime (such as information packet generation rate, communication frequency, and communication buffer size) and the traffic flow regime (such as traffic speed and density) significantly affect the characteristics of information flow propagation (Kim and Peeta, 2016; Kim et al., 2016; Kim and Peeta, 2017b; Du and Dao, 2015; Du et al., 2016). The density of information flow changes with the traffic dynamics. In addition, due to limited channel capacity, the large number of information packets generated by multiple vehicles in a small space and a short time period can lead to congestion in the information flow regime, even if the traffic flow regime is not congested. For example, information flow congestion can exhibit trailing effects even if the traffic flow congestion dissipates. Second, congested information flow can cause a high degree of mutual interference among the transmitted signals. Any information packet that propagates to other vehicles in the vicinity of a vehicle is subject to signal attenuation over distance and interference imposed by other signals transmitted from surrounding vehicles. Third, a simple broadcasting protocol can lead to an exponential growth of retransmitted messages that congest a network, referred to as a broadcast storm (Tseng et al., 2002; Karagiannis et al., 2011). This phenomenon can cause packet collisions, implying that neither can the information stored in the communication buffer be disseminated nor can other information packets be stored into this buffer. Therefore, the multi-hop dissemination requires a special information relay control strategy to prevent packet collisions, so that information packets can share the limited channel capacity efficiently.

Past studies have addressed specific aspects of information flow propagation, including expected information propagation distance (Jin and Recker, 2006; Wang, 2007; Wang et al., 2010; Wang et al., 2011; Wang et al., 2012; Yin et al., 2013; Wang et al., 2015), connectivity of inter-vehicle communication (Jin et al., 2006; Ukkusuri and Du, 2008; Jin and Recker, 2010),

and throughput of information packets to be transmitted to a given distance (Chen et al., 2010). However, all of these analytical studies are based on the assumption of instantaneous spatial propagation of information; that is, they do not consider the time dimension. They simply assume that the vehicles' locations are known based on some space headway distribution. Thereby, these approaches lack realism as they do not consider the impact of traffic flow dynamics on information propagation. Further, most of these analytical approaches oversimplify the wireless communication constraints (e.g., communication range, communication frequency, channel capacity, signal interference, etc.). Therefore, while the assumption of instantaneous information propagation can be analytically convenient, it has limitations in characterizing information flow propagation in the real world.

To facilitate the analysis of information propagation characteristics and the impacts of traffic dynamics at an aggregate level, Kim et al. (2015, 2017b) introduce the concept of an information flow propagation wave (IFPW). When an information packet is generated in a V2V-based traffic system, it spreads through the relay process of multi-hop communications. From a macroscopic perspective, an IFPW "front" forms a moving boundary that separates the traffic flow into informed and uninformed regions, and moves towards the uninformed region (Kim et al., 2017b). This IFPW can be characterized by the direction and speed of the moving boundary. The quantification of speed and position of the IFPW front provides the macroscopic characteristics of information flow propagation. Kim et al. (2017b) proposed an analytical model to describe information propagation in space and time, to quantify the speed and position of the IFPW front. It incorporates modeling realism by not assuming instantaneous information propagation, and by considering V2V communication constraints that are consistent with the real world. They apply different approaches to solve the proposed two-layer model for homogeneous conditions and heterogeneous conditions. The homogeneous conditions refer to the situation (i.e., unidirectional traffic flow with uniform traffic flow density) when the impacts of the traffic flow dynamics on the IFPW speed are uniform in space and time. The heterogeneous conditions refer to situation where the impacts of the traffic flow dynamics on the IFPW speed are not uniform in space and time. The analytical model provides capabilities to: (i) derive a closed-form solution of IFPW speed under homogeneous conditions, (ii) develop a numerical method to estimate the IFPW speed for heterogeneous conditions, (iii) incorporate the effects of congested traffic, such as the backward traffic propagation wave, on information flow propagation, (iv) capture V2V

communication constraints in a realistic manner using a communication kernel, (v) illustrate the linkage between information flow propagation and the underlying traffic dynamics, and (vi) factor the impacts of communication constraints on the success rate of V2V communication, which is consistent with real-world V2V communications.

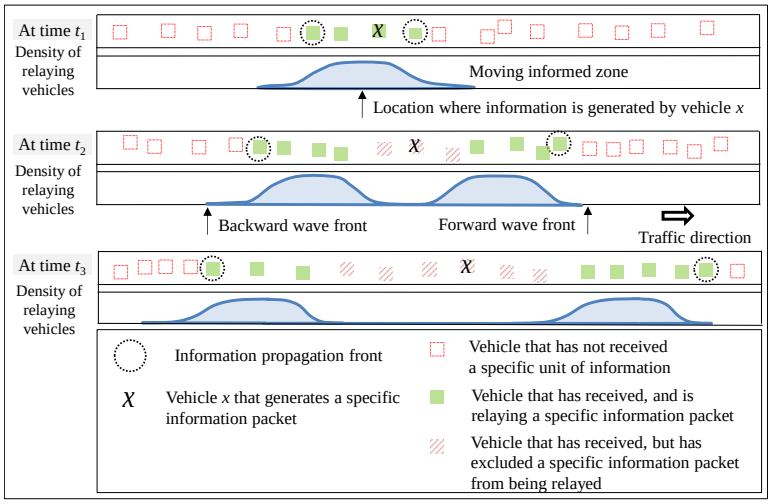


Figure 2.1 Information flow propagation wave under the information relay control strategy

The Kim et al. (2017b) model considers the case where any V2V information can always be transmitted and retransmitted, implying the absence of information flow congestion. The proposed study models the more complex, and realistic case that involves information flow congestion effects, arising from the broadcast storm problem, whereby an equipped vehicle receives a large amount of information and retransmits it to other vehicles. It proposes an information relay control strategy to address these congestion effects, and models the IFPW under this strategy. To the best of our knowledge, this is the first study to analytically model and characterize V2V-based information flow propagation under congestion effects. Additionally, it also factors the effects of traffic flow dynamics.

As stated earlier, an information relay control strategy is essential to prevent the endless broadcasting of the same message and collisions of information packets. This study assumes that all information packets have the same priority to be propagated in the information flow regime. Under this assumption, we propose an "information exclusion" relay control strategy, whereby if the communication buffer is full when new information packets arrive, the first-in-first-out (FIFO) queue discipline is applied to remove the earlier information packets from the buffer, and exclude them from being relayed.

The proposed information relay control strategy leads to special characteristics for the IFPW under information flow congestion. As illustrated in Figure 2.1, some informed vehicles (i.e., those that have received the information of interest) cannot retransmit the specific information packet as it is removed from their communication buffer. Thereby, only a subset of informed vehicles may serve as relay vehicles to retransmit the information packet of interest. This causes the formation of two information-relaying waves, one in the direction of traffic and the other opposite to it.

This study conceptually extends the IFPW proposed by Kim et al. (2017b), and proposes a new analytical modeling approach, to address information flow propagation under congestion effects while factoring traffic flow dynamics. It enables the exploration of the following three questions in a congested V2V communication environment. First, under what traffic and information flow-related conditions can the specific information packet propagate in the network? These conditions relate to the density of V2V-equipped vehicles and the rate at which event-driven information packets are generated in the traffic stream. This question addresses the necessary conditions for the formation of an IFPW. Second, how fast can the specific information packet of interest be transmitted? This question is addressed by determining the asymptotic IFPW speed. Third, what is the number of equipped vehicles that can receive the specific information packet when the proposed information relay control strategy is applied? This question seeks to determine the asymptotic density of informed vehicles, which is the number of vehicles that can receive the specific information packet if the IFPW exists. By comparison, Kim et al. (2017b) address only the second question, and for the uncongested information flow regime.

The proposed two-layer model consists of an information dissemination model in the upper layer and a traffic flow model in the lower layer. A Susceptible-Relay-Excluded (SRX) model, which conceptually extends a Susceptible-Infected-Recovered model (Kendall, 1957) for disease spreading in epidemiology, is proposed to characterize the spatiotemporal information propagation flow in the upper layer. It uses integro-differential equations (IDEs) to model the information flow propagation under the proposed information relay control strategy. The lower layer adopts the Lighthill-Whitham-Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) to describe the traffic flow dynamics. The two layers are linked through the density of the V2V-equipped vehicles.

The contributions of this study are as follows. First, we develop an analytical model to characterize the spatiotemporal propagation of an information packet under information flow congestion. This model factors the information relay control strategy and constraints related to V2V communication and traffic flow dynamics. Thereby, the proposed model addresses the realism issues identified earlier, in characterizing information flow propagation for V2V-based traffic systems. Second, the necessary conditions for the formation of an IFPW are derived under the homogeneous conditions. When the necessary conditions are satisfied, solutions for the asymptotic IFPW speed and the asymptotic density of informed vehicles are analytically obtained. These analytical solutions facilitate the exploration of the impact of traffic and information flow-related factors, including traffic flow density and average duration for which a specific information packet is relayed, on the IFPW speed and the density of informed vehicles. It provides valuable insights for the design of V2V communication-based applications, especially under dense V2V communication environments. Third, the numerical solution method for the proposed two-layer model enables the analysis of information propagation under heterogeneous conditions. This enables characterizing information flow propagation for general traffic conditions.

The remainder of this paper is organized as follows. The next section discusses the modeling framework of the proposed model to characterize the IFPW. Section 3 formulates a two-layer model to characterize the IFPW in space and time under the proposed information relay control strategy. Section 4 analytically solves for the asymptotic IFPW speed and the asymptotic density of informed vehicles under homogeneous conditions. Section 5 presents the numerical solution method for the proposed two-layer model for heterogeneous conditions. Results from numerical experiments are discussed in Section 6, to demonstrate the effectiveness of the proposed two-layer model. Section 7 provides some concluding comments.

2.2 Preliminaries

2.2.1 Modeling framework

The IFPW is the combination of two propagation waves in two regimes: the information dissemination wave in the information flow regime and the traffic flow propagation wave in the traffic flow regime. As illustrated in Figure 2.2, the upper layer is used to capture the information

dissemination wave and the lower layer is used to capture the traffic flow propagation wave. In particular, the lower layer model describes the traffic flow dynamics by adapting the hydrodynamic LWR model. The LWR model has been used extensively due to its capability for capturing key real-world traffic flow phenomena such as shock waves and spillbacks. The traffic flow propagation is formulated as a system of partial differential equations (PDEs). Vehicles in this layer are grouped into two mutually exclusive classes: equipped vehicles *E* and unequipped vehicles *U*, based on the capability for V2V communication. The traffic layer is coupled with the information (upper) layer through the density of equipped vehicles. As shown in Figure 2.2, while the equipped vehicles are associated with both information flow propagation (through the communication success rate) and traffic flow dynamics, the unequipped vehicles are associated only with the traffic flow dynamics.

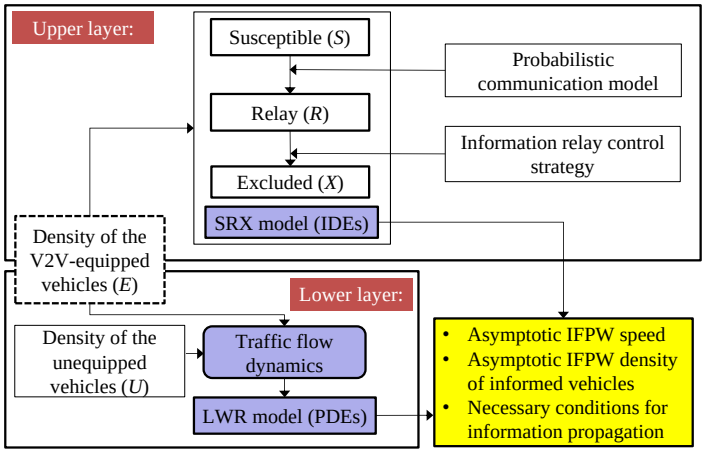


Figure 2.2 The modeling framework for information flow propagation

The upper layer model describes the information dissemination dynamics among equipped vehicles in time and space. To characterize the information dissemination dynamics, the equipped vehicle set *E* is divided into three disjoint sets, denoted by *S*, *R*, and *X*. Notation *S* denotes the set of information-susceptible vehicles; they are the V2V-equipped vehicles that have not received the information of interest. The set of information-relay vehicles, denoted by *R*, refers to equipped vehicles that receive the information and store it in communication buffer so as to relay it to susceptible vehicles. To prevent information packet collisions, information relay vehicles may exclude some information packets from the communication buffer based on the information relay control strategy. If an information relay vehicle excludes the specific

information packet from its communication buffer, implying that this equipped vehicle can no longer relay the information of interest, then it becomes an information-excluded vehicle, denoted by *X*. Applying this concept to the upper layer modeling, the information dissemination dynamics is formulated as a spatial Susceptible-Relay-Excluded (SRX) model, which is motivated by the analogy between the dissemination of information among equipped vehicles and the spread of an infectious disease among individuals.

As stated earlier, the communication signal decays over distance due to the mutual interference from equipped vehicles. Kim et al. (2017b) propose a probabilistic communication model that incorporates the success rate of single-hop communication based on the distance between equipped vehicles. This study adopts the communication kernel that represents the success rate of single-hop communication as a function of distance between two equipped vehicles. After an equipped vehicle receives information, it acts as a relay vehicle that stores information in the communication buffer and retransmits it to surrounding susceptible vehicles. Note that while the information relay control strategy can prevent information packet collisions, it impacts the success rate of information dissemination as some susceptible vehicles may not receive the information packet before all surrounding relay vehicles exclude the information packet. The next section presents the information relay control strategy that will be adopted in the modeling framework.

2.2.2 Information relay control strategy

Consider a highway with a stream of traffic flow consisting of equipped and unequipped vehicles. Different event-driven information packets are generated and propagated by equipped vehicles through multi-hop V2V communications. Each equipped vehicle broadcasts its received information packets to surrounding susceptible vehicles within communication range. When an equipped vehicle receives multiple packets, it filters the information packets to identify those that have not been received before. It then moves them into the communication buffer, where information packets are temporarily stored for retransmission, as shown in Figure 2.3. This study applies a first-in, first out (FIFO) information relay control strategy that excludes the information packet stored for the longest time in the communication buffer from being transmitted, when a new information packet is moved into it.

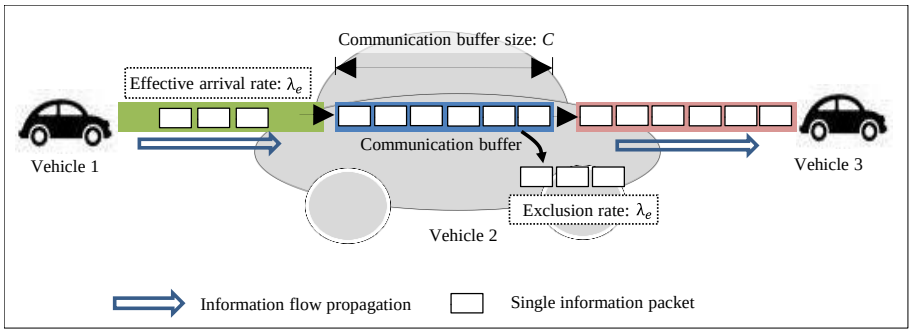


Figure 2.3 Illustration of information relay control strategy from communication buffer

Let parameter λ_e be the *effective* information arrival rate, which represents the number of unduplicated information packets moved into the communication buffer per unit time. The effective information arrival rate is affected by unsuccessful V2V communication and removal of duplicated information packets (i.e., information packets that have already been received before). Because such events are random and independent, this study assumes that the arrival of effective information packets into the communication buffer follows a Poisson process. Denote C as the communication buffer size, which is defined as the maximum number of information packets that can be stored in the communication buffer. This study assumes that each equipped vehicle equally shares the channel capacity, and is able to send all information packets in the communication buffer through each V2V communication. Assume that the communication buffer is full under the situation of congested information flow. When the FIFO queue discipline is applied, the information exclusion follows a Poisson process with parameter λ_e for each equipped vehicle. Under this information relay control strategy, the duration of information packet in the communication buffer follows an exponential distribution with the mean C/λ_e. The mean duration of an information packet in the communication buffer is estimated under different densities of equipped vehicles in Section 6.1. A detailed description of the SRX model is provided in the next section.

2.3 Modeling the spatiotemporal propagation of information flow

2.3.1 Modeling information dissemination flow in the upper layer under uncongested V2V communication

The IFPW is the combination of two propagation waves in two layers: the information dissemination wave in the upper layer and the traffic flow propagation wave in the lower layer.

Consider that a specific information packet of interest is generated initially by an equipped vehicle located at location 0 at time 0. This vehicle disseminates the information packet to the surrounding vehicles through the multi-hop broadcasting process. For an uncongested V2V communication environment (that is, no information congestion), where any V2V information can always be transmitted and retransmitted, Kim et al. (2015, 2017b) proposed a spatial susceptible-informed model to describe the dissemination of a specific piece of information between equipped vehicles as follows:

$$\frac{\partial S(x,t)}{\partial t} = -\beta \cdot S(x,t) \cdot \int_{\Omega} K(x,y) \cdot I(y,t) dy \tag{2.1a}$$

$$\frac{\partial I(x,t)}{\partial t} = -\frac{\partial S(x,t)}{\partial t}, \tag{2.1b}$$

where $I(x,t)$ and $S(x,t)$ denote the density of informed vehicles and the density of susceptible vehicles at time t and location x, respectively. Parameter β denotes the communication frequency and Ω denotes the spatial domain. Function $K(x,y)$ is a communication kernel which represents the probability that a susceptible vehicle at location y successfully receives an information packet sent from an information-relay vehicle at location x. It is a decreasing function with respect to the distance between the two equipped vehicles. The communication kernel factors the impacts of communication constraints (e.g. communication range, communication frequency, channel capacity, signal interference, etc.) on the success rate of V2V communication. The term $\int_{\Omega} K(x,y) \cdot I(y,t) dy$ denotes the aggregate probability that a susceptible vehicle located at point x successfully receives the specific information from an informed vehicle over the space domain Ω.

2.3.2 Information dissemination flow in the upper layer under congested V2V communication

This study models the information dissemination flow with the information relay control strategy in a congested V2V communication; that is, it considers information flow congestion unlike Kim et al. (2017b). In a congested V2Vcommunication environment, an equipped vehicle receives and relays a large amount of information packets through the limited channel capacity. The communication buffer of the equipped vehicle can be full through accumulating information packets. To prevent the endless broadcasting of the same message and collisions of information

packets, the information relay control strategy is applied, whereby if an effective information packet is received and enters into the communication buffer, the FIFO rule would be used to exclude the information packet stored for the longest time in the communication buffer from being transmitted.

Let t be the current time. Denote w as the time step for the discrete-time model. Without loss of generality, suppose $t = nw$, where $n \in \mathbb{N}^+$ is a positive integer. As stated previously, under the information control strategy, informed vehicles consist of information-relay vehicles and information-excluded vehicles. Suppose an informed vehicle located at x receives this information packet at $\tau = iw$ time units ago, $i \in \mathbb{N}^+$. At the current time t, the probability for an informed vehicle to broadcast this specific information packet to a nearby susceptible vehicle, denoted by $H(x, t, \tau)$, is equal to the probability that the information packet received τ time units ago is not excluded from the communication buffer. The information exclusion rate is $\lambda = \lambda_e / C$. Based on the assumption that the information packet exclusion follows the Poisson process, the probability that an informed vehicle remains an information-relay vehicle in the past τ time units satisfies:

$$H(x, t, \tau) = e^{-\lambda\tau}. \tag{2.2}$$

Denote $\Delta S(x, t) = S(x, t) - S(x, t - w)$, which represents the number of susceptible vehicles at location x becoming informed in time interval $[t - w, t]$. Let $\Delta I(y, t - \tau)$ be the density change of informed vehicles at location y at time $t - \tau$. Note that $\Delta I(y, t - \tau) = -\Delta S(y, t - \tau)$, because any equipped vehicle is either susceptible or informed. Conceptually adapting and modifying from epidemiology, $\Delta S(x, t)$ is formulated as:

$$\begin{aligned} \Delta S(x, t) &= -S(x, t - \tau) \cdot \sum_{i=1}^{n} \int_{\Omega} \Delta I(y, t - iw) \cdot K(x, y) \cdot w\beta \cdot H(y, t, iw) dy \\ &= -S(x, t - \tau) \cdot \sum_{i=1}^{n} \int_{\Omega} \Delta I(y, t - iw) \cdot K(x, y) \cdot w\beta \cdot e^{-\lambda \cdot iw} dy, \end{aligned} \tag{2.3}$$

where $w\beta$ denotes the number of transmissions occurring in w time units. The term $\int_{\Omega} \Delta I(y, t - \tau) \cdot K(x, y) \cdot w\beta \cdot e^{-\lambda\tau} dy$ denotes the probability that susceptible vehicles at location x receive the information packet sent by vehicles that received this packet τ time units ago, over the space domain Ω. It can be analytically shown that if $H(y, t, iw) \equiv 1$ and $w \to 0$, equation (2.3) converts into equation (2.1a).

We derive a continuous-time formulation by taking $w \to 0$ in equation (2.3). Divide both sides of equation (2.3) by w, and take the limit as $w \to 0$. Then, we have:

$$\begin{aligned}\lim_{w\to 0}\frac{\Delta S(x,t)}{w} &= \frac{\partial S(x,t)}{\partial t}\\ &= \lim_{w\to 0} -S(x,t-w)\cdot\sum_{i=1}^{n}\int_{\Omega}\frac{\Delta I(y,t-iw)}{w}\cdot K(x,y)\cdot w\beta\cdot e^{-\lambda\cdot iw}dy.\end{aligned} \tag{2.4}$$

Note that

$$\Delta I(y,t-iw) = I(y,t-iw) - I(y,t-(i-1)w) \approx \frac{\partial I(y,t-iw)}{\partial t}w. \tag{2.5}$$

Equation (2.4) can be written as:

$$\frac{\partial S(x,t)}{\partial t} = \lim_{w\to 0} -S(x,t-w)\cdot\sum_{i=1}^{n}\int_{\Omega}\frac{\partial I(y,t-iw)}{\partial t}\cdot K(x,y)\cdot w\beta\cdot e^{-\lambda\cdot iw}dy. \tag{2.6}$$

Recall that $\tau = iw$. Then $(\partial I(y,t-iw)/\partial t)\cdot w$ denotes the number of equipped vehicles that become informed during time interval $[\,t-\tau-w, t-\tau]$. Note that $(\partial I(y,t-\tau)/\partial t)K(x,y)e^{-\lambda\tau}$ is continuous and bounded in the time domain. Therefore, it is Riemann integrable. Equation (2.6) can be reformulated as:

$$\frac{\partial S(x,t)}{\partial t} = -\beta\cdot S(x,t)\cdot\int_{\Omega}\int_{0}^{t}\frac{\partial I(y,t-\tau)}{\partial t}\cdot K(x,y)\cdot e^{-\lambda\tau}d\tau dy. \tag{2.7}$$

As $\Delta I(y,t-\tau) = -\Delta S(y,t-\tau)$ implies that $\partial I(x,t-\tau)/\partial t = -\partial S(x,t-\tau)/\partial t$. Therefore,

$$\frac{\partial S(x,t)}{\partial t} = \beta\cdot S(x,t)\cdot\int_{\Omega}\int_{0}^{t}\frac{\partial S(y,t-\tau)}{\partial t}\cdot K(x,y)\cdot e^{-\lambda\tau}d\tau dy. \tag{2.8}$$

Let $R(y,t) = -\int_0^t \frac{\partial S(y,t-\tau)}{\partial t}\cdot e^{-\lambda\tau}d\tau$, which denotes the density of information-relay vehicles at location y at time t, i.e., vehicles with the information packet in their communication buffer. Let $u = t-\tau$, then

$$R(y,t) = -\int_0^t \frac{\partial S(y,u)}{\partial t}e^{-\lambda(t-u)}du. \tag{2.9}$$

The partial derivative of $R(y,t)$ with respect to t has the following form

$$\begin{aligned}\frac{\partial R(y,t)}{\partial t} &= -\frac{\partial S(y,t)}{\partial t} + \int_0^t \frac{\partial S(y,u)}{\partial t}\cdot\lambda\cdot e^{-\lambda(t-u)}du\\ &= -\frac{\partial S(y,t)}{\partial t} - \lambda\cdot R(y,t).\end{aligned} \tag{2.10}$$

Let $X(x,t)$ denote the density of information-excluded vehicles at location x at time t. As informed vehicles include the information-relay vehicles and information-excluded vehicles, $I(x,t) = R(x,t) + X(x,t)$. Based on this definition, we have:

$$\frac{\partial R(x,t)}{\partial t} + \frac{\partial X(x,t)}{\partial t} = \frac{\partial I(x,t)}{\partial t} = -\frac{\partial S(x,t)}{\partial t}. \tag{2.11}$$

Substituting equation (2.11) into equation (2.10) yields:

$$\frac{\partial X(x,t)}{\partial t} = \lambda \cdot R(x,t). \tag{2.12}$$

Equation (2.12) specifies the rate at which the information-relay vehicles exclude the specific information packet from the communication buffer at time t and become information-excluded vehicles. Substituting $R(y,t) = -\int_0^t \frac{\partial S(y,t-\tau)}{\partial t} \cdot e^{-\lambda\tau} d\tau$ into equation (2.8) yields:

$$\frac{\partial S(x,t)}{\partial t} = -\beta \cdot S(x,t) \cdot \int_\Omega R(y,t) \cdot K(x,y) dy. \tag{2.13}$$

Hence, the information dissemination dynamics under the given information relay control strategy can be formulated as the following integro-differential equations (IDEs)

$$\begin{cases} \dfrac{\partial S(x,t)}{\partial t} = -\beta \cdot S(x,t) \cdot \displaystyle\int_\Omega R(y,t) \cdot K(x,y) dy & \text{(2.14a)} \\ \dfrac{\partial R(x,t)}{\partial t} = \beta \cdot S(x,t) \cdot \displaystyle\int_\Omega R(y,t) \cdot K(x,y) dy - \lambda \cdot R(x,t) & \text{(2.14b)} \\ \dfrac{\partial X(x,t)}{\partial t} = \lambda \cdot R(x,t). & \text{(2.14c)} \end{cases}$$

We label the IDE system (2.14) as the Susceptible-Relay-Excluded (SRX) model.

The IDE system (2.14) conceptually extends the susceptible-infected-removed epidemic model that was originally proposed by Kendall (1957, 1965) to describe the spatiotemporal spreading of a disease. If $R(x,t) \geq 0$ and $\lambda > 0$ in equation (2.14c), then $X(x,t)$ is a non-decreasing function with respect to time. As time passes, the specific information packet becomes dated in the communication buffer, and the probability of excluding it from the communication buffer increases. Given a sufficiently long time, all the information-relay vehicles at a location will become information-excluded vehicles if $\lambda > 0$. That is, $X(x,t) \to I(x,t)$ and $R(x,t) \to 0$, as $\lambda > 0$ and $t \to \infty$. Note that if $\lambda = 0$, equation (2.14) is equivalent to equation (2.1), denoting an uncongested V2V communication environment.

2.3.3 Traffic flow model

The evolution of IFPW depends not only on the characteristics of communication in the information layer but also on the traffic flow dynamics in the traffic layer for the following reasons. First, the communication kernel (i.e., $K(x, y)$) is associated with traffic flow density. A high equipped-vehicle density can induce high communication interference. Second, the spatiotemporal vehicle dynamics impacts the IFPW speed. This study applies the first-order LWR model to describe the evolution of traffic flow in the space and time, which is able to reproduce some essential features of traffic flow, such as the formation and propagation of traffic waves. It assumes that the behavior of traffic at a given point in space and time is only affected by the state of the system in a neighborhood of that point (Daganzo, 1995). The model consists of the flow conservation law and an explicit density-flow relationship known as the fundamental diagram of traffic flow. The flow conservation law and the fundamental diagram can be expressed as the following PDE model:

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{2.15}$$

$$u(x,t) = F(k,x,t), \tag{2.16}$$

where $k(x,t)$ is the traffic flow density at location x at time t, $q(x,t)$ is the instantaneous flow, and $F(k,x,t)$ is the fundamental diagram in which $k(x,t)$ and the space mean speed $u(x,t)$ are related by a continuous and piecewise differentiable equation. The PDE system is able to model the backward and forward traffic shock waves induced by perturbation of the traffic system.

2.3.4 Solutions for homogeneous and heterogeneous conditions

For homogeneous conditions, the traffic flow layer (equations (2.15) and (2.16)) is a single class traffic flow model that is sufficient to characterize the impacts of traffic flow dynamics on the IFPW. This is because when traffic flow is unidirectional and the traffic flow density is uniform, the traffic flow dynamics only shift the IFPW towards the direction of traffic flow and do not change the densities of vehicles of different classes (see Kim et al. (2017b) for more details). Thereby, the impacts of the traffic flow dynamics on the IFPW speed and density of informed vehicles are uniform in space and time. The two layers are part of a coupled system in which dynamics occur in each system simultaneously, consisting of the information

dissemination dynamics in the upper layer and the traffic flow dynamics in the lower layer. The two-layer model (equations (2.14), (2.15) and (2.16)) can be solved sequentially under continuous time setting. The closed-form solution for the IFPW speed and asymptotical density of informed vehicles will be derived in section 4 based on the two-layer model.

When traffic flow is heterogeneous, the densities of vehicles of different classes can be changed by the moving traffic flow (e.g., bidirectional traffic flow). The impacts of the traffic flow dynamics on the IFPW speed are not uniform in space and time (Kim et al., 2017b). This makes it necessary to track the density of vehicles in each class in each cell in the lower layer, which requires discretizing the two-layer model. A numerical solution method will be proposed in section 5 to capture the interactions between the upper and lower layers sequentially under discrete time settings.

2.4 Asymptotic solutions of IFPW speed and density of informed vehicles under homogeneous conditions

2.4.1 Asymptotic IFPW speed and asymptotic density of informed vehicles

The asymptotic IFPW speed is the sum of the asymptotic information dissemination wave speed and the traffic flow propagation wave speed. Let $c > 0$ denote the information dissemination wave speed. In epidemiology literature, the existence of traveling wave solution of the spatial Susceptible-Infected-Recovered model has been well studied. Kendall (1965) and Mollison (1972) proved that if the kernel function $K(x, y)$ satisfies some conditions (discussed later), the traveling waves exist when c is greater than a minimum value. Atkinson and Reuter (1976) analyzed Kendall's model for a general class of averaging kernels and obtained a criterion for the existence of the minimum traveling wave speed. Aronson (1977) proved that the minimum traveling wave speed is the asymptotic speed, and derived the traveling wave solution for any time. In the following, we aim to derive the conditions for existence of the IFPW, the corresponding asymptotic information dissemination wave speed and asymptotic density of informed vehicles under a given traffic flow density and the proposed relay control strategy.

As discussed in Section 3.1, the communication kernel $K(x, y)$ significantly impacts the traveling wave solutions. By factoring communication constraints, Kim et al. (2017b) showed that communication is subject to attenuation over distance. The communication kernel $K(x, y)$ is a decreasing function of communication distance. This study adopts the Gaussian

communication kernel which is a special form of the Nakagami model with Nakagami shape parameter of 0.5 (Naveen and Rajeswari, 2011). We use the Gaussian communication kernel based on Kim et al., (2017b) who found it to be a good fit for the one-hop success rate of V2V communication.

$$K(x,y) = \frac{b}{a\sqrt{\pi}} e^{\frac{-(x-y)^2}{a^2}},\ a > 0,\ 0 < b \leq 1, \tag{2.17}$$

where a and b are parameters to be estimated using real-world V2V communication data under given communication constraints and traffic flow density. Let

$$\overline{K}(x,y) = \frac{1}{a\sqrt{\pi}} e^{\frac{-(x-y)^2}{a^2}},\ a > 0. \tag{2.18}$$

Substituting equation (2.17) and (2.18) into equation (2.14) yields:

$$\begin{cases} \dfrac{\partial S(x,t)}{\partial t} = -\beta b \cdot S(x,t) \cdot \displaystyle\int_{\Omega} R(y,t) \cdot \overline{K}(x,y)dy \\ \dfrac{\partial R(x,t)}{\partial t} = \beta b \cdot S(x,t) \cdot \displaystyle\int_{\Omega} R(y,t) \cdot \overline{K}(x,y)dy - \lambda R(x,t) \\ \dfrac{\partial X(x,t)}{\partial t} = \lambda \cdot R(x,t)\,. \end{cases} \tag{2.19}$$

Suppose information propagation starts at time 0. Then, $I(x,0) = R(x,0) + X(x,0) = 0$. All equipped vehicles at different locations are susceptible vehicles. As the traffic flow is homogeneous, the initial density of susceptible vehicles at different locations is the same. Let σ denotes the density of susceptible vehicles, i.e., $S(x,0) = \sigma$. Note that σ is also the density of equipped vehicles. Denote the initial value for $R(x,0)$ as $R_0(x)$. Aronson (1977) derived the solution for the spatial Susceptible-Infected-Recovered model with population density normalized to one. To apply the Corollary developed by Aronson (1977) for deriving the solution of the SRX model (3.14), $S(x,t)$, $R(x,t)$, $X(x,t)$ and t are rescaled as follows:

$$\bar{S}(x,t) = \frac{1}{\sigma} S(x,t) \tag{2.20a}$$

$$\bar{R}(x,t) = \frac{1}{\sigma} R(x,t) \tag{2.20b}$$

$$\bar{X}(x,t) = \frac{1}{\sigma} X(x,t) \tag{2.20c}$$

$$\bar{t} = \frac{1}{\sigma\beta b} t. \tag{2.20d}$$

Denote $\gamma = \lambda/(\sigma\beta b)$. Substituting equations (2.20) into equation (2.14), we have the following scaled SRX model:

$$\begin{cases} \dfrac{\partial \bar{S}(x,\bar{t})}{\partial \bar{t}} = -\bar{S}(x,\bar{t}) \cdot \displaystyle\int_{\Omega} \bar{R}(x,\bar{t}) \cdot \bar{K}(x,y)dy \\ \dfrac{\partial \bar{R}(x,\bar{t})}{\partial \bar{t}} = \bar{S}(x,\bar{t}) \cdot \displaystyle\int_{\Omega} \bar{R}(x,\bar{t}) \cdot \bar{K}(x,y)dy - \gamma \bar{R}(x,\bar{t}) \\ \dfrac{\partial \bar{X}(x,\bar{t})}{\partial \bar{t}} = \gamma \cdot \bar{R}(x,\bar{t}), \end{cases} \tag{2.21}$$

with initial conditions: $\bar{S}(x,0) = 1, \bar{R}(x,0) = R_0(x)/\sigma$, and $\bar{X}(x,0) = 0$.

Recall that communication kernel $K(x,y)$ is a function of the distance between the sender and receiver vehicles. It can be written as the convolution type, i.e., $\bar{K}(z)$ $(z = |x-y|)$. To derive an analytical solution for equation (2.21), Aronson (1977) proved that kernel $\bar{K}(z)$ must satisfy four conditions:

(C1) $\bar{K}(z)$ is a nonnegative even function defined in $\mathbb{R}$ with $\int_{-\infty}^{\infty} \bar{K}(z)\, dz = 1$.

(C2) There exists a $\nu \geq 0$, such that $\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\, dz < \infty$ for all $\mu \in [0,\nu]$.

(C3) Define $A_\gamma(\mu) = \frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\, dz - \gamma\right]$. For each $\gamma < 1$, there exists a sufficiently large $\nu > 0$ and a $\mu^* = \mu^*(\gamma)$, such that $0 < \bar{c}^* \equiv A_\gamma(\mu) = \inf\{A_\gamma(\mu): \mu \in [0,\nu]\}$, $\partial_\mu A_\gamma(\mu) < 0$ for $\mu \in [0,\mu^*]$ and $\partial_\mu A_\gamma(\mu) > 0$ for $\mu \in (\mu^*,\nu)$, where $\partial_\mu A_\gamma(\mu)$ is the first order partial derivative of $A_\gamma(\mu)$ with respect to μ.

(C4) For any $\tilde{\mu} \in [0,\nu]$, there exists a $r = r(\tilde{\mu}) \geq 0$ such that $e^{\mu x}\bar{K}(x) = \min\{e^{\mu y}\bar{K}(y): y \in [0,x]\}$ for all $\mu \in [0,\tilde{\mu}]$ and $x \geq r(\tilde{\mu})$.

Theorem 2.1: Communication kernel function $\bar{K}(z)$, defined by equation (2.18), satisfies conditions (C1)-(C4).

Proof: Step 1: $\bar{K}(z)$ satisfies condition (C1).

Proof: $\int_{-\infty}^{\infty} \bar{K}(z)\, dz = \int_{-\infty}^{\infty} \frac{1}{a\sqrt{\pi}} e^{\frac{-z^2}{a^2}} dz = 1$; condition (C1) holds.

Step 2: $\bar{K}(z)$ satisfies condition (C2).

Proof: $\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\, dz = \int_{-\infty}^{\infty} e^{\mu z} \frac{1}{a\sqrt{\pi}} e^{\frac{-z^2}{a^2}} dz = \frac{1}{a\sqrt{\pi}} e^{0.25a^2\mu^2} \times \left(a\sqrt{\pi}\right) = e^{0.25a^2\mu^2}$.

Thereby, for any given $\nu \geq 0$, $\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\, dz < \infty$ for all $0 \leq \mu \leq \nu$.

Step 3: $\overline{K}(z)$ satisfies condition (C3).

Proof: Note that

$$\partial_\mu \mathrm{A}_\gamma(\mu) = -\frac{1}{\mu^2}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz - \gamma\right] + \frac{1}{\mu}\int_{-\infty}^{\infty} z\cdot e^{\mu z}\overline{K}(z)\,dz$$

$$= -\frac{1}{\mu^2}\left[\sqrt{2}e^{0.25a^2\mu^2} - \gamma\right] + \frac{1}{a\sqrt{\pi}}\frac{1}{\mu}e^{0.25a^2\mu^2}\int_{-\infty}^{\infty} z\cdot e^{-\left(\frac{z}{a}-0.5a\mu\right)^2}\,dz.$$

Let $h = \frac{z}{a} - 0.5a\mu$, then z $= ah + 0.5a^2\mu$

$$\partial_\mu \mathrm{A}_\gamma(\mu) = -\frac{1}{\mu^2}\left[\sqrt{2}e^{0.25a^2\mu^2} - \gamma\right] + \frac{1}{\sqrt{\pi}\mu}e^{0.25a^2\mu^2}\int_{-\infty}^{\infty}(ah + 0.5a^2\mu)\cdot e^{-h^2}dh$$

$$= -\frac{1}{\mu^2}\left[\sqrt{2}e^{0.25a^2\mu^2} - \gamma\right] + e^{0.25a^2\mu^2}\cdot\frac{1}{2}a^2. \tag{2.22}$$

According to equation (2.22), $\partial_\mu \mathrm{A}_\gamma(\mu) \to -\infty$ as $\mu \to 0_+$, and $\partial_\mu \mathrm{A}_\gamma(\mu) \to +\infty$ as $\mu \to +\infty$. Note that $\partial_\mu \mathrm{A}_\gamma(\mu)$ is a continuous function of μ. There must exist a positive ν such that $\partial_\mu \mathrm{A}_\gamma(\mu) < 0$ for $\mu \in [0, \mu^*]$ and $\partial_\mu \mathrm{A}_\gamma(\mu) > 0$ for $\mu \in (\mu^*, \nu)$, where $\mu^* = \mu^*(\gamma)$ is the point at which $\partial_\mu \mathrm{A}_\gamma(\mu^*) = 0$.

Step 4: $\overline{K}(z)$ satisfies condition (C4).

Proof: $e^{\mu\tilde{y}}\overline{K}(\tilde{y}) = \frac{1}{a\sqrt{\pi}}e^{\frac{-\tilde{y}^2}{a^2}+\mu\tilde{y}} = \frac{1}{a\sqrt{\pi}}e^{0.25a^2\mu^2}e^{-\left(\frac{\tilde{y}}{a}-0.5a\mu\right)^2}$. (2.23)

Equation (2.23) implies that $e^{\mu\tilde{y}}\overline{K}(\tilde{y})$ increases monotonically with respect to $\tilde{y} \in (0,0.5a^2\mu]$, and decreases monotonically with respect to $\tilde{y} \in (0.5a^2\mu, +\infty)$. For $\tilde{y} = 0$, we have $e^{\mu\tilde{y}}\overline{K}(\tilde{y}) = \frac{1}{a\sqrt{\pi}}$. For $\tilde{y} = a^2\mu$, $e^{\mu\tilde{y}}\overline{K}(\tilde{y}) = \frac{1}{a\sqrt{\pi}}$. Note that $e^{\mu\tilde{y}}\overline{K}(\tilde{y})$ decreases when $\tilde{y} \in [0.5a^2\mu, +\infty)$. Thereby, $e^{\mu\tilde{y}}\overline{K}(\tilde{y}) < \frac{1}{a\sqrt{\pi}} = \min\{e^{\mu\tilde{y}}\overline{K}(\tilde{y}): \tilde{y} \in [0, a^2\mu]\}$ for $\tilde{y} \in (a^2\mu, +\infty)$. Hence, for a fixed $\tilde{\mu} \in [0, \nu]$, $e^{\mu x}\overline{K}(x) = \min\left\{e^{\mu\tilde{y}}\overline{K}(\tilde{y}): \tilde{y} \in [0, x]\right\}$ for all $\mu \in [0, \tilde{\mu}]$ when $x \geq a^2\tilde{\mu} = r(\tilde{\mu})$. Theorem 2.1 is proved.

Definition 2.1: *Asymptotic wave speed* (Aronson and Weinberger, 1978): Speed $\bar{c}^*$ is called the asymptotic wave speed if, for any $\bar{c}_1$ and $\bar{c}_2$ with $0 < \bar{c}_1 < \bar{c}^* < \bar{c}_2$, the density of infected people (i.e., informed vehicles in our context) tends to be zero uniformly in the region $|x| \geq \bar{c}_2 t$, whereas it is bounded away from zero uniformly in the region $|x| \leq \bar{c}_1 t$ for sufficiently large time t.

Definition 2.2: *Asymptotic density of informed vehicles*: The asymptotic density of informed vehicles at location x is defined as $\lim_{t\to\infty} I(x,t)$ if the limit exists.

Theorem 2.2 (Aronson, 1977): Assume $\bar{K}(z)$ satisfies conditions (C1)-(C4) and $0<\gamma<1$. Let $\bar{X}(x,t)$ be the solution of equation (2.21).

(1) If $\bar{c}>\bar{c}^*$, then for every $x\in\mathbb{R}$

$$\lim_{t\to\infty,|x|\geq ct} \bar{X}(x,t)=0. \tag{2.24}$$

(2) If $0<\bar{c}<\bar{c}^*$, then for every $x\in\mathbb{R}$

$$\lim_{t\to\infty,|x|\geq ct} \bar{X}(x,t)=\alpha(\gamma), \tag{2.25}$$

where $\alpha(\gamma)$ is the unique solution of α in $(0,1)$ of the following equation:

$$1-\alpha=e^{-\alpha/\gamma}. \tag{2.26}$$

And $\bar{c}^*$ is the minimum value of $\mathrm{A}_\gamma(\mu)$ for all $0<\mu<\nu$, i.e.,

$$\bar{c}^*\equiv \mathrm{A}_\gamma(\mu)=\inf\left\{\frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\,dz-\gamma\right]: 0<\mu<\nu\right\}.$$

Remark 2.1 (Necessary conditions for existence of the IFPW): By the definition of asymptotic wave speed, Theorem 2.2 implies that for $0<\gamma<1$, the information dissemination wave solution of equation (2.21) exists and the asymptotic wave speed is $\bar{c}^*$. This provides the necessary condition for existence of the IFPW. In particular, $\gamma=\lambda/(\sigma\beta b)<1$ implies $1/\lambda=C/\lambda_e>1/(\sigma\beta b)$. Recall that C/λ_e represents the mean duration of information packet in the communication buffer. Therefore, condition $\gamma<1$ quantifies the time length to be reserved for broadcasting the information packet of interest such that the IPFW exists.

As equation (2.21) is a scaled version of equation (2.14) with $\bar{X}(x,t)=\frac{1}{\sigma}X(x,t)$, and $\bar{t}=\frac{1}{\sigma\beta b}t$, the following corollary holds.

Corollary 2.1 (Asymptotic information dissemination wave speed): Let c^* be the asymptotical information dissemination speed of equation (2.14) and $I(x,t)$ be the corresponding solution. If $0<\gamma<1$, then

$$c^*=\sigma\beta b\bar{c}^*=\sigma\beta b\cdot\inf\left\{\frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\bar{K}(z)\,dz-\gamma\right]: 0<\mu<\nu\right\}. \tag{2.27}$$

And, for any $0<c<c^*$

$$\lim_{t\to\infty,|x|\geq ct} X(x,t)=\sigma\lim_{t\to\infty,|x|\geq ct}\bar{X}(x+ct,t)=\sigma\cdot\alpha(\gamma). \tag{2.28}$$

Remark 2.2 (Asymptotic density of informed vehicles): Let I^* be the asymptotic density of informed vehicles. Recall that $\lim_{t\to\infty} X(x,t) = \lim_{t\to\infty} I(x,t)$ for $\lambda > 0$. Note that vehicles are moving at a uniform speed. Equation (2.26) implies that, given a sufficiently long time, the density of information-excluded vehicles at any place within the wave front would be $\sigma \cdot \alpha(\gamma)$. It implies that, if $\lambda > 0$,

$$I^* = \lim_{t\to\infty} I(x,t) = \lim_{t\to\infty} X(x,t) = \sigma \cdot \alpha(\gamma) \quad \forall x. \tag{2.29}$$

Recall that σ denotes the density of equipped vehicles. Then, the asymptotic proportion of informed vehicles at each location is $I^*/\sigma = \alpha(\gamma)$.

Equations (2.27) and (2.29) show that the asymptotic information dissemination speed and the asymptotic density of informed vehicles, respectively, can be obtained by solving a nonlinear equation under homogeneous conditions with $0 < \gamma < 1$. Equations (2.27) and (2.29) are one-dimensional nonlinear equations that can be solved using the Newton method. This helps to circumvent the more complex numerical solution methods.

It should be noted here that equation (2.27) is valid even if $\gamma = 0$. Recall that, if $\lambda = 0$, equation (2.14) is reduced to the Susceptible-Informed (SI) model (1). The following theorem presents the asymptotic information dissemination speed for equation (2.27).

Theorem 2.3: If $\lambda = 0$, then $c^* = \sqrt{\frac{e}{2}}\sigma\beta ba$.

Proof: Note that $\gamma = \lambda/(\sigma\beta b)$. Therefore, $\gamma = 0$ if $\lambda = 0$. According to equation (2.27):

$$c^* = \sigma\beta b \cdot \inf\left\{\frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz\right]: 0 < \mu < \nu\right\}.$$

Recall that $\mathrm{A}_\gamma(\mu) = \frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz\right]$, if $\gamma = 0$. We have

$$\begin{aligned}\partial_\mu \mathrm{A}_\gamma(\mu) &= -\frac{1}{\mu^2}\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz + \frac{1}{\mu}\int_{-\infty}^{\infty} z\cdot e^{\mu z}\overline{K}(z)\,dz \\ &= -\frac{1}{\mu^2}\sqrt{2}e^{0.25a^2\mu^2} + e^{0.25a^2\mu^2}\cdot\frac{\sqrt{2}}{2}a^2.\end{aligned} \tag{2.30}$$

Note that $\partial_\mu \mathrm{A}_\gamma(\mu) < 0$, for $\mu \in [0,\sqrt{2}/a)$, $\partial_\mu \mathrm{A}_\gamma(\mu) = 0$ for $\mu = \sqrt{2}/a$, and $\partial_\mu \mathrm{A}_\gamma(\mu) > 0$, for $\mu \in (\sqrt{2}/a, \infty)$. In addition, according to Theorem 2.1, ν needs to be greater than $\sqrt{2}/a$. This indicates that the minimum value of $\mathrm{A}_\gamma(\mu)$ is obtained at $\mu = \sqrt{2}/a$. Then,

$$\bar{c}^* = \inf\{\mathrm{A}_\gamma(\mu): 0 < \mu < \nu\} = \mathrm{A}_\gamma(\sqrt{2}/a) = \frac{1}{\mu}\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz = \frac{a}{\sqrt{2}}\cdot e^{0.5}.$$

Hence, $c^* = \sigma\beta b\bar{c}^* = \sqrt{\frac{e}{2}}\sigma\beta ba$.

Theorem 2.3 shows that the closed-form solution of the asymptotic information dissemination speed provided by equation (2.27) at $\lambda = 0$ is the same as that derived in Kim et al. (2017b). Note that as all informed vehicles would repetitively broadcast the specific information packet, all susceptible vehicles will become informed eventually if $\lambda = 0$. Therefore, the SI model (2.1) is a special case of the proposed model when the information packet is propagated in an uncongested information regime with $\lambda = 0$.

Theorem 2.4 (Aronson, 1977): Let $\bar{c} > 0$ be the speed of the traveling wave (i.e., the information dissemination wave in our context) of equation (2.21). If $\bar{K}(z)$ satisfies conditions (C1)-(C4) and $\gamma \geq 1$, then $\lim_{t\to\infty, |x|\geq \bar{c}t} \bar{R}(x,t) = 0$.

As equation (2.21) is a scaled version of equation (2.14) with $\bar{X}(x,t) = \frac{1}{\sigma}X(x,t)$, the following corollary holds based on Theorem 2.4.

Corollary 2.2: Denote $c > 0$ as the information dissemination speed of the SRX model (2.14). If $\gamma \geq 1$, then $\lim_{t\to\infty, |x|\geq ct} X(x,t) = \sigma \lim_{t\to\infty, |x|\geq ct} \bar{X}(x,t) = 0$.

Remark 2.3 (Conditions at which the IFPW does not exist): As discussed earlier, if the information exclusion rate λ is strictly positive, then all informed vehicles at location x, $I(x,t)$, will become information-excluded vehicles gradually. Corollary 2.2 implies that the information dissemination wave does not exist with $\gamma > 1$, i.e., the positive c does not exist. This indicates that the specific information packet only propagates to a limited area and cannot be received by equipped vehicles that are far away from the initial sender vehicle. Note that $\gamma = \lambda/(\sigma\beta b)$. If the information exclusion rate λ is too high, or the density of susceptible vehicles σ or the communication frequency β is too low, any information packet will not be propagated far away. Namely, they are propagated locally and fade out over distance.

2.4.2 Traffic flow propagation wave speed

The traffic flow propagation wave speed, denoted by c_T, is determined by the fundamental diagram which is a continuous and a non-increasing function defined on $[0, k_j]$, where k_j is the jam density. This study adopts the following triangular fundamental diagram (Jin and Recker, 2006):

$$c_T = \begin{cases} u_f, & 0 \le k \le k_c \\ \dfrac{k_c(k_j - k)}{(k_j - k_c)} u_f, & k_c \le k \le k_j, \end{cases} \tag{2.31}$$

where u_f is the free flow speed, and k_c is the critical density.

We label the IFPW speed in the direction of vehicular traversal as the forward IFPW speed. The IFPW speed opposite to the direction of vehicular traversal is defined as the backward IFPW speed. Note that when $\gamma \geq 1$, the specific information packet cannot be propagated to other equipped vehicles that are far away because of the information relay control strategy. Then, both the forward and backward asymptotic IFPW speeds are 0. When $\gamma < 1$, the information propagation wave speed c^* is positive. The specific information packet keeps propagating to susceptible vehicles in the traffic network. As V2V communications and vehicle movements occur simultaneously, the solution of the forward asymptotic IFPW speed is expressed as $c^* + c_T$, while the solution of the backward asymptotic IFPW speed is expressed as $c_T - c^*$ on the unidirectional highway.

2.5 Numerical solution method for the two-layer model under heterogeneous conditions

The asymptotic speed and aympototic density derived in the previous section rely on the assumption of homogeneous conditions. Under heterogeneous conditions, the IFPW cannot reach a stable state as the communication kernel varies with traffic density and impacts of traffic flow dynamics on IFPW are non-uniform. The analytical solutions represented by equations (2.27) and (2.29) cannot be applied for the case for heterogeneous conditions. In this section, we provide a numerical method to solve for the proposed two-layer model formulated by equations (2.14), (2.15) and (2.16). The numerical solution method can obtain the density variations of the four vehicle classes in the space and time domains.

The numerical solution method divides the space and time domains into cells with length Δx and time interval Δt, respectively. To approximate the solutions of IDEs in equation (14), the Runge-Kutta method is applied at each time step. The Runge-Kutta method adopts the Euler method to approximate the partial derivatives (i.e., the derivatives of $S(x,t)$, $R(x,t)$ and $D(x,t)$ with respect to t at each time step). The main difficulty to approximate the partial derivatives is to calculate the convolution term ($\int_\Omega R(y,t) \cdot K(x,y)dy$) in equation (2.14). In this study, the Fast Fourier Transform (FFT) method is used to calculate this term due to its computational

efficiency. Let $h(s) = \int_{\Omega} R(y,t) \cdot K(x,y)dy$. As the space is discretized into cells, according to discrete FFT theory (Brigham, 1974), we have

$$\begin{aligned} \mathcal{F}h(s) &= \frac{2L}{N}\mathcal{F}Y(s) \cdot \mathcal{F}K(s) \\ &= \sum_{l=-N/2}^{N/2} R(l\Delta x,t)e^{2\pi \cdot j \cdot s \cdot l \cdot \Delta x} \cdot \sum_{l=-N/2}^{N/2} K(l\Delta x)e^{2\pi \cdot j \cdot s \cdot l \cdot \Delta x}, \end{aligned} \tag{2.32}$$

where $\mathcal{F}$ denotes the FFT operator, L is the communication range, and N is the number of cells within communication range in the two directions; j is the imaginary unit that satisfies $j^2 = -1$. According to equation (2.32), the inversion of FFT provides $h(s)$.

The generalized cell transmission finite difference equations proposed by Daganzo (1995) are adopted to solve the PDE model (equations (2.15) and (2.16)) as follows:

$$[k(x,t+\Delta t) - k(x,t)]/\Delta t = [q(x-\Delta x,t) - q(x,t)/\Delta x] \tag{2.33}$$

$$q(x,t) = \min\left\{T\big(k(x,t)\big), Q\left(k_j - k(x+\Delta x,t)\right)\right\}, \tag{2.34}$$

where T specifies the maximum flow that can be sent by the upstream cell and Q specifies the maximum flow that can be received by the downstream cell.

To numerically solve the two-layer model for heterogeneous conditions, the single-class discrete traffic flow in the lower layer is solved first at time interval t. Then, the following two steps are applied to update the number of vehicles by vehicle class at time interval t, which connect the upper and lower layers consistently in terms of the number of vehicles by vehicle class.

Step 1: Determine the number of vehicles of each class $z \in \{S,R,X,U\}$ that advance to the downstream cell using the outcomes of the lower layer (equations (2.33) and (2.34)) and the proportion of vehicles of each class in the previous time interval in the upper layer, as follows:

$$q_z(x,t) = \frac{k_z(x,t-\Delta t)}{k(x,t-\Delta t)} \cdot q(x,t), \quad z \in \{S,R,X,U\}, \tag{2.35}$$

where $q_z(x,t)$ represents the vehicular flow of class z leaving cell x per unit time at time t and $k_z(x,t-\Delta t)$ represents the density of class z at cell x and time $t-\Delta t$, for all $z \in \{S,R,X,U\}$. Equation (2.35) indicates that the number of vehicles of each class advancing to the downstream cell is proportional to the previous time interval's proportion of vehicles by vehicle class in the upper layer.

Step 2: Update the number of vehicles by vehicle class in each cell of the upper layer using the discrete multi-class flow conservation law, as follows:

$$[k(x,t+\Delta t)-k(x,t)]/\Delta t=[q(x-\Delta x,t)-q(x,t)/\Delta x],\ z\in\{S,R,X,U\}. \quad (2.37)$$

In equation (2.37), $k_S(x,t)$, $k_R(x,t)$ and $k_X(x,t)$ represents $S(x,t)$, $R(x,t)$ and $X(x,t)$, respectively, that describe the number of vehicles of each class in the upper layer.

This process is repeated to solve for the spatiotemporal propagation of information flow.

2.6 Numerical experiments

2.6.1 Experiment design

Table 2.1 Cell characteristics and experiment parameters

Variables	Units	Value	Variables	Units	Value
Free flow speed (u_f)	km/h	108	Total number of cells	-	2,000
Time interval (Δt)	seconds	0.5	Critical density (k_c)	vehicles /km/lane	42
Cell length (Δx)	meters	15	Jam density (k_j)	vehicles /km/lane	167
Number of lanes	-	1	Communication frequency (β)	Hz	2
Market penetration rate (W)	%	50	Communication range (R)	meters	300

This section presents the experiment setup, and the parametric calibration of the information exclusion rate that describes the duration of retransmission period caused by limited channel capacity. The numerical experiments are conducted on a highway consisting of 2,000 cells, which is equivalent to 30 km of length. The cell characteristics and experiment parameters are provided in Table 2.1.

A simulation-based approach is used to calibrate the information exclusion rate. The aim of the simulation-based approach is to generate realistic synthetic data for the information exclusion rate under various congested V2V communication environments. We first analyze the number of information packets that can be transmitted over a limited channel capacity, which is assumed equivalent to the communication buffer size. The scenario that will be used for the parametric calibration is as follows. Assume that there is one channel reserved for event-driven communication, and that single channel is shared by multiple transmitters. The system is capable of transmitting data with the maximum data rate of 3 Mbps. For a given vehicle density *k*, the expected number of equipped vehicles is $2 \cdot k \cdot R \cdot W$, where *R* is the communication range and W is the market penetration rate of V2V-equipped vehicles. As only one communication per

time interval can be used for communication, the available channel capacity has to be shared by $2 \cdot k \cdot R \cdot W$ vehicles, and one vehicle, at maximum, is allowed to transmit s/($2 \cdot k \cdot \beta \cdot R \cdot W$) bits per second, where s is the channel capacity. For example, with R=0.3 km, k=60 veh./km, β =2Hz, W =50%, the maximum transmission rate for each V2V communication is calculated as:

$$C = \frac{s}{2 \cdot k \cdot \beta \cdot R \cdot W} = \frac{3 \cdot 10^6}{2 \cdot 60 \cdot 2 \cdot 0.3 \cdot 0.5} = 83{,}333\ bit = 10{,}417\ byte.$$

Assume that each packet size is 500 bytes; then, 21 packets per communication is the maximum transmission rate in terms of the number of packets that each V2V communication can transmit. The information exclusion rate is estimated by factoring the calculated communication buffer capacity using the simulation-based approach. The approach is as follows. For a given density k, vehicles are assumed to be randomly distributed along the 30 km highway. The simulation is conducted for 60 minutes. It is performed for the scenarios in which each equipped vehicle is assumed to generate a packet with a generation rate of 1 packet per 20 seconds, and communicate with other equipped vehicles within communication range. Suppose that the other equipped vehicles within communication range receive multiple packets from multiple senders, and retransmit these packets to surrounding vehicles with a given maximum broadcasting capacity. This leads the receiver vehicle to receive some new information packets that need to replace the old packets in the communication buffer. The simulation mimics the communication buffer where new information packets are received from other equipped vehicles and old information packets are excluded from the communication buffer at a time instant. The duration of information retransmission is determined by tracking the time after the information packet is received and before it is excluded from the communication buffer. The communication range R and communication frequency β are set as 300 meters and 2 times/second, respectively. The results of the information exclusion rate are summarized in Table 2.2.

Table 2.2 Communication kernel parameters and information exclusion rate under various densities

Density (veh./km)	a	b	C	λ	Density (veh./km)	a	b	C	λ
10	0.464	0.864	125	0.002	90	0.125	0.195	14	0.559
20	0.410	0.685	62	0.01	100	0.109	0.175	12	0.595
30	0.345	0.548	41	0.042	110	0.101	0.165	11	0.592
40	0.273	0.420	31	0.08	120	0.088	0.145	10	0.629
50	0.224	0.345	25	0.175	130	0.083	0.140	10	0.662
60	0.191	0.292	21	0.309	140	0.081	0.131	9	0.694
70	0.156	0.246	18	0.321	150	0.073	0.121	9	0.709
80	0.138	0.218	15	0.441	160	0.070	0.116	8	0.725

To obtain the communication kernel (2.17), Monte Carlo simulations are performed to mimic V2V communications under different traffic conditions, where successful V2V communication is affected by interference caused by simultaneous information transmissions from several equipped vehicles within the communication range. Successful communication between sender and receiver vehicles is identified by Signal to Interference Plus Noise Ratio (SINR) (Gupta and Kumar, 2000), which is defined as the signal power of the sender vehicle divided by the signal power of all interference from other equipped vehicles and background noise. The receiver vehicle b can receive information from sender vehicle a if

$$\frac{P_a/\|\delta_a - \delta_b\|^2}{\sum_{h\in E, h\neq a} P_h/\|\delta_h - \delta_b\|^2 + \vartheta} \geq \sigma,$$

where δ_h and P_h denote the coordinates and signal power of equipped vehicle h, respectively. ϑ denotes the power of background noise. E denotes the set of equipped vehicles within communication range of vehicle a which simultaneously transmit an information packet at the same time instant to vehicle a. The transmitted information from vehicle a is successfully received by vehicle b if the SINR is greater than or equal to threshold σ. The simulation is repeated 100 times. Each simulation run randomly locates equipped vehicles based on the traffic density to account for the location randomness of equipped vehicles within the communication range of the sender vehicle. The parameters of the calibrated communication kernel are presented in Table 2.2. Kim et al. (2017b) provide details of calibration of the communication kernel.

2.6.2 IFPW under homogeneous conditions

We first compute the analytical solution of the asymptotic IFPW speed under homogeneous conditions. Suppose the traffic flow density is 30 veh./km. According to Table 2.2, with 50% penetration rate of equipped vehicles, the parameters a and b in equation (2.17) are 0.345 and 0.548, respectively. The information exclusion rate is $\lambda = 0.042$. Applying these parameters, the analytical solution of the IFPW speed under the information relay control strategy can be computed as follows. Initially, all equipped vehicles are susceptible at each location, thereby:

$$S(x, 0) = \sigma = 30 \times 0.5 = 15\,\text{veh./km} = 0.225\,\text{veh./(15 m)}.$$

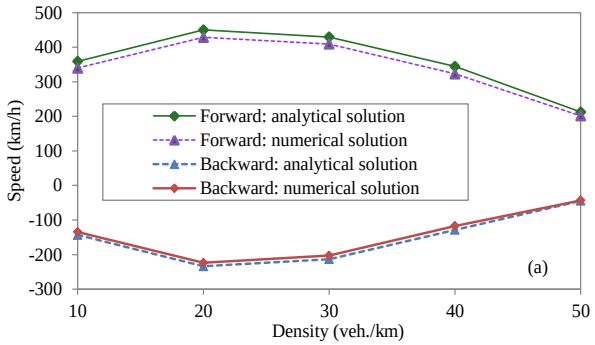
Note that $\gamma = \lambda/(\sigma\beta b) = 0.042/(0.225 \cdot 2 \cdot 0.548) = 0.1703 < 1$. According to Corollary 2.1, the IFPW exists, whose speed can be computed using equation (2.27):

$$c^* = \sigma\beta b\bar{c}^* = 0.225 \cdot 2 \cdot 0.548 \cdot \inf\left\{\frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz - \gamma\right] : 0 < \mu < \nu\right\}.$$

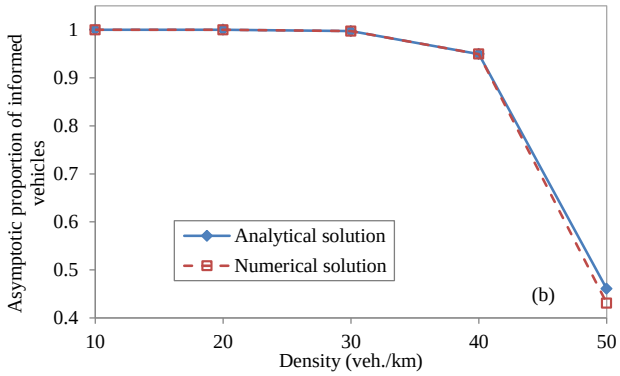
Note that

$$\begin{aligned}\frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz - \gamma\right] &= \frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\overline{K}(z)\,dz - \lambda/(\sigma\beta b)\right] \\ &= \frac{1}{\mu}\left[\int_{-\infty}^{\infty} e^{\mu z}\frac{1}{0.345\sqrt{\pi}}e^{\frac{-z^2}{0.345}}\,dz - 0.1703\right].\end{aligned} \tag{2.37}$$

Using the Newton method, the minimum value of equation (2.37) is 0.3595, which is obtained at $\mu = 3.869$. Then, $c^* = 0.225 \cdot 2 \cdot 0.548 \cdot 0.3595 = 0.0887$ km/s= 319.13 km/h. According to equation (2.31), the traffic flow speed under density 30 veh./km is 108 km/h. Thereby, the forward IFPW speed is $319.13 + 108 = 427.13$ km/h, and the backward IFPW speed is $108 - 319.13 = -211.13$ km/h.



(a) Asymptotic IFPW speed



(b) Asymptotic proportion of informed vehicles.

Figure 2.4 Asymptotic IFPW speed and asymptotic proportion of informed vehicles

Figure 2.4(a) compares the analytical solutions with the numerical solutions of the asymptotic IFPW speeds for different traffic conditions. The numerical solutions of the asymptotic IFPW speeds are computed using the method provided in Section 5. As can be observed from Figure 2.4(a), the analytical solutions match well with the numerical solutions for the forward and backward IFPW speeds. Note that this experiment only compares IFPW speeds under the cases where the traffic density is less than 60 veh./km. If traffic density is higher than 60 veh./km under the current V2V communication setting, then $\gamma > 1$. Based on Remark 3, the IFPW does not exist. Thereby, the asymptotic IFPW speed does not exist for these cases when traffic density is higher than 60 veh./km.

Figure 2.4(a) also shows that the asymptotic IFPW speed reduces when the equipped vehicle density increases, which causes V2V communication congestion and higher information exclusion rate. This comparison clearly shows that the information flow propagation speed may

be overestimated if V2V communication congestion occurs. The overestimation of information flow propagation speed would significantly impact the effectiveness of V2V communication-based applications.

The following example illustrates the analytical and numerical solutions for asymptotic density. If the traffic flow density is 30 veh./km, then the asymptotic density of informed vehicles is calculated using equation (2.29) as:

$$I^* = 0.225 \cdot \alpha(\gamma) = 0.225 \cdot \left\{\alpha | 1 - \alpha = e^{-\alpha/\gamma}; \alpha \in (0,1)\right\}.$$

As $\gamma = 0.042/(0.225 \cdot 2 \cdot 0.548)$, $\alpha = 0.997$ is the solution of $1 - \alpha = e^{-\alpha/\gamma}$ for $\alpha \in (0,1)$. Hence, $I^* = 0.997 \cdot 0.225 = 0.224$ veh./cell. The asymptotic proportion of informed vehicles is calculated by dividing I^* with the density of equipped vehicles in each cell, which is $0.224/0.225 = 0.997$.

Figure 2.4(b) compares the analytical and numerical solutions for the asymptotic proportion of informed vehicles. It shows that the numerical solutions overlap with the analytical solutions, implying that the numerical algorithm proposed in Section 5 can effectively solve the two-layer model for the spatiotemporal propagation of information flow. Figure 2.4(b) also shows that when traffic density is less than 30 veh./km, most equipped vehicles can receive the specific information. The asymptotic proportion of informed vehicles decreases dramatically as the traffic flow density increases. When traffic density is 50 veh./km, only around 55% of the equipped vehicles can receive the specific information packet. This is because the information exclusion rate increases with the traffic density.

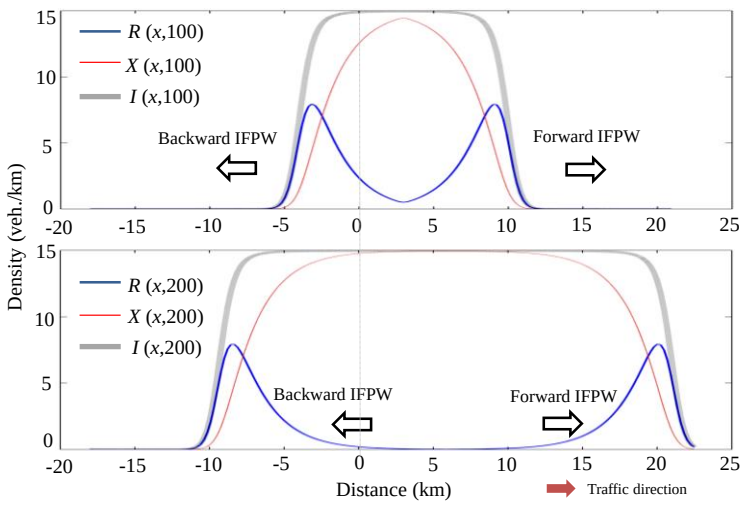


Figure 2.5 The density of relay and excluded vehicles over space, at $t = 100$ and $t = 200$ seconds ($k = 30$ veh./km).

Next, we explore the impact of the information relay control strategy on information dissemination. Suppose a specific information packet is sent by an equipped vehicle at location 0 at time 0. The traffic density is 30 veh./km. Figure 2.5 shows the densities of three classes of vehicles (R, X and I) at $t = 100$ seconds and $t = 200$ seconds over the spatial dimension, respectively. As stated before, $I(x,t) = R(x,t) + X(x,t)$. These results are obtained by solving the IDEs (equation (2.14)) and PDE (equations (2.15) and (2.16))) numerically. The distribution of $I(x,100)$ demonstrates that the specific information packet is disseminated 12 km downstream and 6 km upstream in 100 seconds. Most of the information-relay vehicles (class R) are located close to both fronts of the IFPW. This is because information-relay vehicles have received this specific information packet recently, and have a high probability having this packet in their communication buffers for propagation at the current time. Note that the density of information-relay vehicles (e.g., $R(x,200)$) is relatively low between 0 km and 10 km. As time passes, the density of information-relay vehicles reduces to zero until all informed vehicles exclude the information packet.

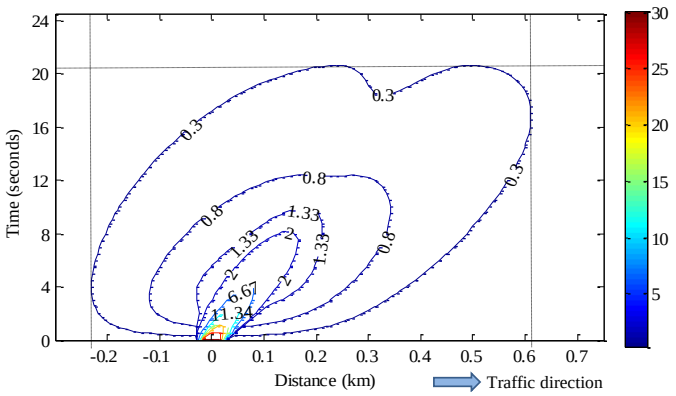


Figure 2.6 Density contour of information-relay vehicles at $k = 60$ veh./km and $\lambda = 0.31$.

2.6.3 Scenario in which IFPW does not exist

According to Corollary 2.2, when $\gamma > 1$, the IFPW does not exist and the information cannot propagate to vehicles far away from the location of the initial informed vehicles. Consider unidirectional flow with uniform traffic density $k = 60$ veh./km. At time 0, only equipped vehicles in the cell located at point 0 are informed vehicles, and all other equipped vehicles are susceptible vehicles. The information exclusion rate is 0.31 packets per second according to Table 2.2. In this context, $\gamma = 1.41 > 1$. Figure 2.6 shows the density contour of information-relay vehicles. It demonstrates that the density of information-relay vehicles has the highest value in the region close to point 0 at time 0. However, the density of information-relay vehicles decreases dramatically in space and time. After 21 seconds, while the information packet is propagated to 0.22 km upstream and 0.61 km downstream, the density of information-relay vehicles is only 0.3 veh./km, and continues decreasing. Recall that only information-relay vehicles can propagate the specific information packet. A susceptible vehicle located far away from the initial sender vehicle will not receive this information packet as there would be no information-relay vehicles in the vicinity of it.

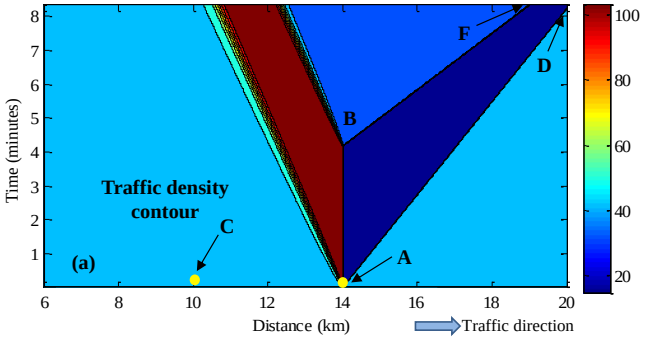
2.6.4 IFPW under heterogeneous conditions

This section investigates the impact of traffic flow dynamics on information propagation under the information relay control strategy in heterogeneous conditions. We consider the traffic density variation due to a traffic incident on a unidirectional highway with a traffic density of 50 veh./km. As illustrated in Figure 2.7(a), the incident occurs at location A at time 0. It reduces the highway capacity by a third of its initial value for 4 minutes, before the incident is cleared. Line

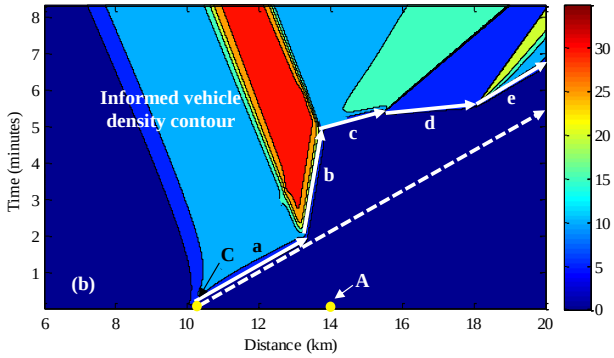
AB in Figure 2.7(a) separates the congested traffic and the free flow traffic departing from the incident occurrence location. Lines AD and BF in Figure 2.7(a) represent two forward propagating traffic waves caused by the occurrence and clearance of the incident, respectively. After the incident occurs, vehicles are jammed at the incident location, leading to a traffic wave propagating backward.

Assume that an information packet is generated at time 0 at location C that is 4 km upstream of the incident occurrence location. The numerical solution method proposed in Section 5 is used to characterize the IFPW under heterogeneous conditions. The impact of the traffic flow dynamics on the success rate of a single-hop communication is captured by the kernel function. The information exclusion rate varies with the traffic density according to Table 2.2.

Figure 2.7(b) illustrates the formation and evolution of the backward and forward IFPWs. After the sender vehicle generates the information packet, the backward IFPW takes some time to stabilize and propagate at the asymptotic speed. The front of the backward IFPW shows a lower density of informed vehicles. By contrast, the forward IFPW is significantly affected by the variation of traffic density. This effect is illustrated by the five stages of the forward IFPW, presented as five arrows with labels a, b, c, d, and e in Figure 2.7(b). First, the IFPW reaches the asymptotic speed and propagates forward, i.e., stage (a). When the forward IFPW enters the traffic jam area, i.e., stage (b), its speed decreases dramatically as the information exclusion rate increases due to the congestion in the information flow regime that is caused by the increased equipped-vehicle density. When the forward IFPW passes through the traffic jam area and reaches the saturation flow rate area, i.e., stage (c), its speed increases due to the reduced communication interference and a smaller information exclusion rate. After stage (c), the forward IFPW enters the free flow traffic area, i.e., stage (d), where the forward IFPW propagates at maximum speed. Later, when the information packet propagates further to catch the traffic flow that is not impacted by the traffic incident, i.e., stage (e), the speed of the forward IFPW reduces to the same value as that in stage (a).



(a) Contour of traffic density



(b) Contour of informed vehicle density under the heterogeneous conditions

Figure 2.7 Contour of traffic density and Contour of informed vehicle density under the heterogeneous conditions

2.7 Concluding remarks

This study explores the spatiotemporal information flow propagation under information congestion effects, by introducing a two-layer macroscopic model and an information packet relay control strategy. An IDE system is established in the upper layer to model the information dissemination in the information flow regime. The LWR model is used in the lower layer to capture the impacts of traffic flow dynamics on information propagation. Under homogeneous conditions, this study derives the necessary conditions for existence of the IFPW. These necessary conditions can be used to quantify the minimum information broadcasting time duration in a complex V2V-based traffic system to ensure the propagation of a specific information packet, especially under congested communication environments. The asymptotic IFPW speed and asymptotic density of informed vehicles can be obtained analytically by solving one-dimensional nonlinear equations derived from the two-layer model. The analytical solutions

help to circumvent computationally-expensive numerical methods to estimate the performance of information propagation in terms of timeliness and informed vehicle density for V2V-based applications. The numerical experiments show that the IFPW speed decreases as the information packet exclusion rate increases, to prevent packet collisions under information congestion conditions. A large information packet exclusion rate can cause information to be propagated only locally. This implies that equipped vehicles far away from the sender vehicle will not be informed. These findings provide valuable insights for the design of effective V2V communication-based applications, especially for trading off the information dissemination duration for each information packet against the number of information packets to disseminate, under dense V2V communication environments. Hence, this study can be leveraged to develop a new generation of information dissemination strategies focused on enabling specific V2V information to reach specific locations at specific points in time.

This work can be extended in several directions. First, the assumption of uniform information exclusion rate can be relaxed as it may dynamically change in time and space. Second, this study assumes all information packets have equal priority to propagate. In real-world applications, different information packets may have different priorities. For example, safety-related information is more urgent than non-safety related information, and should be given a higher priority. Hence, the propagation of information packets with different priorities in space and time can be explored. Third, the IFPW speed and asymptotic density of informed vehicles under the proposed information relay control strategy can be compared with those under other strategies (for example, counter-based, distance-based, and stochastic strategies). Fourth, in addition to communication constraints such as communication range, communication frequency, channel capacity, signal interference, etc., the environment (e.g., weather) and vehicle size (e.g., truck) can also impact V2V communication (Talebpour and Mahmassani, 2016). While these factors are not considered in this study, they can be captured in a straightforward manner by the communication kernel.

CHAPTER 3. MULTICLASS INFORMATION FLOW PROPAGATION CONTROL UNDER VEHICLE-TO-VEHICLE COMMUNICATION ENVIRONMENTS

3.1 Introduction

The rapid development of vehicle-to-vehicle (V2V) communication technologies has motivated their use for a wide spectrum of innovative solutions to enhance transportation safety, efficiency, and sustainability. A V2V communications-based traffic system can potentially be leveraged to enhance traffic safety by more effectively detecting emerging conflict situations, improve traffic efficiency through information-based and other control strategies, and reduce energy consumption and emissions. For example, this entails the communication of a vehicle's status to other vehicles and/or the surrounding infrastructure, and thereby the exchange of information on travel/traffic conditions. Hence, vehicles equipped with such a capability for two-way communications can potentially gain spatio-temporal knowledge on travel-related conditions, which can be used to develop vehicle-level travel strategies and/or network-level traffic management strategies. Further, V2V-based traffic systems enable decentralized information generation and dissemination. Vehicles in a V2V-based system can generate information and relay it to other vehicles through multi-hop processes. Unlike centralized information systems, a V2V-based traffic system can potentially provide timely information in emergency/disaster situations by avoiding delays associated with data collection and communication with control center.

Understanding the characteristics of spatiotemporal information flow propagation in a V2V-based traffic system is important as most applications require timely and reliable information delivery. However, modeling information flow propagation in space and time is challenging. Factors from both traffic flow domain and communications domain significantly affect the reliability of V2V communication and information propagation. The traffic flow dynamics affect occurrence of V2V communications. Communication constraints, such as communication frequency, channel capacity, and communication power, significantly affect the reliability of V2V communications.

In the literature, various models have been proposed to characterize information flow propagation in different traffic flow and communication environments. These models can be

classified into microscopic and macroscopic. Microscopic models address information flow propagation by considering the spatial distribution of traffic flow. They factor the effects of the random positions of equipped vehicles in the traffic stream on information flow propagation. Simulation and analytical models have been developed to analyze the V2V propagation performances related to connectivity of inter-vehicle communication (Ukkusuri and Du, 2008) and the expected information propagation distance (Wang, 2007; Wang et al., 2010; Wang et al., 2011; Wang et al., 2012; Yin et al., 2013; Wang et al., 2015; Du and Dong, 2015; Du et al., 2016), under different traffic flow scenarios. However, these models oversimplify the effects of communication constraints by assuming that information can be retransmitted instantaneously. This assumption neglects the time latency of information flow propagation. Thereby, these models only characterize information dissemination in the space domain, but not in the time domain.

To address the aforementioned gaps, some recent studies have sought to characterize information flow propagation at a macroscopic level (Kim et al. 2017; Wang et al. 2018; Kim et al., 2018) by introducing the concept of information flow propagation wave (IFPW). They use the notion that when information spreads through multi-hop broadcasting communications, from a macroscopic perspective, there is a moving boundary that separates traffic flow into informed and uninformed regions, and moves towards the uninformed region like a wave. By leveraging the analogy of the IFPW with disease spread in epidemiology, analytical models are developed to characterize the IFPW. These macroscopic models relax the assumption of instantaneous information propagation and can describe the spatiotemporal spread of information in the traffic flow. In addition, these models capture the effects of V2V communication constraints realistically using a communication kernel. Further, interactions between V2V communications and traffic flow dynamics are captured by incorporating the effects of congested traffic, such as the backward propagating traffic wave, on information flow propagation.

The models discussed heretofore are descriptive, and seek to describe the spatiotemporal propagation of information to address effects of traffic flow dynamics and/or communication constraints. However, they lack a capability to control the propagation of information flow, which is necessary for traffic management applications in a V2V-based traffic system. For example, real-time traffic/routing information can help travelers choose better routes to reduce travel time. However, congestion can worsen if all travelers receive the same information and

choose the same (or similar) routes in an uncoordinated V2V-based system or receive and choose the same routing suggestions. Hence, the propagation of information flow needs to be controlled so that the spatiotemporal access to information varies across vehicles in such a way as to improve system performance. Similarly, under emergency evacuation, the propagation speed of evacuation information needs to be controlled so that it can reach different areas in the affected region with different impact levels at the desired times to reduce the severe traffic congestion or gridlock that would otherwise occur due to the simultaneous evacuation of all evacuees.

Another common characteristic of previous studies is that they only consider the propagation of a specific information packet of interest or one type of information. In practical applications, information can belong to different classes (e.g., safety information, routing information, work zone information). Hence, a V2V-based system may need to propagate information from different information classes simultaneously. However, the application needs of information from different information classes can be different, in terms of three performance measures: (i) information spread, defined here as the proportion of vehicles informed with a specific information packet, (ii) bounds on time delays for this information to reach specific locations, and (iii) spatial coverage, defined here as the distance this information can be propagated from its point of origin. For example, urgent traffic accident information (e.g., road is blocked by an accident) needs to be delivered to all vehicles in the impacted area with low latency. By comparison, routing information needs to reach only a certain proportion of vehicles to avoid possible congestion arising from the provision of information on the suggested route. Work zone information or sudden hard brake information may need to be propagated in a small area in the vicinity of where they are generated.

This study designs a queuing-based modeling approach to control the propagation of information of different information classes to meet application needs related to information spread, time delay bounds, and spatial coverage. An information class is defined as a type of information which has similar application needs in terms of the three propagation performance measures. To enable control for multiclass information flow propagation, this study assumes that the size of each information packet is the same and the channel capacity is shared equally with all equipped vehicles within communication range of that vehicle (Wang et al., 2018). Under this assumption, an equipped vehicle can send data containing multiple information packets during one transmission, whose number is determined by the size of one information packet, channel

capacity, communication frequency, and the number of equipped vehicles within communication range of that vehicle. This implies that an equipped vehicle can serve (send) multiple information packets simultaneously.

To better characterize the information service (sending) process in our queuing-based approach, we denote a "virtual communication server" (hereafter, referred to as "communication server") as the storage amount in the transmitted data that is equal to the size of an information packet. A communication server can serve at most one information packet at a time. The total number of communication servers is equal to the maximum number of information packets that an equipped vehicle can send simultaneously during one transmission, which is labeled the transmission capacity. We denote communication service time as the time duration an information packet is in the communication server. During the communication service time, the information packet will be repetitively sent by the equipped vehicle where the number of transmissions depends on the communication frequency which is the number of data transmissions per unit time enabled by the V2V device characteristics in the vehicle. We denote the mean communication service rate for a server as the inverse of the mean communication service time of all information packets served by that server.

To enable control for multiclass information flow propagation, for the first time in the literature, a queuing strategy is developed for each V2V-equipped vehicle to propagate the information packets of different information classes that it receives or generates. We assume information packets in different information classes will form different queues. Thereby, when an information packet is received by an equipped vehicle, it will be forwarded to the queue for the information class it belongs to. After being in the queue, the information packet will enter a communication server for this information class to be disseminated. It will be deleted from the server after its assigned communication service time is reached. Based on this conceptual queuing strategy, information propagation control is enabled by assigning different number of communication servers and mean communication service rates to different information classes to send the information. It should be noted that the mean communication service rate for an information class determines the mean communication service time of each information packet in the information class, which impacts the number of transmissions of each information packet in this class. Due to existence of communication failure, an information packet cannot be guaranteed to be received by other vehicles if it is just sent once by an equipped vehicle. The

queuing strategy allows an equipped vehicle to control the number of transmissions of an information packet by leveraging the mean communication service rate so as to control the number of vehicles within communication range of this vehicle that can receive the information packet. Thereby, while the mean communication service rate does not impact the success rate of one V2V communication, it significantly impacts the total success rate of V2V communications by allowing an equipped vehicle to transmit information multiple times. Also, the number of communication servers assigned to an information class significantly impacts the mean waiting time in the queue for information packets in that class, which impacts the information flow propagation speed. Thereby, two control parameters, the number of assigned communication servers and mean communication service rate, can be determined for each information class to achieve the desired propagation performance related to information spread, time delay bounds, and spatial coverage.

This study conceptually extends the macroscopic models developed by Kim et al. (2017) and Wang et al. (2018), and proposes a new two-layer analytical modeling approach to characterize the IFPW under the proposed queuing strategy. An integro-differential equation (IDE) model is derived to characterize the spatiotemporal information propagation flow under the designed queuing strategy in the upper layer. The lower layer adopts the Lighthill-Whitham-Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) to characterize the traffic flow dynamics. The two-layer model enables investigation of the following three questions. First, what is the density of equipped vehicles that can receive a specific information packet under given values of the two control parameters? This question seeks to provide insights on controlling information spread. Second, how do the two control parameters in the queuing system impact the propagation speeds of specific information packets of interest belonging to different information classes? Addressing this question is useful for controlling the time delay of information packets of different information classes in reaching desired locations. Third, what are the conditions that can ensure the specific information packet can form a wave to be propagated over the traffic stream, and how the two control parameters impact the propagation distance of an information packet? This question addresses the necessary conditions for the formation of an IFPW which is related to the spatial coverage of information.

The contributions of this study are fourfold. First, a queuing strategy is developed for an equipped vehicle to propagate the received information. It enables control for multiclass

information flow propagation by determining the values of the two control parameters in the queuing strategy. A two-layer model is developed to characterize the spatiotemporal IFPW based on the designed queuing strategy. It factors the effects of both traffic flow dynamics and communication constraints on information flow propagation. Second, NS-3 simulation (Noori and Olyaei., 2013; Talebpour et al., 2016) is used to investigate the reliability of V2V communications in different traffic flow environments. The communication kernel is calibrated using the data from the NS-3 simulations performed, which enables capturing the effects of communication constraints (e.g., communication frequency, channel capacity, communication power) on information flow propagation more realistically. Third, the necessary conditions for the existence of IFPW and the analytical solution for the asymptotic density of informed vehicle are derived under homogenous conditions (i.e., unidirectional flow with uniform traffic flow density). To the best of our knowledge, the solution of the IDE system analogy to the proposed model has not been studied before even in the epidemiology literature. These analytical expositions quantify the impacts of the two control parameters on the density of informed vehicles and the spatial coverage. Fourth, numerical solutions are designed to solve the two-layer model under homogeneous as well as heterogeneous conditions. They aid in quantifying the relationship between the two control parameters and the asymptotic IFPW speed under homogeneous traffic conditions. Further, they provide valuable insights for controlling multiclass information flow propagation to achieve the desired performance in terms of information spread, the time delay to reach the target locations, and spatial coverage under heterogeneous conditions.

The remainder of the paper is organized as follows. The next section discusses the designed queuing strategy and the framework of the proposed model to characterize the IFPW. Section 3 formulates a two-layer model to characterize the IFPW in space and time under the proposed queuing strategy. In Section 4, the analytically solution for the asymptotic density of informed vehicles and the condition for existence of IFPW under homogeneous traffic conditions are discussed. In addition, the numerical solution method is presented to solve the proposed two-layer model for heterogeneous conditions. Results from numerical experiments are discussed in Section 5, to demonstrate the effectiveness of the proposed model to control the propagation performance of different information classes. Section 6 provides some concluding comments.

3.2 Preliminaries

Consider a highway with a traffic flow stream consisting of V2V-equipped and V2V-unequipped vehicles. Information is generated and broadcasted to other equipped vehicles through multi-hop V2V communications. Each equipped vehicle receives information from other equipped vehicles and broadcasts such information and the information it generates to all other equipped vehicles within communication range. Let information packets relayed in the traffic flow be divided into s classes, each of which has different requirements in terms of information spread, time delay bounds and spatial coverage. Let $\mathcal{L} = \{1,2,\cdots s\}$ denotes the set of information classes. When an equipped vehicle receives multiple packets, it filters the information packets to identify those that have not been received before. It then moves such unduplicated information packets (labeled effective information packets) into the queues for the corresponding information classes to wait to be propagated according to the information class they belong to. The effective information arrival rate is affected by unsuccessful V2V communication and removal of duplicated information packets. Because such events are random and independent, following Wang et al. (2018) and Zhang et al. (2016), this study assumes that the arrival of effective information packets to different information classes follows a Poisson process. Let $\lambda_1, \lambda_2 \cdots, \lambda_s$ be the arrival rate of information packets for information classes $1,2 \cdots, s$, respectively.

Suppose that the size of all information packets disseminated over the traffic flow is identical, and the channel capacity is shared equally with all equipped vehicles within communication range. Let N denote the transmission capacity, which describes the number of information packets that can be delivered by an equipped vehicle through one V2V communication. In this study, we assume the transmission capacity for all equipped vehicles to be the same. It should be noted that N has an upper bound determined by the size of an information packet, the density of the traffic flow, the communication frequency and the channel capacity (Wang et al., 2018).

To control propagation performance for each information class, a queuing strategy for relaying information of different information classes is designed in this study, as shown in Figure 3.1. Note that an equipped vehicle can transmit N information packets during one communication. To better illustrate the queuing strategy, we assume an equipped vehicle has N communication servers each of which can serve one information packet. A communication server represents the storage amount in the transmitted data that is equal to the size of one

information packet (see Figure 3.1). The number of the communication servers assigned to a particular information class determines the maximum number of information packets in this information class that can be transmitted simultaneously by an equipped vehicle. Let n_j be the number of communication servers assigned to information class j, $j \in \mathcal{L}$. To control multiclass information flow propagation, information packets in different classes will form different queues (see Figure 3.1). If one communication server for information class j is empty, the first information packet in the queue for information class j will enter into the server to be sent out. Let u_j be the mean communication service rate (packets/second) for information packets in information class j. The inverse of u_j (i.e., $1/u_j$) is the mean communication service time (i.e., transmission duration) for an information packet in information class j. The information packet in the communication server will be transmitted repetitively until the communication service time is reached. Thereby, the communication service time significantly impact the number of vehicles that can receive the specific information of interest of information class j.

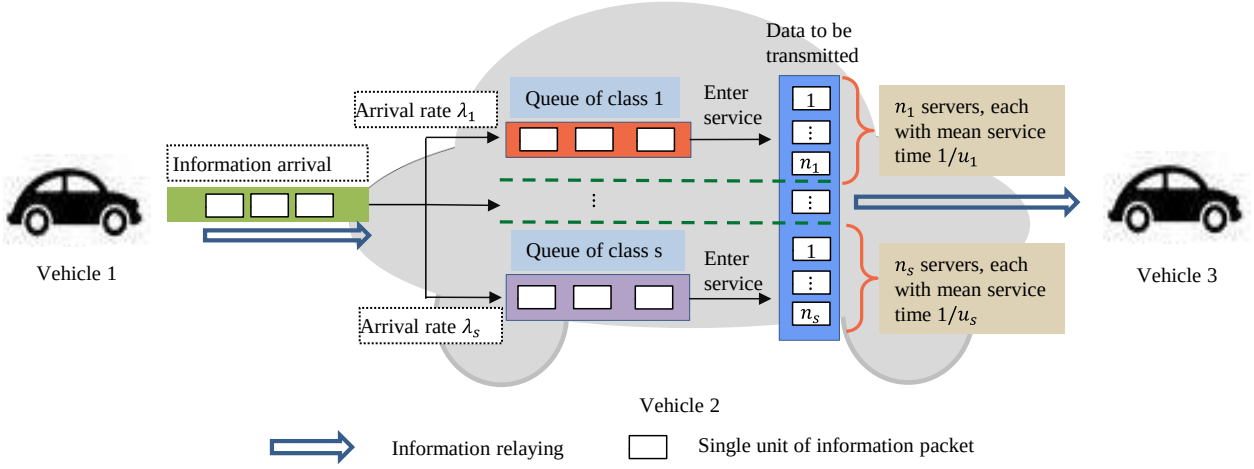


Figure 3.1 Queuing strategy for relaying information of different classes

To facilitate modeling, we assume the communication service time of each information packet in an arbitrary class j, $j \in \mathcal{L}$ follows an exponential distribution with mean $1/u_j$. The communication service time can be generated randomly in advance according to the exponential distribution with mean $1/u_j$. An information packet will be removed from the system if its assigned communication service time is reached. Note that for an arbitrary information class j, the arrival of information packets follows a Poisson process with parameter λ_j and the corresponding communication service time follows an exponential distribution with mean $1/u_j$.

Thereby, propagation of information packets in information class j follows a $M/M/n_j$ queuing process.

Note that the mean communication service rate for information packets in an arbitrary information class j (i.e., u_j) impacts the number of vehicles that can receive information packets from this information class. Further, according to the queueing theory, the mean communication service rate (u_j) and the number of assigned communication servers (i.e., n_j) for information class j determine the mean waiting time of an information packet in the queue. Thereby, the propagation performance of an information packet of information class j in terms of information spread, time delay bounds, and spatial coverage can be controlled by assigning various values to n_j and u_j. It should be noted that the propagation performance of information of different information classes is constrained by the total number of communication servers in an equipped vehicle.

Under the designed queuing strategy, equipped vehicles are divided into four vehicle classes, the susceptible vehicles (labeled S), the information-holding vehicles (labeled H), the information-relay vehicles (labeled R) and the information-excluded vehicles (labeled E). Susceptible vehicles are equipped vehicles that have not received the specific information packet of interest. They become information-holding vehicles if they receive that information packet and are holding it in the queue for transmittal. The information-holding vehicles become information-relaying vehicles if that information packet enters a communication server to be disseminated to the other vehicles. Once the communication service time is reached for that information packet, it will be removed from the vehicle. The information-relying vehicle then becomes an information-excluded vehicle. It is worth noting that the susceptible vehicles can become information-relaying vehicles directly if the specific information packet of interest enters into a communication server without waiting in a queue; that is, when this information packet is received/generated, there is no queue for the corresponding information class.

Similar to Kim et al. (2017) and Wang et al. (2018), the IFPW consists of two waves: the information dissemination wave in the information flow domain and the traffic flow propagation wave in the traffic flow domain. A two-layer model is developed in this study to model the IFPW. The modeling framework is shown in Figure 3.2. In the upper-layer, integro-differential equations (IDEs) will be derived to characterize the information dissemination waves. This layer describes how vehicle densities by vehicle class will change instantaneously through V2V

communications under the designed queuing strategy. The lower layer describes the traffic flow dynamics. In this study, the LWR model will be used to characterize traffic flow dynamics. Based on the two-layer model, the asymptotic IFPW speed, the asymptotic density of informed vehicles and the conditions for existence of IFPW will be investigated in this study.

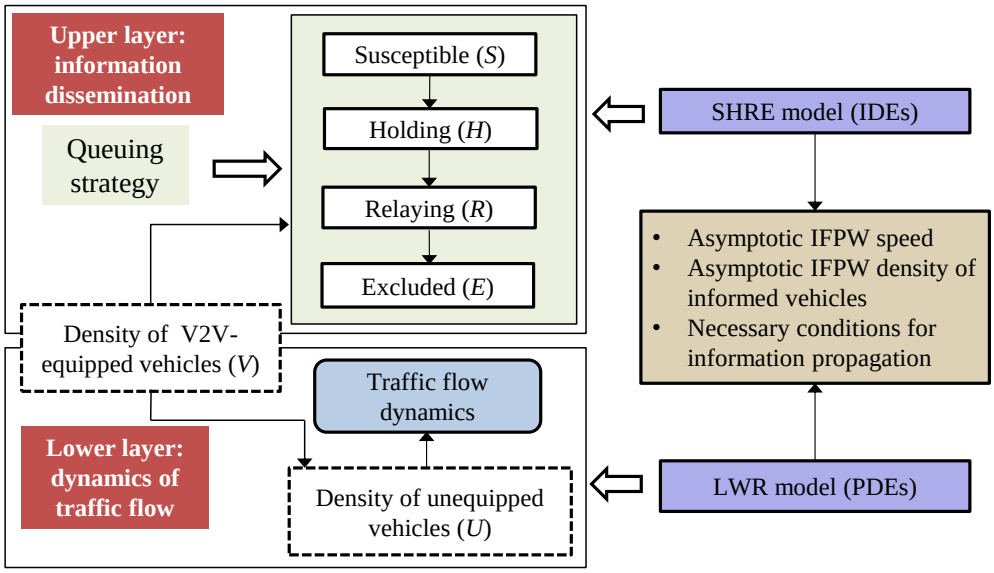


Figure 3.2 Modeling framework of IFPW under the designed queuing strategy

3.3 Modeling the multiclass information flow propagation wave

3.3.1 Modeling the information flow dissemination wave in the upper layer

Assume the specific information packet of interest belongs to an arbitrary information class j, $j \in \mathcal{L}$. This section seeks to model the information dissemination wave in the upper layer under the designed queuing strategy. It describes the instantaneous change in the density of equipped vehicles by vehicle class (i.e., S, H, R, E for information packets in class j) due to V2V communications. The impacts of communication constraints (communication power, communication frequency, signal interference, etc.) on the success of V2V communications is explicitly factored in this model.

Let t be the current time. Divide the time horizon of interest uniformly into consecutive time windows of length w each. Denote $S_j(x,t)$ and $I_j(x,t)$ as the densities of the susceptible vehicles and the vehicles informed with the specific information packet of interest in information class j, respectively, at time t and location x. Note that the informed vehicles consist of vehicles from

classes H, R and E which have received this information packet. Let $\Delta S_j(x,t) = S_j(x,t) - S_j(x,t-w)$ be the density of vehicles informed with the specific information packet of interest during time $[t-w,t]$. Denote $\Delta I_j(y,t-w) = I_j(x,t) - I_j(x,t-w)$ as the density change of informed vehicles at location x during time interval $[t-w,t]$. Note $\Delta I_j(y,t-w) = -\Delta S_j(x,t-w)$ as an equipped vehicle is either susceptible or informed. Let the current time be $t = m \cdot w$; m is a positive integer. Conceptually adapting from epidemiology and modifying, $\Delta S_j(x,t)$ is formulated as:

$$\Delta S_j(x,t) = -S_j(x,t-w) \cdot \int_{\Omega} K(x,y) \cdot w\beta \cdot R_j(y,t)dy \tag{3.1}$$

where Ω denotes the domain of space, β is communication frequency, and $w\beta$ denotes the expected number of transmissions occurring in time length w. Also, $R_j(y,t)$ is the density of vehicles relaying the specific information packet of interest in information class j at time t and location y. Function $K(x,y)$ is a communication kernel which represents the probability that a susceptible vehicle at location y can successfully receive the specific information packet sent from a vehicle at location x under a given communication environment (communication frequency, channel capacity, communication power, etc.) and traffic flow environment (traffic density, etc.). It characterizes the reliability of V2V communication realistically by capturing the impact of factors in both the communication and traffic flow domains. In the study experiments in Section 5, function $K(x,y)$ is calibrated using NS-3 simulation. $\int_{\Omega} K(x,y) \cdot w\beta \cdot R_j(y,t)dy$ denotes the probability that a susceptible vehicle at location x receives the specific information packet of information class j sent by an informed vehicle over the space domain Ω.

Suppose the mean arrival rate of information packets in information class j, $j \in \mathcal{L}$ is λ_j packets/second. Let n_j and u_j be the number of communication servers and the mean communication service rate (packets/second) assigned to information class j, respectively. n_j and u_j are controllable parameters which will be leveraged to control the propagation of the specific information packet of interest of information class j. To ensure the information packets in information class j can be propagated by an equipped vehicle, n_j and u_j are selected such that $\lambda_j < n_j u_j, \forall j \in \mathcal{L}$.

As discussed earlier, under the designed queuing strategy, the arrival and service process of

information packets in information class j follow $M/M/n_j$ queue process. Thereby, at current time t, the vehicles that are relaying the specific information packet of interest of information class j consist of two groups: (1) the vehicles relaying the specific information packet without the information-holding process (i.e., queuing process). This implies when these vehicles receive the specific information packet, there is no queue for information class j in these vehicles. Thereby, the specific information packet of interest can enter into the communication server to be disseminated out directly; and (2) the vehicles relaying the specific information packet with an information-holding process (queue process), i.e., the specific information packet of interest experiences a queuing process before entering into the communication server.

Let T_j^q be the waiting time for the specific information packet of interest of information class j, and $W_j^q(v_j) = \Pr\{T_j^q \le v\}$ be the probability that the waiting time of this information packet in the queue is less than v. According to Gross et al. (2008, page 71), we have

$$W_j^q(v) = \Pr\{T_j^q \le v\} = 1 - \frac{{r_j}^{n_j} P_j^0}{n_j!\,(1-\rho_j)}\, e^{-(n_j u_j - \lambda_j)v_j} \tag{3.2}$$

where $r_j = \lambda_j/u_j$; $\rho_j = \lambda_j/(n_j u_j)$; and P_j^0 is the probability that there is no information packet of information class j in the system, formulated as

$$P_j^0 = \left(\frac{{r_j}^{n_j}}{n_j!\,(1-\rho_j)} + \sum_{l=0}^{n_j-1} \frac{{r_j}^l}{l!} \right)^{-1} \tag{3.3}$$

According to Eq. (3.2),

$$W_j^q(0) = \Pr\{T_j^q \le 0\} = \Pr\{T_j^q = 0\} = 1 - \frac{{r_j}^{n_j} P_j^0}{n_j!\,(1-\rho_j)} \tag{3.4}$$

where W_j^0 is the probability that the specific information packet is received by a vehicle at a time instant when there is no queue for information class j. Let

$$\xi_j = \frac{{r_j}^{n_j} P_0}{n_j!\,(1-\rho_j)} \tag{3.5}$$

Then

$$W_j^q(0) = 1 - \xi_j \tag{3.6}$$

From Eq. (3.2), we have

$$\Pr\{T_j^q > v\} = 1 - W_j^q(v) = \frac{{r_j}^{n_j} P_0}{n_j!\,(1-\rho_j)}\, e^{-(n_j u_j - \lambda_j)v_j} \tag{3.7}$$

Let T_j^s be the service time of the specific information packet of interest in the communication

server, and $\Pr\{T_j^q \le \theta\}$ be the probability that the communication service time is less than θ. Recall the communication service time follows exponential distribution with mean $1/u_j$, then

$$\Pr\{T_j^q \le \theta\} = 1 - e^{-u_j\theta} \tag{3.8}$$

This implies

$$\Pr\{T_j^q > \theta\} = e^{-u_j\theta} \tag{3.9}$$

$R_j(y,t)$ can then be formulated as

$$R_j(y,t) = A_j(y,t) + B_j(y,t) \tag{3.10}$$

where

$$\begin{aligned} A_j(y,t) &= W_j^q(0) \cdot \sum_{i=1}^{t/w} \Delta I_j(y, t-iw) \cdot \Pr\{T_j^s > iw\} \\ &= (1-\xi_j) \cdot \sum_{i=1}^{t/w} \Delta I_j(y, t-iw) \cdot e^{-u_j \cdot iw} \\ B_j(y,t) &= \sum_{i=1}^{t/w} \int_0^{iw} \Delta I_j(y, t-iw) \cdot \frac{\partial\left(\Pr\{T_j^q > v\}\right)}{\partial v} \cdot \Pr\{T_j^s > iw - v\}\, dv \\ &= \sum_{i=1}^{t/w} \int_0^{iw} \Delta I_j(y, t-iw) \cdot (n_j u_j - \lambda_j) \cdot \xi_j \cdot e^{-(n_j u_j - \lambda_j)v} \cdot e^{-u_j \cdot (iw - v)}\, dv \end{aligned}$$

where $A_j(y,t)$ is the accumulated density of vehicles relaying the specific information packet of interest of information class j without queuing process at location y and current time t. $\Delta I_j(y, t-iw) \cdot \Pr\{T_j^s > iw\}$ is the density of vehicles informed at time $t - iw$ and relaying the specific information packet at location y and current time t. $B_j(y,t)$ is the accumulated density of vehicles relaying the specific information packet after queuing process at location y and current time t. The term $\int_0^{iw} \Delta I_j(y, t-iw) \cdot \frac{\partial\left(\Pr\{T_j^q > v\}\right)}{\partial v} \cdot \Pr\{T_j^s > iw - v_j\}\, dv$ denotes the density of vehicles at location y that become informed iw time units ago and are propagating the specific information at current time t after experiencing the queuing process.

To derive the continuous model, let $w \to 0$, and divide both side of Eq. (3.1) by w, then we have

$$\begin{aligned} \lim_{w\to 0} \frac{\Delta S_j(x,t)}{w} &= \frac{\partial S_j(x,t)}{\partial t} \\ &= \lim_{w\to 0} -\beta S_j(x, t-w) \cdot \int_\Omega K(x,y) \cdot w \frac{(A_j(y,t) + B_j(y,t))}{w} dy \end{aligned} \tag{3.11}$$

Note

$$\Delta I_j(y,t-iw) = I_j(y,t-(i-1)w) - I_j(y,t-iw) \approx \frac{\partial I_j(y,t-iw)}{\partial t} w. \tag{3.12}$$

Then

$$\lim_{w\to 0} -\beta S_j(x,t-w)\cdot \int_{\Omega} K(x,y)\cdot w \cdot A_j(y,t)dy$$

$$= \lim_{w\to 0} -\beta(1-\xi_j) S_j(x,t-w)\cdot \sum_{i=1}^{t/w} \int_{\mathbf{\Omega}} \frac{\partial I_j(y,t-iw)}{\partial t}\cdot K(x,y)\cdot w \cdot e^{-u_j\cdot iw} dy. \tag{3.13a}$$

and

$$\lim_{w\to 0} -\beta S_j(x,t-w)\cdot \int_{\Omega} K(x,y)\cdot w \cdot B_j(y,t)dy$$

$$= \lim_{w\to 0} -\beta S_j(x,t-w)\cdot \int_{\Omega} K(x,y)\cdot f_j(y,iw)dy \tag{3.13b}$$

where

$$f_j(y,iw) = \sum_{i=1}^{t/w} \int_0^{iw} \frac{\partial I_j(y,t-iw)}{\partial t} w\cdot (n_j u_j - \lambda_j)\xi_j \cdot e^{-(n_j u_j-\lambda_j)v}\cdot e^{-u_j\cdot(iw-v)}\, dv \tag{3.13c}$$

Note the terms $\big(\partial I_j(y,t-\tau)/\partial t\big)K(x,y)e^{-\lambda\tau}$ and $\int_0^{\tau} \frac{\partial I(y,t-\tau)}{\partial t}\cdot (n_j u_j-\lambda_j)\cdot \xi_j \cdot e^{-(n_j u_j-\lambda_j)v}\cdot e^{-u_j\cdot(\tau-v)}dv$ is continuous and bounded in the time domain. Both of them are Riemann integrable. Thereby, Eq. (3.13a) and Eq. (3.13b) can be written, respectively, as

$$\lim_{w\to 0} -\beta\big(1-\xi_j\big) S_j(x,t-w)\cdot \sum_{i=1}^{t/w} \int_{\mathbf{\Omega}} \frac{\partial I_j(y,t-iw)}{\partial t}\cdot K(x,y)\cdot w \cdot e^{-u_j\cdot iw} dy$$

$$= -\beta\cdot\big(1-\xi_j\big)\cdot S_j(x,t)\cdot \int_{\Omega} \int_0^t \frac{\partial I_j(y,t-\tau)}{\partial t}\cdot K(x,y)\cdot e^{-u_j\tau} d\tau \cdot dy \tag{3.14a}$$

$$\lim_{w\to 0} -\beta S_j(x,t-w)\cdot \int_{\Omega} K(x,y)\cdot f_j(y,iw)dy$$

$$= \beta S_j(x,t)\cdot \int_{\Omega} \int_0^t \int_0^{\tau} \frac{\partial I_j(y,t-\tau)}{\partial t} K(x,y)(n_j u_j-\lambda_j)\xi_j \cdot e^{-(n_j u_j-\lambda_j)v} e^{-u_j\cdot(\tau-v)} dv\cdot d\tau\cdot dy \tag{3.14b}$$

Note $\partial I_j(x,t-\tau)/\partial t = -\partial S_j(x,t-\tau)/\partial t$, submitting Eq. (3.14) into Eq. (3.11), we have

$$\frac{\partial S_j(x,t)}{\partial t} = \beta\cdot\big(1-\xi_j\big)\cdot S_j(x,t)\cdot \int_{\Omega} \int_0^t \frac{\partial S_j(y,t-\tau)}{\partial t}\cdot K(x,y)\cdot e^{-u_j\tau} d\tau dy +$$

$$\beta S_j(x,t)\cdot \int_{\Omega} \int_0^t \int_0^{\tau} \frac{\partial I_j(y,t-\tau)}{\partial t} K(x,y)(n_j u_j-\lambda_j)\xi_j \cdot e^{-(n_j u_j-\lambda_j)v} e^{-u_j\cdot(\tau-v)} dv\cdot d\tau\cdot dy \tag{3.15}$$

According to Eq. (3.10) and Eq. (3.12), in continuous space, the density of information-relaying vehicles can be written as

$$R_j(y,t) = -(1-\xi_j)\int_0^t \frac{\partial S_j(y,t-\tau)}{\partial t}\cdot e^{-u_j\tau}d\tau - \int_0^t\int_0^\tau \frac{\partial S_j(y,t-\tau)}{\partial t}\cdot(n_ju_j-\lambda_j)\xi_j\cdot e^{-(n_ju_j-\lambda_j)v}\cdot e^{-u_j\cdot(\tau-v)}dv\cdot d\tau \quad (3.16)$$

The terms $-(1-\xi_j)\int_0^t \frac{\partial S_j(y,t-\tau)}{\partial t}\cdot e^{-u_j\tau}d\tau dy$ and $-\int_0^t\int_0^\tau \frac{\partial S_j(y,t-\tau)}{\partial t}\cdot(n_ju_j-\lambda_j)\xi_j\cdot e^{-(n_ju_j-\lambda_j)v}\cdot e^{-u_j\cdot(\tau-v)}dv\cdot d\tau$ denote the density of information-relaying vehicles without queuing process and with queuing process, respectively. Let $\eta = t-\tau$, then

$$R_j(y,t) = f_{j,1}(y,t) + f_{j,2}(y,t) \quad (3.17a)$$

where

$$f_{j,1}(y,t) = -(1-\xi_j)\int_0^t \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot e^{-u_j(t-\eta)}d\eta \quad (3.17b)$$

$$f_{j,2}(y,t) = -\int_0^t\int_0^{t-\eta} \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot(n_ju_j-\lambda_j)\xi_j\cdot e^{-(n_ju_j-\lambda_j)v}\cdot e^{-u_j\cdot(t-\eta-v)}dv\cdot d\eta \quad (3.17c)$$

The derivative of $R_j(y,t)$ with respect to t is

$$\frac{\partial R_j(y,t)}{\partial t} = \frac{\partial f_{j,1}(y,t)}{\partial t} + \frac{\partial f_{j,2}(y,t)}{\partial t} \quad (3.18(a)$$

where

$$\frac{\partial f_{j,1}(y,t)}{\partial t} = -(1-\xi_j)\frac{\partial S_j(y,t)}{\partial t} + u_j(1-\xi_j)\int_0^t \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot e^{-u_j(t-\eta)}d\eta$$
$$= -(1-\xi_j)\frac{\partial S_j(y,\eta)}{\partial t} - u_jf_{j,1}(y,t) - u_jR_j(y,t) \quad (3.18b)$$

$$\frac{\partial f_{j,2}(y,t)}{\partial t} = -\int_0^t \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot(n_ju_j-\lambda_j)\xi_j e^{-(n_ju_j-\lambda_j)(t-\eta)}d\eta + u_jf_{j,1}(y,t) \quad (3.18c)$$

Thereby,

$$\frac{\partial R_j(y,t)}{\partial t} = -(1-\xi_j)\frac{\partial S_j(y,\eta)}{\partial t} - \int_0^t \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot(n_ju_j-\lambda_j)\xi_j\cdot e^{-(n_ju_j-\lambda_j)(t-\eta)}d\eta - u_jR_j(y,t) \quad (3.19)$$

Let $H_j(y,t) = -\int_0^t \frac{\partial S_j(y,\eta)}{\partial \eta}\cdot\xi_j\cdot e^{-(n_ju_j-\lambda_j)(t-\eta)}d\eta$. According to Eq. (3.7), the probability that an information packet in information class j is received by a vehicle at time η and is waiting in the queue at current time t $(t>\eta)$ is $\xi_j\cdot e^{-(n_ju_j-\lambda_j)(t-\eta)}$. This implies that $H_j(y,t)$ is the density of information-holding vehicles at location y and time t. Differentiating $H_j(y,t)$ with respect to t, we have

$$\frac{\partial H_j(y,t)}{\partial t} = -\xi_j\frac{\partial S_j(y,t)}{\partial t} - (n_ju_j-\lambda_j)H_j(y,t) \quad (3.20)$$

Let $E_j(x,t)$ denote the density of information-excluded vehicles at location x at time t. As

informed vehicles consist of the information-holding, information-relaying and information-excluded vehicles, $I_j(x,t) = H_j(x,t) + R_j(x,t) + E_j(x,t)$. Thereby, we have

$$\frac{\partial H_j(x,t)}{\partial t} + \frac{\partial R_j(x,t)}{\partial t} + \frac{\partial E_j(x,t)}{\partial t} = \frac{\partial I_j(x,t)}{\partial t} = -\frac{\partial S_j(x,t)}{\partial t}. \tag{3.21}$$

Substituting Eq. (3.19) and Eq. (3.20) into Eq. (3.21), yields

$$\frac{\partial E_j(x,t)}{\partial t} = u_j R_j(x,t) \tag{3.22}$$

According to the above analysis, we have the following IDE system:

$$\begin{cases} \frac{\partial S_j(x,t)}{\partial t} = -\beta S_j(x,t) \int_\Omega R_j(y,t) \cdot K(x,y) dy & \text{(3.23a)} \\ \frac{\partial H_j(x,t)}{\partial t} = \beta \cdot \xi_j \cdot S_j(x,t) \int_\Omega R_j(y,t) \cdot K(x,y) dy - \left(n_j u_j - \lambda_j\right) \cdot H_j(x,t) & \text{(3.23b)} \\ \frac{\partial R_j(x,t)}{\partial t} = \left(1 - \xi_j\right) \cdot \beta S_j(x,t) \int_\Omega R_j(y,t) \cdot K(x,y) dy + \left(n_j u_j - \lambda_j\right) \cdot H_j(y,t) - u_j \cdot R_j(x,t) & \text{(3.23c)} \\ \frac{\partial E_j(x,t)}{\partial t} = u_j \cdot R_j(x,t) & \text{(3.23d)} \end{cases}$$

For simplicity, we will label the IDE system (3.23) as the susceptible-holding-relaying-excluded (SHRE) model. It describes the instantaneous change in densities of vehicles by vehicle class for dissemination of an information packet in information class j. Eq. (3.23) shows that susceptible vehicles become informed vehicles at a rate proportional to the densities of susceptible vehicles and information-relaying vehicles (see Eq. (3.23a)). According to Eq. (3.23b), information-holding vehicles become information-relaying vehicles at a rate inversely proportional to $\left(n_j u_j - \lambda_j\right)$. Thereby, if the assigned number of communication servers (n_j) and the mean communication service rate (u_j) are increased, information-holding vehicles would become information-relaying vehicles faster. This implies that the specific information packet of interest experiences less waiting time in the queue. Hence, it can be propagated in the traffic stream at a higher speed. Eq. (3.23c) indicates that the density change of information-relaying vehicle increases monotonically with respect to number of communication servers (n_j) and the mean communication service rate (u_j). According to Eq. (3.23d), the density change of information-excluded vehicles is proportional to u_j. Note $1/u_j$ denotes the mean communication service time of information packets of information class j. If u_j is smaller, the information packet would stay in the commination server for a longer time, implying that it can be disseminated more times using the repetitive broadcast process. This will impact both the IFPW

speed and asymptotic density of informed vehicles. Thereby, we can control n_j and u_j to meet the application needs of information packets in information class j.

It is worth noting that Eq. (3.23d) can be used to characterize the dissemination wave of information packets in an arbitrary information class j, $j \in \mathcal{L}$. As $\sum_{i=1}^{S} n_i = N$, we can assign different number of communication servers and mean communication service rates for different information classes appropriately to meet their application needs simultaneously. Eq. (3.23) also implies that if $\xi_j \equiv 0$ (i.e., $P_j^0 = 0$, or $n_j u_j \to \lambda_j$), the SHRE model becomes the susceptible-relaying-excluded model studied by Wang et al. (2018). It models the information flow dissemination wave under an information-relay control strategy where there is no queuing delay (i.e., no information-holding vehicles).

3.3.2 Modeling the traffic flow dynamics in the lower-layer

The upper-layer SHRE model describes how the density of vehicles by vehicle class changes instantaneously due to V2V communications. It captures the impacts of communication constraints on success rate of V2V communications, as also factors the effects of the queuing strategy and the distribution of information-relaying vehicles on the IFPW formation. As mentioned before, the IFPW is a combination of the information flow dissemination wave and the traffic flow propagation wave. This section models the traffic flow dynamics to determine the traffic flow propagation wave. The effects of traffic flow dynamics on IFPW are threefold. First, they impact the success rate of V2V communications (i.e., $K(x, y)$) by determining the number of equipped vehicles within communication range of an information-relaying vehicle. Second, the number of the vehicles sending the specific information depends on the spatial distribution of information-relaying vehicles which evolves with the traffic flow dynamics. Third, the traffic flow speed significantly contributes to the IFPW speed. It adds to the IFPW speed in the direction of vehicular traversal and reduces the IFPW speed in direction opposite to that of vehicular traversal.

In this study, the first-order LWR model is used to describe the traffic flow dynamics. It can reproduce some essential features of traffic flow, such as the formation and propagation of traffic flow waves. The model consists of the flow conservation law and an explicit density-flow relationship known as the fundamental diagram of traffic flow. The flow conservation law and the fundamental diagram can be expressed as the following PDE model:

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{3.24}$$

$$q(x,t) = F(k,x,t) \tag{3.25}$$

where $k(x,t)$ is the traffic flow density at location x at time t, $q(x,t)$ is the instantaneous flow, and $F(k,x,t)$ is the fundamental diagram in which the flow and density are related by a continuous and piecewise differentiable equation.

3.4 Analytical and numerical solutions of the two-layer model

This section discusses the analytical and numerical solutions for the two-layer model under homogeneous and heterogeneous traffic conditions, respectively. Homogeneous traffic conditions refer to unidirectional traffic flow with uniform traffic flow density and the heterogeneous traffic conditions refer to the other traffic flow situations (e.g., bi-directional flow). Under homogeneous traffic conditions, the traffic flow dynamics only shift the IFPW towards the direction of traffic flow and do not change the densities of vehicles of different classes. Thereby, the impacts of traffic flow dynamics on the IFPW speed are uniform in space and time. The asymptotic density of informed vehicles and the condition for existence of IFPW can be derived analytically using only the upper-layer SHRE model.

Under heterogeneous traffic flow conditions, the traffic flow dynamics change the densities of vehicles of different classes spatiotemporally. Thereby, the impacts of the traffic flow dynamics on the IFPW speed are non-uniform in space and time. To obtain the solutions of the two-layer model under heterogeneous conditions, the change in density of vehicles of each vehicle class due to V2V communications (in the upper layer) and the traffic flow dynamics (in the lower layer) need to be tracked simultaneously. To do so, a numerical solution method is proposed here to capture the interactions between the upper and lower layers sequentially under discrete time and space settings.

3.4.1 Analytical solutions of the two-layer model under homogeneous traffic flow conditions

The proposed SHRE model conceptually adapts the idea of susceptible-exposed-infected-recovered (SEIR) model that is extensively studied in epidemiology (see e.g., McCluskey, 2012; Li and Muldowney., 1995; Li et al., 1999; Smith et al., 2001). The equilibrium solution and conditions for local stability of the SEIR model are analyzed in these studies. However, the

classical SEIR only address the temporal spreading of a disease among the population at one location. By comparison, the designed SHRE model is a spatial model that seeks to find how a specific information packet of interest will be propagated in both space and time. Thereby, the solutions of SEIR model in previous studies cannot be applied for the SHRE model. To the best of our knowledge, the solution of the IDE system analogy to the SHRE model has not been studied before. In the following, we will derive the analytical solutions for asymptotical density of vehicles by vehicle class of the proposed SHRE model and study the conditions for existence of IFPW.

Let σ be the density of equipped vehicles. Suppose at time 0, all the equipped vehicles are susceptible vehicles in the highway. Thereby, the initial conditions for the SHRE model are $S_j(x,0) = \sigma$, $H_j(x,0) = R_j(x,0) = E_j(x,0) = 0$. Assume the specific information packet of interest in information class j is generated and is propagated by an equipped vehicle at location 0 and time 0. Similar to Kim et al. (2017) and Wang et al., (2017), under homogeneous conditions, the information under the designed queueing strategy will quickly form a wave (if it exists) to propagate backward and forward with a uniform speed. The asymptotical density of vehicles of each vehicle class is the same beyond the location where the wave speed is stable. However, we cannot derive the analytical solutions for asymptotical speed of the IFPW. The IFPW speed will be solved using the numerical method introduced in section 4.2.

As discussed in Section 3.1, the communication kernel $K(x,y)$ characterizes the one-hop success rate of V2V communication. It significantly impacts the traveling wave solutions. By factoring communication constraints (e.g. communication frequency, channel capacity, communication range), Kim et al. (2017) show that the communication is subject to attenuation over distance. In this study, the Gaussian communication kernel proposed by Kim et al. (2017) will be used to characterize the one-hop success rate of V2V communication formulated in Eq. (3.26), in which the parameter a and b will be calibrated using V2V communication data obtained through NS-3 simulation. It is worth mentioning that the communication kernel satisfies

$$\int_{\Omega} K(x,y)dy = \int_{-\infty}^{+\infty} K(x,y)dy = b.$$

$$K(x,y) = \frac{b}{a\sqrt{\pi}} e^{\frac{-(x-y)^2}{a^2}},\ a > 0,\ 0 < b \leq 1, \tag{3.26}$$

The asymptotic density of vehicles by vehicle class and asymptotic density of informed vehicles are defined as follows.

Definition 3.1 (asymptotic density of vehicles by vehicle class): the asymptotic density of vehicles of vehicle class $z_j^*(x)$, $\mathrm{z}_j \in \{S_j, H_j, R_j, E_j\}$ at location x is defined as $z_j^*(x) = \lim_{t\to\infty} \mathrm{z}(x,t)$.

Definition 3.2 (asymptotic density of informed vehicles): let $I_j^*(x)$ be the asymptotic density of informed vehicles at location x, it is defined as $I_j^*(x) = \lim_{t\to\infty} I_j(x,t)$.

The following two theorems will be useful to analyze the asymptotic density of informed vehicles.

Theorem 3.1 (asymptotic density of information-holding vehicles): if $n_j u_j > \lambda_j$, then $H_j^*(x) = \lim_{t\to+\infty} H_j(x,t) = 0$

Proof: let $\omega_j = n_j u_j - \lambda_j$, multiply both sides of Eq. (3.23b) by $e^{\omega_j t}$, we have

$$e^{\omega_j t}\frac{\partial H_j(x,t)}{\partial t} + e^{\omega_j t}\cdot H_j(x,t) = e^{\omega_j t}\beta\cdot\xi_j\cdot S_j(x,t)\int_\Omega R_j(y,t)\cdot K(x,y)dy \tag{3.27}$$

This implies

$$\frac{\partial\left(e^{\omega_j t}H_j(x,t)\right)}{\partial t} = e^{\omega_j t}\beta\cdot\xi_j\cdot S_j(x,t)\int_\Omega R_j(y,t)\cdot K(x,y)dy \tag{3.28}$$

Then

$$\begin{aligned} &e^{\omega_j t}H_j(x,t) - e^{\omega_j t}H_j(x,0) \\ &\qquad = \int_0^t e^{\omega_j \tau}\beta\cdot\xi_j\cdot S_j(x,\tau)\int_\Omega R_j(y,\tau)\cdot K(x,y)dy\cdot d\tau \end{aligned} \tag{3.29}$$

Note $H_j(x,0)\equiv 0$. According to Eq. (3.23d), Eq. (3.29) can be written as

$$\begin{aligned} H_j(x,t) &= e^{-\omega_j t}\beta\cdot\xi_j\cdot\int_0^t e^{\omega_j\tau}S_j(x,\tau)\int_\Omega \frac{1}{u_j}\frac{\partial E_j(y,\tau)}{\partial\tau}\cdot K(x,y)dy\cdot d\tau \\ &= \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_0^t S_j(x,\tau)\,e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy \end{aligned} \tag{3.30}$$

Note that as $t\to+\infty$, the densities of vehicles of different vehicle classes at each location y become stable. Thereby, $\lim_{t\to+\infty}\partial E_j(y,\tau)/\partial t \to 0$. For an arbitrarily small positive value ε, let $g(g<+\infty)$ be the value such that $\partial E_j(y,\tau)/\partial\tau < \varepsilon$ for $\tau > g$. Then

$$\lim_{t\to+\infty} H_j(x,t) = \lim_{t\to+\infty}\frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_0^t S_j(x,\tau)\,e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy \tag{3.31}$$

$$= \lim_{t\to+\infty} \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_0^g S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy$$

$$+ \lim_{t\to+\infty} \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_g^{+\infty} S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy$$

Note $S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}$ is bounded when $\tau\in[0,g]$. Thereby, $\int_0^g S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau$ is bounded. This implies

$$\lim_{t\to+\infty} \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_0^g S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy \to 0 \tag{3.32}$$

As $\partial E_j(y,\tau)/\partial\tau < \varepsilon$ for $\tau > g$. Then

$$\begin{aligned}&\lim_{t\to+\infty} \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\int_\Omega K(x,y)\int_g^{+\infty} S_j(x,\tau)\, e^{\omega_j\tau}\frac{\partial E_j(y,\tau)}{\partial\tau}d\tau\cdot dy\\ &\le \lim_{t\to+\infty} \frac{e^{-\omega_j t}\beta\cdot\xi_j}{u_j}\sigma\, e^{\omega_j t}\cdot\varepsilon\\ &= \frac{\beta\cdot\xi_j}{u_j}\sigma\cdot\varepsilon\end{aligned} \tag{3.33}$$

Note ε is an arbitrarily small positive value. Thereby,

$$H_j^*(x) = \lim_{t\to+\infty} H_j(x,t) = 0 \tag{3.34}$$

Theorem 3.2 (asymptotic density of information-relaying vehicles): if $n_j u_j > \lambda_j$, then $R_j^*(x) = \lim_{t\to+\infty} R_j(x,t) = 0$.

Theorem 3.2 can be proved using the same method used to prove Theorem 3.1; it is omitted here to avoid duplication.

Theorems 3.1 and 3.2 indicate that if $n_j u_j > \lambda_j$, there would be no information-holding and information-relaying vehicles at each location eventually. This is because when $n_j u_j > \lambda_j$, the specific information packet of interest waiting in the queue for information class j will have finite waiting time and finite communication service time. It will enter into the communication server for propagation and is removed from it eventually.

Let $\gamma_j = \beta b\sigma/u_j$, the following theorem discusses the asymptotic density of information-excluded vehicles.

Theorem 3.3 (asymptotic density of information-excluded vehicles): if $n_j u_j > \lambda_j$, and $\gamma_j > 1$, then $R_j^*(x) = \lim_{t\to+\infty} R_j(x,t) = \sigma \cdot \alpha_j^*$, where $\alpha_j^* \in (0,1)$ is the unique solution of the following nonlinear equation

$$e^{-\gamma_j \alpha_j} + \alpha_j - 1 = 0 \tag{3.35}$$

Proof: Note under homogeneous conditions, the traffic flow density is uniform. Thereby, for arbitrary time t and location x, we have

$$S_j(x,t) + H_j(x,t) + R_j(x,t) + E_j(x,t) = \sigma \tag{3.36}$$

Note that $E_j(x,0) = 0$, according to Eq. (3.23d)

$$E_j(x,t) = \int_0^t u_j \cdot R_j(y,t)dt \tag{3.37}$$

Eq. (3.23a) implies that

$$\frac{\partial\left(ln\left(S_j(x,t)\right)\right)}{\partial t} = -\beta \int_\Omega R_j(y,t) \cdot K(x,y)dy \tag{3.38}$$

Integrating both side of Eq. (3.38) from 0 to t

$$ln\left(S_j(x,t)\right) - ln(\sigma) = -\beta \int_\Omega \left[\int_0^t R_j(y,\tau)\, d\tau\right] \cdot K(x,y)dy \tag{3.39}$$

Substituting Eq. (3.37) into Eq. (3.39), yields

$$\begin{aligned} ln\left(S_j(x,t)\right) - ln(\sigma) &= -\frac{\beta}{u_j} \int_\Omega \left[\int_0^t u_j R_j(y,\tau)\, d\tau\right] \cdot K(x,y)dy \\ &= -\frac{\beta}{u_j} \int_\Omega E_j(x,t) \cdot K(x,y)dy \end{aligned} \tag{3.40}$$

Thereby,

$$S_j(x,t) = \sigma \cdot e^{-\frac{\beta}{u_j}\int_\Omega E_j(x,t)\cdot K(x,y)dy} \tag{3.41}$$

Substituting Eq. (3.41) into Eq. (3.36), we have

$$\sigma \cdot e^{-\frac{\beta}{u_j}\int_\Omega E_j(x,t)\cdot K(x,y)dy} + H_j(x,t) + R_j(x,t) + E_j(x,t) = \sigma \tag{3.42}$$

Let $t \to +\infty$, then

$$\sigma \cdot e^{-\frac{\beta}{u_j}\int_\Omega E_j^*(x)\cdot K(x,y)dy} + H_j^*(x) + R_j^*(x) + E_j^*(x) = \sigma \tag{3.43}$$

Note that $\int_\Omega E_j^*(x) \cdot K(x,y)dy = b \cdot E_j^*(x)$, and according to Theorems 3.1 and 3.2, $H_j^*(x) = R_j^*(x) = 0$, then

$$\sigma \cdot e^{-\frac{\beta b}{u_j}E_j^*(x)} + E_j^*(x) = \sigma \tag{3.44}$$

Let $\alpha_j = (1/\sigma) \cdot E_j^*(x)$. As $E_j^*(x) \in [0, \sigma]$, $\alpha_j \in (0,1)$. Eq. (3.44) can be rewritten as

$$e^{-\gamma_j \alpha_j} + \alpha_j = 1 \tag{3.45}$$

For simplicity, denote the function $\varphi(\alpha_j)$ as

$$\varphi(\alpha_j) = e^{-\gamma_j \alpha_j} + \alpha_j - 1 \tag{3.46}$$

Note that

$$\varphi'(\alpha_j) = d\varphi(\alpha_j)/d\alpha_j = -\gamma_j e^{-\gamma_j \alpha_j} + 1 \tag{3.47}$$

According to Eq. (3.47), $d\varphi(\alpha_j)/d\alpha_j\big|_{\alpha_j=0} = -\gamma_j + 1 < 0$ for $\gamma_j > 1$. As $\varphi(0) = 0$, this implies that $\varphi(\alpha_j) < 0$ for α_j sufficiently close to 0. Note that $\varphi(1) = e^{-\gamma_j} > 0$. Hence, there must exist a solution to Eq. (3.45) for $\alpha_j \in (0,1)$. The second-order derivative of $\varphi(\alpha_j)$ with respect to α_j is $\varphi''(\alpha_j) = d^2\varphi(\alpha_j)/d{\alpha_j}^2 = (\gamma_j)^2 e^{-\gamma_j \alpha_j} > 0$. Thereby, $\varphi(\alpha_j)$ is a convex function. There exists at most two solutions for $\varphi(\alpha_j) = 0$. $\alpha_j = 0$ is a solution to $\varphi(\alpha_j) = 0$. This implies that there exists a unique solution to $\varphi(\alpha_j) = 0$ for $\alpha_j \in (0,1)$. Let α_j^* be the corresponding solution. We have $E_j^*(x) = \sigma \cdot \alpha_j^*$. Theorem 3.3 is proved.

Recall $I_j^*(x) = H_j^*(x) + R_j^*(x) + E_j^*(x)$, and $H_j^*(x) = R_j^*(x) = 0$. We have the following corollary.

Corollary 3.1: The asymptotic density of informed vehicles is $I_j^*(x) = E_j^*(x) = \sigma \cdot \alpha_j^*$.

According to Eq. (3.34), α_j^* is determined only by the value of γ_j which equals $\beta b \sigma / u_j$. Thereby, when the communication frequency (β), the parameter b in the communication kernel and the initial density of equipped vehicles (σ) are fixed, the service rate u_j can be leveraged by transportation operators to propagate the specific information in information class j to control the proportion of vehicles that can receive the specific information. Let $P_{I_j}^*$ be the information spread (i.e., proportion of informed vehicles) for the specific information packet of interest in information class j; we have $P_{I_j}^* = \sigma \cdot \alpha_j^* / \sigma = \alpha_j^*$.

The following theorem discusses the existence of the IFPW.

Theorem 3.4 (conditions for existence of IFPW): The IFPW does not exist when $\gamma_j < 1$.

Proof: Note for $\alpha \in [0,1]$,

$$\varphi'(\alpha_j) = d\varphi(\alpha_j)/d\alpha_j = -\gamma_j e^{-\gamma_j \alpha_j} + 1 > -\gamma_j + 1 > 0 \tag{3.48}$$

Thereby, $\varphi(\alpha_j)$ increases monotonically with respect to α_j for $\alpha_j \in (0,1)$. As $\varphi(0) = 0$, there

exists no solution to $\varphi(\alpha_j) = 0$ for $\alpha_j \in (0,1)$. This implies that the vehicles are far from the location (location is labeled as x) where the information packet is generated, and the asymptotic solution of $E_j(x,t)$ is 0. As all informed vehicles will become information-excluded vehicles, this result indicates that no vehicle can receive the specific information packet of interest of information class j if they are far from the location where the information is generated in this case. Thereby, the IFPW does not exist for $\gamma_j < 1$.

Theorem 3.4 shows that if the initial density of equipped vehicles and the service rate are high, the specific information packet of interest can only be propagated locally. Vehicles that are far from the location where the specific information is generated cannot receive it. This property can be used by transportation operators to design effective control strategies to propagate information within a small vicinity (e.g., sudden braking information, lane merge information). It should be noted that u_j will impact the propagation distance of the specific information of interest when the IFPW does not exist. Through a numerical example, we will show that if u_j is set such that γ_j is closer to 1 ($\gamma_j < 1$), the specific information of interest will be propagated further away. If $\gamma_j > 1$, the specific information packet will form a wave to be propagated in the network.

As discussed earlier, we cannot derive an analytical solution for the asymptotic IFPW speed even if it exists. The IFPW speed will be computed using the numerical method introduced in the next section. Note that the asymptotic IFPW speed is significantly impacted by the queuing delay which is determined by the two control parameters (i.e., n_j and u_j) simultaneously. To meet the application needs of information in an arbitrary information class j related to information spread, time delay to reach a target location and spatial coverage, the values for the two control parameters n_j and u_j can be determined as follows: first, choose u_j appropriately according to Theorem 3.4 if the information needs to be only propagated locally. If the information needs to be propagated in the network, then determine u_j appropriately according to Theorem 3.4 and corollary 3.1 so that information spread can be satisfied. Third, determine n_j appropriately to control the IFPW speed so that it can reach the target location in the desired time.

3.4.2 Numerical solution method

The analytical solutions for the various information classes introduced in previous section only

apply to homogeneous traffic conditions. Under heterogeneous conditions, the IFPW may not be stable due to the non-uniform impact of traffic flow dynamics on information dissemination. To analyze how information is spread in space and time under heterogeneous conditions, this section proposes a numerical solution method based on Kim et al. (2017) to solve the two-layer model. The numerical solution method helps to: (1) estimate the IFPW speed under both homogeneous and heterogeneous traffic conditions, and (2) estimate the distance the specific information can be propagated when the IFPW does not exist, under both homogeneous and heterogeneous traffic conditions, and (3) estimate the density of informed vehicles under heterogeneous conditions.

The numerical solution method discretizes space and time into cells of length Δx and time interval Δt, respectively. Let $1,2,3\cdots$ denote the cells in the highway sequentially. The fourth-order Runge-Kutta method will be used to approximate the densities of vehicles by vehicle class (i.e., $S_j(x,t)$, $H_j(x,t)$, $R_j(x,t)$ and $X_j(x,t)$) changed according to the SHRE model in the upper layer. To solve the LWR model in the lower layer, the generalized cell transmission finite difference method proposed by Daganzo (1995) is used to approximate Eq. (3.24) and Eq. (3.25) as follows

$$[k(x,t+\Delta t)-k(x,t)]/\Delta t = [q(x-\Delta x,t)-q(x,t)]/\Delta x \tag{3.49}$$

$$q(x,t) = \min\left\{T\big(k(x,t)\big), Q\left(k_{jam}-k(x+\Delta x,t)\right)\right\}, \tag{3.50}$$

where T specifies the maximum flow that can be sent by the upstream cell and Q specifies the maximum flow that can be received by the downstream cell. k_{jam} is the jam traffic density. Let U denote the unequipped vehicles. The steps to solve the two-layer model numerically are as follows:

Step 1: At time 0 ($t=0$), obtain the initial number of vehicles of each class $z, z \in \{S_j, H_j, R_j, E_j, U\}$ and corresponding density of vehicles of each vehicle class in each cell. Let $t = t+\Delta t$.

Step 2: Solve the lower-layer model to determine the flow in each cell x (i.e., $q(x,t)$) that advances to the downstream cell according to Eq. (3.49) and Eq. (3.50). Update the number of vehicles in each cell.

Step 3: Calculate the number of vehicles of each class $z \in \{S_j, H_j, R_j, E_j, U\}$ that advance to the downstream as follows:

$$q_z(x,t) = \frac{k_z(x,t-\Delta t)}{k(x,t-\Delta t)} \cdot q(x,t), \quad z \in \{S_j, H_j, R_j, E_j, U\}, \tag{3.51}$$

where $q_z(x,t)$ is the traffic flow of class z leaving cell x at time interval t and $k_z(x,t-\Delta t)$ is the density of class z in cell x and time $t-\Delta t$.

Step 4: Update the density of vehicles by vehicle class in each cell of the upper layer using the discrete multiclass flow conservation law, as follows:

$$[k_z(x,t)-k_z(x,t-\Delta t)]/\Delta t=[q_z(x-\Delta x,t)-q_z(x,t)]/\Delta x,\quad z\in\{S_j,H_j,R_j,E_j,U\}. \tag{3.52}$$

where $k_{s_j}(x,t)$, $k_{H_j}(x,t)$, $k_{R_j}(x,t)$ and $k_{E_j}(x,t)$ represents $S_j(x,t)$, $H_j(x,t)$,$R_j(x,t)$ and $E_j(x,t)$, respectively that describe the density of vehicles of each class in the upper layer.

Step 5: Approximate the density of vehicles by vehicle class (i.e., $S_j(x,t), H_j(x,t), R_j(x,t)$ and $E_j(x,t)$) that are changed according to the SHRE model in the upper layer.

Step 6: If the predetermined time length is reached, then stop. Otherwise, let $t=t+\Delta t$, and go to Step 2.

The numerical method solves the discretized LWR model and the SHRE model sequentially to capture the effects of traffic flow dynamics on information dissemination. It is worth noting that to reduce computational load, the convolution term ($\int_\Omega R_j(y,t)\cdot K(x,y)dy$) in Eq. (3.23) can be approximated using the Fast Fourier Transform (FFT) method. More details of the numerical method and the FFT method can be found in Kim et al. (2017). The numerical method can provide the density of vehicles by vehicle class at each cell and time interval.

The numerical method can also be used to verify the analytical solutions of density of informed vehicles and to approximate the IFPW speed under both homogeneous and heterogeneous traffic conditions. Note that IFPW consists of two waves: the forward wave which travels in the direction of vehicular traversal and the backward wave which travels opposite to the direction of vehicular traversal. Correspondingly, there exist two IFPW speeds, the forward and backward IFPW speeds. The method to estimate the two IFPW speeds is as follows. Let t_1 and t_2 be two arbitrary time intervals. Without loss of generality, let $t_1>t_2$. Let $z_{j,0}$ be the density of vehicles in an arbitrary vehicle class z_j. $z_{j,0}$ is set between the minimum and maximum density of vehicle class z_j. Then, at the two time intervals $t_h, h=1,2$, there exist two cells on the two wave fronts, respectively, for which the density of vehicle class z_j is most close to $z_{j,0}$. Denote the two cells as l_{B,t_h} and l_{F,t_h} ($h=1,2$), respectively. Without loss of generality,

let l_{B,t_h} be the cell in the backward IFPW and the l_{F,t_h} be the cell in the forward IFPW. Then, the forward IFPW (labeled as c_F) and the backward IFPW (labeled as c_B) can be approximated as

$$c_F = \frac{\left(l_{F,t_2} - l_{F,t_1}\right)\Delta x}{t_1 - t_2} \tag{3.53}$$

$$c_B = \frac{\left|\left(l_{B,t_2} - l_{B,t_1}\right)\Delta x\right|}{t_1 - t_2} \tag{3.54}$$

3.5 Numerical example

This section discusses several numerical experiments to illustrate the application of the proposed method to control multiclass information flow propagation. Consider a highway with 30 km length. Discretize the highway uniformly into 2000 cells. Table 1 shows other inputs in the experiment.

Table 3.1 Experiment parameters

Traffic flow parameters	Value
Free flow speed (u_f)	108 km/h
Time interval (Δt)	0.5 seconds
Cell length (Δx)	15 meters
Number of lanes	1
Market penetration rate (W)	50%

3.5.1 Calibrate the communication kernel

To calibrate the communication kernel function in Eq. (3.26), NS-3 will be used to simulate the success rate of one-hop V2V communications under different traffic flow densities. NS-3 is a discrete network simulator that can simulate and test a spectrum of communication protocols efficiently. Recently, NS-3 has been used to simulate V2V communications and evaluate the performance of communication protocols for vehicular ad hoc networks (see e.g., Dey et al., 2016; Noori and Olyaei., 2013; Talebpour et al., 2016). The inputs for the V2V communication related parameters in NS-3 are shown in Table 2. The simulation is operated based on IEEE 802.11p protocol in 5.9 GHz band with channel capacity 3 Mbps and communication power 500 m. In V2V communications, whether a receiver vehicle can successfully receive an information packet from a sender vehicle is primarily decided by two factors: the reception signal power, and the noise and interference. The reception signal power determines whether the receiver vehicle can receive the signals from the sender vehicles, and the level of noise and

interference determines the probability of reception error. In this simulation, the Friis propagation loss model (Benin et al., 2012) is used in NS-3 to calculate the reception signal power. It characterizes the impacts of transmission power, distance between receiver and sender, transmission gain, and reception gain on reception signal power. The receiver vehicle receives the information packet only if the reception signal power is larger than the energy detection threshold -96 dBm. To estimate the noise and interference, the signal to (interference and) noise ratio (SINR) model is used in NS-3 simulation. The SINR is the ratio of the power of a certain signal of interest over the sum of the interference power (from all the other interfering signals) and the power of some background noise (for details, see Wang et al. 2018). The threshold of SINR is set as 5 dB (Hisham et al., 2016; Hisham et al., 2017), indicating that the V2V communication is considered to be successful if SINR is larger than 5 dB; otherwise, it will be considered as a communication failure.

Table 3.2 Inputs for NS-3 parameters

Parameters	Value
IEEE 802.11p channel capacity	3 Mbps
Band	5.9 GHz
Communication frequency	2 Hz
Communication power/distance	500 m
Minimum contention window	15 slots
Energy detection threshold	-96 dBm
Noise floor	-99 dBm
SINR threshold	5 dB

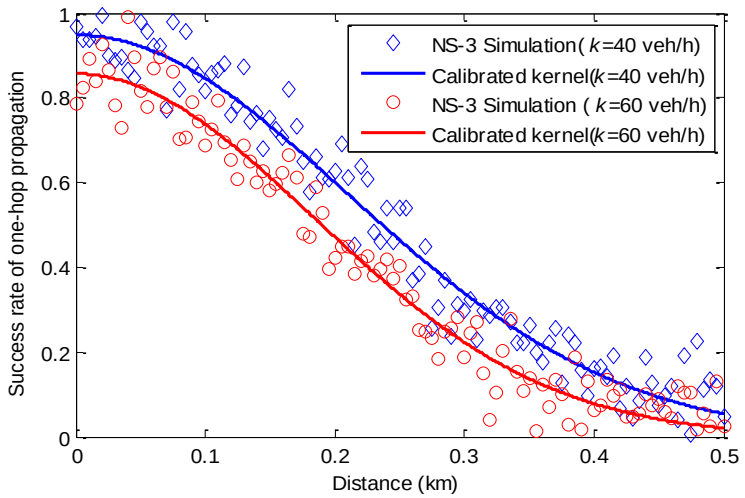


Figure 3.3 Calibrated communication kernel for $k = 40$ veh/h and $k = 60$ veh/h

Table 3.3 Maximum number of communication servers, and calibrated parameters in communication kernel using NS-3 simulation

Density (veh./km)	10	20	30	40	50	60	70	80	90	100
a	0.362	0.351	0.313	0.292	0.267	0.258	0.216	0.199	0.176	0.153
b	0.621	0.576	0.531	0.499	0.434	0.392	0.357	0.291	0.268	0.243
N_{max}	125	62	41	31	25	21	18	15	14	12

Recall that all equipped vehicles within communication range are assumed to share the bandwidth equally. Suppose the single unit of an information packet is 500 bytes. To prevent information congestion effects that would occur if the channel capacity is full, the maximum number of communication servers can be calculated as follows

$$N_{max} = \left[\frac{C}{2 \cdot k \cdot \beta \cdot R \cdot W \cdot \chi}\right].$$

where R is the communication range, C is the channel capacity (3 Mbps), W is the market penetration rate of V2V-equipped vehicles (50% in this study). χ is the information packet size (500 bytes), and k is the traffic flow density. The operator $[h]$ means the largest integer less than h. The calculated N_{max} values for different traffic flow densities are shown in Table 3. They will be used as the total number of communication servers under the corresponding traffic flow densities. To account for the impacts of positions of vehicles on success rate of V2V communications, vehicles are assumed to be randomly distributed along the 30 km highway. The simulation is conducted for 30 minutes, and is repeated 100 times. The calibrated parameters in the communication kernel are presented in Table 2. Figure 3.3 illustrates the simulated success rate of one-hop propagation and calibrated communication kernels at $k = 40$ veh/h and $k = 60$ veh/h. The R-squared values of the calibrated communication kernel at $k = 40$ veh/h and $k = 60$ veh/h are 0.96 and 0.94, respectively, indicating that the calibrated communication kernel robustly captures the relationship between success rate of one-hop propagation and the distance of the sender vehicle to the receiver vehicle. In addition, for the same distance, the success rate of one-hop propagation at $k = 60$ veh/h is less than it is at $k = 40$ veh/h. This because the communication interference increases if more vehicles are located within the communication range of a sender vehicle, causing greater communication failure.

3.5.2 IFPW under homogeneous conditions

3.5.2.1 Asymptotic density of informed vehicles and IFPW speed

The following example shows how to calculate the asymptotic density of informed vehicles analytically under homogeneous conditions. Suppose the traffic flow density is 40 veh/h, and the market penetration rate of equipped vehicles is 50%. Then, for each cell, the number of equipped vehicles is 0.3 veh/cell (σ). According to Table 3.3, the parameters a and b in Eq. (3.26) are 0.292 and 0.499, respectively. Suppose the specific information of interest is from information class j. Assume the number of communication servers (n_j) assigned to send the information packets in information class j is 11 and the corresponding mean communication service rate u_j is 0.05 packet/seconds (i.e., mean service time is 20 seconds). Note $n_j u_j - \lambda_j = 11 \times 0.05 - 0.5 = 0.05 > 0$ and $\gamma_j = \beta b \sigma / u_j = 2 \times 0.499 \times 0.5 * \frac{0.6}{0.05} = 2.994 > 1$. According to Theorem 3.4 and Corollary 3.1, the asymptotic density of informed vehicles exists. Corollary 3.1 indicates that $I_j^*(x) = 0.3 \cdot \alpha_j^*$, where $\alpha_j^*$is the unique solution of the nonlinear solution $e^{-3.384\alpha_j} + \alpha_j - 1 = 0$ for $\alpha_j \in [0,1]$. Using Newton method to solve the nonlinear equation, we have $\alpha_j^* = 0.9613$. Thereby, the asymptotic density of informed vehicles is $I_j^*(x) = 0.3 \cdot 0.9613 = 0.288$ vehicel/ cell $= 19.2\ veh/km$. The information spread (proportion of informed vehicles) is $P_{I_j}^* = \alpha_j^* = 0.9613$.

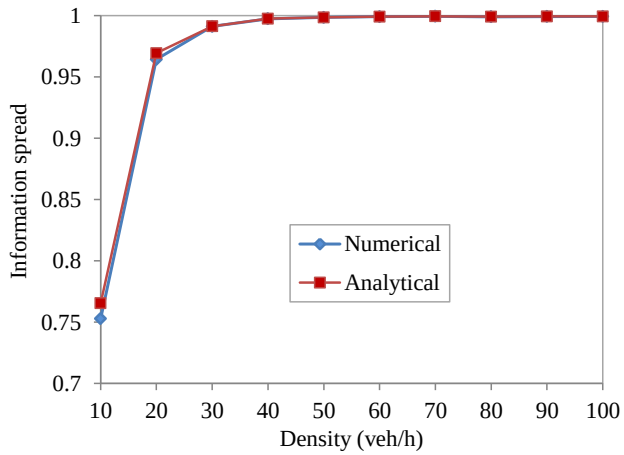


Figure 3.4 Asymptotic proportions of informed vehicles at $n_j = 11$ and $u_j = 0.05$ under different traffic densities

Figure 3.4 compares the information spread of the specific information packet of information class j under different traffic flow densities at $n_j = 11$ and $u_j = 0.05$ packets/second. It shows that the numerical solutions overlap with the analytical solutions, implying that the numerical algorithm proposed in Section 3.4.2 can effectively solve the two-layer model. Figure 3.4 also demonstrates that when traffic flow density increases, the information spread also increases as more vehicles will propagate it in an unit of time. This indicates that under higher traffic flow density scenarios, the mean communication service rate of information packets of class j can be reduced for the same information spread.

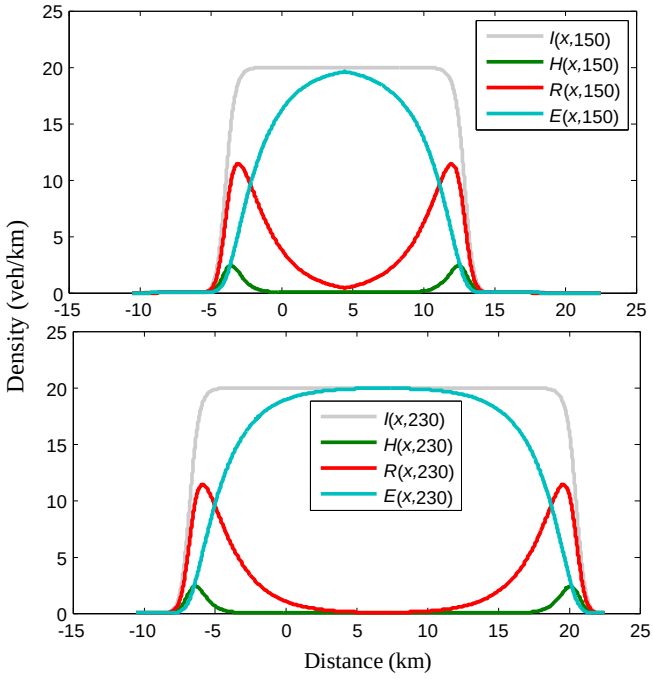


Figure 3.5 Density of vehicles by vehicle class at $t = 150$ seconds and $t = 230$ seconds

Suppose the traffic flow density is 40 veh/h and the specific information packet of interest in information class j is generated by a vehicle at time 0 and location 0. Let $\lambda_j = 0.5$ packets/second, $n_j = 20$ and $u_j = 0.03$ packets/second. Figure 3.5 shows the spatial distribution of density of vehicles by vehicle classes at $t = 150$ seconds and $t = 230$ seconds. It indicates that the IFPW can form the same shape to move forward and backward. Most of the information-holding and information-relaying vehicles are located close to the wave front. This is because when $\lambda_j < n_j u_j$ and $u_j > 0$, the information packet in information class j will experience finite queuing delay and communication service time. Thereby, the vehicles that receive the specific

information packet of interest a long time ago will exclude it from the system. The information-holding and information-relaying vehicles will become information-excluded vehicles eventually.

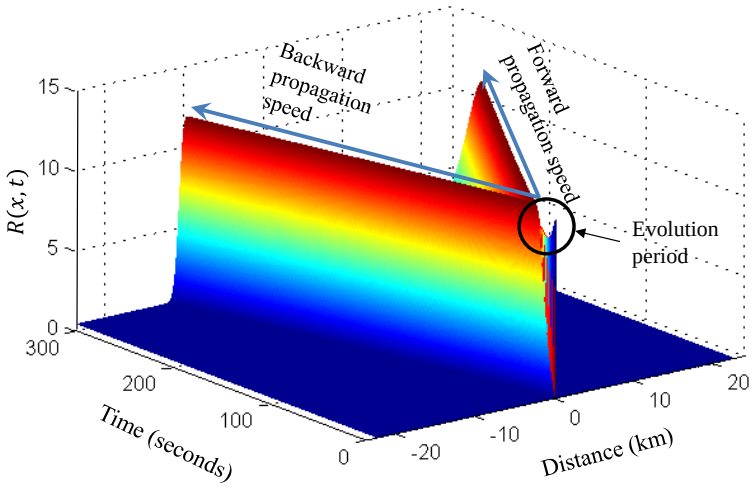


Figure 3.6 Density of information-relaying vehicles in space and time at $k = 40\ veh/km$

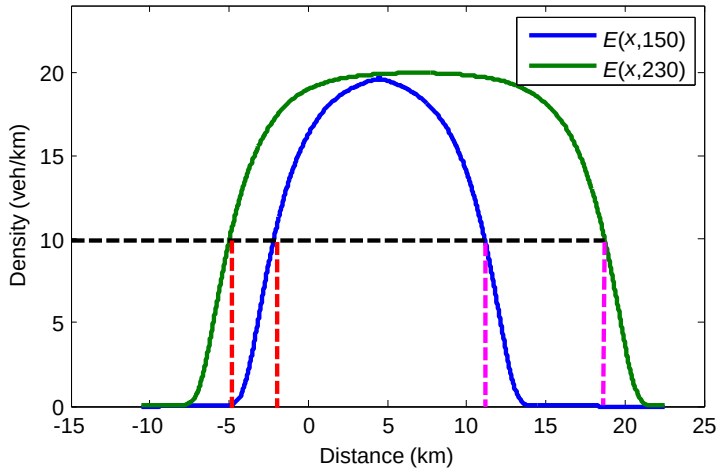


Figure 3.7 Density of information-excluded vehicles at $t = 150$ seconds and $t = 230$ seconds

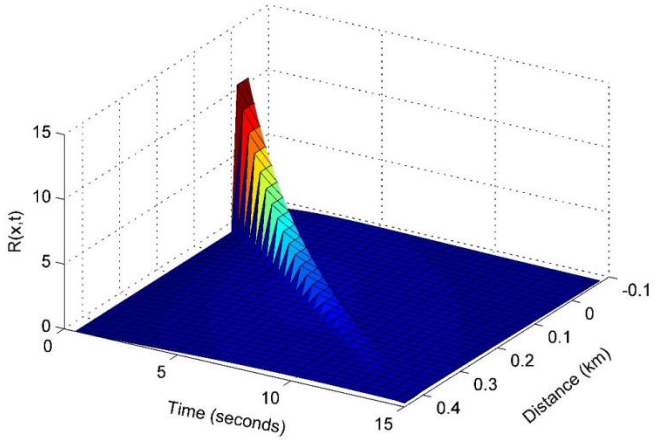
To analyze the asymptotic IFPW speed, Figure 3.6 shows the spatiotemporal distribution of density of information-relaying vehicles. It illustrates that the specific information packet of interest of information class j is propagated backward and forward at a uniform speed reached only a few seconds after it is generated. To numerically estimate the IFPW speed, Figure 3.7 shows the spatial distribution of density of information-excluded vehicles at $t = 150$ seconds and $t = 230$ seconds. Let $R_0 = 10\ veh/km$ be the reference density in Eq. (3.53) and Eq. (3.54).

According to Figure 3.7, the cells whose densities of information-relaying vehicles in the backward and forward IFPWs are most close to R_0 at $t = 150$ seconds are located at -2.085 km and 11.055 km, respectively. The cells whose densities of information-relaying vehicles in the backward and forward IFPWs are most close to R_0 at $t = 230$ seconds are located at -4.875 km and 18.645 km, respectively. According to Eq. (3.53) and Eq. (3.54), the forward and backward IFPW speeds can be estimated as:

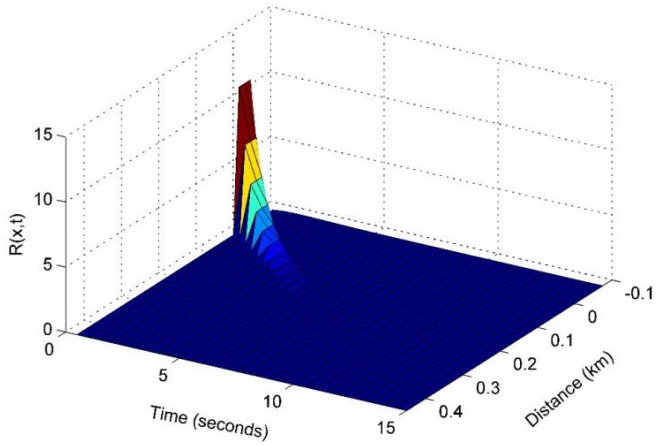
$$c_F = \frac{(l_{F,t_2} - l_{F,t_1})\Delta x}{t_1 - t_2} = \frac{-2.085 + 4.875}{150 - 230} = 0.09488 km/s = 341 km/h$$

$$c_B = \frac{\left|(l_{B,t_2} - l_{B,t_1})\Delta x\right|}{t_1 - t_2} = \frac{|-2.085 + 4.875|}{150 - 230} = 0.03487 km/s = 125 km/h$$

3.5.2.2 Scenarios where the information packet can be propagated only locally



(a) Density of information-relaying vehicle in space and time for $n_j = 20$ and $u_j = 0.323$



(b) Density of information-relaying vehicle in space and time for $n_j = 20$ and $u_j = 0.9$.

Figure 3.8 Scenarios for which the information packets are propagated only locally

As Theorem 3.4 indicates, the IFPW does not exist when the mean communication service rate is high enough such that $\gamma_j < 1$. In this case, the specific information packet in information class j can be propagated only locally. This property can be leveraged to send information packets in a small vicinity of where they are generated. The following example seeks to demonstrate how to control the propagation distance by leveraging communication service rate when information is propagated locally. Suppose the traffic flow density is 40 veh/h, the average arrival rate of information packets of information class j is 2 packets/second, and the number of assigned communication servers for information class j is 20. Let the density of information-relaying vehicles at location 0 and time 0 be 11 veh/km and 0 elsewhere. Figure 3.8(a) and Figure 3.8(b) show the spatiotemporal distribution of density of information-relaying vehicles at $u = 0.323$ and $u = 0.9$, respectively. Note $\gamma_j < 1$ in both cases. Figure 3.8 illustrates that the density of information-relaying vehicles decreases to 0 in space and time. Recall only information-relaying vehicles can propagate the specific information packets of interest. This implies that the specific information packet can only be propagated locally. Vehicles far away from location 0 where the specific information packet is generated will not receive it. It can be noted that the density of information-relaying vehicles decreases to 0 at 350 meters and 150 meters downstream of location 0 at $u = 0.323$ and $u = 0.9$, respectively. This implies the information packets can be propagated further away under a lower communication service rate. Thereby, the mean communication service rate can also be leveraged to propagate information to different distances.

3.5.2.3 Integrated impacts of n_j and u_j on asymptotic IFPW speed and density of informed vehicles

Suppose the traffic flow density is 50 veh/km, and the arrival rate of the information packets in information class j is 1 packet/second. To analyze the impacts of n_j and u_j on IFPW speed and density of informed vehicles, u_j is varied from 0.05 packets/second to 0.25 packets/second. According to M/M/n_j queuing theory, to enable propagation of the specific information packet of information class j, $\lambda_j/(n_j u_j)$ must be less than 1. Thereby, the minimum number of communication servers assigned to send information packets of information class j are 21, 11, 7, 6, 5 for $u_j = 0.05$, 0.1, 0.15, 0.2 ,0.25, respectively. Figure 3.9 shows the asymptotic forward IFPW speed for various values of n_j and u_j. It shows that when u_j is fixed, the asymptotic

forward IFPW speed increases monotonically with respect to n_j. This is because as more communication servers are assigned to information class j, the mean waiting time of the specific information packet in the queue will reduce (see Figure 3.10). Thereby, it can be transmitted faster by the informed vehicles. Figure 3.9 also shows that for a fixed n_j, the IFPW speed decreases monotonically with respect to mean communication service rate u_j in most cases because an increase in mean communication service rate will reduce the transmission duration of the packet. However, in some cases (e.g., $n_j = 22, 23$, etc.), increase in u_j may decrease the forward IFPW speed. This is because for a fixed n_j, increase in u_j will increase the mean waiting time in the queue (see Figure 3.10). Thereby, unlike that of the number of communication servers, the effect of mean communication service rate on the IFPW speed is more intricate. The proposed method in this study aids in determining the appropriate mean communication service rate for each information class to satisfy its application needs in terms of propagation performance.

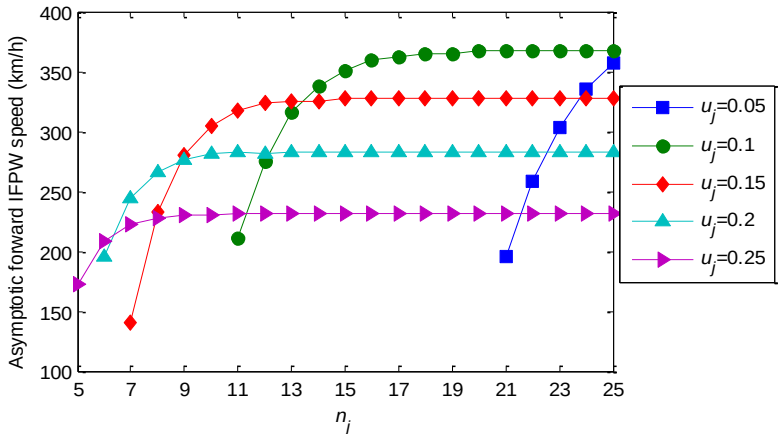


Figure 3.9 Impacts of n_j and u_j on asymptotic forward IFPW speed of an information packets of information class j

Figure 3.11 shows the information spread of the specific information of interest under different communication service rates. As the value of u_j increases, information spread decreases monotonically, implying that less number of vehicles will receive the specific information of interest. This is because an informed vehicle will exclude the specific information packet of interest from the communication servers faster under higher mean communication service rate. It is worth mentioning that the number of communication servers has no effect on information spread. Thereby, to design effective control strategies for propagating information packets in

different classes, the mean communication service rate can be determined first to obtain the desired information spread. Then, the appropriate number of communication servers can be determined to be assigned to different information classes to control their propagation speed. Further, Figure 3.11 shows that the numerical solutions are almost identical to the analytical solutions.

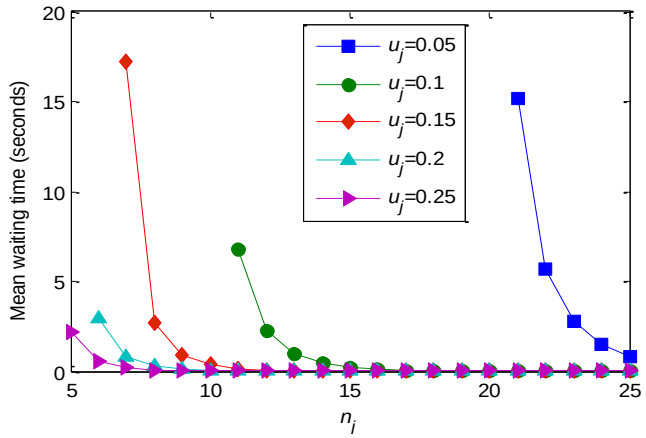


Figure 3.10 Mean waiting time of information packets in the queue for various values of n_j and u_j

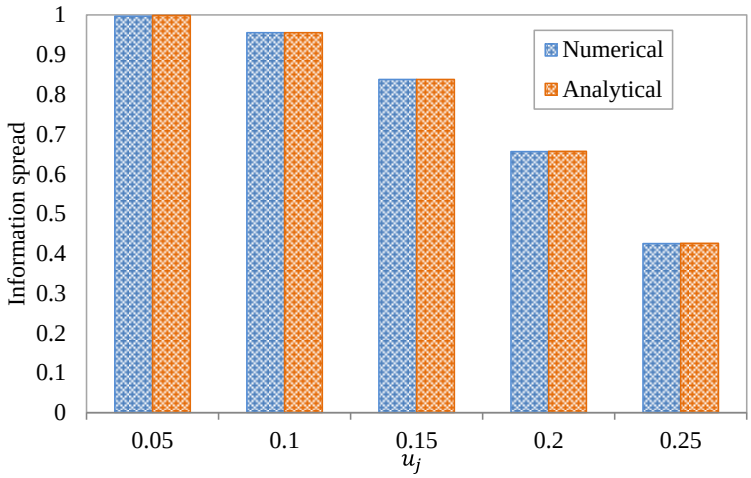


Figure 3.11 Comparison of numerical and analytical solutions of information spread for different values of u_j

3.5.3 Control of multiclass information flow propagation under homogeneous and heterogeneous traffic conditions

3.5.3.1 Control of multiclass information flow propagation under homogeneous traffic conditions

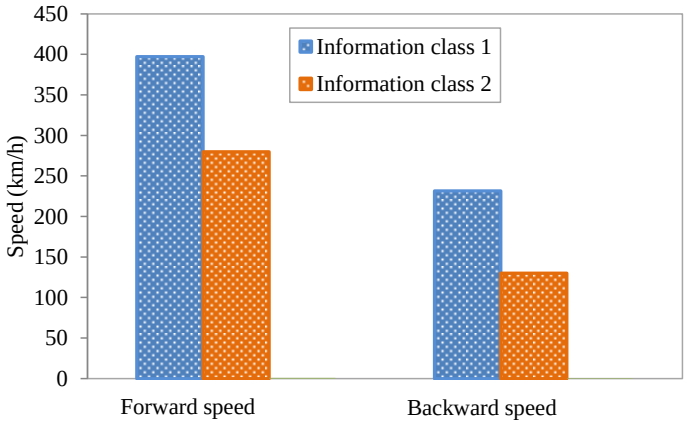


Figure 3.12 Comparison of forward and backward propagation speeds of information classes 1 and 2

This section analyzes the control of multiclass information propagation to meet application needs of various classes simultaneously. Recall that the queuing system for each information class is independent. Thereby, the number of communication servers and the mean communication service rate can be controlled for each information class to achieve desired propagation performance related to information spread (related to density of informed vehicles), time delay bounds (related to IFPW speed) and spatial coverage (related to existence of IFPW). Suppose traffic flow density is 50 veh/km and information from three information classes (labeled information class 1, 2 and 3) is propagated over the traffic stream. Let information class 1 contain "urgent" information (e.g., traffic accident blocks the freeway link fully). It is desirable for this information to reach all upstream and downstream vehicles with low latency. Information class 2 is constituted by less urgent information; for example, routing information. It is delay-tolerant and is expected to reach a lower proportion of equipped vehicles compared to information class 1 to avoid congestion in other routes. Information class 3 contains information with limited impact area, which needs to be propagated locally, for example, information of sudden braking of a vehicle, lane merge information, etc. Suppose the mean arrival rate of information packets of information class 1, 2 and 3 are 0.3, 0.8 and 1.2 packets/second, respectively. According to Table 3.3, the total number of communication servers that can be

assigned under traffic flow density 50 veh/km is 25. Let the number of communication servers assigned to information classes 1, 2 and 3 be 12, 8 and 5, respectively. The mean communication service rates are set correspondingly as $u_1 = 0.05$ packets/second, $u_2 = 0.2$ packets/second, and $u_3 = 0.4$ packets/second. Note that $\gamma_1 > 1, \gamma_2 > 1$, and $\gamma_3 < 1$. According to Theorem 3.4, the IFPW exists for propagation of information packets of information classes 1 and 2 while it does not exist for information packets of class 3.

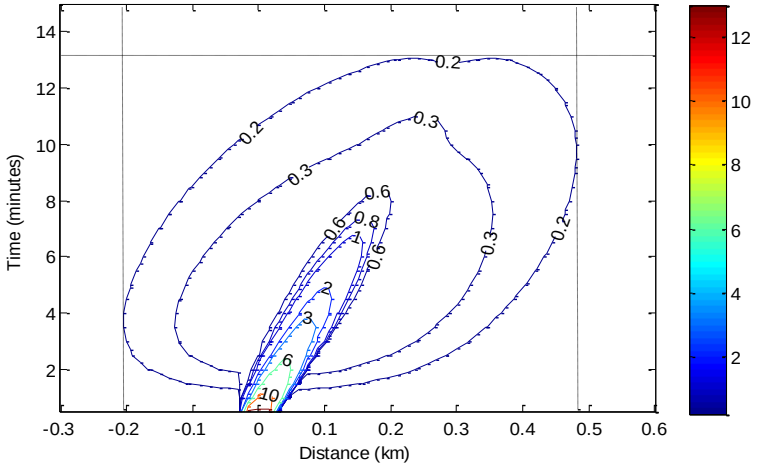


Figure 3.13 Contour of density of information-relaying vehicles of information class 3

Figure 3.12 compares the forward and backward IFPW speed of information classes 1 and 2. It shows that both forward and backward IFPW speeds of information class 1 are greater than those of information class 2. In addition, the proportion of vehicles (information spread) informed with the packets of information class 1 and information class 2 are 99.8% and 65.6%, respectively. Thereby, under the designed control strategy, packets from information class 1 can reach more number of vehicles with lower time delay compared to packets from information class 2. Figure 3.13 shows the contour of the density of information-relaying vehicles. It indicates that vehicles relaying the specific information packet of information class 3 decreases dramatically with space and time. The specific information packet is almost excluded by all vehicles beyond the locations 480 meters downstream and 200 meters upstream of its point of origin (i.e., location 0). Thereby, the information packets of class 3 are only propagated to a small area.

3.5.3.2 Control of multiclass information flow propagation under heterogeneous conditions

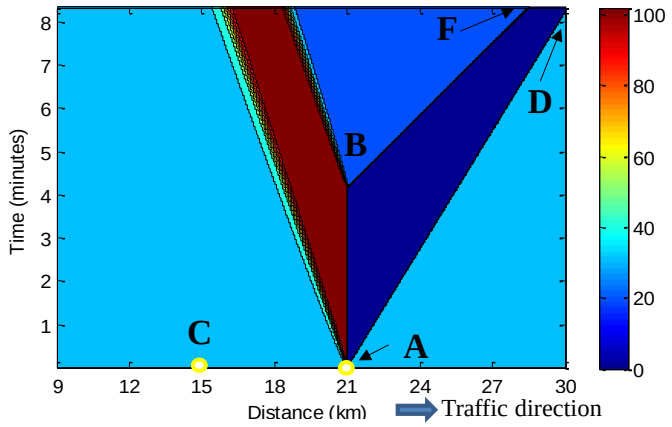


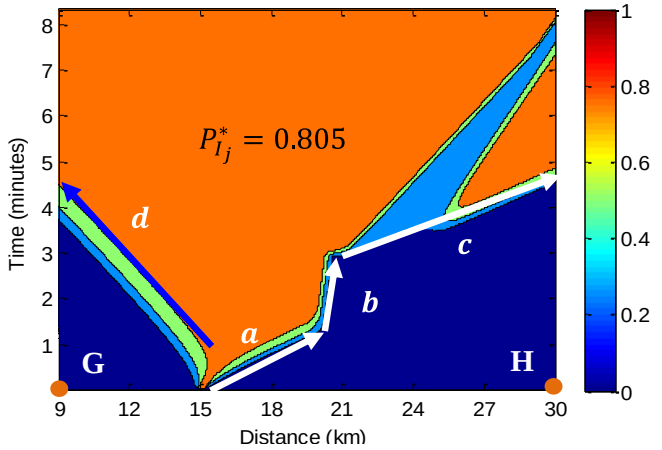
Figure 3.14 Contour of traffic density

This section address the control of information flow propagation under heterogeneous conditions. Similar to Kim et al. (2017), consider that a traffic accident happens at time 0 on a unidirectional highway with a traffic flow density of 50 veh./km. As illustrated by Figure 3.14, the incident occurs at location A. It reduces the link capacity by one third for 4 minutes before it is cleared. The congested traffic and the free flow traffic departing from the incident occurrence location are separated by Line AB. The occurrence and clearance of the incident induce two forward propagating traffic waves denoted by lines AD and BF, respectively. After the incident occurs, vehicles are jammed at the incident location, leading to a traffic wave propagating backward.

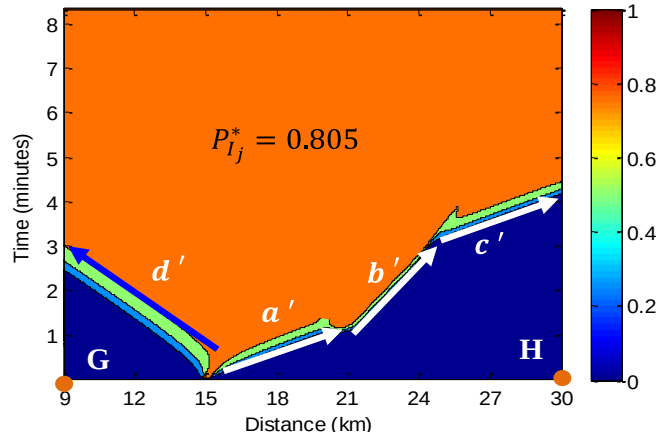
Suppose information packets of three different information classes are generated at location C, and their arrival rates are identical. We label them information classes 1, 2, and 3. Information classes 1 and 2 contain routing information and are expected to reach the same number of equipped vehicles. However, information packet of class 2 is expected to be propagated faster than information packet of class 1 as it contains information related to the traffic accident, which requires more imminent response from the vehicles. Information class 3 contains information related to the level of traffic congestion induced by the traffic accident. Hence, information packets of class 3 are expected to be received by all vehicles in the impacted area.

To achieve these objectives, let the number of communication servers assigned for information classes 1, 2 and 3 be 5, 10 and 10, respectively. The mean communication service rates for the three information classes are $u_1 = 0.15$ packets/second, $u_2 = 0.15$ packets/second,

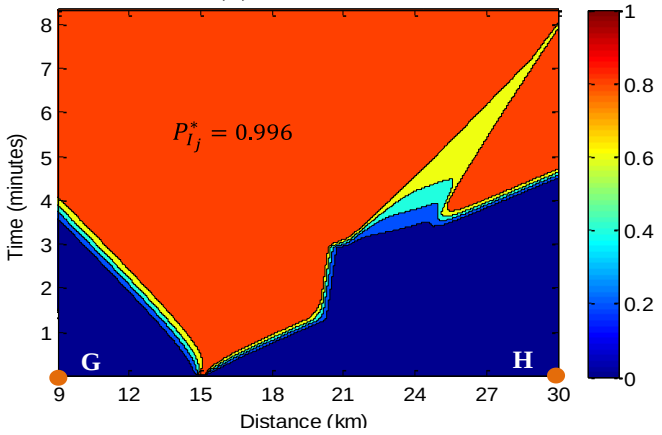
and $u_3 = 0.06$ packets/second. The numerical solution method will be used to calculate the information propagation speed and the proportion of informed vehicles (information spread) for the three information classes.



(a) information class 1



(b) information class 2



(c) information class 3

Figure 3.15 Contour of proportion of information-excluded vehicles of information packets of classes 1, 2 and 3

Figures 3.15(a) and 3.15(b) compare the backward and forward IFPW speeds of information classes 1 and 2, respectively. They illustrate that information packets in both classes are propagated very fast in the uncongested area which is not impacted by the traffic accident (see stages a and a' in Figures 3.15(a) and 3.15(b), respectively). The propagation speed decreases significantly when the information packets arrive at the congested area induced by the traffic accident (stages b and b' in Figures 3.15(a) and 3.15(b), respectively). This is because higher traffic density of vehicles can increase communication interference, causing significant communication failures. The IFPW speed is recovered to the original value when the congested area is passed to catch up with the normal traffic (stages c and c' in Figures 3.15(a) and 3.15(b), respectively). It is important to note that information packets in information class 2 are propagated faster than those in information class 1. For example, the information packets of information class 1 take about 4 minutes and 5 minutes to reach the points G and H located at 9 km and 30 km, respectively. In comparison, it only takes 2 minutes 40 seconds and 4 minutes for information packets of information class 2 to reach the two locations, respectively. These results indicate that under heterogeneous conditions, controlling the number of communication servers assigned to each information class can significantly impact the time delay of the information packets to reach the targeted locations. Figures 3.15(a) and 3.15(b) also reveal that the information spread (i.e., $P_{I_j}^*$) is the same in space and time. This implies that the number of communication servers only impacts the propagation speed, but not the asymptotic proportion of informed vehicles.

Figure 3.15(c) shows that $P_{I_3^*} = 0.996$, implying that almost all equipped vehicles can receive information packets of class 3. Note that the number of assigned communication servers for information classes 2 and 3 are identical. Figures 3.15(b) and Figure 3.15(c) indicate that smaller mean communication service rate will enable more number of vehicles to receive the specific information packet in corresponding information class. However, it may reduce the forward and backward propagation speeds. Figure 3.15(c) shows that it takes longer time for information packets in information class 3 to be delivered at locations G and H, compared to those of information class 2. This is because reducing the mean communication service rate will increase the mean waiting time of information packets in the queue. Thereby, under the designed scenarios, the information propagation speed is reduced due to increased mean waiting time. The proposed method in this study can aid transportation operators to determine the mean

communication service rate and the number of communication servers assigned to each information class to control the information propagation performance under both homogeneous and heterogeneous traffic conditions.

3.6 Conclusions

The traffic information propagated in a V2V-based traffic system can be grouped into different classes based on application needs related to information spread, time delay bounds, and spatial coverage. To meet these needs of multiclass information under different traffic flow and communication environments, this study proposes a queuing strategy for equipped vehicles to propagate the received information packets. The queuing strategy enables control for multiclass information propagation by leveraging two control parameters, the number of communication servers and the mean communication service rate. The spatiotemporal propagation of information in different information classes under the designed queuing strategy is characterized by a two-layer analytical model. The upper layer is an IDE system derived to model the information dissemination under the designed queuing strategy, and a LWR model is used in the lower layer to describe the traffic flow dynamics. An analytical solution of asymptotic density of informed vehicles is developed under homogeneous traffic conditions. It helps to analyze the relationship between the density of informed vehicles and the two control parameters in the queuing strategy. In addition, the necessary conditions for existence of IFPW are derived. It describes the conditions under which the specific information packets will be propagated only locally. A numerical solution is proposed to solve the two-layer model to estimate the IFPW speed, which helps to estimate the time delay for an information packet to reach the target location.

Numerical experiments using the proposed model suggest that the mean communication service rate significantly impacts the asymptotic density of informed vehicles. Also, all else being equal, an increase the number of communication servers assigned to an information class will increase the IFPW speed of the information packets in this information class. In addition, information will be propagated only locally under a high communication service rate because each information packet has little transmission duration. These findings provide valuable insights for controlling the propagation of multiclass information to achieve desired operational performance in a V2V-based traffic system. Hence, this study can be leveraged to develop a new

generation of information dissemination strategies focused on enabling specific V2V-based applications.

This study can be extended in a few directions. First, analytical solutions of the IFPW speed can be derived to provide insights on the relationship between the two control parameters in the queuing strategy and the resulting IFPW speed of information packets in each information class. Second, this study only considers control of information flow propagation in a corridor. The performance of the proposed method on control of network-level information flow propagation needs to be investigated. Third, this study assigns the received information packets into different queues according to the information classes they belong to. It assumes that the number of available communication servers is larger than the number of information classes. This assumption may not hold in scenarios of high traffic density, where the maximum number of information packets (i.e., N_{max}) an equipped vehicle can be transmit in one-hop propagation is small due to information congestion (Wang et al., 2018). To address this, other queuing strategies such as preemptive priority and non-preemptive priority queuing systems will be developed to control the propagation of information of different information classes.

CHAPTER 4. A REAL-TIME DEPLOYABLE MODEL PREDICTIVE CONTROL-BASED COOPERATIVE PLATOONING APPROACH FOR CONNECTED AND AUTONOMOUS VEHICLES

4.1 Introduction

Connected and autonomous vehicle (CAV) technologies provide disruptive and transformational opportunities for innovations toward intelligent transportation systems. Unlike human-driven vehicles, CAVs have shorter reaction times, better knowledge of ambient traffic (in terms of speed, position, acceleration, etc.), and faster information processing speeds. These characteristics enable CAVs to form platoons to drive cooperatively on the road, in which a vehicle maintains a small and nearly constant headway with its preceding vehicle. Past studies suggest that vehicle platooning of CAVs can benefit transportation systems in many ways (Jia et al., 2015). It can increase road capacity, reduce energy consumption and tailpipe emissions, and facilitate vehicle-to-vehicle based applications (involving data sharing and dissemination) due to the relatively fixed positions of vehicles within a platoon.

In the literature, many adaptive cruise control (ACC) models (e.g., VanderWerf et al., 2001; Hasebe et al., 2003; Kesting et al., 2008; Darbha and Rajagopal, 1999) and cooperative ACC (CACC) models (e.g., Ploeg et al., 2014; Zheng et al., 2014; Li et al., 2011; Jia and Ngoduy, 2016; Jin and Orosz., 2014) have been proposed to control longitudinal car-following behavior of vehicles to enable efficient vehicle platooning. The ACC uses information from a preceding vehicle while CACC uses information from a group of preceding vehicles by leveraging connectivity technologies to make decisions to optimize each individual vehicle's performance. It is important to note here that CACC only follows the cooperative sensing concept (Wang et al, 2014b), and the behaviors of vehicles controlled by both ACC and CACC are non-cooperative. That is, the control is not based on viewing a group of vehicles as an integrated system, which can deteriorate system (platoon) performance in terms of safety, mobility, energy consumption, etc. To bridge this gap, recently, model predictive control (MPC)-based cooperative control strategies have been proposed to coordinate the behaviors (accelerations or decelerations) of all of the following vehicles in a CAV platoon (e.g., Wang et al., 2014; Gong and Du, 2018; Zhou et al., 2017). The MPC strategy incorporates an optimal control problem to optimize the control decisions of the following vehicles in the platoon for some future period (labeled prediction

horizon) to maximize the platoon performance based on the vehicles' state information at the current time. Thereby, MPC-based cooperative control strategies enable CAVs in a platoon to collaborate with each other and maneuver under a common goal. Prior studies suggest that these strategies can lead to smoother deceleration behavior and more responsive and agile acceleration behavior compared to non-cooperative controllers (Wang et al., 2014b, Gong and Du, 2018).

While the aforementioned MPC-based cooperative control strategies can coordinate the car-following behaviors of CAVs in a platoon effectively, their real-time deployability requires that at each sampling time instant, the group of CAVs solve the embedded optimal control problem instantaneously (i.e., in much less than 0.1 seconds) to obtain the vehicles' control decisions based on their detected states (e.g., speed and positions) at that instant. These decisions then needs to be executed to control the CAV platoon at the sampling time instant with no delay. However, this requirement cannot be satisfied in practice due to the computational time required by the CAVs to solve the optimal control problem. As pointed by Zhou et al. (2017), the computational time for solving the optimal control problem increases monotonically with the number of vehicles in the platoon and the prediction horizon. It can become intractable in real traffic systems due to the expansion of the dimensionality of state and control input spaces (Wang et al., 2016). Thereby, based on platoon size and prediction horizon length, the computational time of the optimal control problem can cause significant delay (labeled control delay) in the execution of the optimal control decisions for the CAV platoon. As the CAVs' states change dynamically, the control delay can significantly deteriorate performance and even induce vehicle collisions. This precludes these MPC-based cooperative control strategies for a CAV platoon from being applied in real-time.

Some recent studies have sought to reduce the control delay induced by the computational time for solving the optimal control problem embedded in MPC-based cooperative control strategies. Wang et al. (2016) propose a decentralized MPC strategy which considers cooperation among only two vehicles in a decoupled platoon system. It can reduce the computational time substantially by optimizing only two vehicles' control decisions simultaneously. However, the performance of the CAV platoon cannot be enhanced to the fullest under this strategy as only two vehicles' behaviors are coordinated at the same time under a common objective. Further, the computational time for solving the optimal control problem can increase with the prediction horizon for even the decoupled platoon system. Gong and Du (2018) propose a distributed

solution algorithm to reduce computational time by distributing the computational tasks among all CAVs in the platoon. However, the computational time of this algorithm can increase dramatically with platoon size and prediction horizon. Hence, these methods (e.g., Wang et al., 2016; Gong and Du, 2018) alleviate the issue of control delay of MPC-based cooperative control strategies to only a certain extent, but are still limited by platoon size and/or prediction horizon.

This study develops two real-time deployable MPC-based approaches that address the issue of the control delay at a fundamental level. To do so, first, an idealized MPC-based cooperative control strategy is proposed by modifying the strategies proposed by Wang et al. (2014b) and Zhou et al. (2017). It can coordinate the behavior of all of the following CAVs in the platoon to maneuver them efficiently and safely on the idealized assumption that the embedded optimal control problem can be solved instantaneously. To relax this assumption, two deployable approaches, labeled the deployable MPC (DMPC) approach and the DMPC with first-order approximation (DMPC-FOA) approach, are proposed to address the issue of computational delay associated with solving the optimal control problem in the idealized MPC-based strategy. It should be noted that to enable efficient coordination of the car-following behaviors of all CAVs in the platoon, such approaches need to accurately characterize the optimal control decisions of the idealized MPC-based strategy.

The DMPC approach reserves sufficient time before each sampling time instant to solve the optimal control problem so that the optimal control decisions can be obtained in advance to be executed at the corresponding sampling time instant with no delay. However, as the leading vehicle of a platoon needs to respond to the dynamics of the vehicles downstream of it, its behavior cannot be controlled and coordinated with those of the following vehicles in the platoon. Thereby, its position and speed at each sampling time instant need to be predicted ahead of that time, which is determined by the time reserved for computing. Hence, the optimal control decisions of the DMPC approach can deviate from that of the idealized MPC strategy due to error in predicting the leading vehicle's position and speed in advance. To address this problem, the DMPC-FOA approach is proposed to more accurately characterize the optimal control decisions of the idealized MPC strategy. Before each sampling instant, the DMPC-FOA approach reserves sufficient time to determine not only the optimal control decisions using the leading vehicle's predicted position and speed at the sampling time instant, but also the derivatives of the estimated optimal control decisions with respect to the leading vehicle's

position and speed. Thereby, at the sampling time instant when the leading vehicle's actual position and speed are detected, the first-order Taylor approximation method can be applied to correct the estimated optimal control decisions for the following vehicles. Numerical experiments illustrate that the DMPC-FOA approach can address the issue of control delay while accurately estimating the optimal control decisions of the idealized MPC strategy.

The contributions of this study are fourfold. First, an idealized MPC strategy is proposed to coordinate the behaviors of the following vehicles in the platoon by modifying the control strategies proposed by Wang et al. (2014b) and Zhou et al. (2017). Further, a solution algorithm is proposed to solve the optimal control problem with both control constraints and pure state constraints in the idealized MPC strategy. A two-point boundary value problem is derived based on the necessary conditions for optimality to obtain the optimal control decisions to coordinate the behaviors of all vehicles in the platoon to maximize the platoon performance. Second, the study develops the DMPC-FOA approach that simultaneously addresses the control delay issue while accurately characterizing the optimal control decisions of the idealized MPC strategy. Thereby, it can be applied in real-time to efficiently coordinate the car-following behaviors of all CAVs in a platoon. Third, the method for sensitive analysis of the optimal control problem is analytically formulated. It can quantitatively measure the impact of parametric perturbations (e.g., perturbations of initial state of the leading vehicle) on the optimal control decisions and the platoon performance. Fourth, an analytical method is provided for stability analysis of the idealized MPC strategy. It helps to identify the inputs of the parameters in the idealized MPC strategy to better dampen the oscillations in the platoon.

The remainder of this paper is organized as follows. The next section provides the analytical formulation of the idealized MPC cooperative control strategy for a CAV platoon and discusses the framework for the DMPC and DMPC-FOA approaches. Section 3 introduces the solution algorithm to solve the optimal control problem in the idealized MPC strategy. The method for the sensitivity analysis of the optimal control problem is presented in Section 4. Section 5 discusses the conditions for the stability of the idealized MPC strategy without inequality constraints. Section 6 discusses results of numerical experiments to compare the control performance of the idealized MPC strategy and the DMPC and DMPC-FOA approaches. The last section provides some concluding comments.

4.2 MPC approaches for longitudinal control of CAV platoon

4.2.1 An idealized MPC cooperative control strategy for a CAV platoon

This section presents an idealized MPC strategy to control the CAVs in a platoon cooperatively by modifying the control strategies developed by Wang et al. (2014b) and Zhou et al. (2017). It seeks to coordinate the behavior of all following vehicles to: (1) maintain a desired safe spacing (labeled equilibrium spacing) between two consecutive vehicles in a platoon, and reduce traffic flow oscillations in terms of spacing and speed changes, and (2) maximize the comfort of travelers in these vehicles by minimizing deceleration and acceleration. The details of the idealized MPC strategy are as follows.

Consider a stream of CAVs in a single highway lane as shown in Figure 4.1. Let $0,1,2\cdots,n$ represent the CAVs in the platoon sequentially with 0 being the leading CAV and n being the tail CAV. The following assumptions will be used to design the longitudinal control of the CAV platoon:

1. All vehicles in the platoon are CAVs.
2. Two-way V2V communications exist between the leading vehicle and each of the following vehicles in the platoon.
3. All CAVs can sense their kinematic states (speed, position, etc.) accurately and can send that information to the leading vehicle of their platoon instantaneously.
4. The leading CAV computes and sends the optimal control decisions (i.e., accelerations and decelerations) to all of the following CAVs which implement these decisions.
5. The actuator delay is negligible; that is, vehicles can implement the control instantly.
6. The pavement of the highway lane is in good condition and longitudinal slope is negligible.

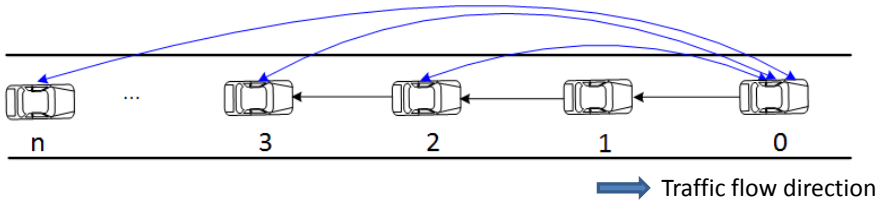


Figure 4.1 A CAV platoon stream.

In this study, we treat a platoon of CAVs as an integrated system, in which vehicles within the platoon are controlled in a coordinated manner. Define the state of a follower vehicle i as $\left(s_i(t)-s_i^*(t), v_i(t)-v_{i-1}(t)\right)$, where $s_i(t)$ is the spacing of vehicle i with its predecessor vehicle at time t, $v_i(t)$ is the speed of vehicle i at time t, and $s_i^*(t)$ is the equilibrium spacing at

time t. This study uses the constant time headway policy to determine the equilibrium spacing. Thereby, $s_i^*(t) = r_i^* \cdot v_i(t) + s_f$, where r_i^* is the constant time headway for vehicle i and s_f is the safe distance to the predecessor vehicle. For simplicity, the constant time headway for each follower vehicle in the platoon is assumed to be the same, i.e., $r_i^* = r^*, \forall i = 1,2,\cdots n$. Let $x_i(t) = s_i(t) - s_i^*(t), \forall i$ be the position error between the desired spacing and actual spacing of vehicle i from its predecessor vehicle at time t. Denote $y_i(t)$ as the speed difference of vehicle i from its predecessor vehicle at time t, i.e., $y_i(t) = v_i(t) - v_{i-1}(t)$. Denote $d_i(t)$ as the longitudinal position of CAV i in the platoon at time t. Then,

$$x_i(t) = d_{i-1}(t) - d_i(t) - r^* \cdot v_i(t) - s_f \tag{4.1}$$

and

$$\dot{x}_i(t) = v_{i-1}(t) - v_i(t) - r^* \cdot u_i(t) \tag{4.2a}$$

$$\dot{y}_i(t) = u_i(t) - u_{i-1}(t) \tag{4.2b}$$

where $\dot{x}_i(t)$ is the first-order derivative of position error of vehicle i from its predecessor vehicle with respect to time t. $\dot{y}_i(t)$ is the first-order derivative of speed difference of vehicle i from its predecessor vehicle with respect to time t. $u_i(t)$ is the acceleration of CAV i at time t.

The spacing of vehicle $i, \forall i = 1,2,\cdots,n$ from its predecessor vehicle can then be expressed as:

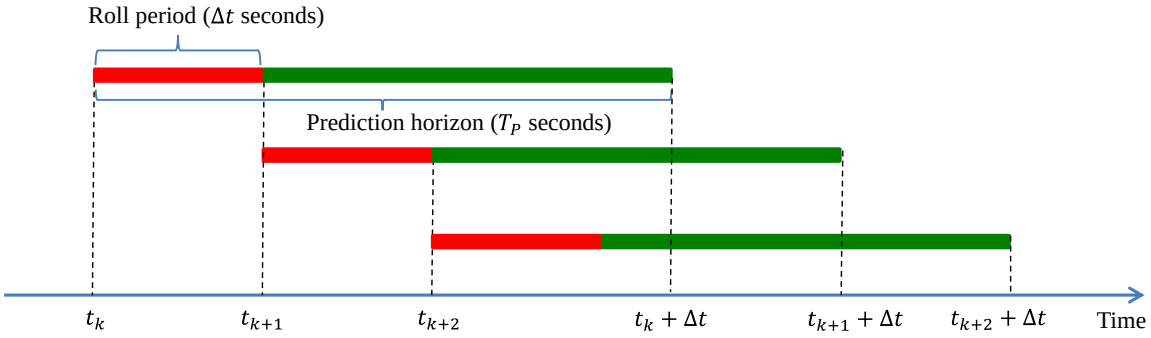
$$s_i(t) = x_i(t) + r^* \cdot \left(v_0(t) + \sum_{j=1}^{i} y_j(t) \right) + s_f, \forall i = 1,2,\cdots,n \tag{4.3}$$

Denote $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]^T$, $\mathbf{y}(t) = [y_1(t), y_2(t), \cdots, y_n(t)]^T$, and $\boldsymbol{u}(t) = [u_1(t), u_2(t), \cdots, u_n(t)]^T$. $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are vectors of state variables. Assume that the leading vehicle 0 travels at a constant speed. Then, the dynamics of the states (i.e., $\mathbf{x}$ and $\mathbf{y}$) are as follows:

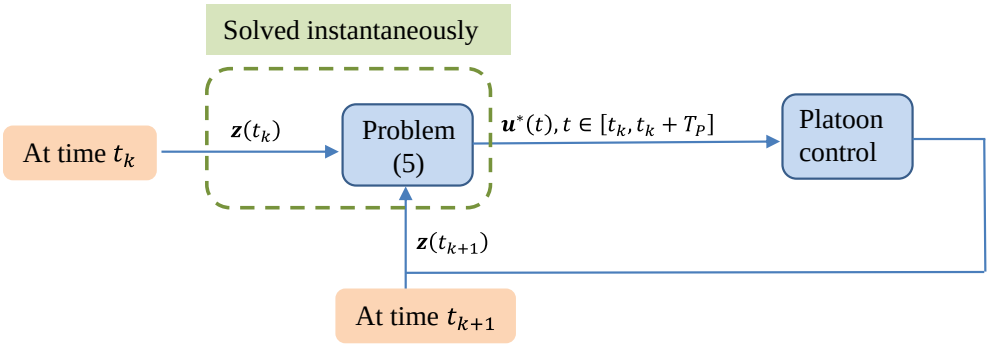
$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{y}}(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{0}_n & -\boldsymbol{E}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{y}(t) \end{bmatrix} + \underbrace{\begin{bmatrix} \boldsymbol{\Lambda} \\ \boldsymbol{S} \end{bmatrix}}_{\boldsymbol{B}} \cdot \boldsymbol{u}(t) \tag{4.4}$$

where $\dot{\mathbf{x}}(t)$ and $\dot{\mathbf{y}}(t)$ are first-order derivatives of $\mathbf{x}(t)$ and $\mathbf{y}(t)$ with respect to time t, $\mathbf{0}_n$ is the n-dimensional zero square matrix, and $\boldsymbol{\Lambda} = -r^* \cdot \boldsymbol{E}_n$, $\boldsymbol{E}_n$ is the n-dimensional identity matrix. Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined in Eq. (4.4). The matrix $\boldsymbol{S}$ is:

$$\boldsymbol{S} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}^{n\times n}$$



(a) Implementation framework



(b) Computational procedure

Figure 4.2 The idealized MPC strategy

Following the elucidation of the state variables, the next step in developing the idealized MPC strategy is the conceptual illustration of its implementation framework and computational procedure, as shown in Figures 4.2(a) and 4.2(b), respectively. In Figure 4.2(a), let $t_k (k = 1,2,3\cdots)$ be the sampling time instant at which new optimal control decisions should be executed to control vehicles in the platoon, T_P be the prediction horizon for which the optimal control decisions are determined, and Δt ($\Delta t \leq T_P$) be the roll period for which these decisions are implemented. Such a rolling horizon framework enables the practical implementation of the control strategy by trading off (solution) computational time with solution accuracy by limiting the prediction horizon size while being responsive to unfolding traffic conditions. Thereby, for a

sampling time instant t_k, the new optimal control decisions are calculated for the prediction horizon $[t_k, t_k + T_P]$, but only implemented for the roll period $[t_k, t_k + \Delta t]$ by the following vehicles in the platoon to control their behavior. Then, at the next sampling time instant t_{k+1} (where $t_{k+1} = t_k + \Delta t$), the procedure is repeated to determine and implement the optimal control decisions for all following CAVs in the platoon for roll period $[t_{k+1}, t_{k+1} + \Delta t]$. This procedure is repeated until the platoon dissipates.

Next, the idealized MPC strategy to determine the optimal control decisions and its computational procedure are exposited. Let $\mathbf{z}(t) = [\mathbf{x}(t)^T, \quad \mathbf{y}(t)^T]^T$. Following Wang et al. (2014b) and Zhou et al. (2017), at each sampling time instant t_k, $\forall k = 0,1,2\cdots$, the optimal control decisions of all of the following vehicles in the platoon can be obtained by solving the following optimal control problem:

$$\min_{\boldsymbol{u}} \int_0^{T_P} \frac{1}{2} e^{-\beta t} L(\boldsymbol{z}(t), \boldsymbol{u}(t))\, dt + \frac{1}{2} e^{-\beta T_P} \phi(\mathbf{z}(T_P)) \tag{4.5a}$$

$$\dot{\mathbf{z}}(t) = \boldsymbol{A} \cdot \mathbf{z}(t) + \boldsymbol{B} \cdot \boldsymbol{u}(t) \tag{4.5b}$$

$$s_i(t) = x_i(t) + r^* \cdot \left(v_0(0) + \sum\nolimits_{j=1}^{i} y_j(t) \right) + s_f \geq s_{min}; i = 1,2,\cdots,n \tag{4.5c}$$

$$u_{min} \leq u_i \leq u_{max}; i = 1,2,\cdots,n \tag{4.5d}$$

$$\mathbf{z}(0) = [\mathbf{x}_0^T \quad \mathbf{y}_0^T]^T \tag{4.5e}$$

where

$$L(\boldsymbol{z}(t), \boldsymbol{u}(t)) = \mathbf{z}(t)^T \begin{bmatrix} \boldsymbol{R}_1 & \\ & \boldsymbol{R}_2 \end{bmatrix} \mathbf{z}(t) + \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t) \tag{4.5f}$$

$$\phi(\mathbf{z}(T_P)) = \mathbf{z}(T_P)^T \begin{bmatrix} \boldsymbol{R}_4 & \\ & \boldsymbol{R}_5 \end{bmatrix} \mathbf{z}(T_P) \tag{4.5g}$$

In problem (4.5), for expository convenience, we consider a generic prediction horizon and ignore the sampling time instant t_k. So, $t \in [0, T_P]$ without loss of generality in (5). Here, $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_3$ $\boldsymbol{R}_4$, and $\boldsymbol{R}_5$ are weight matrices; $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_4$, and $\boldsymbol{R}_5$ are symmetric positive definite matrices; and $\boldsymbol{R}_3$ is a positive definite diagonal matrix (Zhou et al., 2017). $L(\boldsymbol{z}(t), \boldsymbol{a}(t))$ is the running cost which is the cost incurred during an infinitesimal period (Wang et al. 2014b). It consists of two terms. The first term $\mathbf{z}(t)^T \begin{bmatrix} \boldsymbol{R}_1 & \\ & \boldsymbol{R}_2 \end{bmatrix} \mathbf{z}(t)$ seeks to minimize the position errors and the relative speed of all adjacent vehicle pairs. The second component (i.e., $\boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t)$) is to maximize comfort by reducing hard braking and acceleration. $e^{-\beta t}$ is a term to weight the

running cost at different times and β is the discount coefficient. This term provides higher weight for the running cost for the near-term future than for the longer-term future as the uncertainty in running cost increases with time (Wang et al., 2014b). $\phi(\mathbf{z}(T_P))$ is the terminal cost which is used to penalize the value of objective function if the values of the state variables at the end of the prediction horizon deviate from the equilibrium point (i.e., 0). Eq. (4.5b) describes the dynamics of the state variables (i.e., position errors and relative speeds of all adjacent vehicle pairs in the platoon). Eq. (4.5c) is a safety constraint to ensure that the spacing between two consecutive CAVs in the platoon is always larger than a positive lower bound s_{min}, $s_{min} > 0$. Note $v_0(t) \equiv v_0(0)$ under the assumption $u_0(t) \equiv 0$ for $t \in [0, T_P]$. Eq. (4.5d) specifies the upper bound (u_{max}) and lower bound (u_{min}) of the acceleration. Eq. (4.5e) specifies the initial inputs for the state variables. Hence, for example, for any sampling time instant t_k, $\mathbf{x}_k = [x_1(t_k), x_2(t_k), \cdots, x_n(t_k)]$ and $\mathbf{y}_k = [y_1(t_k), y_2(t_k), \cdots, y_n(t_k)]$ are values of $\mathbf{x}_0^T$ and $\mathbf{y}_0^T$, respectively.

There are primarily two differences between optimal control problem (4.5) and the ones developed by Wang et al. (2014b) and Zhou et al. (2017). First, a term $e^{-\beta t}$ is added to the objective function to weight the running costs at different times. Second, a terminal cost $\phi(\mathbf{z}(T_P))$ is added to penalize the objective function if the state variables deviate from the equilibrium point 0 at the end of the prediction horizon. These two terms will be useful to analyze the stability of the idealized MPC strategy. In addition, for convenience of stability analysis, the weight matrices $\boldsymbol{R}_i (i = 1,2,4,5)$ are assumed to have the following forms:

$$\boldsymbol{R}_1 = \boldsymbol{\Lambda}^T \boldsymbol{D}_a \boldsymbol{\Lambda}, \boldsymbol{R}_2 = \boldsymbol{\Lambda}^T \boldsymbol{D}_b \boldsymbol{\Lambda}, \boldsymbol{R}_4 = \boldsymbol{\Lambda}^T \boldsymbol{D}_c \boldsymbol{\Lambda}, \text{ and } \boldsymbol{R}_5 = \boldsymbol{\Lambda}^T \boldsymbol{D}_e \boldsymbol{\Lambda} \tag{4.6}$$

where $\boldsymbol{\Lambda}$ is an orthogonal matrix, $\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T = \boldsymbol{E}_n$, and $\boldsymbol{D}_a$, $\boldsymbol{D}_b$, $\boldsymbol{D}_c$ and $\boldsymbol{D}_e$ are positive definite diagonal matrices. The inputs of these weight matrices will be determined by the stability analysis in Section 5. Eq. (4.6) shows that if $\boldsymbol{\Lambda} = \boldsymbol{E}_n$, then $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_4$, and $\boldsymbol{R}_5$ are positive definite diagonal matrices.

Let $\mathbf{z}(t_k)$ be the actual values of the state variables at the sampling time instant $t_k, (k = 1,2, \cdots)$, $\mathbf{z}(t_k) = [\mathbf{x}_k^T \quad \mathbf{y}_k^T]^T$. The computational procedure of the idealized MPC strategy is summarized in Figure 4.2(b). At each sampling time instant $t_k (k = 1,2, \cdots)$, the leading vehicle obtains the value of $\mathbf{z}(t_k)$ through V2V communications. It solves the optimal control problem (4.5) to determine the optimal control decisions (i.e., $\boldsymbol{u}^*(t)$) during the prediction horizon $[t_k, t_k + T_P]$ by inputting the value of $\mathbf{z}(t_k)$ into Eq. (4.5e). The optimal control decisions are

sent by the leading vehicle to the following vehicles to control their behaviors only for the roll period $[t_k, t_k + \Delta t]$, (i.e., $[t_k, t_{k+1}]$). Then, at the sampling time instant t_{k+1}, the optimal control problem (4.5) is solved again to obtain the optimal control decisions $\boldsymbol{u}^*(t)$ for the prediction horizon $[t_{k+1}, t_{k+1} + T_P]$, and is implemented to control the CAV platoon for the roll period $[t_{k+1}, t_{k+1} + \Delta t]$. These steps are repeated at each sampling time instant.

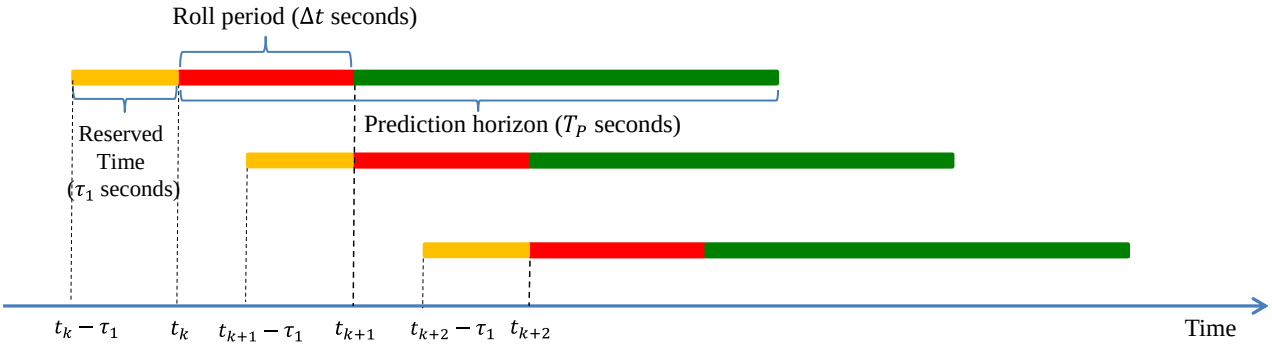
As can be noted, the idealized MPC strategy computes the optimal control decisions by solving the optimal control problem (4.5) at each sampling time instant and implements it to control the CAVs for the roll period starting at that instant. To achieve this, it is assumed that the leading vehicle can solve the optimal control problem (4.5) of the idealized MPC strategy instantaneously at each sampling time instant t_k. However, in practice, the computational time for solving the optimal control problem (4.5) increases with platoon size and prediction horizon size. It can cause significant delays in executing the control decisions, which can deteriorate the performance and even lead to vehicle collisions. Thereby, while the idealized MPC strategy can coordinate the behavior of the following vehicles in the platoon to maneuver them efficiently and safely, it cannot be deployed to control the CAV platoon in real-time.

4.2.2 DMPC approach framework

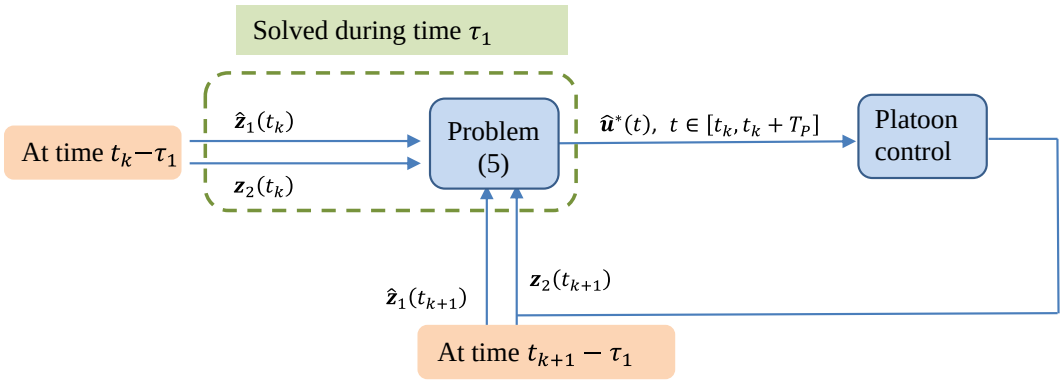
The leading vehicle of a CAV platoon needs to respond to the dynamics of the vehicles downstream of it. Thereby, its behavior is not known in advance. However, the behavior of all following vehicles in the platoon for each roll period can be estimated at the corresponding sampling time instant through the known optimal control decisions of the previous roll period (i.e., $\boldsymbol{u}^*(t), t \in [t_{k-1}, t_{k-1} + \Delta t]$). To account for this difference, we divide $\mathbf{z}(t)$ into two parts, $\boldsymbol{z}_1(t)$ and $\boldsymbol{z}_2(t)$. We denote the vector of position error and speed difference of vehicle 1 from that of the leading vehicle 0 as $\boldsymbol{z}_1(t) = [x_1(t), y_1(t)]^T$, and the vector of state variables for the other following vehicles as $\boldsymbol{z}_2(t) = [x_2(t), x_3(t), \cdots, x_n(t), y_2(t), y_3(t), \cdots, y_n(t)]^T$. At each sampling time instant t_k, the value of $\boldsymbol{z}_1(t_k)$ cannot be computed in advance due to the unknown position and speed of the leading vehicle at that instant. However, $\mathbf{z}_2(t_k)$ can be estimated in advance at a short time before the sampling time instant t_k.

We propose the DMPC approach to address the strong assumption of the idealized MPC strategy that the optimal control problem (4.5) can be solved instantaneously. The implementation framework for the DMPC approach is shown in Figure 4.3(a). Unlike the

idealized MPC strategy, the DMPC approach reserves a sufficient amount of time, labeled reserved time (denoted as τ_1), before each sampling time instant t_k ($k = 1,2,\cdots$) to solve the optimal control problem (4.5) so that the optimal control decisions are available at t_k for the corresponding roll period. It is important to note that the roll period Δt should be larger than τ_1 to enable the real-time implementation of the DMPC approach.



(a) Implementation framework



(b) Computational procedure.

Figure 4.3 The DMPC approach

The DMPC computational procedure is illustrated in Figure 4.3(b). The DMPC approach starts to solve the optimal control problem at time $t_k - \tau_1$ to predict the values of all state variables at time t_k (i.e., $\mathbf{z}(t_k)$). As stated in the assumptions, the leading vehicle can obtain the actual states of all following vehicles at time instant $t_k - \tau_1$ through V2V communications. Also, as discussed earlier in this section, it knows the control decisions of all following vehicles in the time period $[t_k - \tau_1, t_k]$ as they are determined at the beginning of the previous roll period. The DMPC approach leverages these two sets of inputs to predict $\mathbf{z}_2(t_k)$ with low error. This is because in the context of the CAV platooning application, τ_1 is much smaller than the roll period,

in the order of a fraction of a second. Hence, as the actual states are available close to t_k, and prior control decisions are known, we assume that the error in estimating $\mathbf{z}_2(t_k)$ is negligible.

As discussed earlier, the leading vehicle's behavior is not known in advance. Thereby, $\mathbf{z}_1(t_k)$ cannot be estimated with low error unlike $\mathbf{z}_2(t_k)$. Hence, the value of $\mathbf{z}_1(t_k)$ needs to be predicted at time instant $t_k - \tau_1$. To do so, the leading vehicle's behavior at t_k needs to be predicted at $t_k - \tau_1$. As τ_1 is much smaller than the roll period, we assume the acceleration of the leading vehicle 0 during the small time interval $[t_k - \tau_1, t_k]$ remains the same as at time instant $t_k - \tau_1$. Then,

$$\hat{v}_0(t_k) = v_0(t_k - \tau_1) + u_0(t_k - \tau_1) \cdot \tau_1 \tag{4.7a}$$

$$\hat{d}_0(t_k) = d_0(t_k - \tau_1) + v_0(t_k - \tau_1) \cdot \tau_1 + 0.5 \cdot u_0(t_k - \tau_1) \cdot (\tau_1)^2 \tag{4.7b}$$

where $\hat{v}_0(t_k)$ and $\hat{d}_0(t_k)$ are the predicted speed and predicted position of the leading vehicle at time instant t_k, respectively. Here, $v_0(t_k - \tau_1)$, $d_0(t_k - \tau_1)$ and $u_0(t_k - \tau_1)$ are the actual speed, position and acceleration of the leading vehicle at $t_k - \tau_1$, respectively, that are detected through onboard sensors. The position error and relative speed of vehicle 1 from that of the leading vehicle 0 at time instant t_k can then be predicted as:

$$\hat{x}_1(t_k) = \hat{d}_0(t_k) - d_1(t_k) - r^* \cdot v_1(t_k) - s_f \tag{4.8a}$$

$$\hat{y}_1(t_k) = v_1(t_k) - \hat{v}_0(t_k) \tag{4.8b}$$

where $\hat{x}_1(t_k)$ and $\hat{y}_1(t_k)$ are the predicted position error and speed difference of vehicle 1 with respect to the leading vehicle 0 at time t_k, respectively. Note that the speed and position of vehicle 1 at time instant $t_k - \tau_1$ are detected through the onboard sensors, and the corresponding control decision $u_1(t)$, $t \in [t_k - \tau_1, t_k]$ is known. Then, $d_1(t_k)$ and $v_1(t_k)$ can be computed as:

$$v_1(t_k) = v_1(t_k - \tau_1) + \int_{t_k - \tau_1}^{t_k} u_1(t)\, dt \tag{4.9a}$$

$$\begin{aligned} d_1(t_k) &= d_1(t_k - \tau_1) + \int_{t_k - \tau_1}^{t_k} v_1(t)\, dt \\ &= d_1(t_k - \tau_1) + \int_{t_k - \tau_1}^{t_k} \left[v_1(t_k - \tau_1) + \left(\int_{t_k - \tau_1}^{t} u_1(\varsigma) d\varsigma \right) \right] dt \end{aligned} \tag{4.9b}$$

Note that the predicted value $\hat{\mathbf{z}}_1(t_k)$ ($\hat{\mathbf{z}}_1(t_k) = [\hat{x}_1(t_k), \hat{y}_1(t_k)]$) is different from the actual value $\boldsymbol{z}_1(t_k)$ due to the error in predicting the leading vehicle's position and speed. Thereby, the estimated control decisions of the DMPC approach (i.e., $\hat{\boldsymbol{u}}(t)$) are different from the optimal control decisions computed by the idealized MPC strategy (i.e., $\boldsymbol{u}^*(t)$). In the numerical

experiments, we will show that the estimated control decisions of the DMPC approach will deviate significantly from those of the idealized MPC strategy when the error in predicting $\boldsymbol{z}_1(t_k)$ is large. This will deteriorate the efficiency of the CAV platoon and can cause vehicular collisions.

4.2.3 DMPC-FOA approach framework

The DMPC approach circumvents the strong assumption of the idealized MPC strategy at the cost that the estimated control decisions may deviate significantly from those of the idealized MPC strategy due to the error in predicting $\boldsymbol{z}_1(t_k)$. To address this problem, we propose the DMPC-FOA approach which simultaneously addresses the control delay issue of the idealized MPC strategy while more accurately characterizing the optimal control decisions.

Let τ_2 be the reserved time for computing the optimal control decisions for the DMPC-FOA approach. Also, let $\tilde{\boldsymbol{z}}_1(t_k) = [\tilde{x}_1(t_k) \quad \tilde{y}_1(t_k)]$ be the predicted value of $\boldsymbol{z}_1(t_k)$ for the DMPC-FOA approach at time instant $t_k - \tau_2$ by replacing τ_1 with τ_2 in Eqs. (4.7) and (4.8). Here, $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ are the predicted position error and speed difference of vehicle 1 with respect to the leading vehicle at time instant t_k, respectively. Similar to the DMPC approach, we assume the error in estimating $\mathbf{z}_2(t_k)$ is negligible as the actual states (i.e., $\mathbf{z}_2(t_k - \tau_2)$) are available close to t_k, and prior control decisions are known.

Denote $\boldsymbol{\gamma}(t)$ as the vector of costate variables associated with the state equations (5b). The costate variables indicate the change in the objective function value for a unit change in the corresponding state variable at the optimal state (Gaimon, 2002). The computational procedure for the DMPC-FOA approach is illustrated in Figure 4.4, where $\tilde{\boldsymbol{z}}^*(t)$ and $\tilde{\boldsymbol{\gamma}}^*(t), t \in [t_k, t_k + T_P]$ are the solutions for the state and costate variables obtained by solving optimal control problem (4.5) with initial inputs $[\tilde{\boldsymbol{z}}_1(t_k), \boldsymbol{z}_2(t_k)]$). The optimal control decisions for the idealized MPC strategy, $\varphi\big(\boldsymbol{z}^*(t), \boldsymbol{\gamma}^*(t)\big)$ (denoted as $\boldsymbol{u}^*(t)$), are analytically derived in Section 3 (see Eq. (4.23)) which discusses the solution algorithm. Then, $\boldsymbol{u}^*(t), t \in [t_k, t_k + T_P]$ can be approximated by $\varphi\big(\tilde{\boldsymbol{z}}^*(t), \tilde{\boldsymbol{\gamma}}^*(t)\big)$ (denoted as $\tilde{\boldsymbol{u}}^*(t)$). Note that the difference between $[\tilde{\boldsymbol{z}}^*(t), \tilde{\boldsymbol{\gamma}}^*(t)]$ and $[\boldsymbol{z}^*(t), \boldsymbol{\gamma}^*(t)]$ significantly impacts the accuracy of the estimated control decisions $\tilde{\boldsymbol{u}}^*(t)$. To reduce the difference between $\tilde{\boldsymbol{u}}^*(t)$ and $\boldsymbol{u}^*(t), t \in [t_k, t_k + T_P]$, sensitivity analysis of the optimal control problem (4.5) is performed to determine the derivatives of

$\partial\tilde{\boldsymbol{z}}^*(t)/\partial\,\tilde{\boldsymbol{z}}_1(t_k)$ (i.e., $\left[\frac{\partial\,\tilde{\boldsymbol{z}}^*(t)}{\partial\tilde{x}_1(t_k)},\frac{\partial\,\tilde{\boldsymbol{z}}^*(t)}{\partial\tilde{y}_1(t_k)}\right]$) and $\partial\tilde{\boldsymbol{\gamma}}^*(t)/\partial\,\tilde{\boldsymbol{z}}_1(t_k)$ (i.e., $\left[\frac{\partial\,\tilde{\boldsymbol{\gamma}}^*(t)}{\partial\tilde{x}_1(t_k)},\frac{\partial\,\tilde{\boldsymbol{\gamma}}^*(t)}{\partial\tilde{y}_1(t_k)}\right]$). These two terms can quantitatively measure the changes in the optimal solutions for $\tilde{\boldsymbol{z}}^*(t)$ and $\tilde{\boldsymbol{\gamma}}^*(t)$ for a unit increase in $\tilde{\boldsymbol{z}}_1(t_k)$. Thereby, at sampling time instant t_k when the actual value of $x_1(t_k)$ and $y_1(t_k)$ are detected through onboard sensors, the first-order Taylor's approximation is applied to better estimate the solutions of $\boldsymbol{z}^*(t)$ and $\boldsymbol{\gamma}^*(t)$, as follows:

$$\bar{\boldsymbol{z}}^*(t) = \tilde{\boldsymbol{z}}^*(t) + \frac{\partial\,\tilde{\boldsymbol{z}}^*(t)}{\partial\tilde{x}_1(t_k)}\left(x_1(t_k)-\tilde{x}_1(t_k)\right) + \frac{\partial\,\tilde{\boldsymbol{z}}^*(t)}{\partial\tilde{y}_1(t_k)}\left(y_1(t_k)-\tilde{y}_1(t_k)\right), t\in[t_k,t_k+T_P] \tag{4.10a}$$

$$\bar{\boldsymbol{\gamma}}^*(t) = \tilde{\boldsymbol{\gamma}}^*(t) + \frac{\partial\,\tilde{\boldsymbol{\gamma}}^*(t)}{\partial\tilde{x}_1(t_k)}\left(x_1(t_k)-\tilde{x}_1(t_k)\right) + \frac{\partial\,\tilde{\boldsymbol{\gamma}}^*(t)}{\partial\tilde{y}_1(t_k)}\left(y_1(t_k)-\tilde{y}_1(t_k)\right), t\in[t_k,t_k+T_P] \tag{4.10b}$$

where $\bar{\boldsymbol{z}}^*(t)$ and $\bar{\boldsymbol{\gamma}}^*(t)$ are the values of $\boldsymbol{z}^*(t)$ and $\boldsymbol{\gamma}^*(t)$ estimated by the DMPC-FOA approach, respectively.

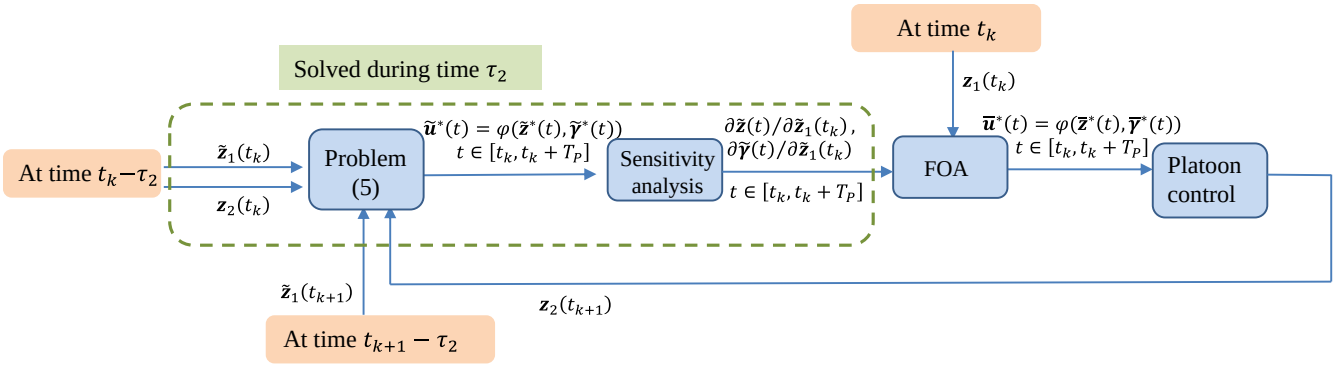


Figure 4.4 Computational procedure of the DMPC-FOA approach.

When compared to $[\,\tilde{\boldsymbol{z}}^*(t),\,\tilde{\boldsymbol{\gamma}}^*(t)]$, $[\,\bar{\boldsymbol{z}}^*(t),\,\bar{\boldsymbol{\gamma}}^*(t)]$ are closer to $[\,\boldsymbol{z}^*(t),\,\boldsymbol{\gamma}^*(t)\,]$. Thereby, in Figure 4.4, the estimated control decisions $\bar{\boldsymbol{u}}^*(t)=\varphi\left(\bar{\boldsymbol{z}}^*(t),\,\bar{\boldsymbol{\gamma}}^*(t)\right)$ are closer to $\boldsymbol{u}^*(t)$ compared to $\tilde{\boldsymbol{u}}^*(t), t\in[t_k,t_k+T_P]$, . It is important to note here that Eq. (4.10) can be calculated instantaneously if the derivatives are obtained before the sampling time instant t_k. In addition, $\varphi\left(\bar{\boldsymbol{z}}^*(t),\,\bar{\boldsymbol{\gamma}}^*(t)\right)$ can also be calculated instantaneously due to the closed-form formulation (Eq. (4.23)). Thereby, the DMPC-FOA approach can be applied for real-time control of the CAV platoon with no control delay.

As can be noted, before each sampling time instant t_k, the DMPC-FOA approach needs to solve the optimal control problem (4.5) and conduct sensitivity analysis. Hence, the reserved time $\tau_2 \geq \tau_1$. Nevertheless, we will show using numerical examples that the gap between $\bar{\boldsymbol{u}}^*(t)$ and $\boldsymbol{u}^*(t)$ is negligible even for large prediction errors of $\boldsymbol{z}_1(t_k)$ at every sampling time instant

t_k.

4.3 Solution algorithm for optimal control problem (4.5)

To solve optimal control problem (4.5), a two-point boundary value problem is developed in this section based on the necessary conditions for optimality, the solution of which determines the optimal control decisions for all following CAVs. The two-point boundary value problem can be solved efficiently using methods such as the shooting method (Keller, 1976), method of steepest descent (Kirk, 2012), and iterative algorithm (Wang et al., 2014a).

Optimal control problem (4.5) contains control constraints (Eq. (4.5d)) and pure state variable inequality constraints (4.5c). The presence of pure state variable inequality constraints increases the difficulty in designing an effective solution algorithm as these constraints depend on the control history. To address this problem, optimal control problem (4.5) is converted to an equivalent optimal control problem without pure state variable inequality constraints. To do so, we define a variable z that has the following functional relationship

$$\dot{z}(t) = \sum_{i=1}^{n} \{[s_i(t) - s_{min}]^2 I(s_i(t) - s_{min})\} \tag{4.11}$$

where

$$I(s_i(t) - s_{min}) = \begin{cases} 0, if \;\; s_i(t) - s_{min} \geq 0 \\ 1, \quad otherwise \end{cases}$$

According to Eq. (4.11), $\dot{z}(t) \geq 0$. Thereby, $z(t)$ is a non-decreasing function of time t. If $z(0) = z(T_P) = 0$, then $z(t) \equiv 0, t \in [0, T_P]$, implying $s_i(t) - s_{min} \geq 0$ for $t \in [0, T_P]$. Then, optimal control problem (4.5) can be rewritten as the following equivalent problem:

$$\min_{\boldsymbol{u}} \int_0^{T_P} \frac{1}{2} e^{-\beta t} [\mathbf{z}(t)^T \boldsymbol{Q}_1 \mathbf{z}(t) + \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t)]\, dt + \frac{1}{2} e^{-\beta T_P} \mathbf{z}(T_P)^T \boldsymbol{Q}_2 \mathbf{z}(T_P) \tag{4.12a}$$

$$\dot{\mathbf{z}}(t) = \boldsymbol{A} \cdot \mathbf{z}(t) + \boldsymbol{B} \cdot \boldsymbol{u}(t) \tag{4.12b}$$

$$\dot{z}(t) = \sum_{i=1}^{n} \{[s_i(t) - s_{min}]^2 I(s_i(t) - s_{min})\} \tag{4.12c}$$

$$u_{min} \leq u_i \leq u_{max}; \; i = 1,2, \cdots n \tag{4.12d}$$

$$\mathbf{z}(0) = [\mathbf{x}_0^T \quad \mathbf{y}_0^T]^T; z(0) = z(T_P) = 0 \tag{4.12e}$$

where

$$\boldsymbol{Q}_1 = \begin{bmatrix} \boldsymbol{R}_1 & \\ & \boldsymbol{R}_2 \end{bmatrix}; \boldsymbol{Q}_2 = \begin{bmatrix} \boldsymbol{R}_4 & \\ & \boldsymbol{R}_5 \end{bmatrix}$$

To develop a two-point boundary value problem based on the necessary conditions for optimality of optimal control problem (4.5), the terminal condition $z(T_P) = 0$ is removed from Eq. (4.12e). To ensure $z(T_P) \to 0$, the term $\mathrm{M} \cdot \left(z(T_P)\right)^2$ is added to the objective function, where M is a sufficiently large number. If $z(T_P) \neq 0$, the objective function is penalized. The optimal control problem (4.12) can then be re-written as:

$$\min_{\boldsymbol{u}} \int_0^{T_P} \frac{1}{2} e^{-\beta t} [\mathbf{z}(t)^T \boldsymbol{Q}_1 \mathbf{z}(t) + \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t)]\, dt + \frac{1}{2} e^{-\beta T_P} \mathbf{z}(T_P)^T \boldsymbol{Q}_2 \mathbf{z}(T_P) + \mathrm{M} \cdot \left(z(T_P)\right)^2 \tag{4.13a}$$

$$\dot{\mathbf{z}}(t) = \boldsymbol{A} \cdot \mathbf{z}(t) + \boldsymbol{B} \cdot \boldsymbol{u}(t) \tag{4.13b}$$

$$\dot{z}(t) = \sum_{i=1}^{n} \{[s_i(t) - \mathrm{s}_{min}]^2 I(s_i(t) - \mathrm{s}_{min})\} \tag{4.13c}$$

$$u_{min} \leq u_i \leq u_{max};\ i = 1,2, \cdots n \tag{4.13d}$$

$$\mathbf{z}(0) = [\mathbf{x}_0^T \quad \mathbf{y}_0^T]^T; z(0) = 0 \tag{4.13e}$$

Optimal control problem (4.13) is equivalent to problem (4.5). It contains only control constraints. Define the vector of functions $\mathbf{f}_1(\boldsymbol{z}(t), \boldsymbol{u}(t))$ and the function $\mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t))$ as follows:

$$\dot{\mathbf{z}}(t) = \begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{y}}(t) \end{bmatrix} = \mathbf{f}_1(\boldsymbol{z}(t), \boldsymbol{u}(t)) = \boldsymbol{A} \cdot \mathbf{z}(t) + \boldsymbol{B} \cdot \boldsymbol{u}(t) \tag{4.14a}$$

$$\dot{z}(t) = \mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t)) = \sum_{i=1}^{n} \{[s_i(t) - \mathrm{s}_{min}]^2 I(s_i(t) - \mathrm{s}_{min})\} \tag{4.14b}$$

Then, the Hermitian function for optimal control problem (4.13) is written as:

$$\begin{aligned} &\boldsymbol{H}\left(\boldsymbol{z}(t), \boldsymbol{\lambda}_A(t), \boldsymbol{u}(t)\right) \\ &= e^{-\beta t} L\left(\boldsymbol{z}(t), \boldsymbol{u}(t)\right) + \boldsymbol{\lambda}(t)^{T^T} \cdot \mathbf{f}_1\left(\boldsymbol{z}(t), \boldsymbol{u}(t)\right) + \lambda(t) \cdot \mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t)). \end{aligned} \tag{4.15}$$

where $\boldsymbol{\lambda}(t) = [\lambda_1(t) \quad \cdots \quad \lambda_{2n}(t)]^T$ and $\lambda(t)$ are the costate variables associated with $\mathbf{f}_1\left(\boldsymbol{z}(t), \boldsymbol{u}(t)\right)$ and $\mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t))$, respectively. Let $\boldsymbol{\lambda}_A(t) = [\boldsymbol{\lambda}(t)^T, \lambda(t)]^T$, and $\boldsymbol{z}_A(t) = [\boldsymbol{z}(t)^T, z(t)]^T$. According to Pontryagin's minimum principle, the necessary conditions for $\boldsymbol{u}^*(t)$ to be an optimal solution for problem (4.13) are

$$\dot{\boldsymbol{\lambda}}_A(t) = -\left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{z}_A(t)}\right) \tag{4.16a}$$

$$\begin{bmatrix} \dot{\mathbf{z}}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\boldsymbol{z}(t), \boldsymbol{u}(t)) \\ \mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t)) \end{bmatrix} \tag{4.16b}$$

with the initial conditions given in Eq. (4.13e) and the terminal conditions as:

$$\boldsymbol{\lambda}(T_P) = \partial\left(\frac{1}{2}e^{-\beta T_P}\mathbf{z}(t)^T\boldsymbol{Q}_2\mathbf{z}(t)\right)/\partial\mathbf{z}(t)\Big|_{t=T_P} = e^{-\beta T_P}\cdot\boldsymbol{Q}_2\cdot\mathbf{z}(T_P); \tag{4.16c}$$

$$\lambda(T_P) = \partial(\mathrm{M}\cdot z(t)^2)/\partial z(t)|_{t=T_P} = 2\,\mathrm{M}\cdot z(T_P). \tag{4.16d}$$

In addition, the optimal state trajectory $\boldsymbol{z}^*(t)$, the optimal costate trajectory $\boldsymbol{\lambda}_A^*(t)$ and the optimal control decisions $\boldsymbol{u}^*(t)$ should satisfy

$$\boldsymbol{H}(\boldsymbol{z}^*(t),\boldsymbol{\lambda}_A^*(t),\boldsymbol{u}^*(t)) \le \boldsymbol{H}\big(\boldsymbol{z}^*(t),\boldsymbol{\lambda}_A^*(t),\boldsymbol{u}(t)\big);\ \ \boldsymbol{u}(t),\boldsymbol{u}^*(t)\in\boldsymbol{\mathcal{U}} \tag{4.16e}$$

where $\boldsymbol{\mathcal{U}} = \{\boldsymbol{u}|u_{min}\le u_i\le u_{max};\ i=1,2,\cdots n\}$. To convert these necessary conditions for optimality into a two-point boundary value problem, we define the current-value Hamiltonian function as follows:

$$\boldsymbol{H_c} = e^{\beta t}\boldsymbol{H} = L(\boldsymbol{z}(t),\boldsymbol{u}(t)) + \boldsymbol{\gamma}(t)^T\mathbf{f}_1(\boldsymbol{z}(t),\boldsymbol{u}(t)) + \gamma(t)\mathrm{f}_2(\boldsymbol{z}(t),\boldsymbol{u}(t)). \tag{4.17}$$

where $\boldsymbol{\gamma}(t) = \boldsymbol{\lambda}(t)e^{\beta t}, \gamma = \lambda(t)e^{\beta t}$ are the costate variables for the current-value Hamiltonian function. Since the discount factor $e^{-\beta t}$ does not depend on the control variables, the optimal control $\boldsymbol{u}^*$ that minimizes the Hamiltonian function $\boldsymbol{H}$ must also minimize the current-value Hamiltonian function (Eq. (4.17)). Let $\boldsymbol{\gamma}_A = [\boldsymbol{\gamma}(t)^T,\gamma(t)]^T$**.** Then,

$$\dot{\boldsymbol{\lambda}}_A(t) = -\beta e^{-\beta t}\boldsymbol{\gamma}_A(t) + e^{-\beta t}\dot{\boldsymbol{\gamma}}_A(t). \tag{4.18a}$$

$$\frac{\partial\boldsymbol{H}}{\partial\boldsymbol{z}_A(t)} = \frac{\partial\boldsymbol{H}_c}{\partial\boldsymbol{z}_A(t)}e^{-\beta\cdot t} \tag{4.18b}$$

Eqs. 4.18(a) and 4.18(b) imply

$$\dot{\boldsymbol{\gamma}}_A(t) = \dot{\boldsymbol{\lambda}}_A(t) + \beta\boldsymbol{\gamma_A}(t) \tag{4.19}$$

Thereby,

$$\dot{\boldsymbol{\gamma}}(t) = -\frac{\partial\boldsymbol{H}_c}{\partial\boldsymbol{z}} + \beta\boldsymbol{\gamma}(t) = -\frac{\partial\,\mathbf{f}_1(\boldsymbol{z},\boldsymbol{u})}{\partial\boldsymbol{z}}\boldsymbol{\gamma}(t) - \frac{\partial\mathrm{f}_2(\boldsymbol{z},\boldsymbol{u})}{\partial\boldsymbol{z}}\gamma(t) - \frac{\partial\,L(\boldsymbol{z},\boldsymbol{u})}{\partial\boldsymbol{z}} + \beta\boldsymbol{\gamma}(t) = -\boldsymbol{A}\cdot\boldsymbol{\gamma}(t) - \begin{bmatrix}\boldsymbol{C}_x\\ \boldsymbol{C}_y\end{bmatrix}\gamma(t) - \boldsymbol{Q}_1\mathbf{z}(t) + \beta\boldsymbol{\gamma}(t). \tag{4.20a}$$

$$\dot{\gamma}(t) - \beta\cdot\gamma(t) = -\partial\boldsymbol{H}_c/\partial z(t) = 0 \tag{4.20b}$$

where

$$\boldsymbol{C}_x = \frac{\partial \mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t))}{\partial \mathbf{x}(t)} = \begin{bmatrix} 2 \cdot [s_1(t) - s_{min}] \cdot I(s_1(t) - s_{min}) \\ 2 \cdot [s_2(t) - s_{min}] \cdot I(s_2(t) - s_{min}) \\ \vdots \\ 2 \cdot [s_n(t) - s_{min}] \cdot I(s_n(t) - s_{min}) \end{bmatrix}$$

$$\boldsymbol{C}_y = \frac{\partial \mathrm{f}_2(\boldsymbol{z}(t), \boldsymbol{u}(t))}{\partial \mathbf{y}(t)} = [C_{1,y} \quad C_{2,y} \quad \cdots \quad C_{n,y}]$$

$$C_{i,y} = \sum_{j=1}^{i} 2 \cdot \left[s_j(t) - s_{min}\right] \cdot I\left(s_j(t) - s_{min}\right), \forall i = 1,2, \cdots n.$$

The terminal conditions in Eq. (4.16c) and Eq. (4.16d) imply that

$$\boldsymbol{\gamma}(T_P) = \boldsymbol{Q}_2 \cdot \mathbf{z}(T_P)^T, \gamma(T_P) = e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot z(T_P) \tag{4.21}$$

Let $\boldsymbol{\gamma}_A^*(t) = \boldsymbol{\lambda}_A^*(t) e^{\beta t}$. Since $e^{\beta t} > 0$, according to Eq. (4.17), at time t , minimizing $\boldsymbol{H}(\boldsymbol{z}^*(t), \boldsymbol{\lambda}_A^*(t), \boldsymbol{u}(t))$ with respect to $\boldsymbol{u}(t)$ is equivalent to minimizing $\boldsymbol{H_c}(\boldsymbol{z}^*(t), \boldsymbol{\gamma}_A^*(t), \boldsymbol{u}(t))$ with respect to $\boldsymbol{u}(t)$ **.** This indicates that if the optimal control $\boldsymbol{u}^*$ minimizes $\boldsymbol{H_c}(\boldsymbol{z}^*(t), \boldsymbol{\lambda}_A^*(t), \boldsymbol{u}(t))$**,** it is the solution to inequality (4.16e). Thereby, $\boldsymbol{u}^*(t)$ can be found by solving the following minimization problem

$$min_{\boldsymbol{u}(t)} \boldsymbol{H_c}(\boldsymbol{z}^*(t), \boldsymbol{\gamma}_A^*(t), \boldsymbol{u}(t)); \quad \boldsymbol{u}(t), \boldsymbol{u}^*(t) \in \boldsymbol{\mathcal{U}} \tag{4.22}$$

Proposition 4.1. *Let* $[p_1(t) \quad p_2(t) \quad \cdots \quad p_n(t)]^T = -(\boldsymbol{R}_3)^{-1}(\boldsymbol{B}^T \boldsymbol{\gamma}^*(t))$*; if* $\boldsymbol{R}_3$ *is a diagonal positive definite matrix, then the optimal control decisions* $\boldsymbol{u}^* = [u_1^* \quad u_2^* \quad \cdots \quad u_n^*]$ *that minimizes* $\boldsymbol{H_c}(\boldsymbol{z}^*, \boldsymbol{\gamma}_A^*, \boldsymbol{u})$ *is unique and can be formulated as*

$$u_i^*(t) = \varphi\left(\boldsymbol{z}^*(t), \boldsymbol{\gamma}^*(t)\right) = \begin{cases} u_{min}, if\ p_i(t) < u_{min} \\ u_{max}, if\ p_i(t) > u_{max} \\ p_i(t), if u_{min} \leq\ p_i(t) \leq u_{max} \end{cases} \tag{4.23}$$

Proof. If $\boldsymbol{u}^* = [u_1^* \quad u_2^* \quad \cdots \quad u_n^*]$ minimizes $\boldsymbol{H_c}(\boldsymbol{z}^*(t), \boldsymbol{\gamma}_A^*(t), \boldsymbol{u}(t))$**,** then we have

$$\begin{aligned} &L\left(\boldsymbol{z}^*(t), \boldsymbol{u}^*(t)\right) + \boldsymbol{\gamma}^*(t)\mathbf{f}_1\left(\boldsymbol{z}^*(t), \boldsymbol{u}^*(t)\right) + \gamma^*(t)\mathrm{f}_2\left(\boldsymbol{z}^*(t), \boldsymbol{u}^*(t)\right) \\ &\leq L(\boldsymbol{z}^*(t), \boldsymbol{u}(t)) + \boldsymbol{\gamma}^*(t)\mathbf{f}_1(\boldsymbol{z}^*(t), \boldsymbol{u}(t)) + \gamma^*(t)\mathrm{f}_2(\boldsymbol{z}^*(t), \boldsymbol{u}(t)) \end{aligned} \tag{4.24}$$

Eq. (4.24) indicates

$$\begin{aligned} &0.5 \cdot \boldsymbol{u}^*(t)^T \boldsymbol{R}_3 \boldsymbol{u}^*(t) + \left(\boldsymbol{\gamma}^*(t)\right)^T \cdot \boldsymbol{B} \cdot \boldsymbol{u}^*(t) \\ &\leq 0.5 \cdot \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t) + \left(\boldsymbol{\gamma}^*(t)\right)^T \cdot \boldsymbol{B} \cdot \boldsymbol{u}(t) \end{aligned} \tag{4.25}$$

Let $\boldsymbol{p}^*(t) = (\boldsymbol{R}_3)^{-1}(\boldsymbol{B}^T \boldsymbol{\gamma}^*(t)) = -[p_1 \quad p_2 \quad \cdots \quad p_n]^T$. Then

$$\left(\boldsymbol{\gamma}^*(t)\right)^T \cdot \boldsymbol{B} \cdot \boldsymbol{u}^*(t) = \left(\boldsymbol{u}^*(t)\right)^T \boldsymbol{B}^T \boldsymbol{\gamma}^*(t) = \left(\boldsymbol{u}^*(t)\right)^T \boldsymbol{R}_3 \boldsymbol{p}^*(t) \tag{4.26a}$$

$$\left(\boldsymbol{\gamma}^*(t)\right)^T \cdot \boldsymbol{B} \cdot \boldsymbol{u}(t) = (\boldsymbol{u}(t))^T \boldsymbol{B}^T \boldsymbol{\gamma}^*(t) = \left(\boldsymbol{u}(t)\right)^T \boldsymbol{R}_3 \boldsymbol{p}^*(t) \tag{4.26b}$$

Substituting Eq. (4.26) into Eq. (4.25), we have

$$\begin{aligned} &0.5 \cdot \boldsymbol{u}^*(t)^T \boldsymbol{R}_3 \boldsymbol{u}^*(t) + \left(\boldsymbol{u}^*(t)\right)^T \boldsymbol{R}_3 \boldsymbol{p}^*(t) \\ &\leq 0.5 \cdot \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t) + (\boldsymbol{u}(t))^T \boldsymbol{R}_3 \boldsymbol{p}^*(t) \end{aligned} \tag{4.27}$$

Adding $0.5 \cdot \boldsymbol{p}^*(t)^T \boldsymbol{R}_3 \boldsymbol{p}^*(t) = 0.5\left(\boldsymbol{\gamma}^*(t)\right)^T \cdot \boldsymbol{B}(\boldsymbol{R}_3)^{-1} \boldsymbol{B}^T \boldsymbol{\gamma}^*(t)$ to both sides of inequality (4.27), we have

$$0.5[\boldsymbol{u}^*(t) + \boldsymbol{p}^*(t)]^T \boldsymbol{R}_3 [\boldsymbol{u}^*(t) + \boldsymbol{p}^*(t)] \leq 0.5[\boldsymbol{u}(t) + \boldsymbol{p}^*(t)]^T \boldsymbol{R}_3 [\boldsymbol{u}(t) + \boldsymbol{p}^*(t)] \tag{4.28}$$

Inequality (4.28) implies that if $\boldsymbol{u}^*$ minimizes $\boldsymbol{H_c}(\boldsymbol{z}^*(t), \boldsymbol{\gamma}_A^*(t), \boldsymbol{u}(t))$, it must minimize inequality (4.27) and vice versa. Thereby

$$\boldsymbol{u}^*(t) = min_{\boldsymbol{u} \in \boldsymbol{U}} [\boldsymbol{u}(t) + \boldsymbol{p}^*(t)]^T \boldsymbol{R}_3 [\boldsymbol{u}(t) + \boldsymbol{p}^*(t)] \tag{4.29}$$

Note $\boldsymbol{R}_3$ is a diagonal positive definite matrix; without loss of generosity, let $\boldsymbol{R}_3 = diag([w_1, w_2 \cdots, w_n])$, $r_i > 0, \forall i = 1,2, \cdots, n$. Then, inequality (4.29) can be written as

$$\begin{aligned} \boldsymbol{u}^*(t) &= min_{\boldsymbol{u} \in \boldsymbol{U}} \sum_{i=1}^{n} w_i [u_i(t) - p_i]^2 \\ &= \sum_{i=1}^{n} min_{u_{min} \leq u_i \leq u_{max}} w_i [u_i(t) - p_i]^2 \end{aligned} \tag{4.30}$$

The only solution to the above inequality is

$$u_i^*(t) = \begin{cases} u_{min}, if\ p_i(t) < u_{min} \\ u_{max}, if\ p_i(t) > u_{max} \\ p_i(t), if u_{min} \leq\ p_i(t) \leq u_{max} \end{cases} ; \forall i = 1,2, \cdots, n \tag{4.31}$$

This completes the proof. ■

Eq. (4.13b), Eq. (4.13c), Eq. (4.20a), Eq. (4.20b) and Eq. (4.23) form a two-point boundary value problem as follows with initial conditions and terminal conditions provided by Eq.(4.12e) and Eq. (4.21), respectively.

$$\dot{\mathbf{z}}(t) = \boldsymbol{A} \cdot \mathbf{z}(t) + \boldsymbol{B} \cdot \varphi\left(\boldsymbol{z}(t), \boldsymbol{\gamma}(t)\right) \tag{4.32a}$$

$$\dot{z}(t) = \sum_{i=1}^{n} \{[s_i(t) - \mathrm{s}_{min}]^2 I(s_i(t) - \mathrm{s}_{min})\} \tag{4.32c}$$

$$\dot{\boldsymbol{\gamma}}(t) = -\boldsymbol{A} \cdot \boldsymbol{\gamma}(t) - \begin{bmatrix} \boldsymbol{C}_x \\ \boldsymbol{C}_y \end{bmatrix} \gamma(t) - \boldsymbol{Q}_1 \mathbf{z}(t) + \beta \boldsymbol{\gamma}(t) \tag{4.32d}$$

$$\dot{\gamma}(t) = \beta \cdot \gamma(t) \tag{4.32e}$$

$$\mathbf{z}(0) = [\mathbf{x}_0^T \quad \mathbf{y}_0^T]^T; z(0) = 0; \tag{4.32f}$$

$$\boldsymbol{\gamma}(T_P) = \boldsymbol{Q}_2 \mathbf{z}(T_P), \gamma(T_P) = e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot z(T_P) \tag{4.32g}$$

The two-point boundary value problem can be solved using many existing solution

algorithms. A review of these algorithms is provided in Kirk (2012). In this study, the shooting method is used to solve the two-point boundary value problem (4.32). The details of implementing the shooting method can be found in Keller (1976). The main advantage of the shooting method is that it converges very fast if the algorithm starts to converge (Keller, 1976). Note, $\partial^2 \boldsymbol{H}_c/\partial(\boldsymbol{u}(t))^2 = \boldsymbol{R}_3$ is a positive definite matrix. Thereby, the solution $(\boldsymbol{z}^*(t), z^*(t), \boldsymbol{\gamma}^*(t), \gamma^*(t))$ of the two-point boundary value problem (4.32) is a minimum solution of optimal control problem (4.5). The optimal control $\boldsymbol{u}^*(t)$ can be obtained by inputting $\boldsymbol{\gamma}^*(t)$ into Eq. (4.23).

4.4 Sensitivity analysis of the optimal control problem

For the DMPC approach, at each sampling time instant t_k, the control decisions are determined by solving the two-point boundary value problem (4.32) with the predicted spacing error and relative speed of vehicle 1 with respect to the leading vehicle (i.e., $\hat{x}_1(t_k)$ and $\hat{y}_1(t_k)$). The resulting control decisions may deviate significantly from those of the idealized MPC strategy due to errors in predicting f $x_1(t_k)$ and $y_1(t_k)$, which can decrease the platoon performance and cause collisions. To address this issue, the DMPC-FOA approach corrects the estimated control decisions of the DMPC approach using first-order Taylor approximation. To do so, the main step is to obtain the derivatives of the optimal solution of the state and costate variables with respect to $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ in the DMPC-FOA approach, respectively.

The sensitivity analysis of an optimal control problem quantitatively measures the change in the optimal solution of the state and costate variables induced by a unit change in the perturbed parameters (i.e., $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ in this study). Parametric sensitivity of optimal problem has been extensively studied. Dorato (1963) developed an analytical model to study the variation of the objective function with respect to parametric perturbations. Malanowski (1984, 1987) discussed the conditions for directional differentiability of the solutions for an optimal control problem with nonlinear ordinary dynamics. Maurer and Pesch (1984) developed an analytical method for sensitivity analysis of optimal control problems with no constraints. This method is further extended to study the sensitivity analysis of optimal control problems with control constraints (Maurer and Pesch, 1995; Malanowski and Maurer, 1996), and pure state variable constraints (Augustin and Maurer, 2001; Malanowski, 2011). Here, the analytical method for sensitivity analysis of the optimal control decisions with respect to $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ will be

derived by modifying the method developed by Maurer and Pesch (1995) for a general optimal control problem.

Denote $\tilde{\boldsymbol{u}}^*(t) = [\tilde{u}_1^*(t), \cdots, \tilde{u}_n^*(t)]$ as the control decisions obtained by solving Eq. (4.32) using $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ predicted by the DMPC-FOA approach. The corresponding solutions for the state variables (i.e., $\mathbf{z}(t)$, $z(t)$) and costate variables (i.e., $\boldsymbol{\gamma}(t)$, $\gamma(t)$) are denoted as $\tilde{\mathbf{z}}^*(t)$, $\tilde{z}^*(t)$, $\tilde{\boldsymbol{\gamma}}^*(t)$ and $\tilde{\gamma}^*(t)$, respectively. Let the derivatives of the optimal solutions for the state and costate variables with respect to $\tilde{x}_1(t_k)$ be defined as follows:

$$\mathbf{h}_{\tilde{x}_1}(t) = \frac{\partial \tilde{\mathbf{z}}^*(t)}{\partial \tilde{x}_1(t_k)};\ h_{\tilde{x}_1}(t) = \frac{\partial \tilde{z}^*(t)}{\partial \tilde{x}_1(t_k)}$$

$$\boldsymbol{\eta}_{\tilde{x}_1}(t) = \frac{\partial \tilde{\boldsymbol{\gamma}}^*(t)}{\partial \tilde{x}_1(t_k)};\ \eta_{\tilde{x}_1}(t) = \frac{\partial \tilde{\gamma}^*(t)}{\partial \tilde{x}_1(t_k)}.$$

According to $\tilde{u}_i^*(t)$, we can obtain the set of time intervals $\Omega_{i,1}$, $\Omega_{i,2}$, and $\Omega_{i,2}$ ($\Omega_{i,1} \cup \Omega_{i,2} \cup \Omega_{i,2} = [0, T_P]$) for each vehicle $i, i = 1,2, \cdots, n$ such that

$$\tilde{u}_i^*(t) = \begin{cases} u_{min}, t \in \Omega_{i,1} \\ u_{max}, t \in \Omega_{i,2} \\ \tilde{p}_i, t \in \Omega_{i,3} \end{cases} \tag{4.33}$$

where $[\tilde{p}_1(t) \quad \tilde{p}_2(t) \quad \cdots \quad \tilde{p}_n(t)]^T = -(\boldsymbol{R}_3)^{-1}(\boldsymbol{B}^T\tilde{\boldsymbol{\gamma}}^*(t))$.

Then, according to Eq. (4.33), we have

$$\frac{d\tilde{u}_i^*(t)}{d\tilde{x}_1(t_k)} = \begin{cases} 0, t \in (\Omega_{i,1} \cup \Omega_{i,2}) \\ m_{\tilde{x}_1,i}(t), t \in \Omega_{i,3} \end{cases} \tag{4.34a}$$

where

$$[m_{\tilde{x}_1,1}(t) \quad m_{\tilde{x}_1,2}(t) \quad \cdots \quad m_{\tilde{x}_1,n}(t)]^T = -(\boldsymbol{R}_3)^{-1}\left(\boldsymbol{B}^T\,\boldsymbol{\eta}_{\tilde{x}_1}(t)\right). \tag{4.34b}$$

Let $\Psi\left(\boldsymbol{\eta}_{\tilde{x}_1}(t)\right) = \left[\frac{d\tilde{u}_1^*(t)}{d\tilde{x}_1(t_k)} \quad \frac{d\tilde{u}_2^*(t)}{d\tilde{x}_1(t_k)} \quad \cdots \quad \frac{d\tilde{u}_n^*(t)}{d\tilde{x}_1(t_k)}\right]^T$. Differentiating both sides of Eqs. (4.32a)-(4.32g) with respect to $\tilde{x}_1(t_k)$, we have

$$\dot{\mathbf{h}}_{\tilde{x}_1}(t) = \boldsymbol{A} \cdot \mathbf{h}_{\tilde{x}_1} + \boldsymbol{B} \cdot \Psi\left(\boldsymbol{\eta}_{\tilde{x}_1}(t)\right) \tag{4.35a}$$

$$\dot{h}_{\tilde{x}_1 1} = \left[\boldsymbol{C}_x^T \quad \boldsymbol{C}_y^T\right] \cdot \mathbf{h}_{\tilde{x}_1} \tag{4.35c}$$

$$\dot{\boldsymbol{\eta}}_{\tilde{x}_1} = -\boldsymbol{A} \cdot \boldsymbol{\eta}_{\tilde{x}_1}(t) - \begin{bmatrix} \boldsymbol{C}_x \\ \boldsymbol{C}_y \end{bmatrix} \eta_{\tilde{x}_1}(t) - \boldsymbol{Q}_1 \mathbf{h}_{\tilde{x}_1} + \beta \boldsymbol{\eta}_{\tilde{x}_1}(t) \tag{4.35d}$$

$$\dot{\eta}_{\tilde{x}_1,}(t) = \beta \cdot \eta_{\tilde{x}_1}(t) \tag{4.35e}$$

with initial and terminal conditions as:

$$\mathbf{h}_{\tilde{x}_1}(0) = \frac{\partial \tilde{\mathbf{z}}(0)}{\partial \tilde{x}_1(t_k)} = \frac{\partial \tilde{\mathbf{z}}(t_k)}{\partial \tilde{x}_1(t_k)} = [1, \mathbf{0}_{1\times 2n-1}]^T \tag{4.35f}$$

$$h_{\tilde{x}_1}(0) = \frac{\partial \tilde{z}(0)}{\partial \tilde{x}_1(t_k)} = \frac{\partial (0)}{\partial \tilde{x}_1(t_k)} = 0 \tag{4.35g}$$

$$\boldsymbol{\eta}_{\tilde{x}_1}(T_P) = \frac{\partial \widetilde{\boldsymbol{\gamma}}_1(T_P)}{\partial \tilde{x}_1(t_k)} = \frac{\partial (\boldsymbol{Q}_2 \tilde{\mathbf{z}}(T_P))}{\partial \tilde{x}_1(t_k)} = \boldsymbol{Q}_2 \cdot \mathbf{h}_{\tilde{x}_1}(T_P) \tag{4.35h}$$

$$\eta_{\tilde{x}_1}(T_P) = \frac{\partial \widetilde{\gamma}(T_P)}{\partial \tilde{x}_1(t_k)} = \frac{\partial \left(e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot \tilde{z}(T_P)\right)}{\partial \tilde{x}_1(t_k)} = e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot h_{\tilde{x}_1}(T_P) \tag{4.35i}$$

where $\mathbf{0}_{1\times 2n-1}$ is a $(2n-1)$-dimensional zero vector. Eqs. (4.35a)-(4.35i) also form a two-point boundary value problem which can be solved using the shooting method.

To obtain the derivatives of the optimal state and costate variables with respect to $\tilde{y}_1(t_k)$, similarly, let

$$\mathbf{h}_{\tilde{y}_1}(t) = \frac{\partial \tilde{\mathbf{z}}^*(t)}{d\tilde{y}_1(t_k)};\ \ h_{\tilde{y}_1}(t) = \frac{\partial \tilde{z}^*(t)}{d\tilde{y}_1(t_k)} \tag{4.36a}$$

$$\boldsymbol{\eta}_{\tilde{y}_1}(t) = \frac{\partial \widetilde{\boldsymbol{\gamma}}^*(t)}{\partial \tilde{y}_1(t_k)};\ \eta_{\tilde{y}_1}(t) = \frac{\partial \tilde{\gamma}^*(t)}{\partial \tilde{y}_1(t_k)}. \tag{4.36b}$$

Differentiating both sides of Eqs. (4.32a)-(4.32f) with respect to $\tilde{y}_1(t_k)$, we can obtain a similar two-point boundary value problem, as follows:

$$\dot{\mathbf{h}}_{\tilde{y}_1}(t) = \boldsymbol{A} \cdot \mathbf{h}_{\tilde{y}_1} + \boldsymbol{B} \cdot \Psi\left(\boldsymbol{\eta}_{\tilde{y}_1}(t)\right) \tag{4.37a}$$

$$\dot{h}_{\tilde{y}_1} = \begin{bmatrix} \boldsymbol{C}_x^T & \boldsymbol{C}_y^T \end{bmatrix} \mathbf{h}_{\tilde{y}_1} \tag{4.37c}$$

$$\dot{\boldsymbol{\eta}}_{\tilde{y}_1} = -\boldsymbol{A} \cdot \boldsymbol{\eta}_{\tilde{y}_1}(t) - \begin{bmatrix} \boldsymbol{C}_x \\ \boldsymbol{C}_y \end{bmatrix} \eta_{\tilde{y}_1}(t) - \boldsymbol{Q}_1 \mathbf{h}_{\tilde{y}_1} + \beta \boldsymbol{\eta}_{\tilde{y}_1}(t) \tag{4.37d}$$

$$\dot{\eta}_{\tilde{y}_1}(t) = \beta \cdot \eta_{\tilde{y}_1}(t) \tag{4.37e}$$

with initial and terminal conditions as:

$$\mathbf{h}_{\tilde{y}_1}(0) = \frac{\partial \tilde{\mathbf{z}}(0)}{\partial \tilde{y}_1(t_k)} = \frac{\partial \tilde{\mathbf{z}}(t_k)}{\partial \tilde{y}_1(t_k)} = [\mathbf{0}_{1\times n}, 1, \mathbf{0}_{1\times n-1}]^T \tag{4.37f}$$

$$h_{\tilde{y}_1}(0) = \frac{\partial \tilde{z}(0)}{\partial \tilde{y}_1(t_k)} = \frac{\partial (0)}{\partial \tilde{y}_1(t_k)} = 0 \tag{4.37g}$$

$$\boldsymbol{\eta}_{\tilde{y}_1}(T_P) = \frac{\partial \widetilde{\boldsymbol{\gamma}}_1(T_P)}{\partial \tilde{y}_1(t_k)} = \frac{d(\boldsymbol{Q}_2 \tilde{\mathbf{z}}(T_P))}{\partial \tilde{y}_1(t_k)} = \boldsymbol{Q}_2 \cdot \mathbf{h}_{\tilde{y}_1}(T_P) \tag{4.37h}$$

$$\eta_{\bar{y}_1}(T_P) = \frac{\partial \widetilde{\gamma}(T_P)}{\partial \tilde{y}_1(t_k)} = \frac{\partial \left(e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot z(T_P)\right)}{\partial \tilde{y}_1(t_k)} = e^{\beta T_P} \cdot \mathrm{M} \cdot 2 \cdot h_{\tilde{y}_1}(T_P) \tag{4.37i}$$

where $\mathbf{0}_{1\times n-1}$ is a $(n-1)$-dimensional zero vector. The vector of functions $\Psi\left(\boldsymbol{\eta}_{\tilde{y}_1}(t)\right)$ is similar to $\Psi\left(\boldsymbol{\eta}_{\tilde{x}_1}(t)\right)$. It is formulated by replacing the subscript "$\tilde{x}_1$" in Eq. (4.34) with "$\tilde{y}_1$".

The derivatives of the optimal solutions for the state and costate variables with respect to $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ can be obtained by solving the two-point boundary value problems (35) and (37), respectively. Then, when the actual value of $x_1(t_k)$ and $y_1(t_k)$ are detected at the sampling time instant t_k, the optimal solution of the state and costate variables of the idealized MPC strategy can be estimated using first-order Taylor approximation, as follows

$$\bar{\boldsymbol{z}}^*(t) = \tilde{\boldsymbol{z}}^*(t) + \mathbf{h}_{\tilde{x}_1}(t)\big(x_1(t_k) - \tilde{x}_1(t_k)\big) + \mathbf{h}_{\tilde{y}_1}(t)\left(y_1(t_k) - \tilde{y}_1(t_k)\right) \tag{4.38a}$$

$$\bar{\boldsymbol{\gamma}}^*(t) = \tilde{\boldsymbol{\gamma}}^*(t) + \boldsymbol{\eta}_{\tilde{x}_1}(t)\big(x_1(t_k) - \tilde{x}_1(t_k)\big) + \boldsymbol{\eta}_{\tilde{y}_1}(t)\left(y_1(t_k) - \tilde{y}_1(t_k)\right) \tag{4.38b}$$

Eq. (4.38a) and Eq. (4.38b) can be calculated instantaneously at the sampling time instant t_k as $\mathbf{h}_{\tilde{x}_1}(t)$, $\mathbf{h}_{\tilde{y}_1}(t)$, $\boldsymbol{\eta}_{\tilde{x}_1}(t)$ and $\boldsymbol{\eta}_{\tilde{y}_1}(t)$ are obtained before t_k. Eq. (4.38) indicates that compared to $\left[\tilde{\boldsymbol{z}}^*(t), \tilde{\boldsymbol{\gamma}}^*(t)\right]$, $[\bar{\boldsymbol{z}}^*(t), \bar{\boldsymbol{\gamma}}^*(t)]$ are closer to $[\boldsymbol{z}^*(t), \boldsymbol{\gamma}^*(t)]$ calculated for the idealized MPC strategy using exact $x_1(t_k)$ and $y_1(t_k)$. According to Eq. (4.23), the optimal control decisions of the idealized MPC strategy can be estimated as

$$\bar{u}_i^*(t) = \varphi\big(\bar{\boldsymbol{z}}^*(t), \bar{\boldsymbol{\gamma}}^*(t)\big) = \begin{cases} u_{min}, if\ \bar{p}_i(t) < u_{min} \\ u_{max}, if\ \bar{p}_i(t) > u_{max} \\ \bar{p}_i(t), if u_{min} \le \bar{p}_i(t) \le u_{max} \end{cases} ; \forall i = 1,2,\cdots,n \tag{4.39}$$

where $[\bar{p}_1(t) \quad \bar{p}_2(t) \quad \cdots \quad \bar{p}_n(t)]^T = -(\boldsymbol{R}_3)^{-1}(\boldsymbol{B}^T\bar{\boldsymbol{\gamma}}^*(t))$. Compared to $\tilde{\boldsymbol{u}}^*(t)$, the estimated $\bar{\boldsymbol{u}}^*(t), (\bar{\boldsymbol{u}}^*(t) = [\bar{u}_1^*(t) \quad \bar{u}_2^*(t) \quad \cdots \quad \bar{u}_n^*(t)]^T)$ is closer to $\boldsymbol{u}^*(t)$ calculated using the idealized MPC strategy as $\bar{\boldsymbol{\gamma}}^*(t)$ is closer to $\boldsymbol{\gamma}^*(t)$ compared to $\tilde{\boldsymbol{\gamma}}^*(t)$.

Proposition 4.2: *If the inequality constraints (4.5c) and (4.5d) are not active along the trajectory of the optimal solution* $(\tilde{\boldsymbol{z}}^*(t), \tilde{z}^*(t), \tilde{\boldsymbol{\gamma}}^*(t), \tilde{\gamma}^*(t))$ *obtained with the predicted initial state* $\tilde{x}_1(t_k)$ *and* $\tilde{y}_1(t_k)$*, then the derivatives of optimal solutions for the state and costate variables with respect to* $\tilde{x}_1(t_k)$ *and* $\tilde{y}_1(t_k)$ *are the same for all solutions of* $(\tilde{\boldsymbol{z}}^*(t), \tilde{z}^*(t), \tilde{\boldsymbol{\gamma}}^*(t), \tilde{\gamma}^*(t))$ *for which the inequality constraints (4.5c) and (4.5d) are not active.*

Proof: If the inequality constraints (4.5c) and (4.5d) are not active along the optimal solution, $\tilde{z}^*(t) \equiv 0, t \in [0, T_P]$. According to Eq. (4.16d), $\tilde{\gamma}^*(T_P) = 2\,\mathrm{M} \cdot \tilde{z}^*(T_P) = 2\,\mathrm{M} \cdot 0 = 0$. Based on Eq. (4.32e), $\tilde{\gamma}^*(t) \equiv 0, t \in [0, T_P]$. This indicates that $\eta_{\tilde{x}_1}(t) = \eta_{\tilde{y}_1}(t) \equiv 0, t \in [0, T_P]$. In addition, $\Psi\left(\boldsymbol{\eta}_{\tilde{x}_1}(t)\right) = -(\boldsymbol{R}_3)^{-1}\left(\boldsymbol{B}^T\boldsymbol{\eta}_{\tilde{x}_1}(t)\right)$ and $\Psi\left(\boldsymbol{\eta}_{\tilde{y}_1}(t)\right) = -(\boldsymbol{R}_3)^{-1}\left(\boldsymbol{B}^T\boldsymbol{\eta}_{\tilde{y}_1}(t)\right)$. Thereby, the two-point boundary value problems (4.35) and (4.37) are the same for different optimal solutions under which the inequality constraints (4.5c) and (4.5d)

are not active. This indicates that the derivatives of the optimal solutions for the state and costate variables with respect to $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ are the same for all of these solutions. ■

Proposition 4.2 implies that if under the optimal control decisions, the following vehicles in the platoon do not brake and accelerate at the maximum values, and the spacing between all adjacent vehicle pairs is larger than the minimum spacing during time interval $[t_k, t_k + T_P]$, the changes in the optimal control decisions for a unit change in $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ would be the same for all of these optimal control decisions. It is worth noting that the idealized MPC strategy can coordinate the behaviors of all following vehicles to minimize the objective function efficiently. It can enable smoother deceleration and acceleration behavior of all following vehicles even if the leading vehicle decelerates or accelerates at the maximum value. The following vehicles accelerate or decelerate at the maximum value only when the spacing between two consecutive vehicles is too large or too small. Thereby, according to Proposition 4.2, under normal conditions, the derivatives of the optimal solutions for the state and costate variables, i.e., $(\tilde{\boldsymbol{z}}(t), \tilde{z}(t), \tilde{\boldsymbol{\gamma}}(t), \tilde{\gamma}(t))$ with respect to $\tilde{x}_1(t_k)$ and $\tilde{y}_1(t_k)$ are the same and are independent of these solutions. Let $\mathbf{h}_l^*(t), h_l^*(t), \boldsymbol{\eta}_l^*(t), \eta_l^*(t), l \in \{\tilde{x}_1, \tilde{y}_1\}, t \in [0, T_P]$ be the corresponding derivatives. These derivatives can be obtained offline to avoid solving the two-point boundary value problem (4.35) and (37). Thereby, under normal situations when the inequality constraints (4.5c) and (4.5d) are not active along the optimal solution, the reserved time for computing in DMPC-FOA approach can be the same as that of the DMPC approach. In addition, $\mathbf{h}_l^*(t), h_l^*(t), \boldsymbol{\eta}_l^*(t), \eta_l^*(t), l \in \{\tilde{x}_1, \tilde{y}_1\}, t \in [0, T_P]$ can be used as the initial value to solve the two-point boundary value problems (4.35) and (4.37) when one of the inequality constraints (4.5c) or (4.5d) is active along the optimal solution. This can significantly reduce the computational time for solving the two problems. This property enhances the applicability of the proposed DMPC-FOA approach for controlling the CAV platoon in real-time.

4.5 Stability analysis of the idealized MPC strategy with no inequality constraints

Stability is an important property for a CAV platoon. It indicates the capability of a platoon to recover to a stable state after external disturbances on the platoon formation (e.g., unexpected hard acceleration and deceleration of the leading vehicle). In this study, the condition for asymptotic stability of the idealized MPC strategy is analyzed to ensure that the CAV platoon can dampen traffic oscillations efficiently. This condition also ensures the stability of the DMPC-

FOA approaches as it is proposed to characterize the control decisions of the idealized MPC strategy. Similar to Gong et al., (2016), the stability analysis of the idealized MPC strategy is based on optimal control problem (4.5) with no inequality constraints as they are not active in most traffic flow scenarios. The conditions for asymptotic stability of the idealized MPC strategy with active constraints will be investigated in our future work.

For convenience of stability analysis, in the following, optimal control problem (4.5) without inequality constraints (4.5c) and (4.5d) is transformed into an equivalent form for analyzing stability. The conditions for asymptotic stability of the unconstrained idealized MPC strategy are analyzed using the stability theorem for continuous MPC problems developed by Mayne et.al. (2000). Let

$$\boldsymbol{z}_\beta(t) = e^{-\frac{\beta}{2}t}\boldsymbol{z}(t) \tag{4.40a}$$

$$\boldsymbol{u}_\beta(t) = e^{-\frac{\beta}{2}t}\boldsymbol{u}(t) \tag{4.40b}$$

Then, optimal control problem (4.5) without inequality constraints (4.5c) and (4.5d) can be formulated as

$$\min_{\boldsymbol{u}_\beta} \int_0^{T_P} \left[\boldsymbol{z}_\beta(t)^T \boldsymbol{Q}_1 \boldsymbol{z}_\beta(t) + \boldsymbol{u}_\beta(t)^T \boldsymbol{R}_3 \boldsymbol{u}_\beta(t)\right] dt + \boldsymbol{z}_\beta(T_P)^T \boldsymbol{Q}_2 \boldsymbol{z}_\beta(T_P) \tag{4.41a}$$

$$s.t\ \dot{\boldsymbol{z}}_\beta(t) = \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z}_\beta(t) + \boldsymbol{B}\boldsymbol{u}_\beta(t) \tag{4.41b}$$

$$\boldsymbol{z}_\beta(0) = [\mathbf{x}_k \quad \mathbf{y}_k]^T \tag{4.41c}$$

The following theorem is used to analyze the asymptotical stability of the idealized MPC strategy with no inequality constraints.

Theorem 4.1 (Mayne et.al. 2000): *Consider the following continuous constrained MPC problem*

$$\min_{\boldsymbol{a}} \int_0^{T_P} L(\boldsymbol{z}(t), \boldsymbol{u}(t))\, dt + F(\boldsymbol{z}(T_P))$$

$$s.t \quad \dot{\boldsymbol{z}} = g(\boldsymbol{z}, \boldsymbol{u})$$
$$\boldsymbol{z}(t) \in \mathcal{Z}, \quad for\ t\epsilon[0, T_P]$$
$$\boldsymbol{u}(t) \in \mathcal{A}, \quad for\ t\epsilon[0, T_P]$$
$$\boldsymbol{z}(T_P) \in \mathcal{Z}_f$$

where $\boldsymbol{z}$ and $\boldsymbol{u}$ are vectors of the state variables and control variables, respectively. $\boldsymbol{z}(T_P)$ is the value of $\boldsymbol{z}(t)$ at terminal time T_P. $\mathcal{Z}$, $\mathcal{A}$, and $\mathcal{Z}_f$ are the feasible sets for $\boldsymbol{z}(t)$, $\boldsymbol{u}(t)$ and $\boldsymbol{z}(T_P)$, respectively. If there exists a nominal controller $\kappa(\boldsymbol{z})$ such that the following four conditions hold for the above continuous MPC problem, then it is asymptotic stable.

(1). $0 \in \mathcal{Z}$

(2). $\kappa(\boldsymbol{z}) \in \mathcal{A}, \forall \boldsymbol{z} \in \mathcal{Z}_f$

(3). $g(\boldsymbol{z}, \kappa(\boldsymbol{z})) \in \mathcal{Z}_f$ *for* $\forall \boldsymbol{z} \in \mathcal{Z}_f$

(4). $\left[\dot{F} + L\right](\boldsymbol{z}, \kappa(\boldsymbol{z})) \leq 0$ *for* $\forall \boldsymbol{z} \in \mathcal{Z}_f$

To enable application of Theorem 4.1 for stability analysis of the unconstrained idealized MPC strategy based on optimal control problem (4.41), let

$$\boldsymbol{z}(t) = \boldsymbol{z}_\beta(t) \tag{4.42a}$$

$$\boldsymbol{u}(t) = \boldsymbol{u}_\beta(t) \tag{4.42b}$$

$$\dot{\boldsymbol{z}}(t) = g(\boldsymbol{z}, \boldsymbol{u}) = \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z}(\boldsymbol{t}) + \boldsymbol{B}\boldsymbol{u}(t) \tag{4.42c}$$

$$L\big(\boldsymbol{z}(t), \boldsymbol{u}(t)\big) = \boldsymbol{z}(\boldsymbol{t})^T \boldsymbol{Q}_1 \boldsymbol{z}(\boldsymbol{t}) + \boldsymbol{u}(t)^T \boldsymbol{R}_3 \boldsymbol{u}(t) \tag{4.42d}$$

$$F\big(\boldsymbol{z}(t)\big) = \boldsymbol{z}(t)^T \boldsymbol{Q}_2 \boldsymbol{z}(t) \tag{4.42e}$$

$$\dot{F}\big(\boldsymbol{z}(t)\big) = \dot{\boldsymbol{z}}(t)^T \boldsymbol{Q}_2 \boldsymbol{z}(t) + \boldsymbol{z}(t)^T \boldsymbol{Q}_2 \dot{\boldsymbol{z}}(t) \tag{4.42f}$$

This study chooses a linear nominal controller (Camacho and Alba, 2013) as follows

$$\kappa(\boldsymbol{z}) = \boldsymbol{\mathcal{K}}\boldsymbol{z} \tag{4.43}$$

Let $\boldsymbol{\mathcal{K}} = \boldsymbol{0}_{2n\times n}$. This choice of matrix $\boldsymbol{\mathcal{K}}$ will simplify the analysis of conditions for asymptotic stability of the unconstrained idealized MPC strategy based on optimal control problem (41). Next, we illustrate the conditions for which optimal control problem (4.41) can satisfy the four conditions in Theorem 4.1.

For optimal control problem (4.41), the feasible set of state variables, control variables, and terminal state variables are $\mathcal{Z} = \mathbb{R}^{2n}$, $\mathcal{A} = \mathbb{R}^n$, and $\mathcal{Z}_f = \mathbb{R}^{2n}$, respectively. Thereby, $0 \in \mathcal{Z}$; condition 1 is satisfied. According to Eq. (4.43), $\kappa(\boldsymbol{z}) = \boldsymbol{\mathcal{K}}\boldsymbol{z} = \boldsymbol{0}_{1\times n} \in \mathbb{R}^n = \mathcal{A}$. Hence, condition 2 in Theorem 4.1 is also satisfied. From Eq. (4.42c) and Eq. (4.43), $g\big(\boldsymbol{z}, \kappa(\boldsymbol{z})\big) = \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z}(t) + \boldsymbol{B}\boldsymbol{\mathcal{K}}\boldsymbol{z}(t) = \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z}(t) \in \mathbb{R}^{2n} = \mathcal{Z}_f$. Therefore, condition 3 holds for optimal control problem (4.41).

To illustrate that condition 4 is satisfied, for simplicity, the notation for time *t* is removed. Substituting Eqs. (4.42c)-(4.42f) into the inequality in condition 4, we have

$$\begin{aligned}\left[\left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z} + \boldsymbol{B}\boldsymbol{\mathcal{K}}\boldsymbol{z}\right]^T \boldsymbol{Q}_2 \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{Q}_2 \left[\left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)\boldsymbol{z} + \boldsymbol{B}\boldsymbol{\mathcal{K}}\boldsymbol{z}\right] + \boldsymbol{z}^T \boldsymbol{Q}_1 \boldsymbol{z} \\ + (\boldsymbol{\mathcal{K}}\boldsymbol{z})^T \boldsymbol{R}_3 (\boldsymbol{\mathcal{K}}\boldsymbol{z}) \leq 0\end{aligned} \tag{4.44}$$

Note $\boldsymbol{\mathcal{K}} = \boldsymbol{0}_{2n\times n}$; hence, inequality (4.44) can be simplified as

$$\boldsymbol{z}^T \left[\left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)^T \boldsymbol{Q}_2 + \boldsymbol{Q}_2 \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right) + \boldsymbol{Q}_1\right]\boldsymbol{z} \leq 0 \tag{4.45}$$

Let $\boldsymbol{W} = \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right)^T \boldsymbol{Q}_2 + \boldsymbol{Q}_2 \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_{2n}\right) + \boldsymbol{Q}_1$. Obviously, inequality (4.45) holds if

matrix $\boldsymbol{W}$ is negative semidefinite. According to Eq. (4.6), $\boldsymbol{R}_1 = \boldsymbol{\Lambda}^{\boldsymbol{T}}\boldsymbol{D}_a\boldsymbol{\Lambda}$, $\boldsymbol{R}_2 = \boldsymbol{\Lambda}^{\boldsymbol{T}}\boldsymbol{D}_b\boldsymbol{\Lambda}$, $\boldsymbol{R}_4 = \boldsymbol{\Lambda}^{\boldsymbol{T}}\boldsymbol{D}_c\boldsymbol{\Lambda}$, and $\boldsymbol{R}_5 = \boldsymbol{\Lambda}^{\boldsymbol{T}}\boldsymbol{D}_e\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{\Lambda}^T\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T = \boldsymbol{E}_n$. Let the diagonal positive definite matrices $\boldsymbol{D}_a$, $\boldsymbol{D}_b$, $\boldsymbol{D}_c$ and $\boldsymbol{D}_e$ be $\boldsymbol{D}_a = \text{diag}(a_1, \dots, a_n)$, $\boldsymbol{D}_b = \text{diag}(b_1, \dots, b_n)$, $\boldsymbol{D}_c = \text{diag}(c_1, \dots, c_n)$, and $\boldsymbol{D}_e = \text{diag}(e_1, \dots, e_n)$, respectively, where $a_i > 0$, $b_i > 0$, $c_i > 0$, and $e_i > 0$ for $i = 1, \dots, n$. The following proposition discusses the sufficient conditions for matrix $\boldsymbol{W}$ to be negative semidefinite

Proposition 4.3. $\boldsymbol{W}$ ($\boldsymbol{W} \in \mathbb{R}^{2n\times 2n}$) *is a negative semidefinite matrix if matrices* $\boldsymbol{D}_a$, $\boldsymbol{D}_b$, $\boldsymbol{D}_c$ *and* $\boldsymbol{D}_e$, *and the discount parameter* β *are set such that*

$$a_i < \beta c_i, \forall i = 1,2,\cdots,n \tag{4.46a}$$

$$e_i \geq \frac{-c_i^2}{\beta(a_i - \beta c_i)}, \forall i = 1,2,\cdots,n \tag{4.46b}$$

$$b_i \leq \frac{c_i^2 + \beta e_i(a_i - \beta c_i)}{a_i - \beta c_i}, \forall i = 1,2,\cdots,n \tag{4.46c}$$

for $\forall i$.

Proof. Matrix $\boldsymbol{W}$ can be expanded as

$$\begin{aligned}\boldsymbol{W} &= \left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_n\right)^T \boldsymbol{Q}_2 + \boldsymbol{Q}_2\left(\boldsymbol{A} - \frac{\beta}{2}\boldsymbol{E}_n\right) + \boldsymbol{Q}_1 \\ &= \begin{bmatrix} \boldsymbol{0}_n & \boldsymbol{0}_n \\ -\boldsymbol{E}_n & \boldsymbol{0}_n \end{bmatrix}\begin{bmatrix} \boldsymbol{R}_4 & \\ & \boldsymbol{R}_5 \end{bmatrix} + \begin{bmatrix} \boldsymbol{R}_4 & \\ & \boldsymbol{R}_5 \end{bmatrix}\begin{bmatrix} \boldsymbol{0}_n & -\boldsymbol{E}_n \\ \boldsymbol{0}_n & \boldsymbol{0}_n \end{bmatrix} - \begin{bmatrix} \beta\boldsymbol{E}_n\boldsymbol{R}_4 & \\ & \beta\boldsymbol{E}_n\boldsymbol{R}_5 \end{bmatrix} \\ &\quad + \begin{bmatrix} \boldsymbol{R}_1 & \\ & \boldsymbol{R}_2 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{R}_1 - \beta\boldsymbol{R}_4 & -\boldsymbol{R}_4 \\ -\boldsymbol{R}_4 & \boldsymbol{R}_2 - \beta\boldsymbol{R}_5 \end{bmatrix}\end{aligned} \tag{4.47}$$

Denote $\widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Lambda} & \\ & \boldsymbol{\Lambda} \end{bmatrix}$, then

$$\begin{aligned}\widehat{\boldsymbol{W}} = \widetilde{\boldsymbol{\Lambda}}\boldsymbol{W}\widetilde{\boldsymbol{\Lambda}}^T &= \begin{bmatrix} \boldsymbol{\Lambda} & \\ & \boldsymbol{\Lambda} \end{bmatrix}\begin{bmatrix} \boldsymbol{R}_1 - \beta\boldsymbol{R}_4 & -\boldsymbol{R}_4 \\ -\boldsymbol{R}_4 & \boldsymbol{R}_2 - \beta\boldsymbol{R}_5 \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda}^T & \\ & \boldsymbol{\Lambda}^T \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^T\boldsymbol{D}_a\boldsymbol{\Lambda} - \beta\boldsymbol{\Lambda}^T\boldsymbol{D}_c\boldsymbol{\Lambda})\boldsymbol{\Lambda}^T & -\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\boldsymbol{D}_c\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T \\ -\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\boldsymbol{D}_c\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T & \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^T\boldsymbol{D}_b\boldsymbol{\Lambda} - \beta\boldsymbol{\Lambda}^T\boldsymbol{D}_e\boldsymbol{\Lambda})\boldsymbol{\Lambda}^T \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{D}_b - \beta\boldsymbol{D}_c & -\boldsymbol{D}_c \\ -\boldsymbol{D}_c & \boldsymbol{D}_b - \beta\boldsymbol{D}_e \end{bmatrix}\end{aligned} \tag{4.48}$$

According to Eq. (4.47), the eigenvalues of matrix $\widehat{\boldsymbol{W}}$ and $\boldsymbol{W}$ are identical. Let $\check{\boldsymbol{z}}_\beta = \left(x_{1,\beta}, y_{1,\beta}, x_{2,\beta}, y_{2,\beta}, \dots, x_{n,\beta}, y_{n,\beta}\right)^T$; $\check{\boldsymbol{z}}_\beta$ is a vector of variables obtained by changing the order of variables in $\boldsymbol{z}_\beta$. Then,

$$\left(\boldsymbol{z}_\beta\right)^T \widehat{\boldsymbol{W}} \cdot \boldsymbol{z}_\beta = \left(\check{\boldsymbol{z}}_\beta\right)^T \underbrace{\begin{bmatrix} \widetilde{\boldsymbol{W}}_1 & & & \\ & \widetilde{\boldsymbol{W}}_2 & & \\ & & \ddots & \\ & & & \widetilde{\boldsymbol{W}}_n \end{bmatrix}}_{\widetilde{\boldsymbol{W}}} \cdot \check{\boldsymbol{z}}_\beta \tag{4.49}$$

where $\widetilde{\boldsymbol{W}}$ is a block diagonal matrix defined above, in which $\widetilde{\boldsymbol{W}}_i$ $(\forall i = 1,2\cdots,n)$ is

$$\widetilde{\boldsymbol{W}}_i = \begin{bmatrix} a_i - \beta c_i & -c_i \\ -c_i & b_i - \beta e_i \end{bmatrix} \tag{4.50}$$

Note $\widetilde{\boldsymbol{W}}_i$ is a symmetric matrix. It is negative semidefinite if

$$a_i - \beta c_i \leq 0 \tag{4.51a}$$

and

$$(a_i - \beta c_i)(b_i - \beta e_i) - c_i^2 \geq 0 \tag{4.51b}$$

Obviously, inequality (4.51a) holds if $a_i < \beta c_i$. According to Eq. (4.51b), we have

$$(a_i - \beta c_i)(b - \beta e_i) - c_i^2 = (a_i - \beta c_i)b_i - \beta e_i(a_i - \beta c_i) - c_i^2 \geq 0 \tag{4.52}$$

Note $a_i < \beta c_i$, inequality (4.52) implies that

$$b_i \leq \frac{c_i^2 + \beta e_i(a_i - \beta c_i)}{a_i - \beta c_i} \tag{4.53}$$

As $b_i \geq 0$, the right-hand side of inequality (4.53) holds only if

$$c_i^2 + \beta e_i(a_i - \beta c_i) \leq 0 \tag{4.54}$$

This implies

$$c_i^2 + \beta e_i(a_i - \beta c_i) \leq 0 \tag{4.55}$$

Thereby,

$$e_i \geq \frac{-c_i^2}{\beta(a_i - \beta c_i)} \tag{4.56}$$

The above discussion shows that if inequalities (4.53), (4.56), and $a_i < \beta c_i$ hold, $\widetilde{\boldsymbol{W}}_i$ is a negative semidefinite matrix. Similarly, we can infer that the block diagonal matrix $\widetilde{\boldsymbol{W}}$ is negative semidefinite if inequalities (4.46a)-(4.46c) hold. This implies that $\widehat{\boldsymbol{W}}$ is negative semidefinite. Note that matrix $\widehat{\boldsymbol{W}}$ is similar to the symmetric matrix $\boldsymbol{W}$**.** Thereby, $\boldsymbol{W}$ is negative semidefinite if inequalities (4.46a)-(4.46c) hold. Proposition 4.3 is proved. ■

It is worth mentioning that Proposition 4.3 only provides a sufficient condition to ensure the asymptotic stability of the unconstrained idealized MPC strategy. There exist other conditions under which the unconstrained idealized MPC strategy is also asymptotically stable. According to Proposition 4.3, the method to determine the diagonal positive definite matrices $\boldsymbol{D}_a$, $\boldsymbol{D}_b$, $\boldsymbol{D}_c$ and $\boldsymbol{D}_e$ and the discount parameter β to ensure asymptotic stable of the unconstrained idealized MPC strategy can be summarized as follows. First, set an arbitrary positive value for β and a

diagonal positive definite matrix $\boldsymbol{D}_c$**.** Second, obtain the matrix $\boldsymbol{D}_a$ such that inequality (4.56a) is satisfied. Then, obtain matrices $\boldsymbol{D}_e$ and $\boldsymbol{D}_b$ according to inequalities (4.56b) and (4.56c), respectively.

4.6 Numerical experiments

This section discusses four numerical experiments to demonstrate the motivation for this study and to illustrate the effectiveness of the proposed DMPC-FOA approach. The first numerical experiment analyzes the computational time required for the leading vehicle to solve optimal control problem (4.5) for different initial inputs, prediction horizons, and the number of following vehicles. The second numerical experiment illustrates the detailed steps for sensitivity analysis of the optimal control problem. The first-order Taylor approximation method is then applied to estimate the solution of state variables, costate variables, and the optimal control decisions when the leading vehicle's initial speed and position are changed. The estimated solution and the exact solution (computed using the solution algorithm in Section 3) are compared. The third numerical experiment compares the control performance of the DMPC-FOA approach with that of the DMPC approach assuming the movement of the leading vehicle is predetermined according to NGSIM field data. The last numerical experiment shows a traffic flow scenario where the DMPC approach fails to control the CAV platoon safely due to poor estimation of the optimal control decisions of the idealized MPC strategy. However, the DMPC-FOA approach can control the CAV platoon effectively and is able to characterize the optimal control decisions of the idealized MPC strategy accurately in this scenario.

4.6.1 Computational time for solving optimal control problem (4.5)

Table 4.1 Input parameters for optimal control problem (4.5)

Variables	Default value
Minimum acceleration (u_{min})	$-5\ m/s^2$
Maximum acceleration (u_{max})	$3\ m/s^2$
Minimum spacing (s_{min})	$5\ m$
Safety space (s_f)	$10\ m$
Time headway (r^*)	$1\ s$

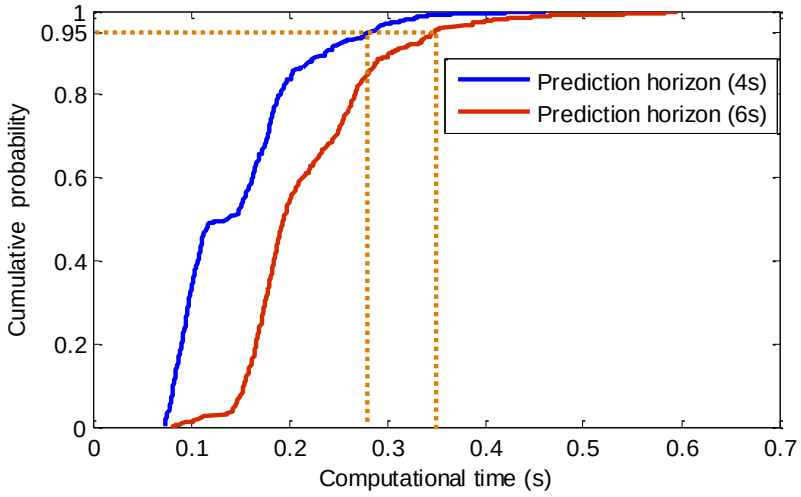


Figure 4.5 Cumulative probability of computational time for solving optimal control problem (4.5) with different initial inputs (i.e., $\mathbf{x}(0)$ and $\mathbf{y}(0)$) at $n = 8$ and $T_P = 4s$ and $6s$.

This section analyzes the computational time for solving optimal control problem (4.5). The computational time is important to determine the reserved times for computing in the DMPC and DMPC-FOA approaches (i.e., τ_1 and τ_2, respectively). From Section 3, the solutions for the state variables and control decisions for optimal control problem (4.5) can be obtained by solving the two-point boundary value problem (32) using the shooting method. Table 1 shows the detailed inputs of the parameters in the optimal control problem (4.5). These inputs are used for all four numerical experiments. The discount parameter β and the matrices $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, $\boldsymbol{R}_4$, and $\boldsymbol{R}_5$ in optimal control problem (4.5) are set as follows: $\beta = 1$, $\boldsymbol{R}_1 = 0.5\boldsymbol{E}_n$, $\boldsymbol{R}_2 = \boldsymbol{R}_3 = \boldsymbol{E}_n$, $\boldsymbol{R}_5 = 3\boldsymbol{E}_n$. These inputs satisfy the inequalities in Proposition 4.3 to ensure that the unconstrained idealized MPC strategy is asymptotic stable.

Without loss of generality, suppose the initial time is 0. To ensure that optimal control problem (4.5) can be solved within τ_1 seconds under different initial inputs of position errors (i.e., $\mathbf{x}(0)$) and speed differences (i.e., $\mathbf{y}(0)$) of all adjacent vehicles pairs, $\mathbf{x}(0)$ is generated randomly in the interval $[-10,100]$ and $\mathbf{y}(0)$ is randomly generated in the interval $[0,20]$. This study generates 1000 different values for $\mathbf{x}(0)$ and $\mathbf{y}(0)$ for which the inequality constraints (Eq. (4.5c)) are satisfied.

The numerical experiments were coded in MATLAB and executed on a computer with an Intel Core i7-4790 3.60-GHz CPU with 8.00 GB of RAM. To analyze the impacts of the number of following vehicles in the platoon (n) and the prediction horizon (T_P) on computational time,

optimal control problem (4.5) is solved 1000 times under different feasible initial inputs for each combination of n and T_P.

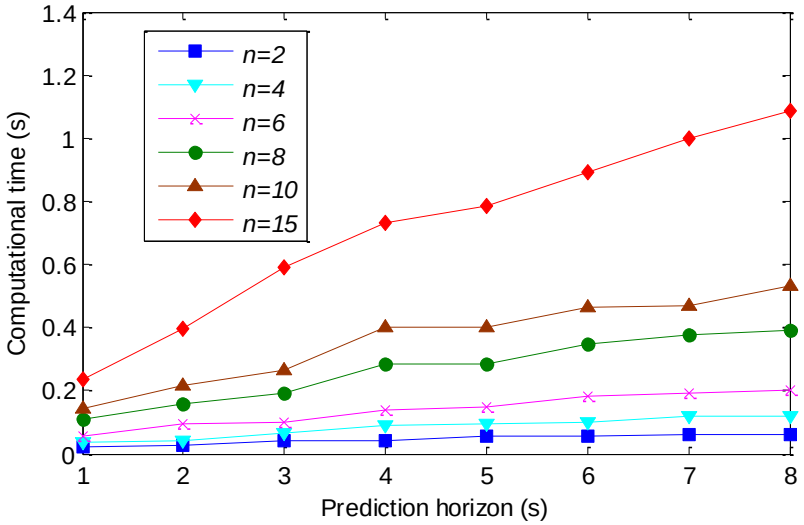
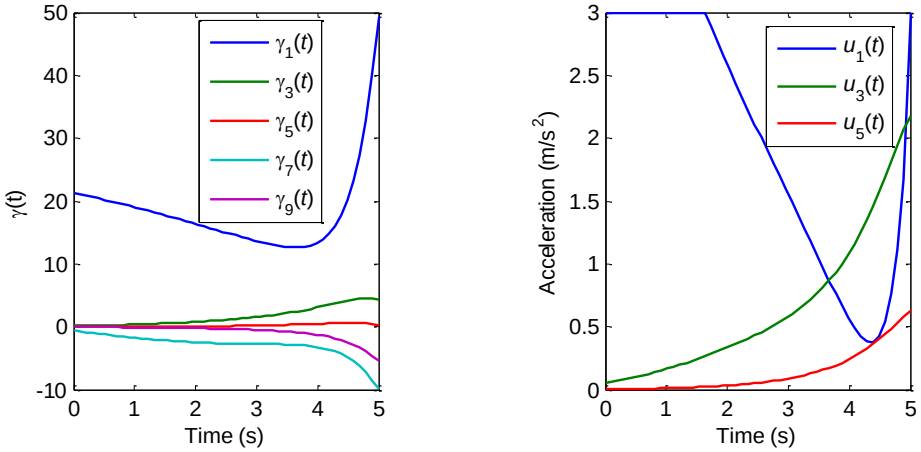


Figure 4.6 Computational time corresponding to 0.95 cumulative probability under different n and T_P.

Figure 4.5 shows the cumulative probability of computational time for solving the optimal control problem (4.5) with different initial inputs (i.e., $\mathbf{x}(0)$ and $\mathbf{y}(0)$) for $n = 8$ and $T_P = 4s$ and $6s$. It shows that the computational time significantly depends on the value of $\mathbf{x}(0)$ and $\mathbf{y}(0)$. The computational time ranges from 0.08s to 0.4s under $n = 8$ and $T_P = 4s$. It is worth noting that computational times are large only when the initial position errors of many adjacent vehicle pairs deviate remarkably from the equilibrium state (i.e., they are close to 100 m), the likelihood of occurrence of which is low in the real world. Hence, this study uses the computational time corresponding to 0.95 cumulative probability as the reference point to determine the reserved time for the DMPC and DMPC-FOA approaches.

Figure 4.6 shows the computational time corresponding to 0.95 cumulative probability under different n and T_P. The computational time corresponding to 0.95 cumulative probability is the time within which 95% of the experimental scenarios can be solved. Figure 4.6 illustrates that the computational time corresponding to 0.95 cumulative probability increases monotonically with the number of following vehicles and the prediction horizon.

4.6.2 Sensitivity analysis of optimal control problem (4.5)



(a) Solutions of costate variables (b) Optimal control decisions.

Figure 4.7 Solutions of costate variables and optimal control decisions at the unperturbed initial state

This section shows the details of the sensitivity analysis method implementation for the optimal control problem (4.5) introduced in section 4. Consider a CAV platoon with 5 following vehicles ($n = 5$). The leading vehicle and all following vehicles drive at a speed of 20 m/s at time 0 (i.e., $\mathbf{y}(0) = 0$). Suppose the initial position errors of vehicle 2 to vehicle 5 are all 0, and the initial position error of vehicle 1 with respect to the leading vehicle is 90 m. This implies that the spacing between vehicle 1 and vehicle 0 is $90 + T \cdot 20 + s_f = 120m$. It indicates a case where the following vehicles seek to catch up with the leading vehicle. Let $T_P = 5\ s$. Figure 4.7(a) shows the optimal solutions of the costate variables obtained using the solution algorithm proposed in Section 3. The optimal control decisions of all following vehicles in the platoon can then be determined according to Eq. (4.23). Figure 4.7(b) shows the optimal control decisions of vehicles 1, 3 and 5. It indicates that vehicle 1 accelerates at the maximum value ($3\ m/s^2$) for the first 1.7 seconds. Then, the acceleration decreases monotonically in the time interval $[1.7\text{s}, 4.3\text{s}]$ and then increases.

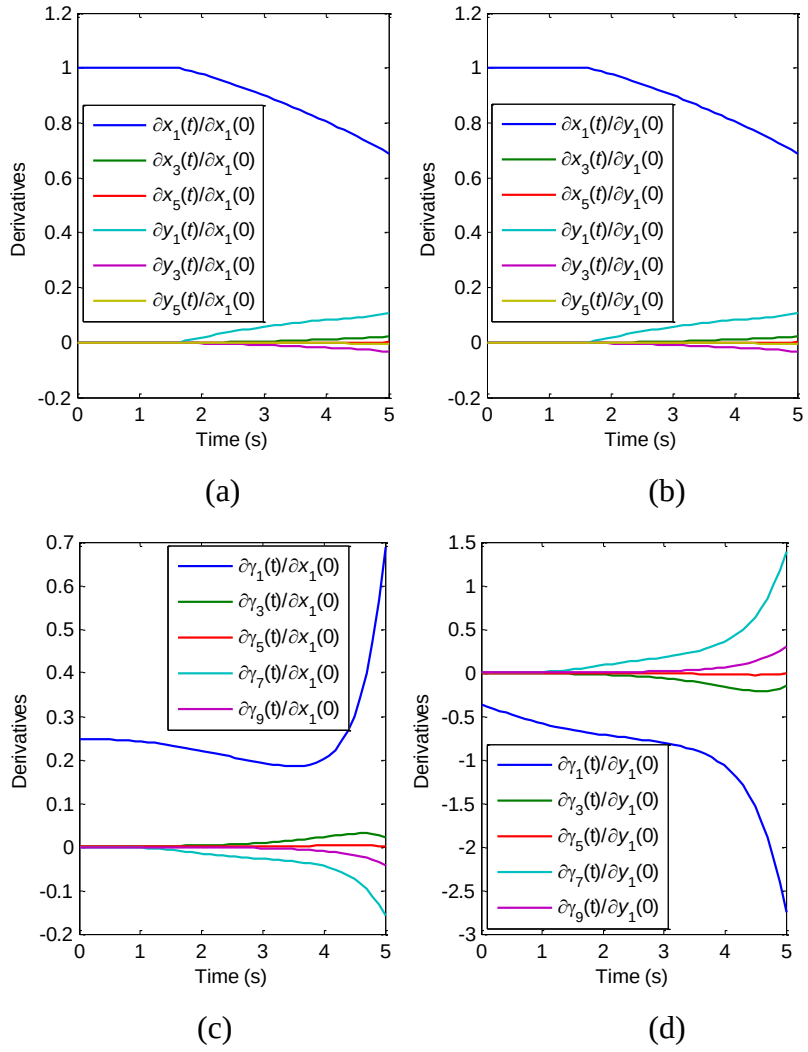
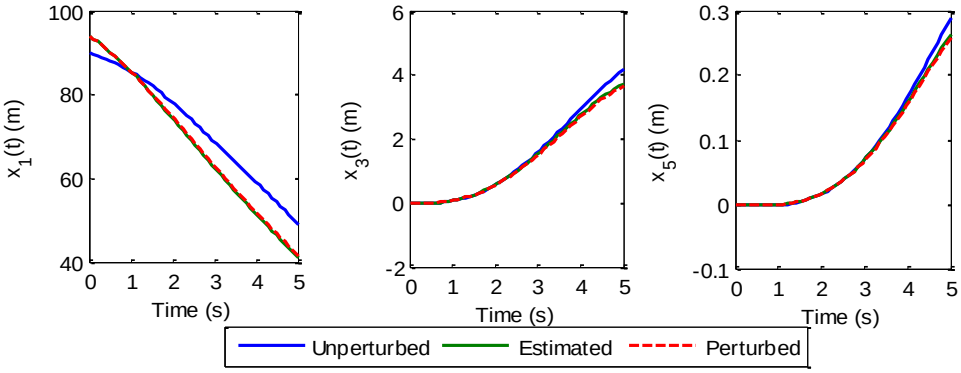
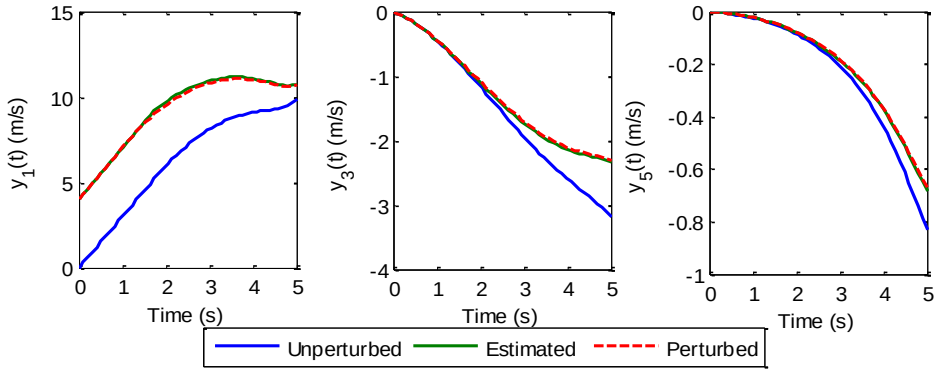


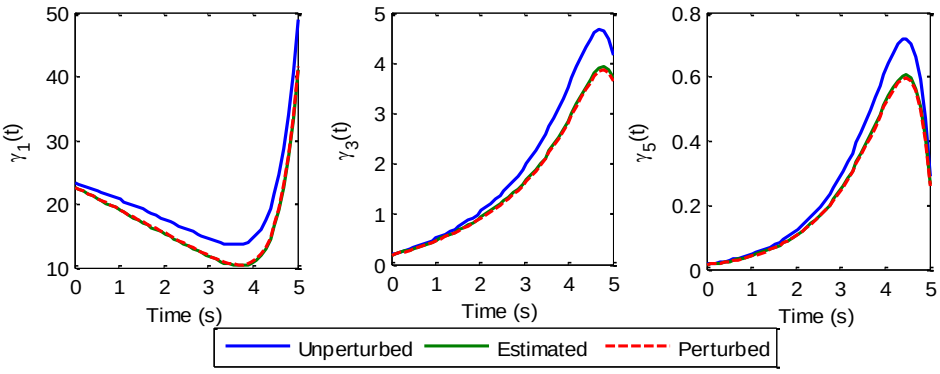
Figure 4.8 Derivatives of the state and costate variables with respect to $x_1(0)$ and $y_1(0)$, respectively, at the unperturbed initial state



(a) Comparison of estimated and perturbed optimal solutions of position errors



(b) Comparison of estimated and perturbed optimal solutions of speed difference for adjacent vehicle pairs



(c) Comparison of estimated and perturbed optimal solutions for the costate variables.

Figure 4.9 Comparison of estimated and perturbed optimal solutions for the state and costate variables.

Suppose the initial position and speed of the leading vehicle at time 0 are perturbed. Then, $x_1(0)$ and $y_1(0)$ change from the unperturbed values 90 and 0, respectively. Figure 4.8 shows the derivatives of solutions for the state and costate variables with respect to $x_1(0)$ and $y_1(0)$,

respectively. They are obtained by solving the two-point boundary value problem (4.35) and (4.37), respectively. Figure 4.8 shows that at the optimal state, a unit change in $x_1(0)$ and $y_1(0)$ will increase the optimal solution of $x_1(t)$ and $y_1(t)$ by 1, respectively, at time interval [0,1.7]. The impacts of variations in $x_1(0)$ and $y_1(0)$ on $x_1(t)$ and $y_1(t)$ decrease after 1.7 seconds.

Suppose both $x_1(0)$ and $y_1(0)$ are increased by 4 units (for example, due to prediction error). Using the first-order Taylor approximation (Eq. (4.38)), Figure 4.9 compares the estimated and perturbed optimal solutions for the state variables and costate variables. The perturbed solutions are obtained using the solution algorithm at the perturbed states of $x_1(0)$ and $y_1(0)$. Figure 4.9 shows that the estimated solutions are very close to those of the perturbed solutions, indicating that the first-order Taylor approximation can accurately characterize the variation in the optimal solutions induced by changes in $x_1(0)$ and $y_1(0)$. Based on the estimated solutions for the costate variables (i.e., $\boldsymbol{\gamma}$), Figure 4.10 compares the optimal control decisions of following vehicles estimated by Eq. (4.39) and the perturbed ones obtained using the solution algorithm in Section 3. It shows that the estimated solutions are also very close to the perturbed ones obtained using the solution algorithm.

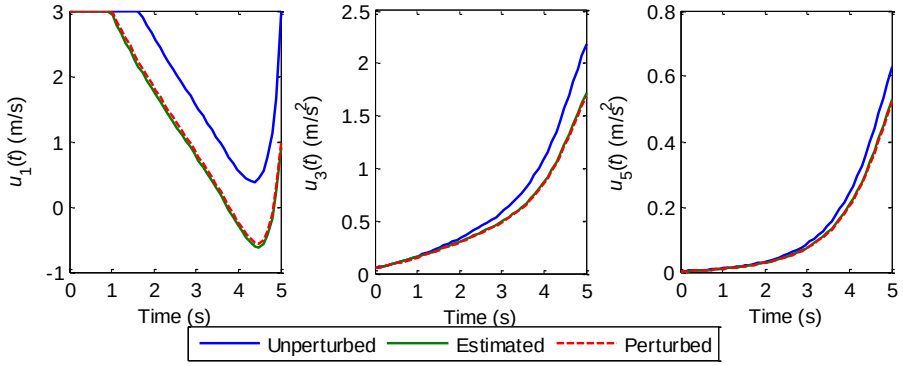


Figure 4.10 Comparison of estimated and perturbed optimal control decisions of the following vehicles.

4.6.3 Control performance of the DMPC and DMPC-FOA approaches

Note that both the DMPC and DMPC-FOA approaches seek to address the issue of control delay and estimate the optimal control decisions of the idealized MPC strategy. This section compares the control decisions of the DMPC approach, the DMPC-FOA approach and the idealized MPC strategy. To do so, we consider a CAV platoon with 8 following vehicles (vehicle IDs 1-8). The acceleration of the leading vehicle is shown in Figure 4.11. It contains a 240-

seconds (with resolution 0.1 second) real-world vehicle control diary collected on eastbound I-80 in the San Francisco Bay area at Emeryville, California. It can be noted that the vehicle decelerated or accelerated mildly most of the time. However, it contains some time slots with hard braking and high acceleration (e.g., the time slots around 110s, 140s and 186s).

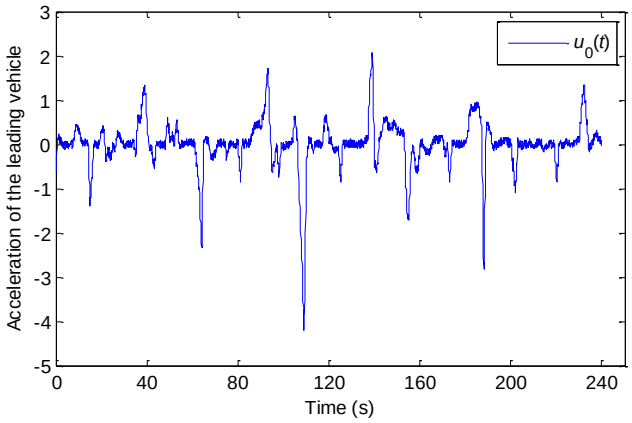
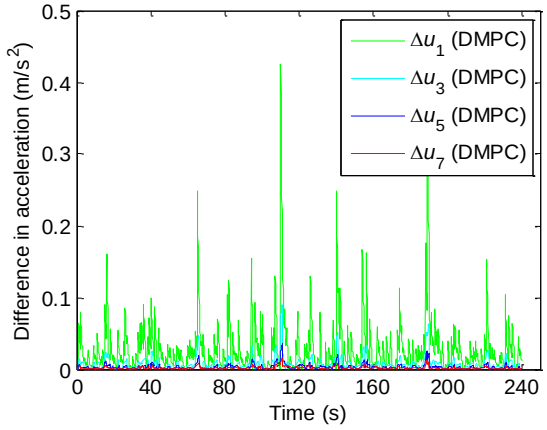


Figure 4.11 Acceleration of the leading vehicle

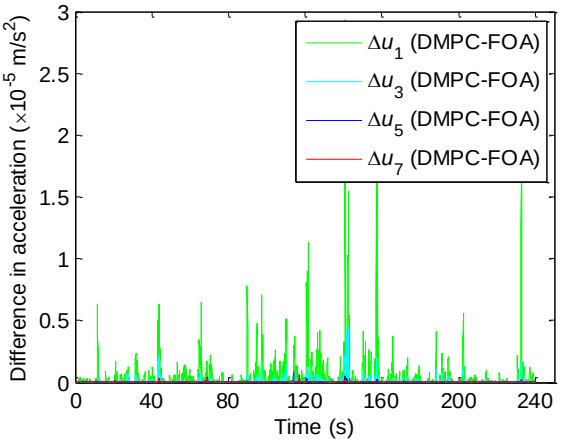
Suppose the prediction horizon and the roll period are $T_P = 5$ seconds and $\Delta t = 1$ second, respectively. According to Figure 4.6, the computational time for solving optimal control problem (4.5) corresponding to 95% cumulative probability with 8 following vehicles is 0.33 seconds. To reserve enough time for solving the optimal control problem, τ_1 is set as 0.4 seconds for the DMPC approach. Note that the DMPC-FOA approach needs to solve optimal control problem (4.5) as well as perform sensitivity analysis of the optimal control problem with respect to $\tilde{x}_1(0)$ and $\tilde{y}_1(0)$. Thereby, $\tau_2 \geq \tau_1$. From 1000 simulations, the total computational time for solving the optimal control problem (4.5) and the two-point boundary value problems (problems (35) and (37)) corresponding to 95% cumulative probability is around 0.56 seconds. Thereby, τ_2 is set as 0.6 seconds. It should be noted that among the 1000 simulations, there are situations where some following vehicles need to brake and accelerate at the maximum rate during the prediction horizon. Thereby, $\tau_2 = 0.6s > \tau_1 = 0.4s$. According to Proposition 4.2, if these situations do not exist and the spacing of each following vehicle is always greater than the minimum value (s_{min}), τ_2 can be set the same as τ_1.

Figure 4.12 shows the difference between the estimated control decisions of the DMPC approach (i.e, $\Delta u_i = \hat{u}_i^*(t) - u_i^*(t), \forall i = 1,2,\cdots n$) and the DMPC-FOA approach (i.e, $\Delta u_i = \bar{u}_i^*(t) - u_i^*(t), \forall i = 1,2,\cdots n$) from those of the idealized MPC strategy. Figure 4.12(a) shows

that the estimated control decisions of the DMPC approach are close to those of the idealized MPC strategy with the maximum difference less than 0.45 m/s^2. The estimation errors of the control decisions of DMPC approach are induced by the prediction error of $x_1(t)$ and $y_1(t)$ at each sampling time instant. However, through first-order Taylor's approximation, the DMPC-FOA approach can significantly improve on the estimation performance of the DMPC approach. As can be seen from Figure 4.12(b), the maximum difference between the control decisions estimated by the DMPC-FOA approach and the idealized MPC strategy is less than $3 \times 10^{-5}\, m/s^2$, indicating that the DMPC-FOA approach can characterize the decisions of the idealized MPC strategy very well.

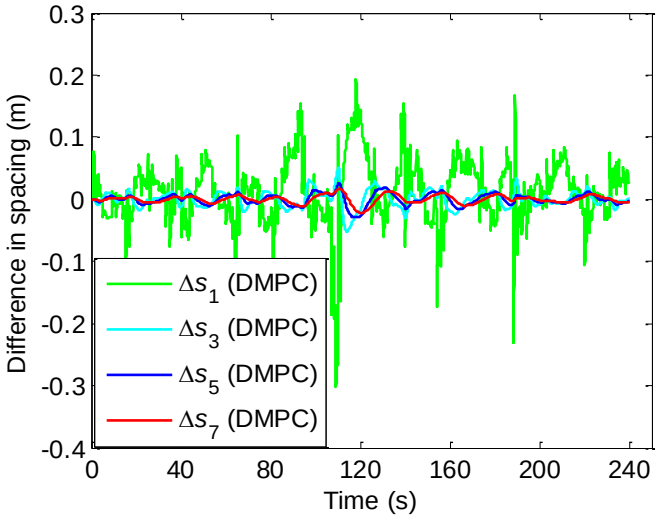


(a) Difference between control decisions of the DMPC approach and those of the idealized MPC strategy

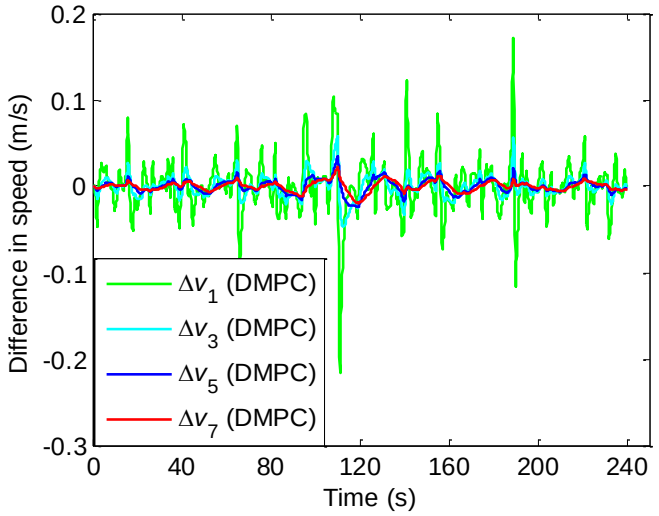


(b) Difference between control decisions of the DMPC-FOA approach and those of the idealized MPC strategy.

Figure 4.12 Differences between the estimated control decisions of the DMPC and DMPC-FOA approaches from those of the idealized MPC strategy.



(a) Difference in optimal spacing



(b) Difference in optimal speed

Figure 4.13 Differences in optimal spacing and speed between the DMPC approach and the idealized MPC strategy.

Figure 4.13 illustrates the differences in optimal spacing and speed between the DMPC approach and the idealized MPC (i.e., Δs_i and Δv_i, respectively, $i = 1,2,\cdots n$). It shows that while the estimated control decisions of DMPC approach deviate from the idealized MPC strategy, the optimal spacing and speed obtained by the DMPC approach are very close to those of the idealized MPC strategy. Hence, the DMPC approach is able to control the CAV platoon efficiently in this case. To investigate the reason for the good control performance of the DMPC approach in this scenario, Figure 4.14 shows the prediction errors of the initial inputs of $x_1(t)$

and $y_1(t)$ at each sampling time instant $t_k, k = 1,2,\cdots$. Recall $\Delta t = 1s$. Hence, $t_k = 1s, 2s, \cdots, 240s$. It shows that the predicted values of $x_1(t_k)$ and $y_1(t_k), k = 1,2,\cdots$ are very close to those of the exact ones as the leading vehicle drives with mild acceleration or deceleration most of the time (see Figure 4.11). The large prediction error occurs at the moments when the leading vehicle has hard acceleration or deceleration (e.g., $t = 110s, 140s, 186s$ etc.). Correspondingly, the DMPC approach also has larger estimation errors in terms of the optimal solutions relative to those of the idealized MPC strategy (see Figure 4.12(a) and Figure 4.13). However, as these "extreme" behaviors of the leading vehicle only last for small time periods, their impacts are small. In addition, if $x_1(t_k)$ and $y_1(t_k)$ are accurately predicted at a time instant t_k, the large difference in optimal solutions between the DMPC approach and the idealized MPC strategy in the previous roll period will be reduced significantly at the current roll period starting from time instant t_k. This can be observed in Figure 4.12(a) and Figure 4.13 where the large differences at time instants $t = 110s, 140s, 186s$ are reduced dramatically in the roll periods following time instants at which $x_1(t)$ and $y_1(t)$ are predicted with low errors at the corresponding sampling time instants (i.e., $t_k = 111s, 141s, 187s$, see Figure 4.14).

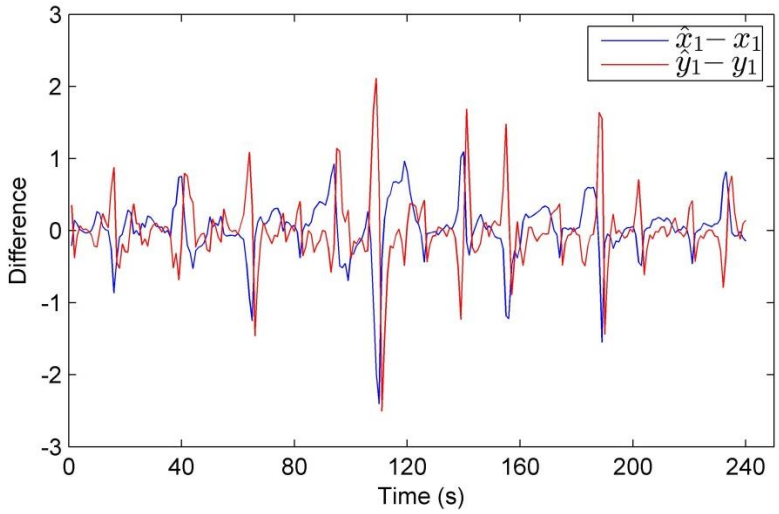


Figure 4.14 Prediction errors of the initial states of $x_1(t_k)$ and $y_1(t_k)$, $t_k = 1s, 2s, \cdots, 240s$.

Figure 4.15 shows the control decisions of the following vehicles estimated by the DMPC-FOA approach. It indicates that when the leading vehicle 0 executes hard acceleration/deceleration, vehicle 1 also executes hard acceleration/deceleration with a magnitude slightly less than that of the leading vehicle 0. The acceleration or deceleration

decreases sequentially in the platoon, indicating that the traffic oscillation is damped sequentially from the head of the platoon to its tail. Figure 4.16 shows the optimal spacing and speed differences of adjacent vehicle pairs in the platoon computed by the DMPC-FOA approach. These results are almost identical to those of the idealized MPC strategy with the maximum absolute error less than 8×10^{-8} due to the high accuracy of the estimated optimal control decisions (see Figure 4.12(b)). As can be seen in Figure 4.16, the oscillation of the optimal spacing and speed difference of adjacent vehicle pairs decreases sequentially in the platoon. These results indicate that the DMPC-FOA approach can lead to smooth deceleration and acceleration behavior of all following vehicles. In addition, it can coordinate the behavior of all following vehicles to dissipate the traffic oscillation to ensure stability of the CAV platoon.

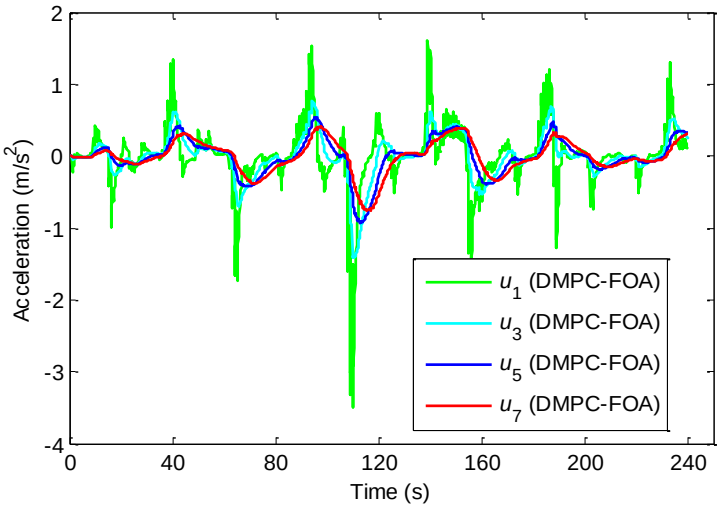
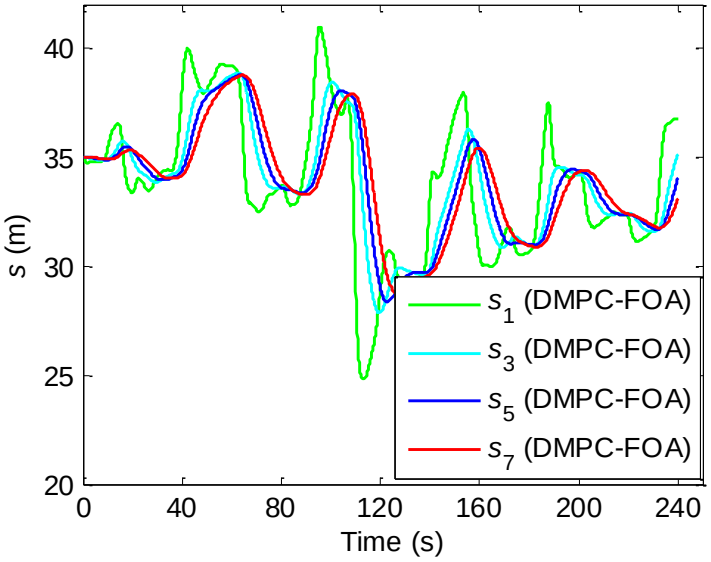
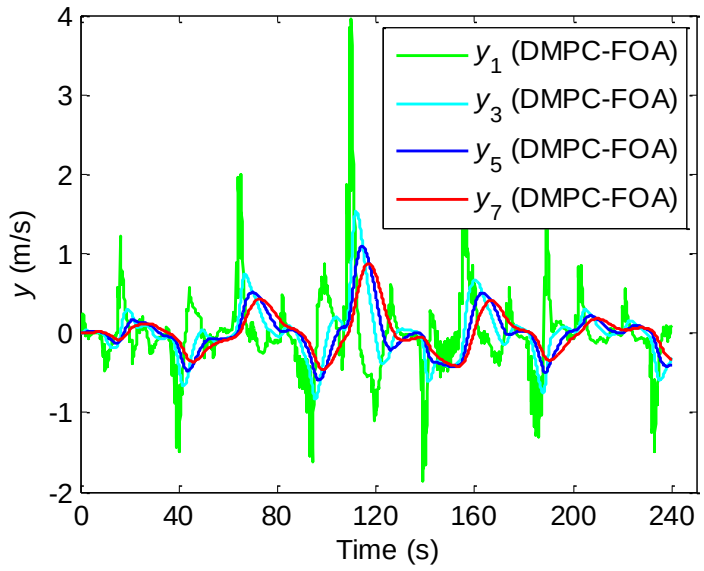


Figure 4.15 Estimated control decisions of the DMPC-FOA approach.



(a) Spacing of adjacent vehicle pairs



(b) Speed difference of adjacent vehicle pairs

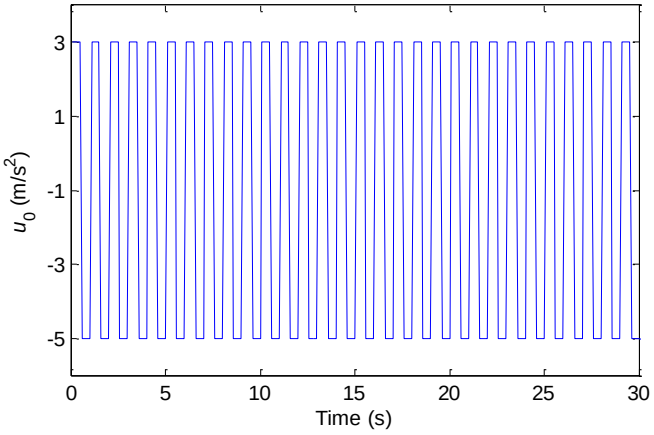
Figure 4.16 Optimal spacing and speed difference for some adjacent vehicle pairs in the platoon computed by DMPC-FOA approach.

4.6.4 Scenario where the DMPC approach fails to control the CAV platoon

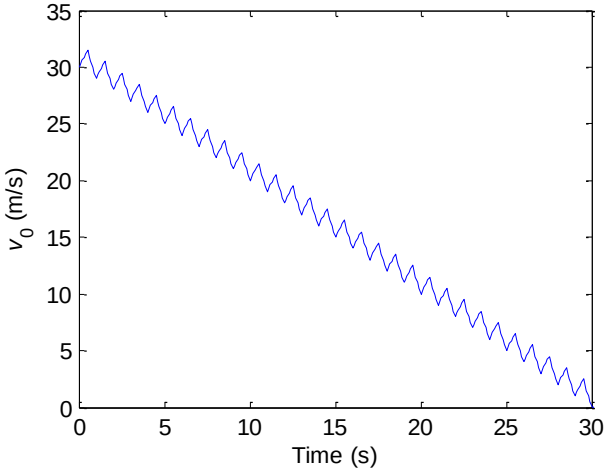
The previous section illustrated a scenario in which the estimated control decisions and the solutions for the state variables of the DMPC approach are very close to those of the idealized MPC strategy. Here, we illustrate a scenario in which when the DMPC approach fails to accurately predict the values of $x_1(t_k)$ and $y_1(t_k)$ at each sampling time instant t_k, the error of the control decisions between the DMPC approach and idealized MPC strategy increases with

each roll period. Then, the car-following behavior of the vehicles controlled by the DMPC approach significantly deviates from that of the idealized MPC strategy. However, as will be illustrated, the DMPC-FOA approach accurately characterizes the optimal control decisions of the idealized MPC strategy.

Consider a CAV platoon with 10 following vehicles. Let $T_P = 5$ seconds and $\Delta t = 1$ second. According to Figure 4.6, the computational time corresponding to 95% cumulative probability is 0.42 seconds. Hence, we set $\tau_1 = 0.5$ seconds for the DMPC approach. By conducting 1000 simulation runs with different initial inputs for $x_1(0)$ and $y_1(0)$, the computational time corresponding to 95% cumulative probability for the DMPC-FOA approach is determined as 0.66 seconds. We will set $\tau_2 = 0.7$ seconds for the DMPC-FOA approach.



(a) Acceleration of the leading vehicle

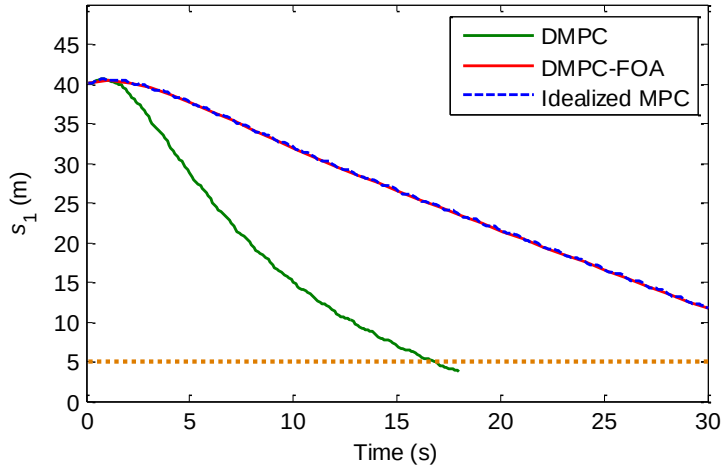


(b) speed of the leading vehicle.

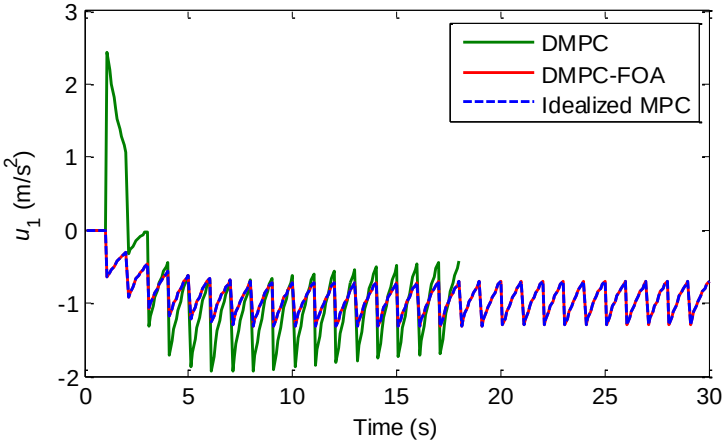
Figure 4.17 Acceleration and speed of the leading vehicle.

Suppose the leading vehicle drives at 30 m/s at time 0. Assume the leading vehicle accelerates at the maximum value 3 m/s^2 for 0.5 seconds and then decelerates at the maximum value $-5\ m/s^2$ for 0.5 seconds. Such behavior will repeat for 30 seconds until the leading vehicle stops. Figure 4.17(a) shows the trajectory of the assumed acceleration of the leading vehicle. The corresponding speed of the leading vehicle is shown in Figure 4.17(b).

As $\Delta t = 1$ second, the sampling time instant $t_k = k$ seconds for $k = 1,2,\cdots$. Under the assumed acceleration behavior of the leading vehicle, the prediction errors of $x_1(t_k)$ and $y_1(t_k)$ using the DMPC approach are $-1\ m$ and $4\ m/s$, respectively, at each sampling time instant t_k. Note that the prediction errors of $x_1(t_k)$ and $y_1(t_k)$ $(k = 1,2,3\cdots)$ for DMPC-FOA are the same as that of DMPC approach.



(a) Comparison of solution for spacing of vehicle 1



(b) Comparison of control decisions for vehicle 1

Figure 4.18 Comparison of solutions for spacing and control decisions of vehicle 1 among the DMPC approach, the DMPC-FOA approach and the idealized MPC strategy.

Figure 4.18 compares the optimal solutions for the DMPC approach, the DMPC-FOA approach and the idealized MPC strategy. It illustrates that both spacing and control decisions of vehicle 1 computed using the DMPC approach deviate significantly from those of the idealized MPC strategy due to the large prediction errors of $x_1(t_k)$ and $y_1(t_k)$ $(k = 1,2,\cdots)$. In addition, the spacing between the leading vehicle 0 and vehicle 1 even reduce to a value less than the minimum allowable spacing $s_{min}(s_{min} = 5\ m)$. Thereby, a collision will occur between leading vehicle 0 and vehicle 1 in the platoon. Note that the DMPC approach stops at $t = 18s$ as the safety constraints (inequality (4.5c)) cannot be satisfied thereafter. Hence, no solution can be found using the DMPC approach. By contrast, the DMPC-FOA approach provides an optimal solution very close to that of the idealized MPC strategy. When the leading vehicle stops at $t = 30s$, the spacing between leading vehicle 0 and vehicle 1 is over $10\ m$ to ensure safety. These results highlight that the DMPC-FOA approach can effectively improve the estimation performance significantly beyond that of the DMPC approach even under extreme scenarios.

4.7 Concluding comments

This study first proposes an idealized MPC-based cooperative control strategy for CAV platooning. Its optimal control decisions can coordinate the behaviors of all following CAVs in the platoon to maneuver them effectively and safely. However, as in existing literature, it is based on the idealized, but unrealistic, assumption that the embedded optimal control problem can be solved instantaneously. To relax this idealized assumption, two deployable strategies, i.e., the DMPC approach and the DMPC-FOA approach, are proposed to address the control delay issue of the idealized MPC strategy and to accurately characterize its optimal control decisions. The DMPC approach addresses the control delay issue by reserving sufficient time before each sampling time instant to solve the embedded optimal control problem. However, the estimated control decisions of the DMPC approach can deviate significantly from those of the idealized MPC strategy due to errors in predicting the leading vehicle's position and speed. By contrast, the DMPC-FOA approach addresses the control delay issue effectively while accurately characterizing the optimal control decisions of the idealized MPC strategy by leveraging the proposed analytical sensitivity analysis method for the embedded optimal control problem. The application of the DMPC-FOA approach for a CAV platoon whose lead vehicle's trajectory is obtained from field data illustrates that it can dampen traffic oscillations efficiently, and can

enable smooth deceleration and acceleration behaviors for all following vehicles. In addition, it can provide control decisions very similar to those of the idealized MPC strategy even under extreme situations where the leading vehicle's speed and position are predicted very poorly at each sampling time instant.

It is important to note that the DMPC-FOA approach concept can also be leveraged to address the issue of control delay for other MPC-based cooperative control strategies (e.g., Wang et al., 2014b) arising from the computational time required to solve the embedded optimal control problem. In addition, the applicability of the DMPC-FOA approach is not constrained by the size of the platoon or prediction horizon. It can be applied for real-time control of large CAV platoons on the condition that the time reserved for computing (i.e., τ_2) is less than the roll period (Δt).

This study can be extended in a few directions. First, the study assumes that the DMPC and DMPC-FOA approaches can estimate the states of all following vehicles at each sampling time instant with low error (negligible) by leveraging the prior control decisions and the actual states at a time close to the corresponding sampling time instant. This assumption is reasonable when pavement conditions are homogeneous during the platooning and the vehicle does not falsely execute the optimal control decisions. If pavement conditions change significantly over the platooning process (for example, due to potholes or weather conditions), or the well-known false execution issue arises occasionally, the control decisions can be suboptimal even for the idealized benchmark MPC strategy. To address these aspects, our ongoing work proposes to capture them using another real-time deployable approach. Second, the application of the DMPC-FOA approach for real-time control of the CAV platoon can be constrained by the reserved time τ_2. We propose to develop a new solution algorithm and discretization technique to solve the two-point boundary value problems faster.

CHAPTER 5. MULTICLASS TRAFFIC ASSIGNMENT MODEL FOR MIXED TRAFFIC FLOW OF HUMAN-DRIVEN VEHICLES AND CONNECTED AND AUTONOMOUS VEHICLES

5.1 Introduction

Communication and automation technologies installed in connected and autonomous vehicles (CAVs) will significantly transform people's travel behavior in the future, mainly due to the following advantages. First, CAVs can enhance travel safety, and reduce traffic accidents caused by human errors (Assidiq et al., 2008). Second, CAVs can enhance roadway capacity by following each other closely. Studies suggest that highway capacity can be doubled if 60% of the vehicles are CAVs, and increased by 4 to 5 times if all vehicles are CAVs (Tientrakool et al., 2011). Third, CAVs require fewer inputs from human beings. Hence, CAV users may have lower value of time (VOT) because they can spend the time during travel on other activities (van den Berg and Verhoef, 2016). Fourth, CAVs can improve energy efficiency by forming platoons. Studies show that CAVs can save 7%-15% of fuel consumption for light-duty CAVs and 15%-21% of fuel consumption for heavy truck CAVs by forming a CAV platoon (Shida et al., 2010; Shida and Nemoto 2009; Bonnet and Fritz, 2000; Eben et al., 2013). CAVs will also lead to other benefits such as fewer parking spaces, reduced vehicle ownership, and reclamation of more green space. An overview of some of the advantages can be found in Fagnant and Kockelman (2015).

The aforementioned advantages can lead CAV users to have different route choice behaviors compared to human-driven vehicle (HDV) users. Here, HDVs are assumed to be not equipped with vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication technologies that can enable seamless connectivity. The differential behaviors result in a mixed traffic flow pattern at the network level with the following three characteristics. First, CAV flows can impact the route choices of HDV users. The travel time of mixed traffic flow is different from that of the single-class HDV flow due to the asymmetry in interactions involving HDVs and CAVs. For example, CAVs may follow HDVs using smaller headways, similar to how they would follow another CAV. By contrast, the driving behavior of HDVs can be different depending on whether they are following an HDV or a CAV. Hence, HDV drivers may perceive different travel times based on the CAV market penetration in the ambient traffic stream. Second, travel cost in this

mixed flow context includes the monetized travel time based on VOT, and the fuel consumption cost. Since VOT can be different for CAV and HDV users, the monetized travel time can also be different for them. Also, since route choice behaviors can be different for CAV and HDV users, fuel consumption costs can be different for them. Further, as CAVs can receive more accurate information on traffic conditions through seamless V2V and V2I connectivity, CAV users are more informed about shortest travel cost paths (which are not necessarily the shortest travel time paths). Third, unlike HDVs, CAVs can have access to dedicated CAV facilities such as autonomous vehicle (AV) dedicated lanes. For example, some previous studies (Chen et al., 2016; Chen et al., 2017) propose the deployment of AV-dedicated lanes to foster CAV usage during the transition period from HDVs to CAVs to better exploit the advantages of CAVs. If AV-dedicated lanes exist in a traffic network, CAVs can use both AV-dedicated lanes and non-AV dedicated lanes (i.e., lanes can be accessed by both CAVs and HDVs) while HDVs can only use the non-AV dedicated lanes. Hence, CAVs can have a larger route choice set. Levin and Boyles (2016) show that even if a small proportion of HDVs are replaced by CAVs, the redistributed network flows can reduce average travel time significantly. In summary, it is critical to estimate the network flows for effective transportation planning during the transition period when both CAVs and HDVs are in use.

To model the network flow under HDVs and CAVs, this study characterizes them as a mixed traffic flow. In transportation literature, such network flows are often estimated by formulating multiclass traffic assignment models, by extending the single-class traffic assignment model (such as the static user equilibrium (UE) model). These extensions can improve modeling realism by integrating the route choices of the users of different user classes using different disutility functions (Dafermos, 1972). Hence, static multiclass traffic assignment has been used to capture the heterogeneity in travel mode (Defermos, 1972; Florian, 1977; Chen et al., 2016; Levin and Boyles, 2015; Jiang et al., 2016), VOT (Yang and Huang, 1994; Nagurney and Dong, 2002), knowledge level of network conditions (Huang and Li, 2007; Huang and Lam, 2004), and risk-taking behavior (Shao et al., 2006; Lo et al., 2006).

Existing multiclass traffic assignment models are insufficient to estimate mixed traffic flows consisting of CAVs and HDVs for the following reasons. First, these models cannot accurately characterize the route choice behavior of HDV users. In particular, most multiclass traffic assignment studies adopt the perfect knowledge assumption and the static user equilibrium

principle for route choices of HDV users (Chen et al., 2016; Levin and Boyles, 2015; Dafermos, 1972; Florian, 1977; Yang and Huang, 1994; Jiang et al., 2016). Some other studies (e.g., Huang and Lam, 2003; Shao et al., 2006; Huang and Li, 2007) adopt the logit-based stochastic user equilibrium (SUE) model to relax the perfect knowledge assumption of UE model for HDV users. Nonetheless, the SUE model suffers from the route overlap issue whereby the flows on routes with overlapping links are overestimated. This may deteriorate the estimation accuracy of mixed traffic flows significantly, raising issues of realism. Since the estimation of HDV flows significantly impacts the estimation of CAV flows, there is the need to more accurately characterize HDVs' route choices. Second, existing multiclass models ignore the asymmetry in driving interactions involving HDVs and CAVs, resulting in inaccurate travel time estimates which could affect flow allocations to routes. CAVs have lower reaction times and can respond to perturbations more quickly than HDVs. Further, CAVs can follow the vehicle in front of them more closely than HDVs, leading to increased link capacity. Thereby, the impact of one unit increase in CAV flow is different from that of a unit of HDV flow. Third, as discussed earlier, the route choice criteria of CAV users differ significantly from those of HDV users due to the reduced VOT, improved fuel economy, and the dedicated AV lanes. These new dimensions of routing criteria have not been considered in the literature, precluding holistic analyses of mixed traffic flows and the corresponding network equilibrium.

To more realistically characterize mixed traffic flows of CAVs and HDVs to support effective transportation planning in the emerging future, this study proposes a variational inequality-based multiclass traffic assignment model in which CAV users choose routes based on the UE principle and HDV users based on the cross nested-logit (CNL) model. The UE model characterizes the CAVs' capability to acquire accurate information on traffic conditions. The CNL model relaxes the strong knowledge-level assumption of the UE model and also overcomes the route overlap issue of logit-based SUE problem (Kitthamkesorn et al., 2016; Prashker and Bekhor, 1999). Remming (2001) illustrates the estimation effectiveness of the CNL model for HDV flows using the eastern Massachusetts network. In our proposed model, the link cost functions integrate the difference in VOT and the asymmetry in interactions involving HDVs and CAVs, and the energy consumption savings through platooning in AV-dedicated lanes. Hence, the proposed multiclass traffic assignment model can enhance realism in characterizing mixed traffic flows.

Due to the complex analytical formulation of the CNL route choice model for HDVs, it is difficult to solve the proposed multiclass traffic assignment model efficiently using variational inequality-based algorithms such as projection-based algorithms (e.g., Nagurney, 2000; Nagurney and Dong, 2002; Jiang et al., 2016) or proximal point methods (e.g., Zhan and Ukkusuri, 2017). To address this problem, this study develops a new solution algorithm by integrating the self-regulated step size choice technique (Liu et al., 2009) into the concept of route swapping strategy (Smith, 1984). One advantage of the proposed algorithm is that, in each iteration, it finds the descent direction according to an analytical model to circumvent solving a computationally-intensive subproblem in the aforementioned algorithms. Another advantage is that the proposed algorithm updates the step size adaptively using information on the descent directions of the current iteration and the previous iteration. We provide rigorous analyses of the convergence of the solution algorithm. The proposed algorithm also resolves the convergence difficulty of the existing route-swapping-based algorithm (Huang and Lam, 2002).

To enhance applicability, this study further derives the analytical formula for the sensitivity analysis of the proposed model. In practice, planned (e.g., road maintenance, construction) or unexpected (e.g., accidents, facility failure, natural disasters) events can impact network flows significantly. Sensitivity analysis can help transportation decision-makers to estimate the impact of these events on network flows, and design better response and recovery strategies to reduce their negative effects. Besides, sensitivity analysis of the traffic assignment models can also be leveraged to design effective solution algorithms for continuous network design problems. There is limited literature for sensitivity analysis for multiclass traffic assignment models although extensive studies are available for single-class traffic assignment models (Yang and Bell, 2005; Tobin and Friesz, 1988; Clark and Watling, 2000). In particular, this study derives sufficient conditions for the existence of derivatives for HDV and CAV equilibrium link flows.

The contributions of the study are threefold. First, a multiclass traffic assignment model is proposed where the routing behaviors of HDVs and CAVs follow the CNL and UE principles, respectively. The proposed model provides enhanced behavioral realism by integrating into the travel cost functions the impacts of heterogeneous VOT and asymmetry in interactions involving HDVs and CAVs. Thereby, this model enables planners to better estimate network flows to support effective transportation planning during the transition to a fully CAV future. Second, this study develops a new solution algorithm based on the route-swapping concept to solve the

proposed multiclass traffic assignment model efficiently. Comparison shows that the proposed algorithm has a convergence speed superior to that of the existing route-swapping-based algorithm (Huang and Lam, 2002) by adaptively updating the step size in each iteration using a modified self-regulated step-size choice technique. Third, an analytical formula is derived for the sensitivity analysis of the proposed model, which enables planners to quickly estimate the perturbed traffic equilibrium and identify critical elements under planned or unexpected disruptive events. It can also be used to solve the continuous network design problem (e.g., to find optimal signal timing or tolling strategy to improve the system performance under mixed traffic flow).

The remainder of the paper is organized as follows. Section 2 introduces the CNL model and an equivalent variational inequality (VI) problem to characterize the equilibrium state. Section 3 presents the proposed multiclass traffic assignment model where HDVs and CAVs choose routes according to the CNL model and UE principle, respectively. Section 4 develops the solution algorithm for the proposed model, and Section 5 provides the sensitivity analysis. Section 6 discusses the results of numerical experiments for the proposed model and the performance of its solution algorithm. Section 7 concludes with the main findings, insights, and potential future research directions.

5.2 Cross-nested logit model and its equivalent VI problem

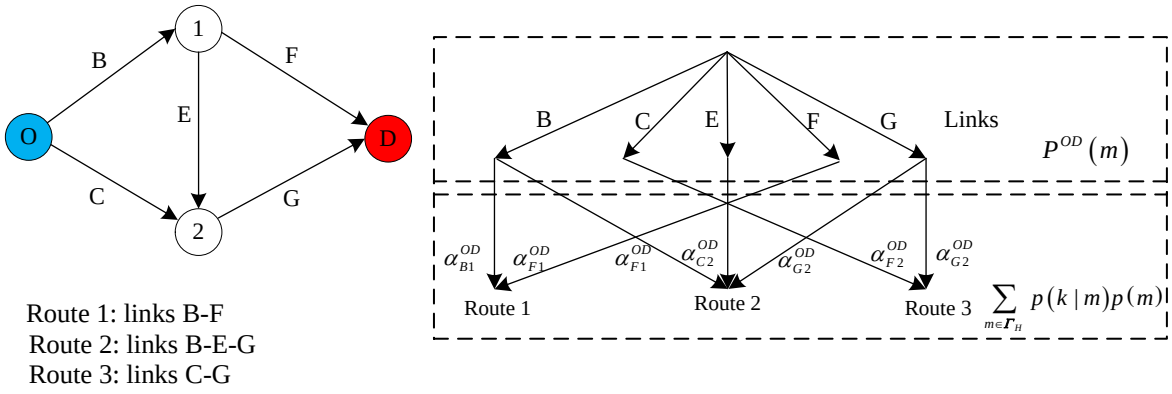


Figure 5.1 Illustration of the hierarchical structure of the CNL model

This section develops a VI-based traffic assignment model for HDV flows that assumes HDV drivers choose routes according to the CNL principle. In this section, all vehicles are assumed to be HDVs. To differentiate from the CAV flows in the multiclass traffic assignment model, this

section contains a subscript H indicating HDVs. Consider a transportation network $\mathrm{G}(\boldsymbol{N}_H, \boldsymbol{\Gamma}_H)$, where $\boldsymbol{N}_H$ represents the set of nodes and $\boldsymbol{\Gamma}_H$ represents the set of links that can be accessed by HDVs. Let W_H denote the set of origin-destination (OD) pairs for HDVs, and R_H^w denote the set of paths connecting OD pair $w \in W_H$. Denote $v_{a,H}$ and $t_{a,H}$ as the HDV flow and HDV travel cost on link $a \in \boldsymbol{\Gamma}_H$, respectively. Let $\mathbf{t}_H$ be the vector of HDV link travel costs and $\mathbf{v}_H$ be the vector of HDV link flows. The HDV flow on path k of OD pair w is denoted by $f_{k,H}^w$, $k \in R_H^w$. Let $\mathbf{f}_H$ be the vector of HDV flows of all paths in the network. Denote q_H^w as the demand of HDVs for OD pair w, and $\mathbf{q}_H$ as the vector of HDV traffic demand for all OD pairs. Let Δ_H and Λ_H denote the link-path and OD-path incidence matrices, respectively. Additional notation will be introduced when necessary.

Vovsha (1997) derived the CNL model based on McFadden's (1981) generalized extreme value (GEV) function for mode choice split. Prashker and Bekhor (1999) extended it to characterize user's route choices. Figure 5.1 shows a small example to demonstrate how the CNL model overcomes the route overlap problem. It contains one OD pair and three routes. Route 2 overlaps with route 1 and route 3 through shared links B and G, respectively. As shown in Figure 5.1, the CNL model consists of two layers, where the upper layer contains all links in the network and the lower layer consists of all potential path choices. Each path in the CNL model is assigned to the upper nests (links) which are used by this path, and each nest in the upper layer groups all paths sharing the specific overlapped link. The CNL model introduces an inclusion coefficient ($\alpha_{m,k}^w, m \in \boldsymbol{\Gamma}_H$) for each path $k \in R_H^w$ and link m to denote the overlapping degree of this path with other paths in nest $m \in \boldsymbol{\Gamma}_H$, $\sum_{k \in R_H^w} \alpha_{m,k}^w = 1$ (Prashker and Bekhor, 1999). Thereby, the CNL model accounts for the covariance between paths under each nest that addresses the route overlap problem. According to the CNL model, the probability of a user choosing path k of OD pair w (labelled as $p^w(k)$) can be written as the product of the marginal probability $p^w(m)$ and conditional probability $p^w(k|m)$:

$$p^w(k) = \sum_{m \in \boldsymbol{\Gamma}_H} p^w(k|m) p^w(m) \tag{5.1a}$$

where

$$p^w(k|m) = \frac{\left[\alpha_{m,k}^w \exp(-\theta c_{k,H}^w)\right]^{1/u}}{\sum_{l \in R_H^w} \left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}} \tag{5.1b}$$

$$p^w(m) = \frac{\left(\sum_{k\in R_H^w}\left[\alpha_{m,k}^w \exp(-\theta c_{k,H}^w)\right]^{1/u}\right)^u}{\sum_{b\in \boldsymbol{\Gamma}_H}\left(\sum_{l\in R_H^w}\left[\alpha_{b,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^u} \quad (5.1c)$$

Here, u is the degree of nesting, $0 < u \leq 1$, θ is the dispersion parameter, $and\ c_{k,H}^w$ is the travel cost of path k of OD pair $w \in W_H$. Cascetta et al. (1996) provide the following formulation for the inclusion coefficient ($\alpha_{m,k}^w, m \in \boldsymbol{\Gamma}_H$):

$$\alpha_{a,k}^w = \left(\frac{l_m}{l_k^w}\right)^\gamma \delta_{a,k}^w \quad (5.2)$$

where l_m and l_k^w are the length of link m and path k of OD pair w, respectively, and $\delta_{a,k}^w = 1$ if path k uses link a and 0 otherwise. Eq. (5.1) denotes that, if the degree of nesting is 0, then the CNL model becomes the logit model.

Prashker and Bekhor (1999) developed a CNL equivalent mathematical program to characterize users' route choices. However, it cannot be extended to model the route choices of mixed traffic flow due to the asymmetric impacts of HDVs and CAVs on link travel cost. Further, beyond path flows, the decision variables in the mathematical program are path flows belonging to different nests. Hence, this program contains a large number of decision variables, which scales the computational complexity for a large-size network. To model multiclass traffic assignment with mixed flow of CAVs and HDVs, this section develops an equivalent VI problem which contains only path flow variables.

In a large network, some paths are less likely to be used. Such paths are excluded from analysis in this study. These paths can be found using methods such as the revised K-shortest path algorithm (De La Barra et al. 1993), the labeling method (Ben-Akiva et al. 1984), and the column generation technique (Ji et al. 2017). Suppose R_H^w contains only paths that are likely to be used. . To characterize the equilibrium condition of the CNL model, let $C_{k,H}^w$ be the generalized travel cost of path k for HDVs for OD pair w, formulated as:

$$C_{k,H}^w = c_{k,H}^w - \frac{u}{\theta} H_{k,H}^w + \frac{u}{\theta} ln\left(\frac{f_{k,H}^w}{q_H^w}\right) \quad (5.3)$$

where

$$H_{k,H}^w = ln\left[\sum_{m\in \boldsymbol{\Gamma}_H}\left(\alpha_{m,k}^w\right)^{1/u}\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^{u-1}\right] \quad (5.4)$$

The following proposition provides the sufficient and necessary conditions for the equilibrium state of the CNL model. This proposition will be used to develop the equivalent VI-based traffic assignment model which will be embedded in the multiclass traffic assignment model in Section 3.

Proposition 5.1: The CNL path flows $\mathbf{f}_H^* = \{f_{k,H}^{w*}, \forall w \in W_H, \forall k \in R_H^w\}$ are at equilibrium if and only if the generalized travel cost of all paths for the corresponding OD pair is the same, i.e., $\mathbf{f}_H^*$ satisfies

$$C_{k,H}^w(\mathbf{f}_H^*) = \tau_H^w, \text{ for } \forall w \in W_H, \forall k \in R_H^w \tag{5.5}$$

where $\mathbf{f}_H^* \in \boldsymbol{\Omega}_{\mathbf{f}_H} = \{\mathbf{f} | \Lambda_H \mathbf{f}_H = \mathbf{q}_H, \mathbf{f}_H \geq 0\}$ and τ_H^w is a real value.

Proof: Based on Eq. (5.3), we have

$$p^w(k) = \frac{f_{k,H}^{w*}}{q_H^w} = \exp\left(\frac{\theta}{u} C_{k,H}^w - \frac{\theta}{u} c_{k,H}^w + H_{k,H}^w\right) \tag{5.6}$$

Note

$$\sum_{j \in R_H^w} p^w(j) = 1 \tag{5.7}$$

Substituting Eq. (5.6) into Eq. (5.7) yields

$$\sum_{j \in R_H^w} p^w(j) = \sum_{j \in R_H^w} \exp\left(\frac{\theta}{u} C_{k,H}^w - \frac{\theta}{u} c_{k,H}^w + H_{j,H}^w\right) = 1 \tag{5.8}$$

Based on Eq. (5.5),

$$\exp\left(\frac{\theta}{u} C_{k,H}^w\right) = \exp\left(\frac{\theta}{u} \tau_H^w\right) = \frac{1}{\sum_{j \in R_H^w} \exp\left(-\frac{\theta}{u} c_{j,H}^w + H_{j,H}^w\right)} \tag{5.9}$$

Substituting Eq. (5.9) into Eq. (5.6), we have

$$\begin{aligned} p^w(k) &= \frac{\exp\left(-\frac{\theta}{u} c_{k,H}^w + H_{k,H}^w\right)}{\sum_{j \in R_H^w} \exp\left(-\frac{\theta}{u} c_{j,H}^w + H_{j,H}^w\right)} \\ &= \frac{\exp\left(-\frac{\theta}{u} c_{k,H}^w + ln\left[\sum_{m \in \Gamma_H} \left(\alpha_{m,k}^w\right)^{1/u} \left(\sum_{l \in R_H^w} \left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^{u-1}\right]\right)}{\sum_{j \in R_H^w} \exp\left(-\frac{\theta}{u} c_{j,H}^w + ln\left[\sum_{b \in \Gamma_H} \left(\alpha_{b,j}^w\right)^{1/u} \left(\sum_{i \in R_H^w} \left[\alpha_{b,i}^w \exp(-\theta c_{i,H}^w)\right]^{1/u}\right)^{u-1}\right]\right)} \\ &= \frac{\exp\left(-\frac{\theta}{u} c_{k,H}^w + ln\left[\sum_{m \in \Gamma_H} \left(\alpha_{m,k}^w\right)^{1/u} \left(\sum_{l \in R_H^w} \left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^{u-1}\right]\right)}{\sum_{j \in R_H^w} \left(\sum_{b \in \Gamma_H} \left(\alpha_{b,j}^w \exp(-\theta c_{j,H}^w)\right)^{1/u} \left(\sum_{i \in R_H^w} \left[\alpha_{b,i}^w \exp(-\theta c_{i,H}^w)\right]^{1/u}\right)^{u-1}\right)} \end{aligned} \tag{5.10}$$

$$= \frac{\sum_{m\in\boldsymbol{\Gamma}_H}\left\{\left[\alpha_{m,k}^{w}\exp(-\theta c_{k,H}^{w})\right]^{1/u}\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^{w}\exp(-\theta c_{l,H}^{w})\right]^{1/u}\right)^{u-1}\right\}}{\sum_{b\in\boldsymbol{\Gamma}_H}\left(\sum_{i\in R_H^w}\left[\alpha_{b,i}^{w}\exp(-\theta c_{i,H}^{w})\right]^{1/u}\right)^{u}}$$

Eq. (5.10) is consistent with the CNL route choice model shown in Eq. (5.1). This completes the proof.

According to Proposition 5.1, the CNL equilibrium state for HDVs can be expressed as follows. At equilibrium, no HDV user can unilaterally change his/her path to reduce the generalized travel cost. Eq. (5.3) also shows that if $u = 1$, $C_{k,H}^{w} = c_{k,H}^{w} + \frac{1}{\theta} ln\left(\frac{f_{k,H}^{w}}{q_H^w}\right)$, which is the generalized travel cost for logit-based SUE model (Guo and Huang, 2016). Thereby, the logit-based SUE condition is a special case of CNL equilibrium. Following Wei et al. (2015), the conditions in proposition 5.1 can further be expressed by a finite-dimensional VI problem as follows:

$$\sum_{w\in W_H}\sum_{k\in R_H^w} C_{k,H}^{w}\left(f_{k,H}^{w} - f_{k,H}^{w*}\right) \geq 0 \tag{5.11}$$

where $\mathbf{f}_H^*$, $\mathbf{f}_H \in \boldsymbol{\Omega}_{\mathbf{f}_H}$.

The CNL equilibrium path flow solution can be obtained by solving the VI problem (5.11). Note that the path flow solution of the CNL model is unique (Prashker and Bekhor, 1999) under the assumption that the link travel cost function is strictly monotonic on traffic flows. Thereby, the path flow solution to the VI problem (5.11) is also unique under the same assumption.

5.3 Multiclass traffic assignment model for mixed traffic flow with HDVs and CAVs

5.3.1 Link travel cost function of HDVs and CAVs

The transition period towards full autonomy will consist of mixed traffic of HDVs and CAVs. Let $\boldsymbol{\Gamma}_1$ be the set of AV-dedicated lanes in the network if such lanes are deployed. To analyze the attractiveness of AV-dedicated lanes and the impact of VOT, we assume that the link travel cost of both CAVs and HDVs consists of two parts: travel time and fuel consumption costs. Both travel time and fuel consumption are converted into equivalent monetary costs. This assumption is also used by Levin and Boyles (2015) to study the effects of AV ownership on trip, mode, and route choice. To determine the link travel times of CAVs and HDVs in the mixed flow, the Bureau of Public Roads (BPR) function is used. It is a strictly monotone function of traffic flows.

$$\bar{t}_a(v_{a,H}, v_{a,A}) = \frac{l_a}{s_a}\left[1 + \left(\frac{v_{a,H} + v_{a,A}}{Q_a}\right)\right]^4, a \in \boldsymbol{\Gamma}_H \tag{5.12a}$$

$$\bar{t}_a(v_{a,A}) = \frac{l_a}{s_a}\left[1 + \left(\frac{v_{a,A}}{Q_{a,A}}\right)\right]^4, a \in \boldsymbol{\Gamma}_1 \tag{5.12b}$$

where $\bar{t}_a(v_{a,H}, v_{a,A})$ is the link travel time of either a CAV or a HDV, and $\boldsymbol{\Gamma}_H$ is the set of regular links (i.e., links not dedicated to AVs) which can be used by both HDVs and CAVs. s_a is the free-flow speed of link a, Q_a is the capacity of link a with mixed traffic flow, $Q_{a,A}$ is the capacity of a link with pure CAV flow, and $v_{a,A}$ is the CAV flow on link a.

As discussed in Section 1, CAVs and HDVs have different driving behavior in mixed traffic. In this study, we assume that CAVs can follow leading vehicles (either a CAV or HDV) with smaller headways than HDVs due to the reduced reaction time. Hence, the capacity of a regular link a (i.e., Q_a) is a function of the proportion of CAVs as the time headway of CAVs is smaller (Levin and Boyles, 2016). According to Levin and Boyles (2016), when the distribution of CAVs is uniform along the link, the capacity of a regular link with mixed traffic flow can be formulated as

$$Q_a = s_a \frac{1}{s_a\left(\frac{v_{a,HAV}}{v_{a,HAV} + v_{a,CAV}}\omega_H + \frac{v_{a,CAV}}{v_{a,HAV} + v_{a,CAV}}\omega_A\right) + \varsigma}, \tag{5.13}$$

where ω_H and ω_A ($\omega_A < \omega_H$) are reaction times of HDVs and CAVs, respectively, and ς is the average vehicle length. According to Eq. (5.13), if all vehicles in a regular link are HDVs, then

$$Q_{a,H} = s_a \frac{1}{s_a\omega_H + \varsigma}, \tag{5.14}$$

where $Q_{a,H}$ denotes the link capacity for pure HDV flow. According to Eq. (5.14),

$$\omega_H = \frac{1}{Q_{a,H}} - \frac{\varsigma}{s_a}, \tag{5.15}$$

Similarly,

$$\omega_A = \frac{1}{Q_{a,A}} - \frac{\varsigma}{s_a}, \tag{5.16}$$

Substitute Eq. (5.15) and Eq. (5.16) into Eq. (5.13),

$$Q_a = \frac{1}{\frac{v_{a,H}}{v_{a,H} + v_{a,A}}\frac{1}{Q_{a,H}} + \frac{v_{a,A}}{v_{a,H} + v_{a,A}}\frac{1}{Q_{a,A}}} \tag{5.17}$$

Note $Q_{a,A} \geq Q_{a,H}$ as the reaction times of CAVs are no larger than those of HDVs. From Eq. (12a) and Eq. (5.17), the increase in link travel time through a unit increase in CAV flow is no larger than that of a unit increase in HDV flow as link capacity can be improved due to increased proportion of CAVs. Thereby, the marginal effects of CAV and HDV flows on link travel time are asymmetric.

To quantify the benefits of energy savings from a CAV platoon on an AV-dedicated lane, the fuel consumption of a vehicle on a link will be estimated and incorporated into the link cost function. We assume all vehicles (either HDVs or CAVs) are light-duty gasoline vehicles. Based on field experiments, Zhang et al. (2014) show that the gasoline consumption rate (per vehicle-mile) of a vehicle is an exponential function of the average traffic speed. Thereby, the total fuel consumption of a vehicle driving on a regular link $a \in \boldsymbol{\Gamma}_H$ can be estimated as

$$E_a = \vartheta_1 \left(\frac{L_a}{\bar{t}_a(v_{a,H}, v_{a,A})} \right)^{-\vartheta_2} l_a, \tag{5.18}$$

where l_a is the length of link a, $l_a/\bar{t}_a(v_{a,H}, v_{a,A})$ is the average travel speed of a vehicle on regular link a, and $\vartheta_1 > 0$ and $0<\vartheta_2 < 1$ are positive coefficients that need to be estimated.

Compared to a regular link, AV-dedicated lanes not only reduce CAV travel times, but also enhance their fuel efficiency by allowing them to form platoons to minimize air resistance. To account for the reduced energy consumption due to platooning, a discount factor σ will be used with E_a to account for energy savings for a CAV on an AV-dedicated lane. Thereby, the link travel cost of HDVs and CAVs can be formulated as follows:

$$t_{a,H} = t_{a,A} = \bar{t}_a(v_{a,H}, v_{a,A}) \cdot VOT_H + \eta \cdot E_a, a \in \boldsymbol{\Gamma}_H \tag{5.19a}$$

$$t_{a,A} = \bar{t}_a(v_{a,A}) \cdot VOT_H + \sigma \cdot \eta \cdot E_a, a \in \boldsymbol{\Gamma}_1 \tag{5.19b}$$

where VOT_H and VOT_A are the value of time for HDV users and CAV users, respectively. η is the price per unit of gasoline, and σ is the percentage fuel consumption savings due to platooning. Note that the link travel cost functions defined by Eq. (5.19a) and Eq. (5.19b) are strictly monotonic with respect to link flow of either HDVs or CAVs.

5.3.2 Multiclass traffic assignment model

Let W_A be the set of OD pairs and R_A^w be the set of routes connecting OD pair $w \in W_A$ for CAVs. Denote $c_{k,A}^w$ and $C_{k,A}^w$ as the travel cost of path k and generalized travel cost of path k for CAVs for OD pair w, respectively. The vector of CAV travel costs and CAV flows of all links and are

denoted by $\mathbf{t}_A$ and $\mathbf{v}_A$, respectively. Denote $f_{k,A}^w$ as the CAV flow of path k for OD pair w and $\mathbf{f}_A$ as the vector of CAV flows of all paths. Let q_A^w be the CAV demand for OD pair w, $\mathbf{q}_A$ be the vector of CAV demands of all OD pairs, and Δ_A and Λ_A be the link-path and OD-path matrices, respectively.

Unlike HDVs, CAVs can obtain information on traffic conditions through vehicle-to-infrastructure and vehicle-to-vehicle communications. Thereby, we assume that CAVs have perfect knowledge of traffic conditions and can always choose paths with the minimum travel cost for the corresponding OD pairs, implying a static UE at equilibrium. In this context, $C_{k,A}^w = c_{k,A}^w, \forall k \in R_A^w, \forall w \in W_A$. Suppose $\mathbf{f}_A^* = \left[f_{k,A}^{w*}, \forall k \in R_A^w, \forall w \in W_A\right]$ is the UE path flow solution for CAVs; then, the generalized path travel cost of CAVs must satisfy:

$$C_{k,A}^w = c_{k,A}^w = \begin{cases} \tau_A^w, \text{if } f_{k,A}^{w*} > 0 \\ \geq \tau_A^w, \text{if } f_{k,A}^{w*} = 0 \end{cases}, \forall k \in R_A^w, \forall w \in W_A, \tag{5.20}$$

where $\mathbf{\Omega}_{\mathbf{f}_A} = \{\mathbf{f}_A | \Lambda_A \mathbf{f}_A = \mathbf{q}_A, \mathbf{f}_A \geq 0\}$, and τ_A^w is the cost of the shortest routes for CAVs. Note that for notational convenience, the generalized route travel cost $C_{k,A}^w$ will be used hereafter to model the multiclass traffic assignment model instead of $c_{k,A}^w$ as they are identical for CAVs.

According to equilibrium conditions (5.5) and (5.20), it can be shown that the path flows $(\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*})$ are the equilibrium state of the CNL model and UE model for HDVs and CAVs, respectively, if and only if they satisfy the following VI problem:

$$\sum_{w \in W_H} \sum_{k \in R_H^w} C_{k,H}^w(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(f_{k,H}^w - f_{k,H}^{w*}\right) + \sum_{w \in W_A} \sum_{k \in R_A^w} C_{k,A}^w(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(f_{k,A}^w - f_{k,A}^{w*}\right) \geq 0, \tag{5.21}$$

where $[\mathbf{f}_H^T, \mathbf{f}_A^T], [\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*}] \in \mathbf{\Omega}_{\mathbf{f}} = \{[\mathbf{f}_H^T, \mathbf{f}_A^T] | \Lambda_H \mathbf{f}_H = \mathbf{q}_H;\ \Lambda_A \mathbf{f}_A = \mathbf{q}_A; \mathbf{f}_H \geq 0; \mathbf{f}_A \geq 0\}$. The superscript ^{T} denotes transpose. The VI problem (5.21) is a multiclass traffic assignment model that characterizes the equilibrium state of the route flow of HDVs and CAVs. The equivalence between VI problem (5.21) and the two equilibrium conditions in Eq. (5.5) and Eq. (5.20) can be shown using the method in Nagurney (2000). It is omitted here to avoid duplication. Let $\mathbf{C}_H$ and $\mathbf{C}_A$ be the vector of generalized costs of all paths for HDVs and CAVs, respectively. Denote $\mathbf{C}_H^*$ and $\mathbf{C}_A^*$ as the generalized costs of all paths for HDVs and CAVs at the equilibrium state $[\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*}]$, respectively. The VI problem (5.21) can be written in vector form:

$$\mathbf{C}_H^*(\mathbf{f}_H - \mathbf{f}_H^*)^T + \mathbf{C}_A^*(\mathbf{f}_A - \mathbf{f}_A^*)^T \geq 0, \tag{5.22}$$

where $[\mathbf{f}_H^T, \mathbf{f}_A^T] \in \mathbf{\Omega}_{\mathbf{f}}$. Note $\sum_{w \in W_A} \sum_{k \in R_A^w} C_{k,A}^w(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(f_{k,A}^w - f_{k,A}^{w*}\right) = \sum_{a \in \Gamma_A} t_{a,A}(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(v_{a,A}^w - v_{a,A}^{w*}\right)$, where $\boldsymbol{\Gamma}_A$ is the set of links for CAVs, $v_{a,A}^{w*}$ is the equilibrium flow of CAVs on link a at

the equilibrium state $(\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*})$. The VI problem (5.21) can be reformulated as follows by defining it on path flows of HDVs and link flows of CAVs:

$$\sum_{w\in W_H}\sum_{k\in R_H^w} C_{k,H}^w(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(f_{k,H}^w - f_{k,H}^{w*}\right) + \sum_{a\in \Gamma_A} t_{a,A}(\mathbf{f}_H^*, \mathbf{f}_A^*)\left(v_{a,A}^w - v_{a,A}^{w*}\right) \geq 0, \quad (5.23)$$

where $[\mathbf{f}_H^T, \mathbf{v}_A^T] \in \boldsymbol{\Omega}_{\mathbf{v}} = \{[\mathbf{f}_H^T, \mathbf{v}_A^T] | \Lambda_H \mathbf{f}_H = \mathbf{q}_H;\ \Lambda_A \mathbf{f}_A = \mathbf{q}_A; \Delta_A \mathbf{f}_A = \mathbf{v}_A; \mathbf{f}_H \geq 0; \mathbf{f}_A \geq 0\}$.

Equivalently, the VI problem (5.23) can be written as

$$\mathbf{C}_H^*(\mathbf{f}_H - \mathbf{f}_H^*)^T + \mathbf{t}_A^*(\mathbf{v}_A - \mathbf{v}_A^*)^T \geq 0, \quad (5.24)$$

where $[\mathbf{f}_H^T, \mathbf{v}_A^T] \in \boldsymbol{\Omega}_{\mathbf{v}}$. $\mathbf{t}_A^*$ is the vector of link travel cost of CAVs at the equilibrium state $[\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*}]$**.** The following proposition discusses the existence of the path flow solutions of VI problem (5.21).

where $[\mathbf{f}_H^T, \mathbf{v}_A^T] \in \boldsymbol{\Omega}_{\mathbf{v}}$. $\mathbf{t}_A^*$ is the vector of link travel costs of CAVs at the equilibrium state $[\mathbf{f}_H^{T*}, \mathbf{f}_A^{T*}]$. The following proposition discusses the existence of path flow solutions of VI problem (5.21).

Proposition 5. 2: The VI problem (5.21) has at least one path flow solution.

*Proof***:** As the link travel cost functions (Eq. (5.19)) for both CAVs and HDVs are continuous with respect to link flows of CAVs and HDVs, the generalized path cost for both HDVs $(C_{k,H}^w(\mathbf{f}_H, \mathbf{f}_A), \forall k, w)$ and CAVs $(C_{k,A}^w(\mathbf{f}_H, \mathbf{f}_A), \forall k, w)$ are continuous with respect to $\mathbf{f}_H$ and $\mathbf{f}_A$**.** Besides, the constraints in the feasible path flow set $\boldsymbol{\Omega}_{\mathbf{f}}$ are affine. Thereby, the feasible path flow set $\boldsymbol{\Omega}_{\mathbf{f}}$ is closed and convex. According to Theorem 1.4 in Nagurney (2013), VI problem (5.21) has at least one path flow solution. ■

Proposition 5.2 implies that there exists at least one solution to VI problem (5.23) or (5.24). Let $\mathbf{f} = [\mathbf{f}_H^T, \mathbf{f}_A^T]^T$, $\mathbf{C} = [\mathbf{C}_H^T, \mathbf{C}_A^T]^T$. Note that the generalized path travel cost vector $\mathbf{C}$ in VI problem (5.22) is not a monotonic function of path flow $\mathbf{f}$ as the link travel cost is asymmetric between HDVs and CAVs (Nagurney, 2000). Thereby, both the VI problems (5.21) and (5.23) can have multiple local solutions.

Proposition 5.3: If $VOT_H = VOT_A$, then the path travel cost of CAVs for an OD pair is no larger than the path travel cost of HDVs for the same OD pair.

Proof: Note that all HDV paths are also potential CAV paths. Thereby, $R_H^w \subset R_A^w$, for $\forall w \in (W_A \cap W_H)$. As CAVs choose the paths with minimum travel cost for an OD pair, the path travel cost of an arbitrary CAV must be no larger than the path travel cost of an arbitrary HDV for the same OD pair. ■

When the level of automation of CAVs is not high, the drivers need to monitor the vehicle frequently and be ready to take over when requested to do so. In this scenario, the VOT of CAV users is close to that of HDV users. Proposition 5.3 suggests that a large proportion of HDV demand and CAV demand will be distributed on the same paths with shortest travel cost in this case, which increases traffic congestion and reduces system performance. It will be shown in the numerical example that the system performance of HDVs will benefit from the condition $VOT_H > VOT_A$, where a large proportion of HDV demand and CAV demand for the same OD pair will be distributed on different paths to reduce the network congestion level.

5.4 Solution algorithm

Many VI-based solution algorithms can be used to solve the proposed multiclass traffic assignment model, such as the projection method (Nagurney, 2000; Nagurney and Dong, 2002; Jiang et al., 2016), Tikhonov regularization method (Tikhonov, 1963), proximal point methods (Bauschke, 2004; Zhan and Ukkusuri, 2017), etc., provided that the corresponding convergence conditions are satisfied. However, these methods need to solve a subproblem to obtain the descent direction at each iteration. This is computationally expensive for the proposed multiclass traffic assignment model due to the presence of the complex generalized travel path cost function for HDVs (see Eq. (5.3)). To circumvent this issue, a route-swapping-based solution algorithm will be developed in this study to solve the multiclass traffic assignment model (21). At each iteration, this algorithm calculates the descent direction using a closed-form formulation to circumvent solving the subproblem in VI-based solution algorithms.

Route-swapping models are usually formulated to characterize the evolution of traffic flow based on drivers' knowledge of traffic conditions (Wang et al., 2016). They address whether and how the flow pattern evolves from a non-equilibrium state toward an equilibrium state. Depending on route choice assumptions, existing route-swapping models can converge to a stationary state equivalent to UE (Smith, 1984; Smith and Wisten, 1995; Huang and Lam, 2002; Peeta and Yang, 2003, Friesz et al., 1994) or logit-based SUE (Guo, 2013; Smith and Watling, 2016). This characteristic enables route-swapping models to obtain a feasible solution algorithm for network equilibrium problems. Huang and Lam (2002) develop a heuristic algorithm based on the projected route-swapping (PRS) dynamic system to solve a VI-based departure time choice equilibrium problem. For convenience, we label it PRS algorithm. At each iteration, the

descent direction of the PRS algorithm is calculated using an analytical model that assumes flow will shift from the more costly paths to the least-cost path at a rate that is proportional to the flow on the more costly paths and the cost difference from the least-cost path. This algorithm is used to solve other VI-based traffic assignment problems (see e.g., Lam and Huang, 2003; Szeto and Lo, 2006). However, many studies report that the algorithm converges slowly (Szeto and Lo, 2006; Ramadurai and Ukkusuri, 2010). This is partly because the PRS algorithm may fail to find a descent direction when it is close to the optimal solution as the flow shifts from all other paths to the least-cost path can be overestimated. For comparison, the steps to solve the multiclass traffic assignment problem (5.21) using PRS algorithm are presented in Table 5.1.

Table 5.1 Steps of PRS algorithm proposed by Huang and Lam (2002) to solve the multiclass traffic assignment problem (5.21)

Step 1: *Initialization*. Choose initial vectors of path flows $\mathbf{f}_{n,H}, \mathbf{f}_{n,A}$. Set the iteration index $n = 1$;

Step 2: *Generalized route travel cost update*. For both HDVs and CAVs, obtain the minimum travel cost between each OD pair and the corresponding path set by
$C_{z,min}^{w} = min\{C_{k,z}^{w}(\mathbf{f}_{n,H}, \mathbf{f}_{n,A}): k \in R_z^w\}$, and
$\tilde{R}_z^w = \{k | C_{k,z}^{w}(n) = C_{z,min}^{w}, k \in R_z^w\}, z \in \{HDV, CAV\}$

Step 3: *Route flow update*. Update path flows for HDVs and CAVs ($f_{k,z}^{w}(n+1), \forall k, w, z \in \{HDV, CAV\}$) by
$$f_{k,z}^{w}(n+1) = f_{k,z}^{w}(n) + \beta_n f_{k,z}^{w}(n)\left[C_{k,z}^{w} - C_{z,min}^{w}\right], \quad k \in R_z^w \backslash \tilde{R}_z^w, z \in \{HDV, CAV\}$$
$$f_{k,z}^{w}(n+1) = f_{k,z}^{rs}(n) + \frac{\sum_{i \in R_z^w \backslash \tilde{R}_z^w} \beta_n f_{i,z}^{w}(n)\left[C_{i,z}^{w} - C_{z,min}^{w}\right]}{\left|\tilde{R}_z^w\right|}, k \in \tilde{R}_z^w, z \in \{HDV, CAV\}$$
where $\left|\tilde{R}_z^w\right|$ is the number of paths in set $\tilde{R}_z^w$; $f_{i,z}^{w}(n)$ is the flow of path on OD pair w for vehicle class z on iteration n, β_n is the step of iteration n.

Step 4. *Convergence check*. If the convergence criteria is satisfied, then step. Otherwise, let $n = n + 1$go to step 2.

This study develops a new route-swapping-based solution algorithm to solve the multiclass traffic assignment problem (5.21). We will show that the proposed algorithm can solve (21) effectively and can converge much faster than the PRS algorithm. Let n denotes the iteration number. At iteration $n + 1$, unlike the PRS algorithm, the path flows $\mathbf{f}_{n+1}$ of both HDVs and CAVs will be updated according to the following model, which is a revised version of the route-swapping model proposed by Smith (1984) to incorporate multi-user classes,

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \beta_n \mathbf{\Phi}(\mathbf{f}_n) = \begin{bmatrix} \mathbf{f}_{n,H} \\ \mathbf{f}_{n,A} \end{bmatrix} + \beta_n \begin{bmatrix} \mathbf{\Phi}_H(\mathbf{f}_n) \\ \mathbf{\Phi}_A(\mathbf{f}_n) \end{bmatrix} \tag{5.25}$$

where $\beta_n > 0$ is a step, $\mathbf{f}_{n,H}$ and $\mathbf{f}_{n,A}$ are vectors of path flows for HDVs and CAVs on iteration n, respectively. $\mathbf{\Phi}_H(\mathbf{f}_n) = \left(\Phi_{k,H}^w(\mathbf{f}_n), \forall k \in R_H^w, w \in W_H\right)$ and $\mathbf{\Phi}_A(\mathbf{f}_n) = \left(\Phi_{k,A}^w(\mathbf{f}_n), \forall k \in R_A^w, w \in W_A\right)$ are updated by:

$$\Phi_{k,H}^w(\mathbf{f}_n) = \sum_{g \in R_H^w} \left[f_{g,H}^w(n) \left(C_{g,H}^w(\mathbf{f}_n) - C_{k,H}^w(\mathbf{f}_n) \right)_+ - f_{k,H}^w(n) \left(C_{k,H}^w(\mathbf{f}_n) - C_{g,H}^w(\mathbf{f}_n) \right)_+ \right] \tag{5.26a}$$

$$\Phi_{k,A}^w(\mathbf{f}_n) = \sum_{g \in R_A^w} \left[f_{g,A}^w(n) \left(C_{g,A}^w(\mathbf{f}_n) - C_{k,A}^w(\mathbf{f}_n) \right)_+ - f_{k,A}^w(n) \left(C_{k,A}^w(\mathbf{f}_n) - C_{g,A}^w(\mathbf{f}_n) \right)_+ \right] \tag{5.26b}$$

where $f_{i,z}^w(n)$ is the flow of path i for vehicle class $z \in \{HDV, CAV\}$ for OD pair w in iteration n.

For simplicity, we label Eq. (5.25) as the revised Smith's route-swapping (RSRS) algorithm. This algorithm finds the descent direction based on an analytical formulation $\mathbf{\Phi}(\mathbf{f}_n)$ to circumvent the subproblem in VI-based algorithms**.** In each iteration, the descent direction can be updated quickly based on the path flow and generalized path travel cost of the previous iteration. Different from the PRS algorithm that swaps flow from a path to the least-cost paths (see Appendix A), the RSRS algorithm swaps flow from a path to all other paths with a lower cost. This prevents the overestimation of flow swaps to the least-cost paths which is likely to occur in the PRS algorithm when the solution is close to the optimum. The following proposition describes the equivalence between the stationary point of Eq. (5.25) and the solution of the multiclass traffic assignment problem (5.21).

Proposition 5.4: If the path flows $\mathbf{f}_n$ determined by the RSRS algorithm converge, then they must converge to the pattern that simultaneously satisfies the CNL equilibrium conditions for HDVs and UE conditions for CAVs.

Proof: Suppose $\mathbf{f}_n$ is a stationary point for Eq. (5.25), $\mathbf{f}_n = \mathbf{f}_{n+1}$. Then for arbitrary OD pair $w, w \in W_H$

$$\begin{aligned} 0 &= \left(\mathbf{C}_H^w(\mathbf{f}_n)\right)^T \left(\mathbf{f}_{n+1,H}^w - \mathbf{f}_{n,H}^w\right) \\ &= \left(\mathbf{C}_H^w(\mathbf{f}_n)\right)^T \left(\mathbf{f}_{n,H}^w + \beta_n \mathbf{\Phi}_H^w(\mathbf{f}_n) - \mathbf{f}_{n,H}^w\right) \\ &= \beta_n \sum_{k \in R_H^w} C_{k,H}^w \sum_{g \in R_H^w} \left[f_{g,H}^w(n) \left(C_{g,H}^w(\mathbf{f}_n) - C_{k,H}^w(\mathbf{f}_n) \right)_+ - f_{k,H}^w(n) \left(C_{k,H}^w(\mathbf{f}_n) - C_{g,H}^w(\mathbf{f}_n) \right)_+ \right] \end{aligned} \tag{5.27}$$

where $\mathbf{f}_{n,H}^w$ is the vector of flow of all paths for OD pair $w \in W_H$ at iteration n. $\mathbf{C}_H^w(\mathbf{f}_n)$ is the vector of generalized costs of all paths for HDVs for OD pair $w \in W_H$ at iteration n. $\mathbf{\Phi}_H^w(\mathbf{f}_n) = \left(\Phi_{k,H}^w(\mathbf{f}_n), \forall k \in R_H^w\right)$. Let $|R_H^w|$ denote the number of paths for HDVs for OD pair $w, w \in W_H$.

Without loss of generality, suppose $C_{1,H}^{w}(\mathbf{f}_n) \geq C_{2,H}^{w}(\mathbf{f}_n) \geq \cdots \geq C_{|R_H^w|,H}^{w}(\mathbf{f}_n)$. Then, Eq. (5.27) can be simplified as:

$$\begin{aligned} 0 &= \left(\mathbf{C}_H^w(\mathbf{f}_n)\right)^T(\mathbf{f}_{n+1}^w - \mathbf{f}_n^w) \\ &= \beta_n \sum_{k=1}^{|R_H^w|} \sum_{g=k+1}^{|R_H^w|} f_{k,H}^w(n)\left(C_{k,H}^w(\mathbf{f}_n) - C_{g,H}^w(\mathbf{f}_n)\right)^2 \end{aligned} \tag{28}$$

Eq. (5.28) holds only when it satisfies Eq. (5.5). Using the same method, it can be shown that the stationary path flows for HDVs and CAVs for all OD pairs obtained by Eq. (5.25) satisfy the CNL equilibrium condition and UE condition in Eq. (5.5) and Eq. (5.20), respectively. Proposition 5.4 is proved. ■

Let SOL($\mathbf{\Omega_f}, \mathbf{C}$) be the solution set of VI problem (5.22). According to Proposition 5.4, SOL($\mathbf{\Omega_f}, \mathbf{C}$) $= \{\mathbf{f}|: \mathbf{\Phi}(\mathbf{f}) = \mathbf{0}\}$. The following theorem based on Mounce and Carey (2015) provides sufficient conditions for the convergence of the RSRS algorithm.

Theorem 5.1. Suppose the generalized path travel cost $\mathbf{C}(\mathbf{f})$ is monotonic in $\mathbf{\Omega_f}$. Then, the RSRS algorithm (42) converges to the CNL equilibrium state and UE state for HDVs and CAVs, respectively, if β_n satisfy $\lim_{n\to\infty} \beta_n = 0, \sum_{n=1}^{\infty} \beta_n = \infty$.

The proof of Theorem 5.1 can follow the same method proposed by Mounce and Carey (2015). We omit the proof here to avoid duplication. Note that the monotonicity of $\mathbf{C}(\mathbf{f})$ is not guaranteed by the VI problem (5.22). We make the following assumption to analyze the convergence of the RSRS algorithm.

Assumption 5.1: The generalized path travel cost $\mathbf{C}(\mathbf{f})$ is monotonic in a small vicinity around a local optimal solution $\mathbf{f}^*$.

Assumption 5.1 implies that while the $\mathbf{C}(\mathbf{f})$ may not be monotonic in the feasible set $\mathbf{\Omega_f}$, it can be monotonic in a small vicinity around a local optimal solution. Huang and Lam (2002) also made the same assumption to ensure the local stability of the PRS algorithm. It is also often used for designing solution algorithms for non-convex traffic assignment problems (e.g., Zhan and Ukkusuri, 2017; Shao et al., 2006).

To ensure the convergence of the RSRS algorithm, Mounce and Carey (2015) use predetermined step sizes ($\beta_n = 1/n$) to solve the UE problem. However, the RSRS algorithm converges slowly using these predetermined step sizes, especially for large-size network problems (Mounce and Carey, 2015). Huang and Lam (2002) propose another predetermined

step size choice strategy. It assumes the step sizes are the same for a large number of consecutive iterations and non-increasing with respect to iterations. For example, $\beta_n = 10^{-4} \times \left(1^{(1\to1000)}, 1/2^{(1001\to2000)}, 1/3^{(2001\to3000)} \cdots\right)$ where the step size remains the same for every 1000 iterations. However, the PRS algorithm still converges slowly though it is better than that using $\beta_n = 1/n$ (Ramadurai and Ukkusuri, 2010). In summary, it is difficult to determine effective step sizes offline as small step sizes make the algorithm converge slowly while large step sizes can preclude it from converging.

To enhance the convergence efficiency of the RSRS algorithm, a modified self-regulated average (MSRA) method is proposed in this study to determine the step size in each iteration based on the work of Liu et al. (2009). This method adaptively updates the step size in each iteration using information from both the current iteration and the previous iteration. Eq. (5.29) represents the step size determination rule. If the norm descent direction in iteration n is larger than that in iteration $n-1$, implying that the RSRS algorithm tends to diverge, a large value of Υ_1 is used to shrink the current step size. The opposite case implies that the RSRS method tends to converge, and the MSRA method attempts to apply a large step size at the current iteration by setting a small value of Υ_2. In addition, to ensure the path flow at each iteration is within the feasible path flow set, the term $1/h_n$ is introduced by the MSRA method. It is formulated to ensure the total flow swapped from an arbitrary path to other paths by the RSRS algorithm is no larger than the flow of this path. It is also an adaptive term determined by the maximum value of the summation of differences in the generalized costs of path pairs and the minimum generalized path cost for the corresponding OD pair.

$$\beta_n = \frac{1}{h_n} \cdot \frac{1}{\chi_n} \tag{29a}$$

$$\chi_n = \begin{cases} \chi_{n-1} + \Upsilon_1; if\ \|\boldsymbol{\Phi}(\mathbf{f}_n)\| \geq \|\boldsymbol{\Phi}(\mathbf{f}_{n-1})\| \\ \chi_{n-1} + \Upsilon_2; if\ \|\boldsymbol{\Phi}(\mathbf{f}_n)\| < \|\boldsymbol{\Phi}(\mathbf{f}_{n-1})\| \end{cases} \tag{29b}$$

where $h_n = \max\left(h_{z,i}^w(n) \middle| h_{z,i}^w(n) = \sum_{j \in R_z^w \setminus i} \left(C_{j,z}^w(\mathbf{f}_n) - C_{i,z}^w(\mathbf{f}_n)\right)_+, i \in R_z^w; \forall w \in W_z, \forall z \in \{HAV, CAV\}\right)$; and Υ_1 and Υ_2 are predetermined values, $\Upsilon_1 > 1; \Upsilon_2 \in (0,1)$. $\chi_1 = 1$.

Let $dist(\mathbf{f}, \text{SOL}(\boldsymbol{\Omega}_\mathbf{f}, \mathbf{C}))$ denote the Euclidean distance from $\mathbf{f}$ to the path flow solution set $\text{SOL}(\boldsymbol{\Omega}_\mathbf{f}, \mathbf{C})$. The following proposition discusses the convergence of the RSRS algorithm with the step sizes provided by the MSRA method (Eq. (5.29), labeled the "RSRS-MSRA algorithm".

Proposition 5.5: For arbitrary positive value ζ, the RSRS-MSRA algorithm converges to the set $\{\mathbf{f} \in \boldsymbol{\Omega}_\mathbf{f} | dist(\mathbf{f}, \text{SOL}(\boldsymbol{\Omega}_\mathbf{f}, \mathbf{C})) < \zeta\}$ given Assumption 5.1.

Proof: Let $\mathbf{f}_n, n = 1,2,\cdots$ be the path flow obtained using the RSRS-MSRA algorithm at each iteration n. Suppose the proposition does not hold. Then, we have $dist(\mathbf{f}_n, \text{SOL}(\mathbf{\Omega_f}, \mathbf{C})) \geq \zeta$ for $n = 1,2,\cdots$. This implies that at each iteration, the value of h_n is strictly larger than 0. Define the function h as follows:

$$h = \max\left(h_{z,i}^w \middle| h_{z,i}^w = \sum_{j \in R_z^w \setminus i}\left(C_{j,z}^w(\mathbf{f}) - C_{i,z}^w(\mathbf{f})\right)_+, i \in R_z^w; \forall w \in W_z, \forall z \in \{HAV, CAV\}; \mathbf{f} \in \mathbf{\Omega_f}\right)$$

Without loss of generality, let ϑ_{min} and ϑ_{max} be the minimum and maximum values of h, respectively, with the constraint $dist(\mathbf{f}, \text{SOL}(\mathbf{\Omega_f}, \mathbf{C})) \geq \zeta$. Note $\vartheta_{min} > 0$ under this constraint, and ϑ_{max} is bounded as the generalized travel cost is bounded. By definition, $\vartheta_{max} \geq h_n \geq \vartheta_{min} > 0$, for $\forall n$. According to Eq. (5.29b), $n \cdot \Upsilon_2 \leq \vartheta_n \leq n \cdot \Upsilon_1$. Then $\frac{1}{\vartheta_{max}}\frac{1}{n \cdot \Upsilon_2} \leq \beta_n \leq \frac{1}{\vartheta_{min}}\frac{1}{n \cdot \Upsilon_1}$. As both Υ_1 and Υ_2 are fixed positive values, $\lim_{n\to\infty} \beta_n \leq \lim_{n\to\infty} \frac{1}{\vartheta_{min}}\frac{1}{n \cdot \Upsilon_1} \to 0$, $\sum_{n=1}^{\infty} \beta_n \geq \sum_{n=1}^{\infty} \frac{1}{\vartheta_{max}}\frac{1}{n \cdot \Upsilon_2} \to \infty$. According to Theorem 5.1, the path flow determined by the RSRS-MSRA algorithm must eventually enter into the set $\{\mathbf{f} \in \mathbf{\Omega_f} | dist(\mathbf{f}, \text{SOL}(\mathbf{\Omega_f}, \mathbf{C})) < \zeta\}$ under assumption 5.1. Proposition 5.5 is proved. ■

According to above discussion, the steps to implement the RSRS-MSRA algorithm are summarized as follows:

Step 1: *Initialization*. Choose initial vectors of feasible path flows $\mathbf{f}_{n,H}, \mathbf{f}_{n,A}$. Set the iteration index $n = 1$;

Step 2: *Route flow update*. Update path flows for HDVs and CAVs ($f_{k,z}^w(n+1), \forall k, w, z \in \{HDV, CAV\}$) according to Eq. (5.25) and Eq. (5.29);

Step 3. *Convergence check*. If the convergence criterion is satisfied, then stop. Otherwise, let $n = n + 1$ and go to step 2.

The RSRS-MSRA solution algorithm is developed upon route flows. To find the routes that are likely to be used for each OD pair, the revised K-shortest path method developed by De La Barra et al. (1993) is used in this study. Note that only acyclic paths are considered.

5.5 Sensitivity analysis

Planned (e.g., road maintenance, construction) or unexpected (e.g., accidents, facility failure, nature disasters) events can impact network flows significantly. When an event occurs, it is often difficult for planners and decision-makers to evaluate its impact on network performance (e.g., congestion, OD travel cost of CAVs) and to design effective strategies to mitigate negative effects. To address these issues, this study develops an analytical model for sensitivity analysis

of the multiclass traffic assignment model. The sensitivity analysis determines the change in the objective function value for a unit change in the value of an explanatory variable. In literature, analytical formulations have been explored for sensitivity analysis of single-class traffic assignment models, including the UE problem (Tobin and Friesz, 1988; Wang et al., 2016; Boyles et al., 2012; Jafari and Boyles, 2016), the logit-based and probit-based SUE models (Clark and Watling, 2000), the paired combinatorial logit model (Wang et al., 2018a), and the elastic demand model (Yang, 1997). However, there are no studies that develop a sensitivity analysis method for multiclass traffic assignment model. The rest of this section derives the analytical model for sensitivity analysis of the proposed multiclass traffic assignment model (22) to obtain the gradients of equilibrium link flows of HDVs and CAVs with respect to perturbed parameters (e.g., link capacity, free-flow travel time, etc.). The analytical model can be used for applications such as: (i) constructing approximation methods to quickly estimate the perturbed mixed traffic equilibrium, (ii) identifying critical parameters (link capacity, signal splits) impacting network performance, (iii) risk analysis to provide insights on network performance reliability, (iv) analyzing impacts of traffic control methods on road network equilibrium, and (v) constructing solution algorithms for continuous network design problem (e.g., to find optimal signal timing or tolling strategy to improve the system performance under mixed traffic flow). This study explores the first application. Next, we will show the sufficient conditions for uniqueness of a local solution for link flows of HDVs and CAVs. The analytical model to obtain the gradients of the optimal link flow solution of HDVs and CAVs with respect to perturbed parameters is developed in Section 5.2.

5.5.1 Uniqueness of local solution of link flows of HDVs and CAVs

The uniqueness of a local solution of link flows of HDVs and CAVs is a necessary condition for the existence of gradients of the equilibrium solution of link flows of HDVs and CAVs with respect to the perturbations (Facchinei and Pang, 2007). This section determines sufficient conditions for the uniqueness of a local solution of the VI problem (5.24). We will show that these sufficient conditions are mild, so that the local solution of path flows for HDVs and link flows for CAVs is unique in general.

Let $\boldsymbol{\nu} = [\mathbf{f}_H, \mathbf{v}_A]$, and $\tilde{\mathbf{C}} = [\mathbf{C}_H, \boldsymbol{t}_A]$. Denote $\boldsymbol{\nu}^*$ as the solution to VI problem (5.24). Let SOL($\boldsymbol{\Omega}_{\mathbf{f}}, \mathbf{C}$) and SOL($\boldsymbol{\Omega}_{\boldsymbol{\nu}}, \tilde{\mathbf{C}}$) be the set of solutions of VI problems (22) and (24), respectively.

Denote $\mathbf{f}^*$ ($\mathbf{f}^* \in \mathrm{SOL}(\boldsymbol{\Omega}_\mathbf{f}, \mathbf{C})$) and $\boldsymbol{\nu}^*$ ($\boldsymbol{\nu}^* \in \mathrm{SOL}(\boldsymbol{\Omega}_\boldsymbol{\nu}, \tilde{\mathbf{C}})$) as one of the solutions of VI problems (22) and (24), respectively. The following definition will be used to analyze whether $\mathbf{f}^*$ and $\boldsymbol{\nu}^*$ are locally unique.

Definition 1: Uniqueness of local solution (Facchinei and Pang, 2007). $\mathbf{f}^*$ is said to be a locally unique solution for VI problem (5.22) if there exists a neighborhood $\mathcal{H}_{\mathbf{f}^*}$ of $\mathbf{f}^*$ such that $\mathrm{SOL}(\boldsymbol{\Omega}_\mathbf{f}, \mathbf{C}) \cap \mathcal{H}_{\mathbf{f}^*} = \mathbf{f}^*$; Similarly, $\boldsymbol{\nu}^*$ is said to be a locally unique solution for VI problem (5.24) if there exists a neighborhood $\mathcal{H}_{\boldsymbol{\nu}^*}$ of $\boldsymbol{\nu}^*$ such that $\mathrm{SOL}\left(\boldsymbol{\Omega}_\boldsymbol{\nu}, \tilde{\mathbf{C}}\right) \cap \mathcal{H}_{\boldsymbol{\nu}^*} = \boldsymbol{\nu}^*$.

As CAVs choose paths according to the UE principle, any path flow solution $\mathbf{f}^*$ of the VI problem (5.22) may be locally non-unique because there may exist multiple equilibrium path flow solutions for CAVs in the neighborhood $\mathcal{H}_{\mathbf{f}^*}$ of $\mathbf{f}^*$. According to Facchinei and Pang (2007), if Jacobian matrix of $\tilde{\mathbf{C}}$ with respect to $\boldsymbol{\nu}$ at the local optimal solution $\boldsymbol{\nu}^*$ (denoted as $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$) is positive definite, then $\boldsymbol{\nu}^*$ is a locally unique solution for VI problem (5.24). In the following, we will analyze the sufficient conditions for $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$ being a positive definite matrix. Note

$$\frac{\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)}{\partial\boldsymbol{\nu}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{f}_H} & \dfrac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{v}_A} \\ \dfrac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{f}_H} & \dfrac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{v}_A} \end{bmatrix}; \tag{5.30}$$

where

$$\frac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{f}_H} = \nabla_{\mathbf{v}_H} \boldsymbol{t}_A^* \cdot \Delta_H \tag{5.31}$$

To characterize $\partial \boldsymbol{C}_H^*/\partial \mathbf{f}_H$ and $\partial \boldsymbol{C}_H^*/\partial \mathbf{v}_A$ analytically. Let j be an arbitrary path for HDV between an arbitrary OD pair w_1, $w_1 \in W_H$, according to Eq. (5.3)

$$\frac{dC_{k,H}^w}{df_{j,H}^{w_1}} = \frac{dc_{k,H}^w}{df_{j,H}^{w_1}} - \frac{u}{\theta}\frac{dH_{k,H}^w}{df_{j,H}^{w_1}} + \frac{u}{\theta}\frac{1}{f_{k,H}^w}\frac{df_{k,H}^w}{df_{j,H}^{w_1}} \tag{5.32}$$

where $df_k^w/df_j^{w_1} = 1$, if $w_1 = w$ and $k = j$. Otherwise $df_k^w/df_j^{w_1} = 0$. According to Eq. (5.32)

$$\frac{dH_{k,H}}{df_{j,H}^{w_1}} = \frac{\sum_{m\in\Gamma_H}\left(\alpha_{m,k}^w\right)^{\frac{1}{u}}(u-1)\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{\frac{1}{u}}\right)^{u-2}}{\sum_{m\in\Gamma_H}\left(\alpha_{m,k}^w\right)^{1/u}\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^{u-1}} \sum_{l\in R_H^w}\left(-\frac{\theta}{u}\right)\left(\alpha_{m,l}^w\right)^{\frac{1}{u}}\exp\left(-\frac{\theta}{u}c_{l,H}^w\right)\frac{dc_{l,H}^w}{df_{j,H}^{w_1}}$$

Let

$$\rho_{k,l}^w = \frac{\sum_{m\in\Gamma_H}\left(\alpha_{m,k}^w\right)^{\frac{1}{u}}\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{\frac{1}{u}}\right)^{u-2}}{\sum_{m\in\Gamma_H}\left(\alpha_{m,k}^w\right)^{1/u}\left(\sum_{l\in R_H^w}\left[\alpha_{m,l}^w \exp(-\theta c_{l,H}^w)\right]^{1/u}\right)^{u-1}}\left(\alpha_{m,l}^w\right)^{\frac{1}{u}}\exp\left(-\frac{\theta}{u}c_{l,H}^w\right) \tag{5.33}$$

where $i \in R_H^w$. According to Eq. (5.33), $\rho_{k,l}^w \geq 0$ and

$$\sum_{l \in R_H^w} \rho_{k,l}^w = 1 \tag{5.34}$$

$$\frac{dH_k}{df_j^{w_1}} = \left(-\frac{\theta}{u}\right)(u-1)\left[\rho_{k,1}^w \quad \rho_{k,2}^w \quad \cdots \quad \rho_{k,|R_H^w|}^w\right]\begin{bmatrix} \frac{dc_{1,H}^w}{df_{j,H}^{w_1}} & \frac{dc_{2,H}^w}{df_{j,H}^{w_1}} & \cdots & \frac{dc_{|R_H^w|,H}^w}{df_{j,H}^{w_1}} \end{bmatrix}^T \tag{5.35}$$

To characterize the matrix $\partial \boldsymbol{C}_H^*/\partial \mathbf{f}_H$ analytically, let $\boldsymbol{\rho}^w$ be a square matrix with dimension $|R_H^w|$,

$$\boldsymbol{\rho}^w = \begin{bmatrix} \rho_{1,1}^w & \rho_{1,2}^w & \cdots & \rho_{1,|R_H^w|}^w \\ \rho_{2,1}^w & \rho_{2,2}^w & \cdots & \rho_{2,|R_H^w|}^w \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{|R_H^w|,1}^w & \rho_{|R_H^w|,2}^w & \cdots & \rho_{|R_H^w|,|R_H^w|}^w \end{bmatrix} \tag{5.36}$$

Let $\boldsymbol{\rho}$ be a diagonal block matrix with dimension equals the number of paths for HDVs in the network

$$\boldsymbol{\rho} = \begin{bmatrix} \boldsymbol{\rho^1} & 0 & 0 & 0 \\ 0 & \boldsymbol{\rho^2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \boldsymbol{\rho}^{|W_H|} \end{bmatrix} \tag{5.37}$$

where $|W_H|$ is the number of all OD pairs for HDVs in the network. According to Eq. (5.3), the gradient of $\mathbf{C}_H$ with respect to $\mathbf{f}_H$ at $\mathbf{f}^*$ is

$$\begin{aligned} \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{f}_H} &= \nabla_{\mathbf{f}_H}\mathbf{c}_H^* - (1-u)\boldsymbol{\rho}^*\nabla_{\mathbf{f}_H}\mathbf{c}_H^* + \frac{u}{\theta} diag\left(\frac{\mathbf{1}}{\mathbf{f}_H^*}\right) \\ &= \Delta_H^T \cdot \nabla_{\mathbf{v}_H}\boldsymbol{t}_H^* \cdot \Delta_H - (1-u)\boldsymbol{\rho}^*\Delta_H^T \cdot \nabla_{\mathbf{v}_H}\boldsymbol{t}_H^* \cdot \Delta_H + \frac{u}{\theta} diag\left(\frac{\mathbf{1}}{\mathbf{f}_H^*}\right) \end{aligned} \tag{5.38}$$

where $\boldsymbol{\rho}^*$ denotes the value of matrix $\boldsymbol{\rho}$ at $\mathbf{f}^*$. Similarly,

$$\frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{v}_A} = \nabla_{\mathbf{v}_A}\mathbf{c}_H^* - (1-u)\boldsymbol{\rho}^*\,\nabla_{\mathbf{v}_A}\mathbf{c}_H = \Delta_H^T \cdot \nabla_{\mathbf{v}_A}\boldsymbol{t}_H^* - (1-u)\boldsymbol{\rho}^*\Delta_H^T \cdot \nabla_{\mathbf{v}_A}\boldsymbol{t}_H^* \tag{5.39}$$

Let $\mathbf{M}$ be a symmetric matrix defined as follows

$$\mathbf{M} = \frac{\partial \tilde{\mathbf{C}}(\mathbf{v}^*)}{\partial \mathbf{v}} + \left[\frac{\partial \tilde{\mathbf{C}}(\mathbf{v}^*)}{\partial \mathbf{v}}\right]^T = \begin{bmatrix} \mathbf{M}_1 + \frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right) & \mathbf{M}_2 \\ \mathbf{M}_3 & 2\nabla_{\mathbf{v}_A}\boldsymbol{t}_A^* \end{bmatrix} \tag{5.40}$$

where

$$\mathbf{M}_1 = \nabla_{\mathbf{f}_H}\mathbf{c}_H^* + \left(\nabla_{\mathbf{f}_H}\mathbf{c}_H^*\right)^T - (1-u)\boldsymbol{\rho}^*\,\nabla_{\mathbf{f}_H}\mathbf{c}_H^* - \left[(1-u)\boldsymbol{\rho}^*\,\nabla_{\mathbf{f}_H}\mathbf{c}_H^*\right]^T \tag{5.41a}$$

$$\mathbf{M}_2 = \Delta_H^T \cdot \nabla_{\mathbf{v}_A}\boldsymbol{t}_H^* - (1-u)\boldsymbol{\rho}^*\Delta_H^T \cdot \nabla_{\mathbf{v}_A}\boldsymbol{t}_H^* + \left(\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^* \cdot \Delta_H\right)^T \tag{5.41b}$$

$$\mathbf{M}_3 = \mathbf{M}_2^T \tag{5.41c}$$

The following proposition will be used to show the sufficient conditions for uniqueness of local solution of VI problem (5.24).

Proposition 6.6: $\mathbf{M}_1 + \frac{2u}{\theta} diag(\mathbf{f}_H) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3$ is a non-singular matrix if the matrix $\frac{\theta}{2u} diag(\mathbf{f}_H^*) \cdot \left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3 - \mathbf{M}_1\right)$ has no eigenvalue 1.

Proof: suppose $\mathbf{M}_1 + \frac{2u}{\theta} diag(\mathbf{f}_H) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3$ is not a non-singular matrix, then there exists a nonzero vector $\boldsymbol{x}$ such that

$$\left(\mathbf{M}_1 + \frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3\right)\boldsymbol{x} = \mathbf{0} \tag{5.42}$$

This implies

$$\frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right)\boldsymbol{x} = \left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3 - \mathbf{M}_1\right)\boldsymbol{x} \tag{5.43}$$

Note $\frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right)$ is invertible, then

$$\boldsymbol{x} = \frac{\theta}{2u} diag(\mathbf{f}_H^*)\left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3 - \mathbf{M}_1\right)\boldsymbol{x} \tag{5.44}$$

As $\frac{\theta}{2u} diag(\mathbf{f}_H^*)\left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3 - \mathbf{M}_1\right)$ has no eigenvalue 1, Eq. (5.44) cannot hold. Proposition 5.6 is proved.

Proposition 5.7: If $\partial\mathbf{C}(\mathbf{f}^*)/\partial\mathbf{f}$ is positive semidefinite and the matrix $\frac{\theta}{2u} diag(\mathbf{f}_H^*)\left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3 - \mathbf{M}_1\right)$ has no eigenvalue 1, then $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$ is positive definite.

Proof: Note

$$\frac{\partial\mathbf{C}(\mathbf{f}^*)}{\partial\mathbf{f}} = \begin{bmatrix}\mathbf{E}_H & \\ & \Delta_A^T\end{bmatrix}\frac{\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)}{\partial\boldsymbol{\nu}}\begin{bmatrix}\mathbf{E}_H & \\ & \Delta_A\end{bmatrix} \tag{5.45}$$

where $\mathbf{E}_H$ is an identity matrix with dimension equal to the number of paths for HDVs in the network. As $\partial\mathbf{C}(\mathbf{f}^*)/\partial\mathbf{f}$ is positive semidefinite, both the matrices $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$ and $\mathbf{M}$ are positive semidefinite. According to Eq. (5.40), $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$ is positive definite if and only if $\mathbf{M}$ is positive definite. Note that $\mathbf{M}$ is a symmetric matrix and $\nabla_{\mathbf{v}_A}\boldsymbol{t}_A^*$ is a symmetric positive definite matrix. According to Proposition 16.2 in Gallier (2011), $\mathbf{M}_1 + \frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3$ is a symmetric positive semidefinite matrix, indicating that the eigenvalues of this matrix are nonnegative. Proposition 5.6 shows that $\mathbf{M}_1 + \frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3$ is a non-singular matrix. Thereby, $\mathbf{M}_1 + \frac{2u}{\theta} diag\left(\frac{1}{\mathbf{f}_H^*}\right) - \mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-\mathbf{1}}\mathbf{M}_3$ is positive definite. This implies both the matrices $\mathbf{M}$ and $\partial\tilde{\mathbf{C}}(\boldsymbol{\nu}^*)/\partial\boldsymbol{\nu}$ are positive definite. Proposition 5.7 is proved.

The two conditions in Proposition 5.7 are not strong. First, the condition $\partial \mathbf{C}(\mathbf{f}^*)/\partial \mathbf{f}$ is positive semidefinite is sufficient to ensure $\mathbf{f}^*$ is a stationary optimal solution for VI problem (5.21), that is, $\mathbf{f}^*$ is not a saddle point (Facchinei and Pang, 2007). It can be satisfied by Assumption 5.1 in Section 4. Second, it can verify that if $VOT_A = VOT_H$, and $Q_{a,H} = Q_{a,A}, \forall a \in \boldsymbol{\Gamma}_H$, $\frac{\theta}{2u} diag(\mathbf{f}_H^*)\left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-1}\mathbf{M}_3 - \mathbf{M}_1\right) = \frac{\theta}{2u} diag(\mathbf{f}_H^*) \cdot \quad \mathbf{0} = \mathbf{0}$, in which case all eigenvalues of this matrix are 0. Note the eigenvalues of this matrix vary continuously with respective to VOT_A, and $Q_{a,H}, \forall a \in \boldsymbol{\Gamma}_H$. Thereby, the matrix $\frac{\theta}{2u} diag(\mathbf{f}_H^*) \cdot \left(\mathbf{M}_2\left(2\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^*\right)^{-1}\mathbf{M}_3 - \mathbf{M}_1\right)$ having an eigenvalue 1 is a very special case, and the likelihood of its occurrence in a real-world network is low for given values of VOT_A, and $Q_{a,H}, \forall a \in \boldsymbol{\Gamma}_H$. Thereby, $\partial \tilde{\mathbf{C}}(\mathbf{v}^*)/\partial \mathbf{v}$ is positive definite at the equilibrium state in general. This indicates that while the solution of VI problems (5.22) is not locally unique, the solution of VI problem (5.24) is locally unique provided the conditions in Proposition 5.7 are satisfied. This property is a necessary condition for sensitivity analysis of the multiclass traffic assignment model in the next section.

5.5.2 Sensitivity analysis of the multiclass traffic assignment problem

This section presents an analytical model to determine the gradients of the equilibrium link flow of both CAVs and HDVs with respect to the perturbed parameters (e.g., signal splits, link capacity of pure CAVs and HDVs, value of time for CAVs and HDVs, etc). Let $\boldsymbol{\varepsilon}$ be a vector of perturbed parameters in the multiclass traffic assignment problem (5.22), and $\mathbf{f}^* = [\mathbf{f}_H^{T*}(\mathbf{0}), \mathbf{f}_A^{T*}(\mathbf{0})]$ be a local equilibrium path flow solution to the VI problem (5.22) at the unperturbed state (i.e., $\boldsymbol{\varepsilon} = \mathbf{0}$). The KKT conditions at $\mathbf{f}^*$ are

$$\mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) - \boldsymbol{\pi}_H^* - \Lambda_H^T \boldsymbol{\mu}_H^* = \mathbf{0} \tag{5.46a}$$

$$\mathbf{C}_A(\mathbf{f}^*, \mathbf{0}) - \boldsymbol{\pi}_A^* - \Lambda_A \boldsymbol{\mu}_A^* = \mathbf{0} \tag{5.46b}$$

$$\begin{bmatrix} \Lambda_H & \\ & \Lambda_A \end{bmatrix} \begin{bmatrix} \mathbf{f}_H^* \\ \mathbf{f}_A^* \end{bmatrix} - \begin{bmatrix} \mathbf{q}_H(\mathbf{0}) \\ \mathbf{q}_A(\mathbf{0}) \end{bmatrix} = \mathbf{0} \tag{5.46e}$$

$$\boldsymbol{\pi}_H^* \mathbf{f}_H^* = \mathbf{0} \tag{5.46c}$$

$$\boldsymbol{\pi}_A^* \mathbf{f}_A^* = \mathbf{0} \tag{5.46d}$$

$$\mathbf{f}_A^* \geq \mathbf{0}; \mathbf{f}_H^* > \mathbf{0} \tag{5.46f}$$

$$\boldsymbol{\pi}_H^* \geq \mathbf{0}; \boldsymbol{\pi}_A^* \geq \mathbf{0} \tag{5.46g}$$

where $\boldsymbol{\mu}_H^*$ and $\boldsymbol{\mu}_A^*$ are vectors of Lagrange multipliers associated with the constraints $\Lambda_H \mathbf{f}_H = \mathbf{q}_H$ and $\Lambda_A \mathbf{f}_A = \mathbf{q}_A$, respectively. $\boldsymbol{\pi}_H^*$ and $\boldsymbol{\pi}_A^*$ are vectors of Lagrange multipliers associated with the nonnegative path flow constraints for HDVs and CAVs, respectively.

As discussed in Section 5.1, the path flow solution to VI problem (5.22) may not be locally unique. This precludes the existence of gradients for the path flow solution of VI problem (5.22) with respect to the perturbed parameters. To address this problem, a linear equation designed by Yang and Bell (2005) will be used in this study to obtain the path flow solution of CAVs that has the desired uniqueness. The method is developed upon the following assumption.

Assumption 5.2: The equilibrium path flow solution of CAVs of VI problem (5.22) at unperturbed state ($\boldsymbol{\varepsilon} = 0$) is not degenerate. That is, there exists a path flow solution such that the flow of equilibrated paths (i.e., paths with minimum travel cost for the corresponding OD pair) for CAVs is positive.

Assumption 5.2 is not strong in the sense that the degenerate points are isolated points when $\boldsymbol{\varepsilon}$ is nonzero. The likelihood of occurrence is low at the unperturbed state (i.e., $\boldsymbol{\varepsilon} = 0$) in the real network.

Let $\hat{R}_A$ be the set of equilibrated paths for CAVs at $[\mathbf{f}_H^*, \mathbf{f}_A^*]$, and $\hat{\mathbf{f}}_A^*$ be the flow vector for all equilibrated paths in set $\hat{R}_A$. Note that $\hat{\mathbf{f}}_A^*$ is not unique as $\mathbf{f}_A^*$ is not unique. Let $\begin{bmatrix}\hat{\Delta}_A \\ \hat{\Lambda}_A\end{bmatrix}$ be the link-path and OD-path matrix for paths in set $\hat{R}_A$. Denote $\begin{bmatrix}\bar{\Delta}_A \\ \bar{\Lambda}_A\end{bmatrix}$ as a full column matrix constituted by column vectors in $\begin{bmatrix}\hat{\Delta}_A \\ \hat{\Lambda}_A\end{bmatrix}$ that has the same rank as $\begin{bmatrix}\hat{\Delta}_A \\ \hat{\Lambda}_A\end{bmatrix}$. Let $\bar{R}_A$ be the set of paths with link-path and OD-path matrix $\begin{bmatrix}\bar{\Delta}_A \\ \bar{\Lambda}_A\end{bmatrix}$. The paths in set $\bar{R}_A$ are labeled equilibrated and linearly independent (ELI) paths, and the equilibrated paths for CAVs not in set $\bar{R}_A$ are labeled equilibrated and linearly dependent (ELD) paths. Let $\bar{\mathbf{f}}_A$ and $\tilde{\mathbf{f}}_A$ be the vector of all ELI and ELD path flows for CAVs, respectively. According to Assumption 5.2, an equilibrium path flow solution exists for CAVs such that flow of all equilibrated paths is positive at $\boldsymbol{\varepsilon} = 0$. Let $\begin{bmatrix}\bar{\mathbf{f}}_A^* \\ \tilde{\mathbf{f}}_A^*\end{bmatrix}$ ($\bar{\mathbf{f}}_A^* > 0; \tilde{\mathbf{f}}_A^* > 0$) be such a path flow solution for CAVs. Then

$$\begin{bmatrix}\hat{\Delta}_A \\ \hat{\Lambda}_A\end{bmatrix} \hat{\mathbf{f}}_A^* = \begin{bmatrix}\bar{\Delta}_A & \tilde{\Delta}_A \\ \bar{\Lambda}_A & \tilde{\Lambda}_A\end{bmatrix} \begin{bmatrix}\bar{\mathbf{f}}_A^*(\mathbf{0}) \\ \tilde{\mathbf{f}}_A^*(\mathbf{0})\end{bmatrix} = \begin{bmatrix}\mathbf{v}_A(\mathbf{0}) \\ \mathbf{d}_A(\mathbf{0})\end{bmatrix} \tag{5.47}$$

where $\begin{bmatrix}\tilde{\Delta}_A \\ \widetilde{\Lambda}_A\end{bmatrix}$ is the link-path and OD-path matrix for all paths in $\tilde{\mathbf{f}}_A$. According to Eq. (5.47),

$$\begin{bmatrix}\bar{\Delta}_A \\ \overline{\Lambda}_A\end{bmatrix}\bar{\mathbf{f}}_A^*(\varepsilon) = \begin{bmatrix}\mathbf{v}_A(\varepsilon) \\ \mathbf{q}_A(\varepsilon)\end{bmatrix} - \begin{bmatrix}\bar{\Delta}_A \\ \overline{\Lambda}_A\end{bmatrix}\tilde{\mathbf{f}}_A^*(0) \tag{5.48}$$

When ε changes, only $\bar{\mathbf{f}}_A^*(\varepsilon)$ changes, and the flow of ELD paths for HDVs are fixed at $\tilde{\mathbf{f}}_A^*(0)$. As $\bar{\mathbf{f}}_A^*(0) > 0$, $\bar{\mathbf{f}}_A^*(\varepsilon) > 0$ for a small perturbation of $\boldsymbol{\varepsilon}$. Note that as $\begin{bmatrix}\bar{\Delta}_A \\ \overline{\Lambda}_A\end{bmatrix}$ is a full column matrix, $\bar{\mathbf{f}}_A^*(\varepsilon)$ is unique. Thereby, when $\boldsymbol{\varepsilon}$ varies for a sufficiently small value from 0, a unique path flow solution $\begin{bmatrix}\bar{\mathbf{f}}_A^*(\varepsilon) \\ \tilde{\mathbf{f}}_A^*(0)\end{bmatrix}$ can be found by Eq. (5.48). According to Assumption 5.2, the non-equilibrated paths will remain non-equilibrated for a small perturbation of $\boldsymbol{\varepsilon}$. Since the flows of HDVs are all positive, and $\bar{\mathbf{f}}_A^* > 0, \tilde{\mathbf{f}}_A^* > 0$, all the Lagrange multipliers in vector $\boldsymbol{\pi}_H^*$ and $\boldsymbol{\pi}_A^*$ are nonbinding. Thereby, Eq. (5.46) can be simplified as

$$\mathbf{C}_H(\mathbf{f}^*, 0) - \Lambda_H^T \boldsymbol{\mu}_H^*(0) = \mathbf{0} \tag{5.49a}$$

$$\bar{\mathbf{C}}_A(\mathbf{f}^*, 0) - \Lambda_A^T \boldsymbol{\mu}_A^*(0) = \mathbf{0} \tag{5.49b}$$

$$\begin{bmatrix}\Lambda_H & \\ & \overline{\Lambda}_A\end{bmatrix}\begin{bmatrix}\mathbf{f}_H^*(0) \\ \bar{\mathbf{f}}_A^*(0)\end{bmatrix} + \begin{bmatrix}0 & \\ & \widetilde{\Lambda}_A\end{bmatrix}\begin{bmatrix}\mathbf{f}_H^*(0) \\ \tilde{\mathbf{f}}_A^*(0)\end{bmatrix} - \begin{bmatrix}\mathbf{q}_H(0) \\ \mathbf{q}_A(0)\end{bmatrix} = \mathbf{0} \tag{5.49c}$$

where $\bar{\mathbf{C}}_A(\mathbf{f}^*, 0)$ is the vector of generalized travel cost of all ELI paths for CAVs. As $\tilde{\mathbf{f}}_A^*(0)$ is fixed, the gradients of Eq. (5.49) with respect to the perturbed parameter $\boldsymbol{\varepsilon}$ are

$$\nabla_{\boldsymbol{\varepsilon}}\mathbf{C}_H(\mathbf{f}^*, 0) + \nabla_{\mathbf{f}_H}\mathbf{C}_H(\mathbf{f}^*, 0)\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H(\mathbf{0}) - \Lambda_H^T \nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_H^*(0) = \mathbf{0} \tag{5.50a}$$

$$\nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{C}}_A(\mathbf{f}^*, 0) + \nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A(\mathbf{f}^*, 0)\nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{f}}_A(\mathbf{0}) - \overline{\Lambda}_A^T \nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_A^*(0) = \mathbf{0} \tag{5.50b}$$

$$\begin{bmatrix}\Lambda_H & \\ & \overline{\Lambda}_A\end{bmatrix}\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H^*(0) \\ \nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{f}}_A^*(0)\end{bmatrix} + \begin{bmatrix}0 & \\ & \widetilde{\Lambda}_A\end{bmatrix}\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H^*(0) \\ \nabla_{\boldsymbol{\varepsilon}}\tilde{\mathbf{f}}_A^*(0)\end{bmatrix} - \begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{q}_H(\mathbf{0}) \\ \nabla_{\boldsymbol{\varepsilon}}\mathbf{q}_A(\mathbf{0})\end{bmatrix} = \mathbf{0} \tag{5.50e}$$

Note that $\nabla_{\boldsymbol{\varepsilon}}\tilde{\mathbf{f}}_A^*(0) \equiv 0$ as $\tilde{\mathbf{f}}_A$ is fixed. Thereby

$$\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H^*(0) \\ \nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{f}}_A^*(0) \\ \nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_H^*(0) \\ \nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_A^*(0)\end{bmatrix} = \left(\boldsymbol{J}_{\bar{\mathbf{f}}}^*\right)^{-1}(-\boldsymbol{J}_{\boldsymbol{\varepsilon}}^*) \tag{5.51}$$

where

$$\boldsymbol{J}_{\bar{\mathbf{f}}}^* = \begin{bmatrix}\nabla_{\mathbf{f}_H}\mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A}\mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) & \Lambda_H^T & \\ \nabla_{\mathbf{f}_H}\bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) & & \overline{\Lambda}_A^T \\ \Lambda_H & & & \\ & & \overline{\Lambda}_A & \end{bmatrix} \tag{5.52}$$

$$\boldsymbol{J}_{\boldsymbol{\varepsilon}}^{*}=\begin{bmatrix}\nabla_{\boldsymbol{\varepsilon}}\mathbf{C}_{H}(\mathbf{f}^{*},\mathbf{0})\\ \nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{C}}_{A}(\mathbf{f}^{*},\mathbf{0})\\ \nabla_{\boldsymbol{\varepsilon}}\mathbf{q}_{H}(\mathbf{0})\\ \nabla_{\boldsymbol{\varepsilon}}\mathbf{q}_{A}(\mathbf{0})\end{bmatrix} \tag{5.53}$$

The analytical form of $\nabla_{\mathbf{f}_H}\mathbf{C}_H\left(\bar{\mathbf{f}}^{*},\mathbf{0}\right)$ can be found in Eq. (5.38). The analytical forms of $\nabla_{\bar{\mathbf{f}}_A}\mathbf{C}_H\left(\bar{\mathbf{f}}^{*},\mathbf{0}\right)$, $\nabla_{\mathbf{f}_H}\bar{\mathbf{C}}_A\left(\bar{\mathbf{f}}^{*},\mathbf{0}\right)$ and $\nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A\left(\bar{\mathbf{f}}^{*},\mathbf{0}\right)$ are

$$\nabla_{\bar{\mathbf{f}}_A}\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0})=\Delta_H^T\cdot\nabla_{\mathbf{v}_A}\boldsymbol{t}_H^{*}\cdot\bar{\Delta}_A-(1-u)\boldsymbol{\rho}^{*}\Delta_H^T\cdot\nabla_{\mathbf{v}_A}\boldsymbol{t}_H^{*}\cdot\bar{\Delta}_A \tag{5.54a}$$

$$\nabla_{\mathbf{f}_H}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0})=\bar{\Delta}_A^T\cdot\nabla_{\mathbf{v}_H}\boldsymbol{t}_A^{*}\cdot\Delta_H \tag{5.54b}$$

$$\nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0})=\bar{\Delta}_A^T\cdot\nabla_{\mathbf{v}_A}\boldsymbol{t}_A^{*}\cdot\bar{\Delta}_A \tag{5.54e}$$

The following theorem constructed based on Corollary 3.2.5 in Fiacco (1983) provides the sufficient conditions for existence of gradients $\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H^{*}(0)$, $\nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{f}}_A^{*}(0)$, $\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_H^{*}(0)$ and $\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_A^{*}(0)$ in Eq. (5.51).

Theorem 5.2: The gradients $\nabla_{\boldsymbol{\varepsilon}}\mathbf{f}_H^{*}(0)$, $\nabla_{\boldsymbol{\varepsilon}}\bar{\mathbf{f}}_A^{*}(0)$, $\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_H^{*}(0)$ and $\nabla_{\boldsymbol{\varepsilon}}\boldsymbol{\mu}_A^{*}(0)$ exist if $\boldsymbol{J}_{\bar{\mathbf{f}}}^{*}$ is invertible and the terms $\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0})$, $\mathbf{C}_A(\mathbf{f}^{*},\mathbf{0})$, $\mathbf{q}_H(\mathbf{0})$ and $\mathbf{q}_A(\mathbf{0})$ are first-order differentiable with respect to $\boldsymbol{\varepsilon}$.

According to the link cost functions (Eq. (5.19)), $\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0})$, $\mathbf{C}_A(\mathbf{f}^{*},\mathbf{0})$, $\mathbf{q}_H(\mathbf{0})$ and $\mathbf{q}_A(\mathbf{0})$ are first-order differentiable with respect to $\boldsymbol{\varepsilon}$. Thereby the derivatives in the left-hand side of Eq. (5.41) exists if $\boldsymbol{J}_{\bar{\mathbf{f}}}^{*}$ is invertible. The following proposition discusses the sufficient conditions for invertibility of $\boldsymbol{J}_{\bar{\mathbf{f}}}^{*}$.

Proposition 5.8: if the matrix $\frac{\partial\tilde{\mathbf{C}}(\mathbf{v}^{*})}{\partial\mathbf{v}}=\begin{bmatrix}\frac{\partial\boldsymbol{C}_H^{*}}{\partial\mathbf{f}_H} & \frac{\partial\boldsymbol{C}_H^{*}}{\partial\mathbf{v}_A}\\ \frac{\partial\boldsymbol{t}_A^{*}}{\partial\mathbf{f}_H} & \frac{\partial\boldsymbol{t}_A^{*}}{\partial\mathbf{v}_A}\end{bmatrix}$ is positive definite. Then $\boldsymbol{J}_{\bar{\mathbf{f}}}^{*}$ is invertible**.**

Proof: let $\boldsymbol{z}=[\boldsymbol{z}_1^T\quad\boldsymbol{z}_2^T\quad\boldsymbol{z}_3^T\quad\boldsymbol{z}_4^T]^T$; The dimension of $\boldsymbol{z}_1^T$ and $\boldsymbol{z}_3^T$ equal the number of paths in $\mathbf{f}_H$, and the dimension of $\boldsymbol{z}_2^T$ and $\boldsymbol{z}_4^T$ equal the number of paths in $\bar{\mathbf{f}}_A$. Proposition 5.8 will be proved by showing that the solution to the following equation is $\boldsymbol{z}=\mathbf{0}$.

$$\begin{bmatrix}\nabla_{\mathbf{f}_H}\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A}\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0}) & \Lambda_H^T & \\ \nabla_{\mathbf{f}_H}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0}) & & \bar{\Lambda}_A^T\\ \Lambda_H & & & \\ & \bar{\Lambda}_A & & \end{bmatrix}\begin{bmatrix}\boldsymbol{z}_1\\ \boldsymbol{z}_2\\ \boldsymbol{z}_3\\ \boldsymbol{z}_4\end{bmatrix}=\mathbf{0} \tag{5.55}$$

According to Eq. (5.55)

$$\nabla_{\mathbf{f}_H}\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0})\boldsymbol{z}_1+\nabla_{\bar{\mathbf{f}}_A}\mathbf{C}_H(\mathbf{f}^{*},\mathbf{0})\boldsymbol{z}_2-\Lambda_H^T\boldsymbol{z}_3=\mathbf{0} \tag{5.56a}$$

$$\nabla_{\mathbf{f}_H}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0})\boldsymbol{z}_1+\nabla_{\bar{\mathbf{f}}_A}\bar{\mathbf{C}}_A(\mathbf{f}^{*},\mathbf{0})\boldsymbol{z}_2-\bar{\Lambda}_A^T\boldsymbol{z}_4=\mathbf{0} \tag{5.56b}$$

$$\Lambda_H \boldsymbol{z}_1 = 0 \tag{5.56c}$$

$$\bar{\Lambda}_A \boldsymbol{z}_2 = 0 \tag{5.56d}$$

Multiply Eq. (5.56a) and Eq. (5.56b) by $\boldsymbol{z}_1^T$ and $\boldsymbol{z}_2^T$ from left-hand side, respectively, we have

$$\boldsymbol{z}_1^T \nabla_{\mathbf{f}_H} \mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) \boldsymbol{z}_1 + \boldsymbol{z}_1^T \nabla_{\bar{\mathbf{f}}_A} \mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) \boldsymbol{z}_2 = \mathbf{0} \tag{5.57a}$$

$$\boldsymbol{z}_2^T \nabla_{\mathbf{f}_H} \bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) \boldsymbol{z}_1 + \boldsymbol{z}_2^T \nabla_{\bar{\mathbf{f}}_A} \bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) \boldsymbol{z}_2 = \mathbf{0} \tag{5.57b}$$

Equivalently, Eq. (5.57a) and Eq. (5.57b) can be written in a vector form

$$[\boldsymbol{z}_1^T \quad \boldsymbol{z}_2^T] \begin{bmatrix} \nabla_{\mathbf{f}_H} \mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A} \mathbf{C}_H(\mathbf{f}^*, \mathbf{0}) \\ \nabla_{\mathbf{f}_H} \bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) & \nabla_{\bar{\mathbf{f}}_A} \bar{\mathbf{C}}_A(\mathbf{f}^*, \mathbf{0}) \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} = 0 \tag{5.58}$$

Thereby,

$$[\boldsymbol{z}_1^T \quad \boldsymbol{z}_2^T] \begin{bmatrix} \mathbf{E}_H & \\ & \bar{\Delta}_A^T \end{bmatrix} \begin{bmatrix} \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{f}_H} & \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{v}_A} \\ \frac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{f}_H} & \frac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{v}_A} \end{bmatrix} \begin{bmatrix} \mathbf{E}_H & \\ & \bar{\Delta}_A \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} = 0 \tag{5.59}$$

Note the matrix $\begin{bmatrix} \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{f}_H} & \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{v}_A} \\ \frac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{f}_H} & \frac{\partial \boldsymbol{t}_A^*}{\partial \mathbf{v}_A} \end{bmatrix}$ is positive definite. Eq. (5.59) holds only when $\begin{bmatrix} \mathbf{E}_H & \\ & \bar{\Delta}_A \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} = \mathbf{0}$. This implies $\mathbf{E}_H \boldsymbol{z}_1 = 0$ and $\bar{\Delta}_A \boldsymbol{z}_2 = 0$. Note $\mathbf{E}_H \boldsymbol{z}_1 = 0$ only if $\boldsymbol{z}_1 = 0$. As $\begin{bmatrix} \bar{\Delta}_A \\ \bar{\Lambda}_A \end{bmatrix}$ is a full column matrix, $\begin{bmatrix} \bar{\Delta}_A \\ \bar{\Lambda}_A \end{bmatrix} \boldsymbol{z}_2 = 0$ only if $\boldsymbol{z}_2 = 0$. Thereby, Eq. (5.59) holds only if $\boldsymbol{z}_1 = 0$ and $\boldsymbol{z}_2 = 0$. As both Λ_H^T and $\bar{\Lambda}_A^T$ are full column matrix, $\boldsymbol{z}_3 = 0$, and $\boldsymbol{z}_4 = 0$. The solution to Eq. (5.55) is $[\boldsymbol{z}_1^T \quad \boldsymbol{z}_2^T \quad \boldsymbol{z}_3^T \quad \boldsymbol{z}_4^T] = \mathbf{0}$. $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$ is invertible. Proposition 5.8 is proved.

$\partial \tilde{\mathbf{C}}(\mathbf{v}^*)/\partial \mathbf{v}$ is positive definite is a sufficient condition to ensure that the solution $\mathbf{v}^*$ of VI problem (5.24) is locally unique. The two conditions for $\partial \tilde{\mathbf{C}}(\mathbf{v}^*)/\partial \mathbf{v}$ being a positive definite matrix are discussed in Proposition 5.7. Note that the two conditions are not strong. Thereby, $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$ is invertible in general at the equilibrium state. It is important to note that if $\mathbf{v}^*$ is not locally unique, then the gradients in the left-hand side of Eq. (5.51) do not exist (Tobin, 1986). When applying the sensitivity analysis method in practice, rather than checking whether $\partial \tilde{\mathbf{C}}(\mathbf{v}^*)/\partial \mathbf{v}$ is positive definite, a simple way to check invertibility of matrix $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$ is to determine whether it is a full rank matrix at the equilibrium state. According to Eq. (5.51)

$$\nabla_{\boldsymbol{\varepsilon}} \mathbf{v}_H^*(0) = \Delta_H \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}_H^*(0) \tag{5.60a}$$

$$\nabla_{\boldsymbol{\varepsilon}} \mathbf{v}_A^*(0) = \bar{\Delta}_A \nabla_{\boldsymbol{\varepsilon}} \bar{\mathbf{f}}_A^*(0) \tag{5.60b}$$

Eq. (5.51) is derived based on the ELI path set $\bar{R}_A$ and the assumption of fixed flow of ELD paths when $\boldsymbol{\varepsilon}$ changes. Following the proof of Yang and Bell (2005), it can be shown that the gradients $\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_H^*(0)$ and $\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_A^*(0)$ are independent with respect to the ELI path set $\bar{R}_A$ and the corresponding flows of ELD paths. Thereby, the gradients $\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_H^*(0)$ and $\nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_A^*(0)$ exist if the condition in Proposition 5.8 is satisfied. The perturbed equilibrium link flow for HDVs and CAVs can then be estimated using the first-order approximation (FOA) approach as follows:

$$\hat{\mathbf{v}}_H^* = \mathbf{v}_H^*(0) + \delta_{\boldsymbol{\varepsilon}} \times \nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_H^*(0) \tag{5.61a}$$

$$\hat{\mathbf{v}}_A^* = \mathbf{v}_A^*(0) + \delta_{\boldsymbol{\varepsilon}} \times \nabla_{\boldsymbol{\varepsilon}}\mathbf{v}_A^*(0) \tag{5.61b}$$

where $\delta_{\boldsymbol{\varepsilon}}$ is the scale of perturbation.

5.6 Numerical analysis

5.6.1 Convergence performance of the solution algorithm

This section investigates the convergence performance of the proposed RSRS-MSRA algorithm using the Nguyen-Dupuis (Figure 5.2) and Sioux Falls networks (Figure 5.3). The Nguyen-Dupuis network is a small network with four OD pairs, i.e., W={1-2;1-3;4-2;4-3}. The OD demands for the four OD pairs are 1320, 990, 820, and 990, respectively. The other inputs for the network can be found in Table 5.2. The Sioux Falls network is a larger network consists of 24 nodes, 76 links and 552 O-D pairs. The inputs for the Sioux Falls network can be found in Leblanc (1973).

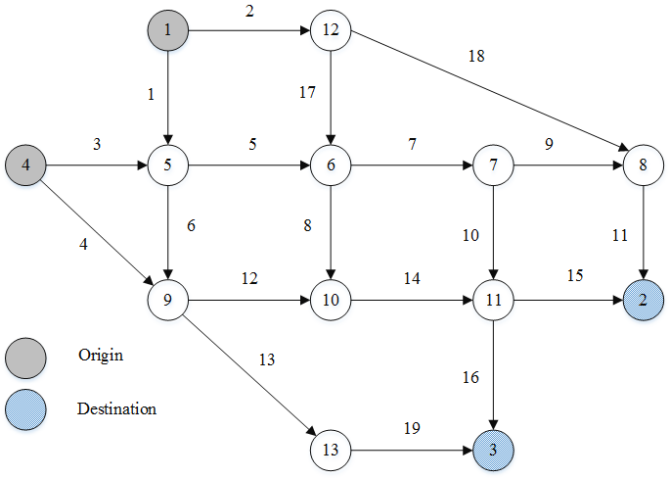


Figure 5.2 Nguyen-Dupuis network

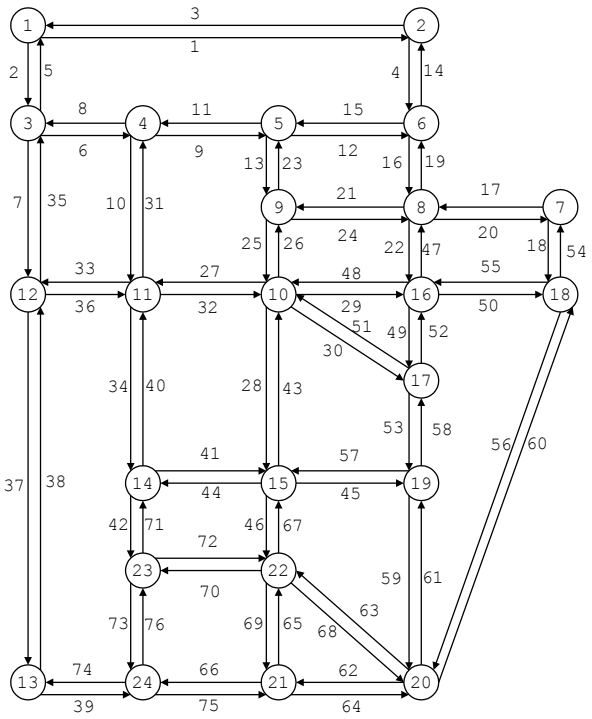


Figure 5.3 Sioux Falls network

The CAV demand for each OD pair for the two networks is assumed to be 50%. The dispersion parameter and the nesting degree for both networks are $\theta = 0.5, u = 0.5$. $Q_{a,A} = 2Q_{a,H}, \forall a \in \boldsymbol{\Gamma}_H$. The VOT for HDVs and CAVs is set as 10 and 5, respectively. The revised K-shortest path method (De La Barra et al., 1993) is used to find the paths that are likely to be used for each OD pair. It identifies about 3400 such routes in the Sioux Falls network. To estimate the gasoline consumption of a vehicle on link $a \in \boldsymbol{\Gamma}_H$, the following model (Zhang et al., 2014) which is calibrated using field data of 40 gasoline vehicles will be used in this study:

$$E_a = 147.92\left(\frac{l_a}{\bar{t}_a(v_{a,H}, v_{a,A})}\right)^{-0.689}\frac{1.609 * l_a}{3.785} = 62.88\frac{(l_a)^{0.311}}{\bar{t}_a(v_{a,H}, v_{a,A})} \tag{5.62}$$

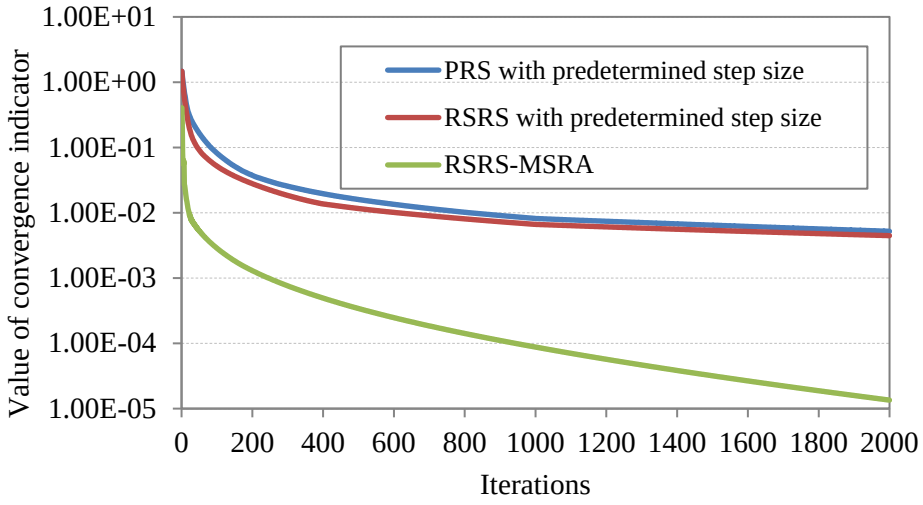
where 1.609 is the conversion rate of mile to kilometer, 3.785 is conversion rate of gallon to liter. The travel cost for HDVs and CAVs on link $a, a \in \boldsymbol{\Gamma}_H$ is

$$t_{a,z} = \bar{t}_a\left(v_{a,A}\right) \cdot VOT_z + 3 \cdot 53.69\frac{(l_a)^{0.311}}{\bar{t}_a\left(v_{a,H}, v_{a,A}\right)}, a \in \boldsymbol{\Gamma}_H, z \in \{HAV, CAV\} \tag{5.63}$$

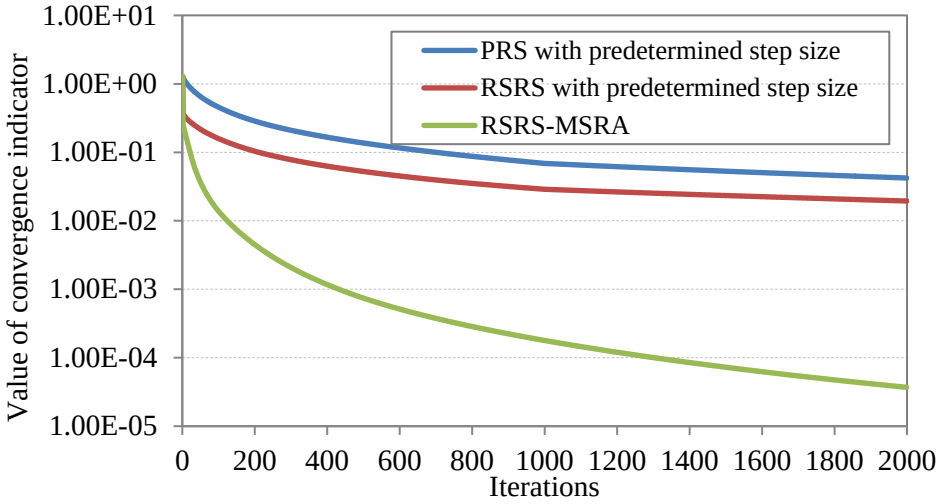
where the link travel time $\bar{t}_a\left(v_{a,H}, v_{a,A}\right)$ is shown in Eq. (5.12a).

Table 5.2 Inputs for Nguyen-Dupuis network

Link	l_a	Q_H	Q_A	Link	l_a	Q_H	Q_A
1	7	300	600	11	9	500	1000
2	9	200	400	12	10	550	1100
3	9	200	400	13	9	200	400
4	12	200	400	14	6	400	800
5	3	350	700	15	9	300	600
6	9	400	800	16	8	300	600
7	5	500	1000	17	7	200	400
8	13	250	500	18	14	300	600
9	5	250	500	19	11	200	400
10	9	300					



(a) Convergence results for Nguyen-Dupuis network



(b) Convergence results for Sioux Falls network

Figure 5.4 Convergence results for the three algorithms

To measure the quality of the solutions, the convergence indictor (denoted as G) is formulated follows

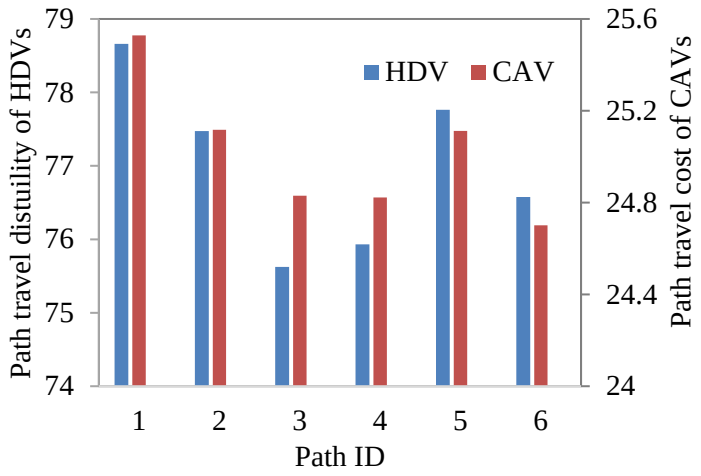
$$G = \frac{\sum_{w\in W_H}\sum_{k\in R_H^w} f_{k,H}^w\left(C_{k,H}^w - C_{min,H}^w\right) + \sum_{w\in W_A}\sum_{k\in R_A^w} f_{k,A}^w\left(C_{k,A}^w - C_{min,A}^w\right)}{\sum_{w\in W_H}\sum_{k\in R_H^w} f_{k,H}^w C_{k,H}^w + \sum_{w\in W_A}\sum_{k\in R_A^w} f_{k,A}^w C_{k,A}^w} \quad (55)$$

where $C_{min,H}^w$ and $C_{min,A}^w$ are minimum generalized path costs for HDVs and CAVs, respectively, for OD pair w. For comparison, the PRS and RSRS algorithm with predetermined step sizes proposed by Huang and Lam (2002) are also used to solve the multiclass traffic assignment model. The predetermined step sizes are set as $0.001 \times (1^{(1\to 499)}, 1/2^{(500\to 999)}, 1/3^{(1000\to 1499)} \cdots)$. These step sizes are carefully chosen as having the best convergence performance. For the RSRS-MSRA method, the two parameters Υ_1 and Υ_2 are set as 2 and 0.01, respectively. The three algorithms start from the same initial point. The experiments were coded in MATLAB and executed on a computer with an Intel Core i7-4790 3.60-GHz CPU with 8GB RAM. Figure 5.4 shows that the convergence performance of the RSRS algorithm with predetermined step sizes is better than that of the PRS algorithm with the same step sizes for both test networks. However, both algorithms performed very poorly in that the stop criterion cannot reach 0.001 after even 2000 iterations. The RSRS-MSRA algorithm substantially improves the convergence performance by adaptively choosing the step size in each iteration. As illustrated in Figure 5.4, the RSRS-MSRA algorithm takes around 1000 iterations for the Nguyen-Dupuis and 1400 iterations for the Sioux Falls network, respectively, to achieve a value of convergence indicator less than 0.0001. As the RSRS-MSRA algorithm updates the path flow in each iteration very efficiently, it only takes about 2.3 seconds and 92.2 seconds to achieve this convergence performance for the Nguyen-Dupuis and Sioux Falls networks, respectively.

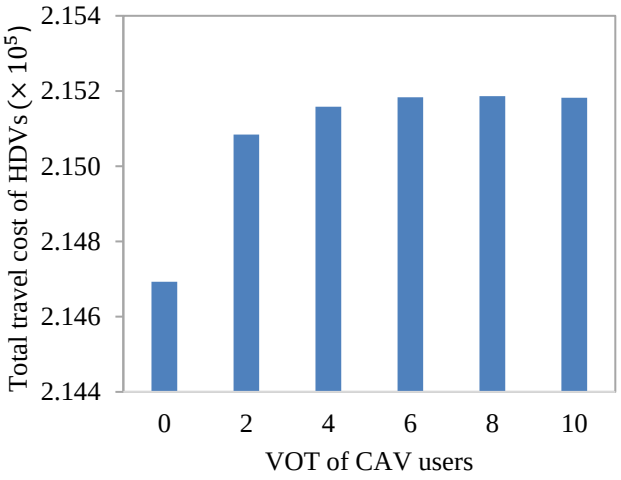
5.6.2 Impacts of CAVs on network performance

The Nguyen-Dupuis network is used to test the impacts of CAVs on network performance. Figure 5.6(a) compares the travel costs of all the paths for OD pair 1-3 (see Table 5.3) at the equilibrium state when the VOT of HDVs and CAVs is 10 and 5, respectively. It indicates that the paths with minimum travel cost for HDVs and CAVs are different (i.e., path 3 for HDVs and path 6 for CAVs). Hence, a large proportion of HDVs and CAVs for OD pair 1-4 are distributed on different paths, which reduces the network congestion level. However, as indicated by proposition 5.3, when the VOT of HDVs and CAVs is equal, the routes with minimum travel

cost for HDVs and CAVs are the same. Then, a large number of HDVs and CAVs for an OD pair would use the same paths, leading to more traffic congestion. Figure 5.5(b) shows that as the VOT of CAVs increases, the total travel cost of HDVs also increases monotonically.



(a) Travel cost of paths between OD pair 1-4 for HDVs and CAVs



(b) Total travel cost of HDVs under different VOT of CAV users

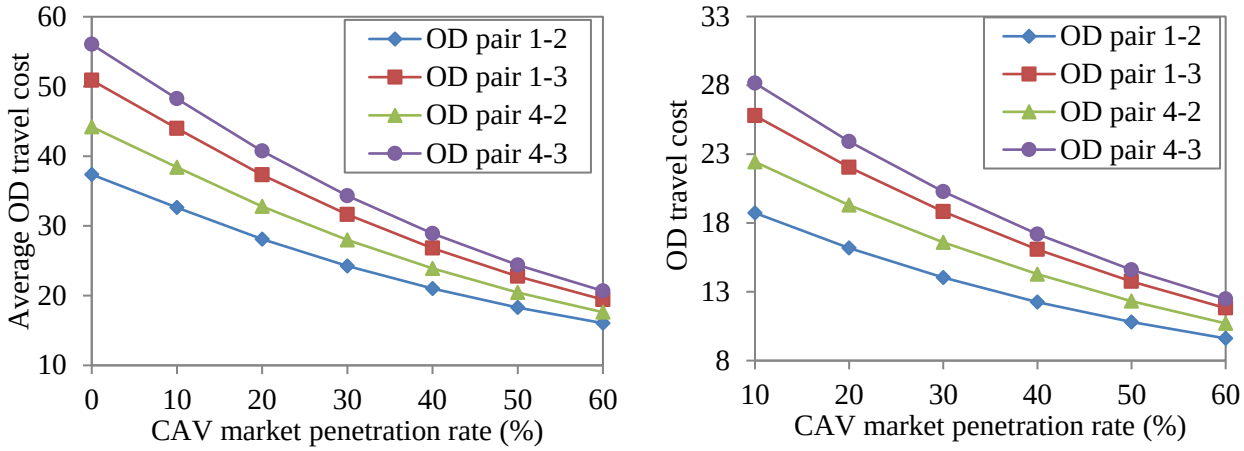
Figure 5.5 Network performance at the equilibrium state

To analyze the impacts of the CAV market penetration rate, Figure 5.6 shows the average OD travel costs of HDVs and CAVs under different CAV market penetration rates. It suggests that even a small percentage of CAVs (e.g., 10%) can significantly reduce the OD travel cost of both HDVs and CAVs. In addition, the average OD travel cost of both CAVs and HDVs

decreases monotonically as the market penetration rate of CAVs increases, because it can increase the link capacity to reduce travel cost.

Table 5.3 Route-link incidence relationship for OD pair 1-3

Path ID	Node sequence
1	1-12-6-10-11-3
2	1-12-6-7-11-3
3	1-5-9-13-3
4	1-5-9-10-11-3
5	1-5-6-10-11-3
6	1-5-6-7-11-3



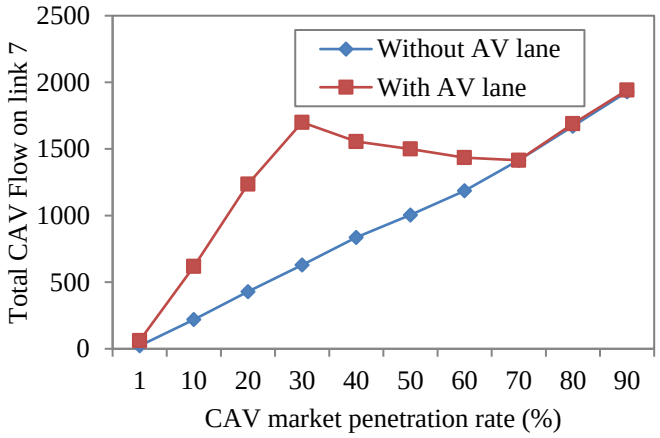
(a) Average OD travel cost for HDVs

(b) OD travel cost for CAVs

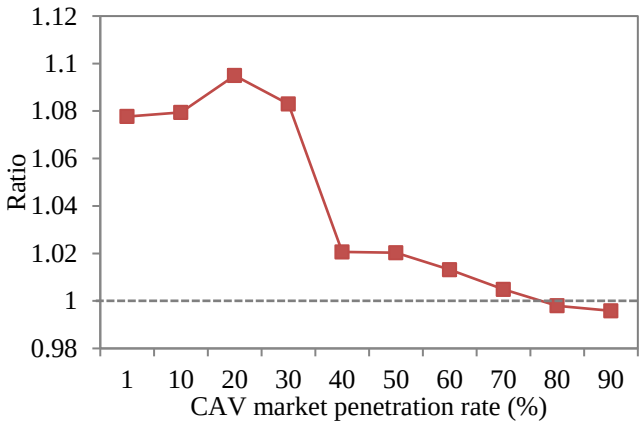
Figure 5.6 Average OD travel cost for HDVs and CAVs

Suppose one of the lanes on link 7 is converted to an AV-dedicated lane. Figure 5.7(a) shows the total CAV flow on link 7 before and after the deployment of AV-dedicated lane. It indicates that the deployment of the AV-dedicated lane can attract more number of CAVs to link 7, especially when the market penetration rate of CAVs is low. To analyze the impact of the AV-dedicated lane on network performance, Figure 5.7(b) shows the ratio of the total travel costs (TTC) (i.e., the summation of travel costs of all CAV and HDV users in the network) after the deployment of the AV-dedicated lane and before the deployment. It suggests that the deployment of an AV-dedicated lane can reduce the network performance when the market penetration rate of CAVs is low as the usage of the AV-dedicated lane is low. Thereby, it is necessary to design an effective strategy to deploy an AV-dedicated lane optimally to improve the system

performance under different CAV market penetration rates. The multiclass traffic assignment model and sensitivity analysis method proposed in this study can help to achieve this goal.



(a) Comparison of total CAV flow on link 7



(b) Ratio of total travel cost (TTC) after the deployment of AV dedicated lane over it is before the deployment of AV dedicated lane.

Figure 5.7 Comparison of total CAV flow and total travel cost before and after deployment of the AV dedicated lane

5.6.3 Sensitivity analysis

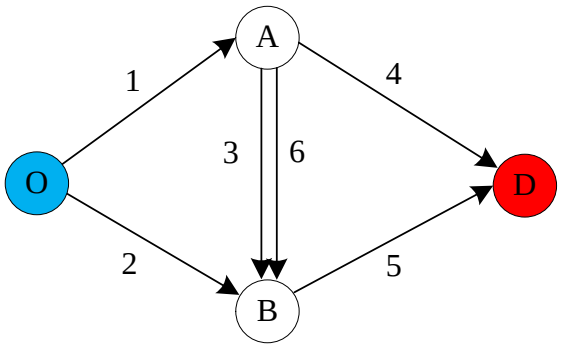


Figure 5.8 Network for demonstrating sensitivity analysis

This section presents two examples to demonstrate the application of sensitivity analysis for the multiclass traffic assignment model. We use the first example, constructed on a small network shown in Figure 5.8, to illustrate the details of and insights into the sensitivity analysis process. It contains one OD pair, 5 regular links (i.e., links 1, 2, 3, 4 and 5) and one AV-dedicated lane (i.e., link 6). There are three paths for HDVs and four paths for CAVs. The paths for HDVs are represented by link chain as: path 1:{1,4}, path 2: {2,5} , path 3: {1,3,5}, and the additional path for only CAVs, i.e., path 4: {1,6,5}. The VOT for HDVs and CAVs is set as 10 and 5, respectively. The OD demand for HDVs and CAVs is 30 and 20, respectively. The dispersion parameter and degree of nesting are $\theta = 1, u = 0.5$, respectively. The other inputs for this network can be found in Table 5.4. It is important to note that the capacity of the AV-dedicated lane (i.e., link 6) for HDVs is set to a very small value to prevent the HDVs from using the AV-dedicated lane.

Table 5.4 Inputs for study network in Figure 5.8

Links	1	2	3	4	5	6
Length of links (l_a)	4	7	1	5	2	1
Speed limit (s_a)	50	50	50	50	50	50
Capacity of HDVs ($Q_{a,H}$)	20	20	10	20	20	0.0001
Capacity of CAVs ($Q_{a,A}$)	40	40	20	20	40	60

The equilibrium path flow and link flow solutions for HDVs and CAVs are computed as

$$\mathbf{f}_H^* = [10.358 \quad 11.249 \quad 8.392]^T; \mathbf{f}_A^* = [6.304 \quad 9.834 \quad 0 \quad 3.861]^T;$$

$$\mathbf{v}_H^* = [18.751 \quad 11.249 \quad 8.392 \quad 10.358 \quad 19.642 \quad 0]^T;$$

$$\mathbf{v}_A^* = [10.166 \quad 9.834 \quad 0 \quad 6.304 \quad 13.696 \quad 3.861]^T$$

Note that path 3 is a non-equilibrated route for CAVs, and the link-path and OD-path incidence vector of the other three routes are linearly independent. Then $\bar{\mathbf{f}}_A^* = [6.304 \quad 9.834 \quad 3.861]^T$, and

$$\bar{\Delta}_A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}; \bar{\Lambda}_A = [1 \quad 1 \quad 1]; \Delta_H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}; \Lambda_H = [1 \quad 1 \quad 1]$$

Suppose the OD demand for HDVs increases by 5. Then $\varepsilon = q_H$, and

$$\boldsymbol{J}_{\boldsymbol{\varepsilon}}^* = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0]^T \tag{5.64}$$

To calculate $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$, the matrix $\boldsymbol{\rho}^*$ is obtained first by Eq. (5.37) as follows

$$\boldsymbol{\rho}^* = \begin{bmatrix} 0.805 & 0 & 0.194 \\ 0 & 0.911 & 0.089 \\ 0.240 & 0.120 & 0.641 \end{bmatrix} \tag{5.65}$$

According to Eq. (5.38) and Eq. (5.54), the gradients of generalized path travel cost for CAVs and HDVs with respect to path flow of CAVs and HDVs are computed as

$$\begin{aligned} \frac{\partial \boldsymbol{C}_H^*}{\partial \mathbf{f}_H} &= \Delta_H^T \cdot \nabla_{\mathbf{v}_H} \boldsymbol{t}_H^* \cdot \Delta_H - (1-u)\boldsymbol{\rho}^* \Delta_H^T \cdot \nabla_{\mathbf{v}_H} \boldsymbol{t}_H^* \cdot \Delta_H + \frac{u}{\theta} diag\left(\frac{\mathbf{1}}{\bar{\mathbf{f}}_H^*}\right) \\ &= \begin{bmatrix} 0.154 & -0.011 & 0.069 \\ -0.007 & 0.151 & 0.047 \\ 0.088 & 0.064 & 0.242 \end{bmatrix} \end{aligned} \tag{5.66}$$

$$\begin{aligned} \frac{\partial \boldsymbol{C}_H^*}{\partial \bar{\mathbf{f}}_A} &= \Delta_H^T \cdot \nabla_{\mathbf{v}_A} \boldsymbol{t}_H^* \cdot \bar{\Delta}_A - (1-u)\boldsymbol{\rho}^* \Delta_H^T \cdot \nabla_{\mathbf{v}_A} \boldsymbol{t}_H^* \cdot \bar{\Delta}_A \\ &= \begin{bmatrix} 0.053 & -0.005 & 0.036 \\ -0.004 & 0.053 & 0.024 \\ 0.044 & 0.032 & 0.081 \end{bmatrix} \end{aligned} \tag{5.67}$$

$$\frac{\partial \boldsymbol{C}_A^*}{\partial \mathbf{f}_H} = \bar{\Delta}_A^T \cdot \nabla_{\mathbf{v}_H} \boldsymbol{t}_A^* \cdot \Delta_H = \begin{bmatrix} 0.121 & 0 & 0.097 \\ 0 & 0.121 & 0.065 \\ 0.097 & 0.065 & 0.163 \end{bmatrix} \tag{5.68}$$

$$\frac{\partial \boldsymbol{C}_A^*}{\partial \bar{\mathbf{f}}_A} = \bar{\Delta}_A^T \cdot \nabla_{\mathbf{v}_A} \boldsymbol{t}_A^* \cdot \bar{\Delta}_A = \begin{bmatrix} 0.061 & 0 & 0.049 \\ 0 & 0.061 & 0.033 \\ 0.049 & 0.033 & 0.081 \end{bmatrix} \tag{5.69}$$

Submit Eq. (5.58) into Eq. (5.51). We can obtain the value of $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$. It can verify that the matrix $\boldsymbol{J}_{\bar{\mathbf{f}}}^*$ is invertible as it is a non-singular square matrix. Thereby, the gradients $\nabla_{\boldsymbol{\varepsilon}} \mathbf{f}_H^*(0)$, $\nabla_{\boldsymbol{\varepsilon}} \bar{\mathbf{f}}_A^*(0)$, $\nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\mu}_H(0)$, $\nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\mu}_A(0)$ exist. According to Eq. (5.51)

$$\begin{bmatrix} \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}_H^*(0) \\ \nabla_{\boldsymbol{\varepsilon}} \bar{\mathbf{f}}_A^*(0) \\ \nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\mu}_H(0) \\ \nabla_{\boldsymbol{\varepsilon}} \boldsymbol{\mu}_A(0) \end{bmatrix} = [0.381 \quad 0.407 \quad 0.212 \quad 0.710 \quad 0.456 \quad -1.166 \quad -0.062 \quad -0.053]^T \tag{5.70}$$

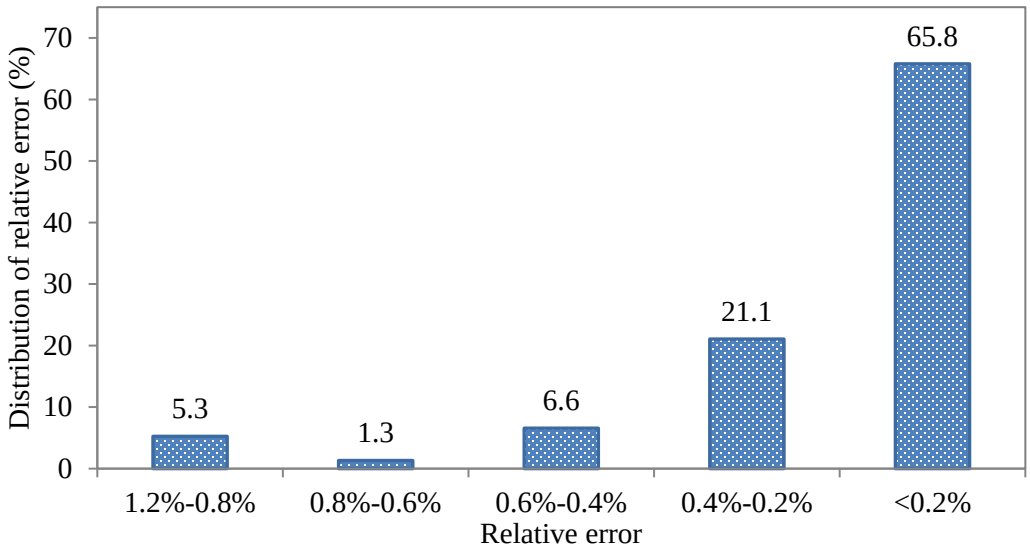
If the OD demand for HDVs is increased by 5, the perturbed equilibrium link flows of HDVs (denoted as $\hat{\mathbf{v}}_H^*(5)$) and CAVs (denoted as $\hat{\mathbf{v}}_A^*$ (5)) can be estimated by the first-order approximation (FOA) method as follows:

$$\hat{\mathbf{v}}_H^*(5) = \mathbf{v}_H^*(0) + 5 \times \Delta_H \cdot \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}_H^*(0)$$
$$= [20.529 \quad 12.471 \quad 9.030 \quad 11.499 \quad 21.501 \quad 0]^T$$
$$\hat{\mathbf{v}}_A^*(5) = \mathbf{v}_A^*(0) + 5 \times \Delta_A \cdot \nabla_{\boldsymbol{\varepsilon}} \mathbf{f}_A^*(0) = [8.798 \quad 11.202 \quad 0 \quad 8.435 \quad 11.565 \quad 0.363]^T$$

The calculated equilibrium link flows of HDVs (denoted as $\mathbf{v}_H^*(5)$) and CAVs (denoted as $\mathbf{v}_A^*(5)$) using the solution algorithm are

$$\mathbf{v}_H^*(5) = [20.514 \quad 12.486 \quad 8.996 \quad 11.518 \quad 21.482 \quad 0]^T$$
$$\mathbf{v}_A^*(5) = [8.964 \quad 11.036 \quad 0 \quad 8.181 \quad 11.819 \quad 0.784]^T$$

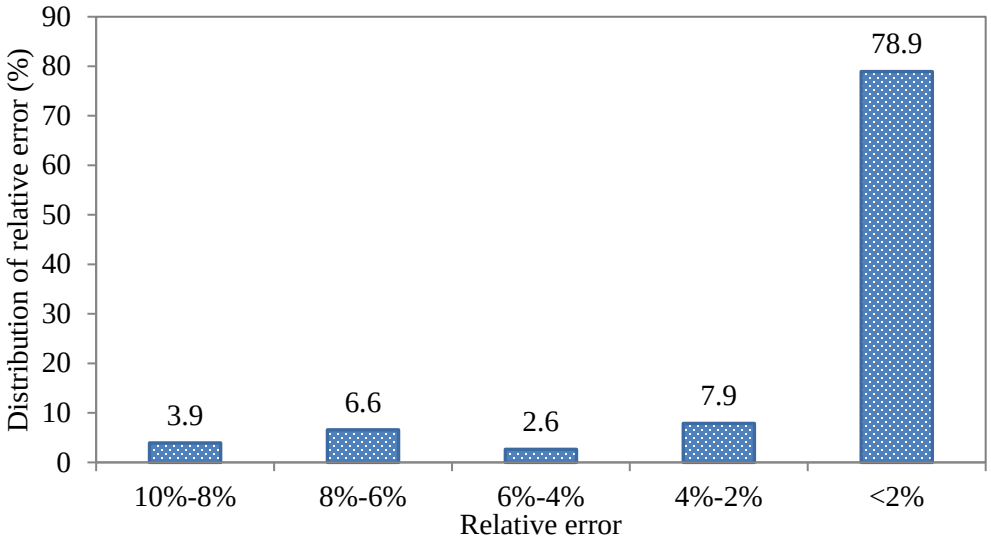
It can be seen that the estimated equilibrium link flows are very close to the calculated ones. Thereby, FOA approach can effectively estimate the perturbed solutions. It should be noted that the accuracy of the FOA approach decreases as the perturbation increases.



(a) The distribution of relative errors of the estimated link flow of HDVs

Figure 5.9 Relative errors of the estimated link flow

Figure 5.9 continued



(b) The distribution of relative errors of the estimated link flow of CAVs

The Sioux Falls network shown in Figure 5.3 is used to demonstrate the estimation performance of equilibrium flows of the FOA approach due to parametric perturbations in a larger network. Assume the demand for both CAVs and HDVs for OD pairs 10-15,15-10,14-15,15-14, 19-15 and 15-19 are increased by 100%. Figure 5.9 shows the distribution of relative errors of estimated link flows for HDVs and CAVs. It illustrates that the FOA approach accurately estimates the HDV and CAV flows for most links with relative errors less than 0.2% and 2%, respectively. Further, the maximum relative errors of the estimated link flows for HDVs and CAVs are less than 1.2% and 10%, respectively. Thereby, the FOA approach accurately captures the flow variation due to the demand increase. Note that the estimation performance of CAV link flows is poorer than that of the HDV link flows. This is because CAVs choose routes based on the UE principle. They are more informed of traffic conditions than HDVs and are more sensitive to the network congestion due to increased demand. Thereby, the variation of link flows for CAVs is larger than that for HDVs, which reduces the estimation performance of the FOA approach.

5.7 Concluding comments

This study proposes a multiclass traffic assignment model in which HDV and CAV users choose routes based on the CNL model and UE principle, respectively. The CNL model captures HDV

users' perception uncertainty due to limited knowledge of traffic conditions while overcoming the route overlap issue of logit-based SUE. The UE model can characterize the CAV's capability for acquiring traffic information more accurately. In addition, the asymmetry in interactions involving HDVs and CAVs is analytically captured by the designed link travel cost functions. Thereby, the proposed multiclass traffic assignment model can enhance realism in characterizing mixed traffic flows. It can aid planners to quantitatively estimate the impacts of the VOT of CAV users, the CAV market penetrate rate, and the deployment of AV-dedicated lanes. A new route-swapping-based solution algorithm, RSRS-MSRA, is developed to solve the multiclass traffic assignment model effectively. It converges much faster than the existing route-swapping-based algorithm (Huang and Lam, 2002) by adaptively determining the step size in each iteration. This solution algorithm can also be used to solve other path-based VI problems (e.g., Huang and Lam, 2004, Szeto and Lo, 2006; Ramadurai and Ukkusuri, 2010). The study also develops an analytical model for sensitivity analysis of the multiclass traffic assignment model.

The study results suggest that when the VOT of CAV users is small, a large proportion of HDV and CAV demand will be distributed on different routes, which reduces the total travel cost of HDVs. The deployment of AV-dedicated lanes can attract more CAV flow than non-AV dedicated lanes. But changing an existing lane into an AV-dedicated lane may decrease system performance under a low CAV market penetration rate. In addition, the average OD travel cost of HDVs can be reduced significantly even with a relatively low CAV market penetration rate. The analytical model for sensitivity analysis of the multiclass traffic assignment model enables planners to quickly estimate the perturbed network flows due to expected or unexpected events. These insights can assist decision-makers to design effective planning and operational strategies that promote the benefits of CAVs and mitigate traffic congestion under mixed traffic flows during the transition to a fully autonomous and connected transportation system.

This study can be extended in a few directions. First, we will leverage the sensitivity analysis method to identify critical parameters (link capacity, signal splits) that impact network performance, and perform risk analysis to generate insights on network performance reliability. Second, a continuous network design problem can be developed upon the multiclass traffic assignment model to determine the optimal signal timing and tolling strategies to maximize system performance under different CAV market penetration rates. Third, a combined modal split and multiclass traffic assignment model can be developed to simultaneously estimate the

network flows and OD demand of both HDVs and CAVs by incorporating factors such as travel cost, price of vehicles, etc.

CHAPTER 6. CONCLUSIONS AND FUTURE WORK

This chapter summarizes the contributions of this dissertation, and suggests directions for future research. Section 6.1 summarizes the research and discusses associated conclusions. Section 6.2 highlights the significance of the research, and Section 6.3 discusses possible extensions for future research.

6.1 Summary and conclusions

This dissertation systematically addresses modeling needs in three connected topics to improve transportation systems by leveraging the advantages of CAVs. The topics are: (1) modeling and controlling information flow propagation, (2) designing a deployable cooperative control strategy for platoons of CAVs, and (3) modeling network equilibrium under mixed traffic flow. This dissertation deepens our understanding of three research questions:

1. How is traffic information propagated in the network spatiotemporally, and how can multiclass information (i.e., routing, safety-related information, work zones, etc.) be controlled so that performance in terms of information spread, time delay bounds and spatial coverage is satisfied for each information class?
2. How can a real-time, deployable cooperative control mechanism be designed for CAV platoons to maximize platoon performance?
3. How can the equilibrium network flow be estimated for CAVs and HDVs by incorporating the characteristics of travel costs and interactions of route choices?

Chapter 2 explores the spatiotemporal information flow propagation under information congestion effects by introducing a two-layer macroscopic model and an information relay control strategy that propagates the received information packets according to the first-in, first-out queue discipline. It trades off the need to enable the dissemination of every information packet as far as possible against the congestion effects that will accrue due to the presence of multiple information packets. An IDE system is established in the upper layer to model information dissemination in the information flow regime. The LWR model is used in the lower layer to capture the impacts of traffic flow dynamics on information propagation. The two-layer model is used to derive the analytical solution for asymptotic IFPW speed and density of

vehicles informed with the specific information of interest. Numerical experiments show that the IFPW speed decreases as the information packet exclusion rate increases, to prevent packet collisions under information congestion conditions. A large information packet exclusion rate can cause information to be propagated only locally, implying that equipped vehicles far from the sender vehicle will not be informed. This study can be leveraged to develop a new generation of information dissemination strategies focused on enabling specific V2V information to reach target locations at desired times.

Chapter 3 proposes a queuing strategy to control the spatiotemporal propagation of multiclass information. The queuing strategy assigns a certain number of virtual "communication servers" for each information class. Two control parameters—the number of assigned communication servers and the communication service rate—are determined for each information class to achieve the desired propagation performance related to information spread, time delay bounds, and spatial coverage. Similarly, a two-layer analytical model is derived to characterize the spatiotemporal propagation of multiclass information under the designed queuing strategy. The two-layer model captures the impact of communication constraints (communication frequency, channel capacity, and communication range, etc.) through a communication kernel which is calibrated using NS-3 simulation. Analytical and numerical solutions are also proposed to solve the two-layer model, quantifying the effects of the two control parameters on IFPW speed, density of informed vehicles, and information propagation distance. The proposed information flow propagation control strategy trades the limited communication resources and the different application needs of multiclass information. Therein, it can be useful for traffic operators to design effective control strategies for delivering multiclass traffic information in the network.

In Chapter 4, an idealized MPC strategy is proposed to coordinate the behavior of the vehicles in a platoon effectively to maneuver them under a common goal on the strong assumption that the embedded optimal control problem can be solved instantaneously. To address the issue of control delay of the idealized MPC approach, a DMPC approach is proposed, which reserves sufficient time to solve the optimal control problem before each sampling time instance. To reduce the deviation of control decisions between DMPC approach and the idealized MPC strategy, a DMPC-FOA approach is developed to improve the estimation performance of the DMPC approach. The DMPC-FOA approach not only addresses the issue of control delay of the idealized MPC strategy effectively, but also can accurately characterize the optimal control

decision of the idealized MPC strategy. Application of the DMPC-FOA approach for CAV platoons with real-world trajectory data for the leading vehicle shows that it can dampen traffic oscillations effectively, leading to smooth deceleration and acceleration behavior for the following vehicles in the platoon.

Chapter 5 proposes a multiclass traffic assignment model in which HDV and CAV users choose routes that follow the CNL and UE principles, respectively. The CNL model captures HDV users perceptual uncertainty associated with limited knowledge of traffic conditions while overcoming the route overlap issue of the logit-based SUE problem. The UE model can characterize the CAVs' capability for acquiring accurate information on traffic conditions. In addition, the asymmetry in interactions involving HDVs and CAVs is analytically captured by the designed link travel cost functions. The study results suggest that when the VOT of CAV users is small, a large proportion of HDV and CAV demand will be distributed on different routes, which reduces the total travel cost of HDVs. The deployment of AV-dedicated lanes can attract more CAV flow than non-AV dedicated lanes. But changing an existing lane into an AV-dedicated lane may decrease system performance under a low CAV market penetration rate. In addition, the average OD travel cost of HDVs can be reduced significantly even with a relatively low CAV market penetration rate. The proposed multiclass traffic assignment model provides behavioral realism in modeling the mixed traffic flow. The multiclass traffic assignment model can be used to help planners to quantitatively estimate the impacts of VOT on CAV users, the market penetration rate of CAVs, and the deployment of AV-dedicated lane on network flows.

6.2 Contributions summary

The primary contributions of this dissertation are as follows:

First, this dissertation develops an analytical macroscopic model to characterize the IFPW under congested V2V communication environments, and the IFPW under an effective queuing strategy for controlling multiclass information flow propagation. Analytical and numerical solutions are derived to solve the macroscopic models under different traffic flow conditions. These solutions help to analyze the impacts of traffic flow dynamics and the communication constraints (e.g., communication frequency, channel capacity, and communication range) on information propagation performance related to IFPW speed, density of informed vehicles and spatial coverage. Further, it also provides useful information on how to leverage the two

parameters in the designed queuing strategy to control the propagation performance of multiclass information. The macroscopic models and the queuing strategy for information propagation can be used by traffic managers to disseminate traffic information to control and guide traffic to improve the system performance under connected and autonomous transportation.

Second, this dissertation develops a real-time deployable control mechanism for platoons of CAVs to coordinate the behaviors of all CAVs effectively to maximize platoon performance. Compared to non-cooperative car-following controllers, our proposed control mechanism for CAV platoons can enhance its stability, leading to smoother acceleration and deceleration behavior for all following CAVs. However, the computational load of the embedded optimal control problem can induce significant control delays which can deteriorate platoon performance and cause traffic accidents. The proposed real-time deployable control mechanism bridges this gap for the first time in the literature by addressing the control delay issue effectively. It can be applied in real-time to coordinate the behaviors of CAVs in a platoon effectively to maneuver them under a common goal. In addition, the method for sensitivity analysis of the optimal control problem is analytically formulated; it can quantitatively measure the impact of parametric perturbations (e.g., perturbations of the initial state of the leading vehicle) on the optimal control decisions and platoon performance. Further, an analytical method is provided for stability analysis of the idealized MPC strategy, helping to identify parameter inputs in the idealized MPC strategy to better dampen the oscillations in the platoon.

Third, this dissertation develops a multiclass traffic assignment model for mixed traffic flow of HDVs and CAVs, which can capture the characteristics of mixed traffic flow, such as the difference in value-of-time between HDVs and CAVs, the asymmetry in their driving interactions, and the impacts of AV-dedicated lanes; thereby enhancing behavioral realism in the modeling. The proposed multiclass traffic assignment model can estimate the distribution of network flow in the transition period for different market penetration rates of CAVs, which can assist transportation decision-makers in designing effective planning and operational strategies to leverage the advantages of CAVs to manage traffic congestion under mixed traffic flows. This dissertation also develops a new solution algorithm based on the route-swapping concept to solve the proposed multiclass traffic assignment model efficiently. It has a convergence speed superior to that of the existing route-swapping-based algorithm (Huang and Lam, 2002). Further, an analytical formula is derived for sensitivity analysis of the proposed model, which enables

planners to quickly estimate the perturbed traffic equilibrium and identify critical elements under planned or unexpected disruptive events. It can also be used to solve the continuous network design problem (e.g., to find optimal signal timing or tolling strategies to improve system performance under mixed traffic flow).

6.3 Directions for future work

In this dissertation, we have developed several models to address both operational and planning needs for CAVs, including information dissemination through V2V communication, cooperative control mechanism for CAV platoons, and network equilibrium modeling for mixed traffic flow. Although the topics addressed in this dissertation represent a first step to exploit the potential of CAVs, we believe they enhance the understanding and shed light on a wide range of related future research directions. Here, we identify some potential directions for future research.

First, this dissertation models and addresses control of information flow propagation in the context of only a corridor. Extensions and new strategies will need to be proposed to control network-level multiclass information propagation. In addition, this dissertation assumes that information is relayed by vehicles through multi-hop broadcasting V2V communication. The modeling framework can be developed for information flow propagation using other broadcasting protocols; for example, unicast V2V communications. This dissertation assumes all information packets have equal priority to propagate. In real-world applications, different information packets may have different priorities. This motivates the exploration of other queuing strategies such as preemptive priority and non-preemptive priority strategies to control propagation of multiclass information.

Second, related to cooperative control mechanisms for CAV platoons, this study assumes that there are no measurement errors of the vehicles' states and false executions of optimal control decisions. In future work, a noise term can be added to the state dynamics equation to account for state uncertainties, and a new deployable robust control approach can be developed to control the CAV platoon. Additionally, new solution algorithms and discretization techniques can be developed to solve the optimal control problems quickly to enhance the real-time applicability of the proposed cooperative control mechanism.

Third, related to multiclass traffic assignment models for mixed traffic flow, the sensitivity analysis method can be leveraged to identify critical parameters (link capacity, signal splits) that

impact network performance, while performing risk analysis to generate insights on network performance reliability. Also, a continuous network design problem can be developed using the multiclass traffic assignment model to determine optimal signal timing and tolling strategies to maximize system performance under different CAV market penetration rates. Further, a combined modal split and multiclass traffic assignment model can be developed to simultaneously estimate the network flows and OD demand of both HDVs and CAVs by incorporating factors such as travel cost, vehicle price, and demand elasticity.

REFERENCES

Aronson, D. G., 1977. The asymptotic speed of propagation of a simple epidemic. In Nonlinear Diffusion, edited by W. E. Fitzgibbon and H. F. Walker, Research Notes in Math, Vol. 14, Pitman, London, 1-23.

Aronson, D. G., Weinberger, H. F., 1978. Multidimensional nonlinear diffusion arising in population genetics. Advances in Mathematics, 30 (1), 33-76.

Assidiq, A.A., Khalifa, O.O., Islam, M.R., Khan, S., 2008. Real time lane detection for autonomous vehicles. Proceedings of the International Conference on Computer and Communication Engineering, July, pp. 82-88, Kuala Lumpur, Malaysia.

Atkinson, C., Reuter, G. E. H., 1976. Deterministic epidemic waves. Mathematical Proceedings of the Cambridge Philosophical Society, 80 (2), 315-330.

Augustin, D., Maurer, H., 2001. Second order sufficient conditions and sensitivity analysis for the optimal control of a container crane under state constraints. Optimization, 49 (4), 351-368.

Bauschke, H.H., Matoušková, E., Reich, S., 2004. Projection and proximal point methods: convergence results and counterexamples. Nonlinear Analysis: Theory, Methods & Applications, 56 (5), 715-738.

Ben-Akiva M., Bergman M.J., Daly A.J., Ramaswamy R., 1984. Modelling inter urban route choice behaviour. Proceedings of the 9th International Symposium on Transportation and Traffic Theory, July, pp. 299-330, Utrecht, Netherlands.

Benin, J., Nowatkowski, M., Owen, H., 2012. Vehicular network simulation propagation loss model parameter standardization in ns-3 and beyond. 2012 Proceedings of IEEE Southeastcon, March, pp. 1-5, Florida, USA.

Bonnet, C., Fritz, H., 2000. Fuel consumption reduction in a platoon: Experimental results with two electronically coupled trucks at close spacing. Future Transportation Technology Conference & Exposition, August, pp. 1-9, California, USA.

Boyles, S.D., 2012. Bush-based sensitivity analysis for approximating subnetwork diversion. Transportation Research Part B: Methodological, 46 (1), 139-155.

Brigham, E. O., 1974. The fast Fourier transform (Vol. 7). Englewood Cliffs, NJ: Prentice-Hall.

Camacho, E. F., Alba, C. B., 2013. Model predictive control. Springer Science & Business Media.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified-logit route-choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. Proceedings of International Symposium on Transportation and Traffic Theory, July, pp. 697-711, Lyon, France.

Chen, R., Jin, W.-L., Regan, A., 2010. Multi-hop broadcasting in vehicular ad hoc networks with shockwave traffic. In 7th IEEE Consumer Communications and Networking Conference, January, pp. 1-5, Irvine, USA,

Chen, Z., He, F., Yin, Y., Du, Y., 2017. 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., 2016. Optimal deployment of autonomous vehicle lanes with endogenous market penetration. Transportation Research Part C: Emerging Technologies, 72, 143-156.

Clark, S., Watling, D., 2000. Probit-based sensitivity analysis for general traffic networks. Transportation Research Record: Journal of the Transportation Research Board, 1733, 88-95.

Dafermos, S.C., 1972. The traffic assignment problem for multiclass-user transportation networks. Transportation science, 6 (1), 73-87.

Daganzo, C. F., 1995. A finite difference approximation of the kinematic wave model of traffic flow. Transportation Research Part B: Methodological, 29 (4), 261–276.

Darbha, S., Rajagopal, K.R., 1999. Intelligent cruise control systems and traffic flow stability. Transportation Research Part C: Emerging Technologies, 7 (6), 329-352.

De La Barra, T., Perez, B., Anez, J., 1993. Multidimensional path search and assignment. 21st PTRC Summer Annual Meeting, pp. 307-320, University of Manchester, United Kingdom.

Dey, K.C., Rayamajhi, A., Chowdhury, M., Bhavsar, P., Martin, J., 2016. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network–Performance evaluation. Transportation Research Part C: Emerging Technologies, 68, 168-184.

Ding, J. W., Wang, C. F., Meng, F. H., Wu, T. Y., 2010. Real-time vehicle route guidance using vehicle-to-vehicle communication. IET Communications, 4 (7), 870-883.

Dorato, P., 1963. On sensitivity in optimal control systems. IEEE Transactions on Automatic Control, 8 (3), 256-257.

Du, L., Dao, H., 2015. Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream. IEEE Transactions on Intelligent Transportation Systems, 16 (1), 66-80.

Du, L., Gong, S., Wang, L., Li, X., 2016. Information-traffic coupled cell transmission model for information spreading dynamics over vehicular ad hoc network on road segments. Transportation Research Part C: Emerging Technologies, 73, 30-48.

Eben Li, S., Li, K., Wang, J., 2013. Economy-oriented vehicle adaptive cruise control with coordinating multiple objectives function. Vehicle System Dynamics, 51 (1), 1-17.

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.

Facchinei, F., Pang, J.S., 2007. Finite-dimensional variational inequalities and complementarity problems. Springer Science & Business Media.

Fiacco, A.V., 1983. Introduction to sensitivity and stability analysis in nonlinear programming. London, UK: Academic Press.

Florian, M., 1977. A traffic equilibrium model of travel by car and public transit modes. Transportation Science, 11 (2), 166-179.

Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin, R.L., Ganjalizadeh, S., 1994. Day-to-day dynamic network disequilibria and idealized traveler information systems. Operations Research, 42 (6), 1120-1136.

Gaimon, C., 2002. Optimal control theory: Applications to management science and economics. JSTOR

Gallier,J., 2011. Geometric methods and applications: for computer science and engineering (Vol. 38). Springer Science & Business Media.

Gross, D., Shortle, J. F., Thompson, J. M., Harris, C. M., 2008. Fundamentals of queuingtheory, 4th Edition. John Wiley & Sons.

Guo, R.Y., Huang, H., 2016. A discrete dynamical system of formulating traffic assignment: Revisiting Smith's model. Transportation Research Part C: Emerging Technologies, 71, 122-142.

Guo, R.Y., Yang, H., Huang, H., 2013. A discrete rational adjustment process of link flows in traffic networks. Transportation Research Part C: Emerging Technologies, 34, 121-137.

Gong, S., Du, L., 2018. Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles. Transportation Research Part B: Methodological, 116, 25-61.

Gong, S., Shen, J., Du, L., 2016. Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. Transportation Research Part B: Methodological, 94, 314-334.

Hasebe, K., Nakayama, A., Sugiyama, Y., 2003. Dynamical model of a cooperative driving system for freeway traffic. Physical Review E, 68 (2). No.026102.

Hisham, A., Ström, E. G., Brännström, F., Yan, L., 2017. Scheduling and power control for V2V broadcast communications with adjacent channel interference. arXiv preprint arXiv:1708.02444.

Hisham, A., Sun, W., Ström, E. G., Brannstrom, F., 2016. Power control for broadcast V2V communications with adjacent carrier interference effects. IEEE International Conference on Communication, May, pp.1-6, Kuala Lumpur, Malaysia.

Huang, H., Lam, W.H., 2002. Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. Transportation Research Part B: Methodological, 36 (3), 253-273.

Huang, H., Lam, W.H., 2003. A multi-class dynamic user equilibrium model for queuing networks with advanced traveler information systems. Journal of Mathematical Modelling and Algorithms, 2 (4), 349-377.

Huang, H., Li, Z., 2007. A multiclass, multicriteria logit-based traffic equilibrium assignment model under ATIS. European Journal of Operational Research, 176 (3), 1464-1477.

Jafari, E., Boyles, S.D., 2016. Improved bush-based methods for network contraction. Transportation Research Part B: Methodological, 83, 298-313.

Jia, D., Lu, K., Wang, J., Zhang, X., Shen, X., 2015. A survey on platoon-based vehicular cyber-physical systems. IEEE Communications Surveys & Tutorials, 18 (1), 263-284.

Jia, D., Ngoduy, D., 2016. Platoon based cooperative driving model with consideration of realistic inter-vehicle communication. Transportation Research Part C: Emerging Technologies, 68, 245-264.

Jiang, Y., Szeto, W.Y., Long, J., Han, K., 2016. Multi-class dynamic traffic assignment with physical queues: intersection-movement-based formulation and paradox. Transportmetrica A: Transport Science, 12 (10), 878-908.

Jin, I.G., Orosz, G., 2014. Dynamics of connected vehicle systems with delayed acceleration feedback. Transportation Research Part C: Emerging Technologies, 46, 46-64.

Jin, W.-L., Recker, W. W., 2006. Instantaneous information propagation in a traffic stream through inter-vehicle communication. Transportation Research Part B: Methodological, 40 (3), 230-250.

Jin, W.-L., Recker, W. W., 2010. An analytical model of multihop connectivity of inter-vehicle communication systems. IEEE Transactions on Wireless Communications, 9 (1), 106–112.

Ji, X., Ban, X.J., Li, M., Zhang, J., Ran, B., 2017. Non-expected route choice model under risk on stochastic traffic networks. Networks and Spatial Economics, 17 (3), 777-807.

Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., Weil, T., 2011. Vehicular networking: a survey and tutorial on requirements, architectures, challenges, standards and solutions. IEEE Transactions on Communications Surveys and Tutorials, 13 (4), 584-616.

Keller, H.B., 1976. Numerical solution of two point boundary value problems (Vol. 24). Philadelphia: Society for Industrial and Applied Mathematics.

Kendall, D. G., 1957. Discussion of 'Measles periodicity and community size' by M. S. Bartlett. Journal of the Royal Statistical Society Series A: General, 120 (1), 64-76.

Kendall, D. G., 1965. Mathematical models of the spread of infection. Mathematics and Computer Science in Biology and Medicine, 213-224.

Kesting, A., Treiber, M., Schonhof, M., Helbing, D., 2008. Adaptive cruise control design for active congestion avoidance. Transportation Research Part C: Emerging Technology, 16 (6), 668–683.

Kim, Y.H., Peeta, S., 2016. Graph-based modeling of information flow evolution and propagation under V2V communications-based advanced traveler information systems. Computer-Aided Civil and Infrastructure Engineering, 31 (7), 499-514.

Kim, Y.H., Peeta, S., 2017a. Modeling of the dynamic flow propagation of multiple units of information under vehicle-to-vehicle communications based advanced traveler information systems. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 21 (4), 310-323.

Kim, Y. H., Peeta, S., He, X., 2015. Macroscopic modeling of spatiotemporal information flow propagation wave under vehicle-to-vehicle communications. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, September, pp. 751-756, Qingdao, China,

Kim, Y.H., Peeta, S., He, X., 2018. An analytical model to characterize the spatiotemporal propagation of information under vehicle-to-vehicle communications. IEEE Transactions on Intelligent Transportation Systems,19 (1): 3-12.

Kim, Y. H., Peeta, S., He, X., 2017b. Modeling the information flow propagation wave under vehicle-to-vehicle communications. Transportation Research Part C: Emerging Technologies, 85, 377-395.

Kirk, D.E., 2012. Optimal control theory: an introduction. Courier Corporation, New York, USA.

Kitthamkesorn, S., Chen, A., Xu, X., Ryu, S., 2016. Modeling mode and route similarities in network equilibrium problem with go-green modes. Networks and Spatial Economics, 16 (1), 33-60.

Lam, W.H., Huang, H., 2003. Combined activity/travel choice models: time-dependent and dynamic versions. Networks and Spatial Economics, 3 (3), 323-347.

Lapuerta, M., J., Rodríguez-Fernández, J., Herreros, M. 2017. Gaseous and Particle Greenhouse Emissions from Road Transport. Environmental Impacts of Road Vehicles, 25-45.

Leblanc, L.J., Abdulaal, M., 1982. Combined mode split-assignment and distribution-model split-assignment models with multiple groups of travelers. Transportation Science, 16 (4), 430-442.

Levin, M.W., Boyles, S.D., 2015. Effects of autonomous vehicle ownership on trip, mode, and route choice. Transportation Research Record: Journal of the Transportation Research Board, 2493, 29-38.

Levin, M.W., Boyles, S.D., 2016. A multiclass cell transmission model for shared human and autonomous vehicle roads. Transportation Research Part C: Emerging Technologies, 62, 103-116.

Lighthill, M. J., Whitham, J. B., 1955. On kinematic waves. I: Flow movement in long river; II: A theory of traffic flow on long crowded roads. Proceedings of Royal Society A 229 (1178), 281-345.

Li, M. Y., Graef, J. R., Wang, L., Karsai, J., 1999. Global dynamics of a SEIR model with varying total population size. Mathematical Biosciences, 160 (2), 191-213.

Li, M. Y., Muldowney, J. S., 1995. Global stability for the SEIR model in epidemiology. Mathematical biosciences, 125 (2), 155-164.

Liu, H.X., He, X., He, B., 2009. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. Networks and Spatial Economics, 9 (4), 485–503.

Li, Y., Sun, D., Liu, W., Zhang, M., Zhao, M., Liao, X., Tang, L., 2011. Modeling and simulation for microscopic traffic flow based on multiple headway, velocity and acceleration difference. Nonlinear Dynamics, 66 (1-2), 15-28.

Lo, H.K., Luo, X.W., Siu, B.W., 2006. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B: Methodological, 40 (9), 792-806.

Mahmassani, H. S., Peeta, S., 1995. System optimal dynamic assignment for electronic route guidance in a congested traffic network. In Urban Traffic Networks, Springer, pp. 3–37.

Malanowski, K., 1984. Differential stability of solutions to convex, control constrained optimal control problems. Applied Mathematics and Optimization, 12 (1), 1-14.

Malanowski, K., 1987. Stability and sensitivity of solutions to optimal control problems for systems with control appearing linearly. Applied Mathematics and Optimization, 16 (1), 73-91.

Malanowski, K., 2011. Sensitivity analysis for state constrained optimal control problems. Control & Cybernetics, 40 (4), 1043-1058.

Malanowski, K., Maurer, H., 1996. Sensitivity analysis for parametric control problems with control-state constraints. Computational Optimization and Applications, 5 (3), 253-283.

Maurer, H., Pesch, H.J., 1994. Solution differentiability for nonlinear parametric control problems. SIAM Journal on Control and Optimization, 32 (6), 1542-1554.

Maurer, H., Pesch, H.J., 1995. Solution differentiability for parametric nonlinear control problems with control-state constraints. Journal of Optimization Theory and Applications, 86 (2), 285-309.

Mayne, D. Q., Rawlings, J. B., Rao, C. V., Scokaert, P. O., 2000. Constrained model predictive control: Stability and optimality. Automatica, 36 (6), 789-814.

McFadden, D., 1981. Econometric models of probabilistic choice. Structural analysis of discrete data with econometric applications, MIT Press, Cambridge.

Mollison, D., 1972. Possible velocities for a simple epidemic. Advances in Applied Probability 4 (2), 233-257.

Mounce, R., Carey, M., 2014. On the convergence of the method of successive averages for calculating equilibrium in traffic networks. Transportation science, 49 (3), 535-542.

Nagurney, A., 2013. Network economics: A variational inequality approach (Vol.10). Springer Science & Business Media.

Nagurney, A., 2000. A multiclass, multicriteria traffic network equilibrium model. Mathematical and Computer Modelling, 32 (3-4), 393-411.

Nagurney, A., Dong, J., 2002. A multiclass, multicriteria traffic network equilibrium model with elastic demand. Transportation Research Part B: Methodological, 36 (5), 445-469.

Naveen, V. J., Rajeswari, K. R., 2011. Generation of Nakagami fading signals with arbitrary correlation and fading parameters. International Journal of Future Generation Communication and Networking, 4 (2), 49-68.

Noori, H., Olyaei, B. B., 2013. A novel study on beaconing for VANET-based vehicle to vehicle communication: Probability of beacon delivery in realistic large-scale urban area using 802.11p. 2013 International Conference on Smart Communications in Network Technologies, June, pp. 1-6, Paris, France.

Paz, A., Peeta, S., 2009. Behavior-consistent real-time traffic routing under information provision. Transportation Research Part C: Emerging Technologies, 17 (6), 642-661.

Peeta, S., Yang, T.H., 2003. Stability issues for dynamic traffic assignment. Automatica, 39 (1), 21-34.

Peeta, S., Yu, J. W., 2002. Data-consistent fuzzy approach for online driver behavior under information provision. Transportation Research Record: Journal of the Transportation Research Board 1803, 76-86.

Ploeg, J., Shukla, D.P., van de Wouw, N., Nijmeijer, H., 2014. Controller synthesis for string stability of vehicle platoons. IEEE Transactions on Intelligent Transportation Systems, 15 (2), 854-865.

Prashker, J., Bekhor, S., 1999. Stochastic user-equilibrium formulations for extended-logit assignment models. Transportation Research Record: Journal of the Transportation Research Board, 1676, 145-152.

Richards, P. I., 1956. Shock waves on the highway. Operations Research 4, 42-51.

Ramadurai, G., Ukkusuri, S., 2010. Dynamic user equilibrium model for combined activity-travel choices using activity-travel supernetwork representation. Networks and Spatial Economics, 10 (2), 273-292.

Ramming, M.S., 2001. Network knowledge and route choice. Ph. D. Thesis, Massachusetts Institute of Technology.

Shao, H., Lam, W.H., Tam, M.L., 2006. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Networks and Spatial Economics, 6 (3-4), 173-204.

Shida, M., Nemoto, Y., 2009. Development of a small-distance vehicle platooning system. 16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services, September, Stockholm , Sweden.

Shida, M., Doi, T., Nemoto, Y., Tadakuma, K., 2010. A short-distance vehicle platooning system (second report): Evaluation of fuel savings by the developed cooperative control. In Proceedings of the 10th International Symposium on Advanced Vehicle Control, Auguest, pp. 719-723, Loughborough, UK.

Smith, H. L., Wang, L., Li, M. Y., 2001. Global dynamics of an SEIR epidemic model with vertical transmission. SIAM Journal on Applied Mathematics, 62 (1), 58-69.

Smith, M.J., 1984. The stability of a dynamic model of traffic assignment: an application of a method of Lyapunov. Transportation Science, 18 (3), 245–252.

Smith, M.J., Wisten, M.B., 1995. A continuous day-to-day traffic assignment model and the existence of a continuous dynamic user equilibrium. Annals of Operations Research, 60 (1), 59-79.

Smith, M.J., Watling, D.P., 2016. A route-swapping dynamical system and Lyapunov function for stochastic user equilibrium. Transportation Research Part B: Methodological, 85, 132-141.

Szeto, W.Y., Lo, H.K., 2006. Dynamic traffic assignment: Properties and extensions. Transportmetrica 2 (1), 31-52.

Talebpour, A., Mahmassani, H. S., Bustamante, F. E., 2016. Modeling driver behavior in a connected environment: Integrated microscopic simulation of traffic and mobile wireless telecommunication systems. Transportation Research Board 95th Annual Meeting. Washington D.C, USA.

Talebpour, A, Mahmassani, H. S., 2016. Influence of connected and autonomous vehicles on traffic flow stability and throughput. Transportation Research Part C: Emerging Technologies, 71, 143-163.

Talebpour, A., Mahmassani, H., Mete, F., Hamdar, S., 2014. Near-crash identification in a connected vehicle environment. Transportation Research Record: Journal of the Transportation Research Board, 2424, 20-28.

Tientrakool, P., Ho, Y.C., Maxemchuk, N.F., 2011, September. Highway capacity benefits from using vehicle-to-vehicle communication and sensors for collision avoidance. In Vehicular Technology Conference (VTC Fall), May, pp. 1-5, San Francisco, California, USA.

Tikhonov, A., 1963. Solution of incorrectly formulated problems and the regularization method. Soviet mathematics - doklady., 4, 1035-1038.

Tobin, R.L., 1986. Sensitivity analysis for variational inequalities. Journal of Optimization Theory and Applications, 48 (1), 191-204.

Tobin, R.L., Friesz, T.L., 1988. Sensitivity analysis for equilibrium network flow. Transportation Science, 22 (4), 242-250.

Tseng, Y., Ni, S., Chen, Y., Shen, J., 2002. The broadcast storm problem in a mobile ad hoc network. Wireless networks, 8, 153-167.

Ukkusuri, S., Du, L., 2008. Geometric connectivity of vehicular ad hoc networks: analytical characterization. Transportation Research Part C: Emerging Technologies, 16 (5), 615–634.

Van den Berg, V.A., Verhoef, E.T., 2016. Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity. Transportation Research Part B: Methodological, 94, 43-60.

VanderWerf, J., Shladover, S. E., Kourjanskaia, N., Miller, M., Krishnan, H., 2001. Modeling effects of driver control assistance systems on traffic. Transportation Research Record, 1748, 167–174.

Vovsha, P., 1997. The cross-nested logit model: application to mode choice in the Tel-Aviv metropolitan area. Presented at the 76th Transportation Research Board Annual Meeting, Washington, D.C.

Wang, J., He, X., Peeta, S., 2016. Sensitivity analysis based approximation models for day-to-day link flow evolution process. Transportation Research Part B: Methodological, 92, 35-53.

Wang, J., Kim, Y. H., He, X., Peeta, S., 2018a. Analytical model for information flow propagation wave under an information relay control strategy in a congested vehicle-to-vehicle communication environment. Transportation Research Part C: Emerging Technologies, 94, 1-18.

Wang, J., Peeta, S., He, X., Zhao, J., 2018b. Combined multinomial logit modal split and paired combinatorial logit traffic assignment model. Transportmetrica A: Transport Science. 14 (9), 737-760.

Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014a. Rolling horizon control framework for driver assistance systems. Part I: Mathematical formulation and non-cooperative systems. Transportation Research Part C: Emerging Technologies, 40, 271-289.

Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2014b. Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. Transportation Research Part C: Emerging Technologies, 40, 290-311.

Wang, M., Daamen, W., Hoogendoorn, S.P., van Arem, B., 2016. Cooperative car-following control: Distributed algorithm and impact on moving jam features. IEEE Transactions on Intelligent Transportation Systems, 17 (5), 1459-1471.

Wang, X., 2007. Modeling the process of information relay through inter-vehicle communication. Transportation Research Part B: Methodological, 41 (6), 684-700.

Wang, X., Adams, T. M., Jin, W.-L., Meng, Q., 2010. The process of information propagation in a traffic stream with a general vehicle headway: A revisit. Transportation Research Part C: Emerging Technologies, 18 (3), 367-375.

Wang, X., Yin, K., Qin, X., 2011. An approximate Bernoulli process in information propagation through inter-vehicle communication along two parallel roads. Transportation Research Part C: Emerging Technologies, 19 (3), 469-484.

Wang, X., Yin, K., Yan, X., 2015. Vehicle-to-vehicle connectivity on parallel roadways with large road separation. Transportation Research Part C: Emerging Technologies, 52, 93-101.

Wang, X., Yin, K., Zhang, Y., 2012. A Markov process for information propagation via inter-vehicle communication along two parallel roads. IEEE Transactions on Wireless Communication, 11 (3), 865-868.

Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transportation Research Part B: Methodological, 31 (1), 55-70.

Yang, H., Bell, M.G., 2005. Sensitivity analysis of network traffic equilibrium revisited: the corrected approach. In 4th IMA International Conference on Mathematics in Transport Institute of Mathematics and its Applications, September, London, United Kingdom.

Yang, H., Huang, H., 2004a. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. Transportation Research Part B: Methodological, 38 (1), 1-15.

Yang, X., Liu, L., Vaidya, N. H., Zhao, F., 2004b. A vehicle-to-vehicle communication protocol for cooperative collision warning. In Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, August, pp. 114-123, Massachusetts, USA.

Yeo, H., Shladover, S., Krishnan, H., Skabardonis, A., 2010. Microscopic traffic simulation of vehicle-to-vehicle hazard alerts on freeway. Transportation Research Record: Journal of the Transportation Research Board, 2189, 68-77.

Yin, K., Wang, X., Zhang, Y., 2013. Vehicle-to-vehicle connectivity on two parallel roadways with a general headway distribution. Transportation Research Part C: Emerging Technologies 29, 84-96.

Zhang, J., Han, G., Qian, Y., 2016. Queuingtheory based co-channel interference analysis approach for high-density wireless local area networks. Sensors, 16 (9), 1348.

Zhang, S., Wu, Y., Liu, H., Huang, R., Un, P., Zhou, Y., Fu, L., Hao, J., 2014. Real-world fuel consumption and CO2 (carbon dioxide) emissions by driving conditions for light-duty passenger vehicles in China. Energy, 69, 247-257.

Zhan, X., Ukkusuri, S.V., 2017. Multiclass, simultaneous route and departure time choice dynamic traffic assignment with an embedded spatial queuing model. Transportmetrica B: Transport Dynamics. (In publish).

Zheng, Y., Li, S.E., Wang, J., Wang, L.Y., Li, K., 2014. Influence of information flow topology on closed-loop stability of vehicle platoon with rigid formation. 17th International Conference on Intelligent Transportation Systems, October, pp. 2094-2100, Qingdao, China.

Zhou, Y., Ahn, S., Chitturi, M., Noyce, D.A., 2017. Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty. Transportation Research Part C: Emerging Technologies, 83, 61-76.