Network Traffic Signal Control with Short-Term Origin-Destination Demand in a Connected Vehicle Environment via Mobile Edge Computing

by

Can Zhang

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

Doctor of Philosophy

in

TRANSPORTATION ENGINEERING

Department of Civil and Environmental Engineering

University of Alberta

ABSTRACT

This thesis develops and analyzes centralized and decentralized network-level traffic signal control system under in a connected vehicle (CV) environment with mobile edge computing (MEC). The goal is to provide a framework of decentralized signal control (DSC) system especially for real-time control and large-scale traffic network. Short-term origin-destination (OD) demand is used as an input given that the technological paradigm assumed is within the CV environment, unlike most previous works that look at network control but in a current technological paradigm.

Considering short-term OD demand as inputs, a queue-based dynamic traffic assignment (DTA) model is proposed to predict traffic dynamics in traffic networks with signal control. Although DTA has been an effective tool to describe traffic dynamics for traffic optimization, and many researchers have considered traffic signal control in their models, signal timings have been simplified without considering complex, but realistic, phase sequence and duration restrictions. This work formulates traffic signal timing as a component of the link performance function with three control variables: cycle length, phase split, and offset. In addition, both user-optimal (UO) and system-optimal (SO) DTA problems are solved within a single corridor network.

Finally, this thesis provides a simulation-based framework of both centralized and decentralized signal control to solve the network-level traffic signal control optimization problem. For the centralized system, this work solves the issue of optimal control using a three-step naïve method. Because the optimization of large-scale network traffic signals is a Nondeterministic Polynomial Time (NP)-complete problem, the centralized system is further

decomposed into a decentralized system where the network is divided into subnetworks. – Each subnetwork has its own agent that optimizes signals within the subnetwork. The proposed control systems are applied to a set of test scenarios constructed using different demand levels in different grid networks. This work also investigates the impact of network decomposition strategy on the signal control system performance. Results show that network decomposition with smaller subnetworks results in less Computational Time (CT), but also increased Average Travel Time (ATT) and Total Travel Delay (TTD).

This thesis contributes to the literature by a queue-based DTA model for traffic network with real traffic signal timing plan, a simulation-based framework of DSC system within the MEC-enabled CV environment, and a scalable and extendable decomposition method for a DSC system.

Keywords: network-level traffic signal control, mobile edge computing, decentralized signal control, short-term origin-destination demand, network decomposition.

PREFACE

Published

1. **Zhang, C.**, Bie, Y., & Qiu, T. Z. (2017). Short-Term Demand Prediction at Freeway Bottleneck Under VSL Control. *Transportation Research Board 96th Annual Meeting, 2017.*
2. Bie, Y., Qiu, T. Z., **Zhang, C.**, & Zhang, C. (2017). Introducing weather factor modelling into macro traffic state prediction. *Journal of advanced transportation*, *2017*.
3. Bie, Y., Seraj, M., **Zhang, C.**, & Qiu, T. Z. (2018). Improving traffic state prediction model for variable speed limit control by introducing stochastic supply and demand. *Journal of Advanced Transportation*, *2018*.
4. Wang, X., Li, J., **Zhang, C.**, & Qiu, T. Z. (2019). Active warning system for highway-rail grade crossings using connected vehicle technologies. *Journal of advanced transportation*, *2019*.
5. **Zhang, C.**, Qiu, C., & Qiu, T. Z. (2019). Reliability Constrained Cordon Pricing in Traffic Networks Under Stochastic Demand. *Transportation Research Board 98th Annual Meeting, 2019.*

Under Review

1. **Zhang, C.**, Qiu, T. Z., & Kim, A. Centralized and Decentralized Signal Control with Short-Term Origin-Destination Demand for Network Traffic. Submitted to *Journal of Advanced Transportation.*

ACKNOWLEDGEMENT

I feel lucky to conduct my PhD program under the co-supervision of Dr. Tony Qiu and Dr. Amy Kim. I would like to thank them for their guidance and patience, especially for the recent year, which enables me to successfully complete the doctoral program and master the high-level qualities required for the transportation research.

In addition, my thanks to my officemates and colleagues at University of Alberta for their support and companionship, especially Huiyu Chen, Kaizhe Hou, Yuwei Bie, and Chen Qiu for the teamwork and technical discussion; Shuoyan Xu, Fan Wu, Bryan Tran, Matt Woo, and Jianjing Jin for their encourage and support; also, Sharon Harper, Alice Da Silva, and Rachelle Foss for their professional work on proof-reading and admin.

I would also like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada, City of Edmonton, Transport Canada, TELUS, and Compute Canada for providing resources for my research in University of Alberta.

As well, thanks to Dr. Tony Qiu, Dr. Amy Kim, Dr. Tae Kwon, Dr. Michael Buro, Dr. Bob Koch, and Dr. Huazhou Li for participating in my candidacy exam. And thanks to Dr. Tony Qiu, Dr. Amy Kim, Dr. Tae Kwon, Dr. Bob Koch, Dr. Bruce Hellinga, and Dr. Yong Li for attending my final defense.

Finally, many thanks to my wife, Lingfei Guan, who unconditionally supports and encourages me, to my parents, Cunjie Zhang and Huiying Cao, who give me constant love, and to God for everything in my life.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ATDM	active traffic and demand management
ACO	ant colony optimization algorithm
ADP	approximate dynamic programming
AFT	adaptive fine-turning
AI	artificial intelligence
ASC	adaptive signal control
ATT	average travel time
BA	bisection algorithm
BP	backpressure algorithm
BSM	basic safety message
CA	clustering algorithm
CSC	centralized signal control
CT	computational time
CTM	cell transmission model
CV	connected vehicle
CV2X	connected vehicle to everything
DSC	decentralized signal control
DSRC	dedicated short range communications
DTA	dynamic traffic assignment
FFT	free flow travel time
FIFO	first-in-first-out
GA	genetic algorithms
IoV	internet of vehicles
IS	infinite server
ITS	intelligent transportation system

LIFO-PR	last in first out with preemptive
MEC	mobile edge computing
MFD	macroscopic fundamental diagram
MILP	mixed integer linear programming
NP-complete	nondeterministic polynomial-time complete
OBE	onboard equipment
OD	origin-destination
PQ	point queue
PS	processor sharing
QNM	queueing network model
RL	reinforcement learning
RSE	roadside equipment
SO	system-optimal
SPF	shortest path first algorithm
TMC	traffic management center
TTD	total travel delay
UO	user-optimal
VI	variable inequality
VSL	variable speed limit

1 INTRODUCTION

1.1 Background

The fundamental principle of active traffic and demand management (ATDM) is reducing traffic congestion before breakdown conditions occur. ATDM embodies proactive management rather than reactive, and consists of strategies to prevent traffic congestion by re-balancing traffic demand and supply through traffic signal control, dynamic lane use control, variable speed limit, and adaptive ramp metering. This thesis focusses on optimizing network traffic signal control. Network traffic signal control aims to improve network mobility and traffic safety by adjusting directional capacity at each signalized intersection. However, network-level traffic signal control is hard to implement, and suffers from issues of data quality, computation capacity, and control latency (i.e., the sum of computational latency and data transmission latency).

Rapidly developing Connected Vehicle (CV) technologies that produce increasingly numerous types of vehicle data, personal travel data, and geometric data offer the potential to resolve this data quality issue. For example, the 2017 SAE J2735 Dedicated Short Range Communications (DSRC) report summarizes 17 messages, 156 data frames, 230 data elements, and 58 external data element definition references (Perry, 2017). In addition, more types of data will be readily available in the near future thanks to the rapid development of Connected Vehicle (CV) and communication technologies. For example, Origin-Destination (OD) demand data can be generated from a variety of sources such as navigation systems (e.g., Google Maps), ride-sharing applications (e.g., Uber), and GPS-equipped vehicles, and can include vehicle origin, destination, and start time. However, many recent studies for network traffic signal control still use data from fixed sensors as inputs, such as vehicle counts from loop detectors. Traffic data collected by induction loop detector and other common sensors can only be used to understand traffic states in a short period locally. OD demand, however, which is widely used as an input for network modeling, can help predict traffic dynamics over some given period on network level. Moreover, in congested conditions when demand exceeds capacity, actual flow is lower than demand (or capacity), which indicates that using flow as input has fundamental limitations on demand management.

However, computational capacity limitations remain despite excellent algorithms, Artificial intelligence (AI) technology, and large-capacity computers. The primary issue rests with the fact that network-level traffic signal control is highly complex, nonlinear, and affected by many external and internal factors due to the nature of the traffic system. The network traffic signal optimization problem has been proven to be an Nondeterministic Polynomial Time (NP)-complete problem (Adacher et al., 2014). Computation complexity will increase exponentially with respect to the number of nodes in the network (i.e., network size). In a centralized network-level traffic signal control system, the computational latency caused by the data analysis required to determine an optimized solution is a too large to be practical, especially for real-time control.

With the help of mobile edge computing (MEC), a decentralized signal control (DSC) system can be applied to solve the issues of computation capacity and control latency. MEC allows more data resources, ensures low data transmission latency, and enables distributed computation. Traditional network traffic signal control systems have a central traffic management center (TMC) to manage traffic. Although data can be transferred from each end of the traffic network in CV environment, the traditional framework is hampered by bandwidth limitations during data processing. And data transmission latency worsens as more devices are connected to the central TMC. By adding MEC technology, data can be analyzed at the location where it is collected. Thus, processed traffic information can be transferred between each end of the traffic network easily with low data transmission latency due to its small size, normalized format, and accurate information. Moreover, in the DSC system, the network is divided into subnetworks, and each subnetwork has its own agent (regional MEC controller) to control signals inside it. Each agent is designed to receive messages from surrounding agents, make decisions for traffic signal control of intersections inside the subnetwork, and send messages to surrounding intersections. As such, decentralized systems reduce data analysis latency by breaking the information down into smaller parcels, making them easier to deliver and analyze. In other words, a DSC system decomposes network traffic signal control problem in the CSC system, making it a distributed structure, thereby reducing the computational latency and need for huge computation power. Hence, this research is directed towards investigating and developing decentralized traffic signal control.

1.2 Research Questions

This thesis proposes a simulation-based decentralized framework to solve the network-level traffic signal optimization problem. Short-term OD demand is used as an input, given that the technological paradigm assumed is within the connected, MEC-enabled environment. Establishing an efficient decentralized signal control system requires answering the following research questions:

1. Coordinated network traffic signal optimization is based on prediction of traffic dynamics. How do we predict traffic dynamics in a large-size traffic network with signalized intersections?

2. In the MEC-enabled CV environment, how do we design the structure of a decentralized signal control (DSC) system? And how do we evaluate the performance of these systems?

3. There are many characteristics which impact DSC system performance, such as network size, traffic demands, and how a network is decomposed with respect to the traffic signal control system. How does each of these affect traffic network performance?

Addressing the above questions will help to bring the benefits of a decentralized control system to urban traffic signal control in the foreseeable future, through more efficient use of network computing capacity and improved network traffic performance.

1.3 Thesis Objectives & Contributions

The main objective of this thesis is to develop a simulation-based framework to model and test different configurations of a decentralized signal control (DSC) system, considering short-term OD demand as inputs in a MEC-enabled CV environment. The performances of networks of different sizes, decompositions, and with different traffic demand levels are compared and assessed.

This thesis contributes to the literature by developing the following: 1) A queue-based DTA model for traffic network with real traffic signal timing plan; 2) A simulation-based

framework of DSC system within the MEC-enabled CV environment; and 3) A scalable and extendable decomposition method for a DSC system.

The significant reduction of computational time in a DSC system, compared to the centralized signal control (CSC) system, makes network traffic signal optimization via DSC potentially applicable for real-time control. The results can also provide guidance to transportation engineers and planners about how to optimize DSC system via network decomposition.

1.4 Thesis Structure

The remainder of this thesis is organized into the following six chapters.

Chapter 2 presents a literature review, focusing on four areas of the research literature: MEC, network traffic signal control, dynamic traffic assignment (DTA), and DTA based network traffic signal control.

Chapter 3 describes basic assumptions and inputs used in this thesis research. A MEC-based CV environment ensures the accessibility of short-term OD demand, low data transmission latency, and distributed computation ability. Four assumptions on DTA are used to predict traffic dynamics. In addition, discussions are provided why short-term demand is chosen as inputs instead of local counts, and definitions are given for different short-term demands used in this thesis.

Chapter 4 develops and analyzes a queue-based DTA model for traffic network with real traffic signal timing plan considering short-term OD demand, which answers first question in Section 1.2.

Chapter 5 develops and compares centralized and decentralized systems for optimal traffic signal control with short-term OD demand as inputs. Performance of both systems from computational time (CT), average travel time (ATT), total travel delay (TTD) is investigated via numerical simulations within 5 test scenarios. Results answer question 2 & 3 in Section 1.2.

Chapter 6 concludes the whole thesis, including an overview, main findings, contributions, and a brief discussion of future research needs.

Figure 1-1 illustrates the relationship between the literature review (Ch. 2), assumptions and inputs (Ch.3), two major chapters (Ch. 4 & 5), and contributions of this thesis. Again, this research proposes a simulation-based framework of DSC considering short-term OD demand as inputs to optimize network signals. Chapter 3 describe main assumptions and definitions used in this thesis, explains why and what short-term demand is used. Chapter 4 takes short-term OD demand as inputs and introduces how to predict traffic dynamics with real traffic signal timing plan. And Chapter 5 integrates the DTA model with network traffic signal control system and analyzes CSC and DSC system in five scenarios.

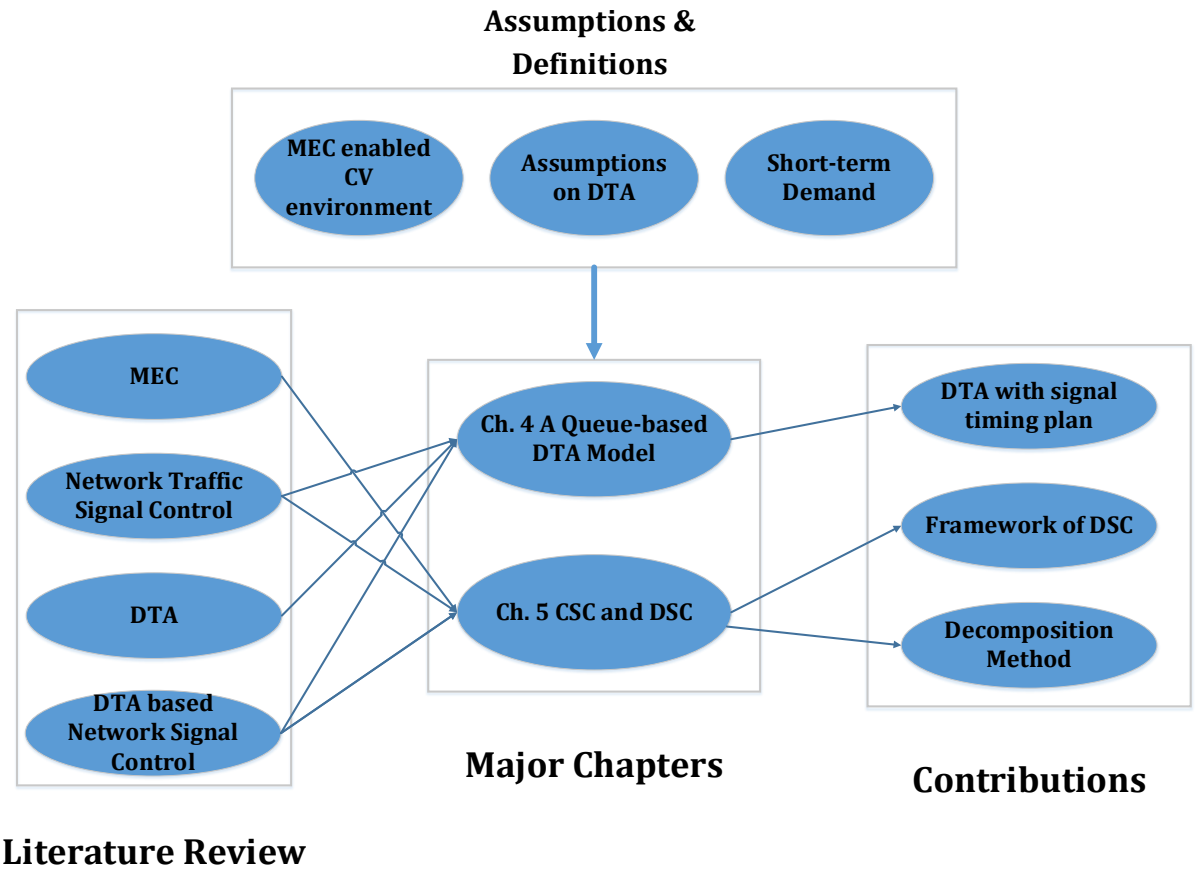


Figure 1-1 Relation of literature review, chapters, and research contributions.

2 LITERATURE REVIEW

The earliest literature on traffic signal control is Webster's fundamental work from 1958 (Webster, 1958). The need for modeling large-scale networks has motivated many researchers to develop innovative control strategies, related models, and problem-solving algorithms. The focus of my proposal is to develop a decentralized traffic signal control system given that the technological paradigm assumed is within the connected, MEC-enabled environment. And given the different components of research needs, this thesis reviews five areas of the research literature: Mobile Edge Computing (MEC), Dynamic Traffic Assignment (DTA), network traffic signal control, and DTA based network traffic signal control.

2.1 Mobile Edge Computing (MEC)

MEC is also referred to multi-access edge computing, which is defined as a network architecture concept that enable cloud computing capacities and an IT service environment at the edge of cellar network, and more general at the edge of any network (Filali et al., 2020). In traffic network, edges can be vehicles, infrastructures, or even the mobile phones of system users. In the Introduction (Ch. 1), it was stated that network-level traffic signal control is difficult to implement, suffering from issues of data quality, computation capacity, and control latency. MEC adds data quality and computational capacity, which will solve at least the issue of data quality in the near future.

Currently, the vehicle is not only the edge of data reception but also data collection. Researchers have already introduced a deep learning method for vehicular platoon control based on MEC analysis (Ferdowsi et al., 2017). A recent study applied MEC to a new framework of Internet of Vehicles (IoV) and developed a resource allocation algorithm to process high-dimensional data (G. Wang & Xu, 2020). In addition, a MEC-based public vehicle system is proposed by researcher recently to improve traffic efficiency and vehicle occupancy ratios by scheduling ridesharing among travelers and reduce the delay of decision making by leveraging edge computing(Lin et al., 2020). MEC also delivers opportunities to design an edge computing ecosystem for autonomous vehicles is to deliver enough computing power, redundancy, and

security so as to guarantee the safety of autonomous vehicles (Liu et al., 2019). MEC is more and more popular among fields of transportation research.

Finally, the introduction of 5G will enable MEC to perform efficiently in these contexts. The inherent features of 5G such as high connectivity, low latency, and large bandwidth will facilitate the demanding computational and communication requirements of MEC (Hu et al., 2015). A new vehicular network architecture integrated with 5G mobile communication technologies and software defined networking is proposed within a CV environment, enabling vehicle to talk with roadside units (Ge et al., 2017). The combination of CV, MEC and 5G technologies will provide high-quality data, stronger computational hardware, low-latency and high-bandwidth communication technology, the basis upon which helps develop more efficient signal control systems for network-level traffic and real-time control. This study seeks to understand the performance of different network decompositions when fed OD demand as inputs, motivated by the possibilities of these new technologies.

2.2 Dynamic Traffic Assignment (DTA)

Short-term Origin-Destination (OD) demand is generally defined as time-dependent travel demand between origin and destination nodes in the network, which is used as inputs in DTA models. DTA describes the path choice or departure time decisions of travelers under time-dependent demand and capacity constraints, rather than static (Static Traffic Assignment, STA). Since the seminal work by Merchant and Nemhauser (M-N model) in 1978, DTA has evolved substantially (Merchant & Nemhauser, 1978a, 1978b). The original formulation was pre-determined as fixed-demand, single-destination, and single-commodity. Subsequent to their work, the research has been developed in two directions: analytical approaches and simulation-based DTA.

Taking an analytical approach, Carey formulated convex programming for DTA in 1987 (Carey, 1987). Then, in 2000, Carey and Subrahmanian illustrated the “First In First Out” (FIFO) and holding-back issues (Carey & Subrahmanian, 2000). In addition, their group introduced additional objectives or variables such as System Optimal (SO), marginal costs, and tolls into their DTA models (Carey & Watling, 2012).

Other researchers have focused on the Variable Inequality (VI) form of DTA, which is a method that is more computable than general mathematical programming but more specific as well. The general VI form of DTA was formulated in 1980 (Dafermos, 2008). Because multiple forms developed from that point forward, Nagurney provided a summary of all the VI forms related to traffic assignment in 2003 (Nagurney, 2003). In a specific case, Chen and Hsueh solved a link-based VI for the User Optimal (UO)-DTA model (H. K. Chen & Hsueh, 1998).

In addition to the analytical approaches to DTA, some researchers have integrated the DTA model into optimal control. Friesz et al. discussed a link-based optimal control formulation in the original M-N model for both SO and UO objectives (Friesz & Tobin, 1989). They also introduced a VI-based formulation (Friesz et al., 1993). Ran et al. used the optimal control approach to formulate a convex model for instantaneous UO-DTA problem-solving considering link flows as variables (Ran et al., 1993).

For simulation-based DTA, simulators, such as DYNASMART (Jayakrishnan et al., 1994), CONTRAM (https://catalogue.nla.gov.au/Record/2870210), DynaMIT (Ben-Akiva et al., 1998), have been used. Much simulation has been done in GIS and MATLAB based on the cell transmission model (CTM), where the fundamental work was completed by Daganzo (Daganzo, 1994).

DTA has been developed extensively over the past decades, proving an effective tool for determining traffic control strategies despite their complexity and computational burdens. Analytical approaches are hindered by some strong and unrealistic assumptions, which is particularly problematic for real-time, large-scale urban network traffic signal control. As such, simulation-based DTA has been chosen for prediction of traffic dynamics in this research.

2.3 Network Traffic Signal Control

2.3.1 Adaptive signal control

Traffic signal control in urban transportation networks is optimized to increase network capacity utilization, improve mobility (including reducing traveler delays), and reduce vehicle-based emissions, amongst other aims. Adaptive Signal Control (ASC) systems are widely used in many modern cities, from the earliest SCATS (A. G. Sims & Dobinson, 1980) and SCOOT (Y. T. Wu

& Ho, 2009) systems, to intermediate typical OPAC (Gartner, 1983), RHODES (Mirchandani & Head, 2001), UTOPIA (Ben-Akiva et al., 2003), ASC Lite systems (Luyanda et al., 2003), and more recent adaptive fine-tuning (AFT) systems (Manolis et al., 2016) and deep learning-based system (Gao et al., 2017). However, these approaches all rely on traffic volume data from upstream links, or queue lengths on each leg of an intersection. Traffic conditions predicted from these types of data only reflect the local traffic state over a short period.

2.3.2 Centralized signal control (CSC)

CSC optimizes the parameters of each intersection in the network simultaneously to find an optimal control solution. Much of the research literature has focused on the algorithms used in different networks towards a range of objectives. For instance, Genetic Algorithms (GA) and Approximate Dynamic Programming (ADP) approaches have been proposed for traffic signal control in oversaturated networks (Hajbabaie, 2012). Another algorithmic solution for oversaturated networks has been the ant colony optimization algorithm (ACO) (Putha et al., 2012). Other researchers have used heuristic algorithms – for example, Beard and et al. (2007) used a mixed integer linear programming (MILP) model to optimize traffic signal control.

The above CSC cases were studied for small networks. Even escalating to a medium-size network means that the complexity (NP-complete) and increased time and topological scale caused greater difficulties (Adacher et al., 2014). Thus, strong assumptions are made to avoid complex variables and traffic dynamics. For example, Gregoire et al. (2015) applied a backpressure algorithm (BP) to solve the network traffic signal control problem, ignoring travel times for each link and turning movement to keep the problem tractable. Prashanth and Bhatnagar (2011) used Reinforcement Learning (RL) to generate and train data towards minimizing average cost for ASC. However, the application of the method results in the curse of dimensionality, where network traffic signal control with RL remains an NP-complete problem. Although recent advances in computing offer some resources to solve real-time, large-scale data collection and processing, the NP-complete problem of network-wide real-time signal control remains unresolved.

2.3.3 Decentralized Signal Control (DSC)

There are limitations for CSC systems to handle systems of high dimension and with many inputs and outputs. As discussed in the previous section, the optimization problem is still very complex. In addition, CSC systems require high levels of connectivity. Network performance is quite sensitive to small failures within each part of the network (McKenney & White, 2013). Prompted by the difficulties involved in CSC, research attention turned to distributed systems.

Distributed control is also referred to as decentralized control. Large scale network systems have been a topic of heavy interest in the control system field since the 1970s (S. H. Wang & Davison, 1973); since this time, there has been much investigation into both the hardware and software design requirements for decentralized control architectures. Generally, systems with distributed devices, distributed computing, and/or distributed computing can be regarded as distributed systems. The distributed control in this thesis can be defined as a spatially interconnected system, in which physical control devices are distributed inside the traffic network and can communicate with their neighbors (D'Andrea & Dullerud, 2003).

In transportation area, there are many researchers working on decentralized signal control. Gokulan et al. (2010) developed a distributed, multi agent-based approach for a traffic-responsive signal control system. The distributed signal control was achieved by multi-agents at each subnetwork. This method significantly reduced the complexity issue of network-level traffic signal control. Chow et al. (2020) recently developed and compared centralized and decentralized signal control systems with BP model for a road network in Central London, UK. Their results showed that total travel delay was lower with a centralized system while a decentralized system reduced computational time by 40%.

Reinforcement Learning (RL) also plays an important role in DSC systems. Pol & Oliehoek (2016) considers explicit coordination mechanisms between learning agents using coordination graphs. Researchers have used individual RL agents to control traffic signals in a multi-intersection network from two angles: without communication between RL agents as "game theory" (Nowé et al., 2012), and with communication between agents as "neighborhoods" (Nishi et al., 2018, Wei et al., 2019). The RL method is powerful as it can be applied in real-world scenarios (2,510 traffic lights in Manhattan) (C. Chen et al., 2020). However, there are

several challenges identified by Wei et al. (2019), two of which are critical: learning costs are too high for complex problems, and risk management is not widely adopted causing potential traffic safety issues in a real world setting.

Network decomposition topology is an important consideration in some decentralized signal control research. Cell-based decomposition (Mehrabipour & Hajbabaie, 2017) and intersection-based decomposition (Withanawasam & Karunananda, 2018; Adacher et al., 2014) are two common approaches, but both result in subnetworks with only one intersection each. Adacher and Tiriolo (2017) addressed this by working with subnetworks containing more than one intersection. Their recent research using a Clustering Algorithm (CA) to solve the DSC problem based on a Cell Transmission (CTM) model highlighted the importance of spatial decompositions (Adacher & Tiriolo, 2020). However, they tested one network that includes a specific group of subnetworks with a simplified signal phase, thereby offering limited results. A grouping method was developed to decrease delay and number of stops while minimizing traffic operators' subjective decision (Lei Zhang et al., 2017), and was applied to a one-way corridor network with 21 intersections in Montgomery County, Maryland. The question of how to decompose a network to optimize signal control remains to be solved.

2.4 DTA-Based Network Traffic Signal Control

Combining DTA with network traffic signal control is not an innovative pairing. As early as 1977, Allsop and Charlesworth had combined traffic signal control and traffic assignment in an analytic way (Allsop & Charlesworth, 1977). Later, in 1995, Yang and Yagar looked at the relationship between DTA and traffic signal control in a saturated network as a bi-level problem (Yang & Yagar, 1995). Sun et al. used GA by adjusting cycle, offset, and phase split in a bi-level model (Sun et al., 2006).

Control objectives for DTA-based network traffic signal control have varied as well. Minimizing average travel time (ATT) and total travel delay (TTD) is the most common objective of DTA-based traffic signal control. Reverse capacity is considered a measurement of network mobility, and Chiou combined DTA and traffic signal control to achieve maximum reverse capacity (Chiou, 2007). Cai et al. applied adaptive traffic signal control in approximate dynamic programming within a corridor network to minimize the ATT of the network (Cai et al.,

2009). In a CV environment, objectives such as minimizing TTD and increasing average travel speed are considered as well (Lee et al., 2013). Li et al. have addressed the problems of simultaneous route guidance and traffic signal optimization to minimize TTD (Li et al., 2015). Some researchers expand network capacity via traffic signal control using DTA models (Karoonsoontawong & Waller, 2010).

What is most salient for this research here is that, although many researchers have considered traffic signal control in their DTA model, traffic signal timing representations have been simplified without considering complex, but realistic, phase sequence and duration restrictions.

2.5 Summary

Considering that local counts are inputs for most existing ASC systems, new technologies such as CV and MEC will enable the collection of high-quality data and local data processing thereby offering new possibilities for enriched data inputs.

Given that the technological paradigm assumed is within the connected, MEC-enabled environment, short-term OD demand (already an input to DTA models) is chosen as inputs to optimize network traffic signal control. DTA is a well-accepted approach to model network dynamics, which can predict short-term demand for links. However, although DTA has been integrated with signal control for decades, traffic signal timing representations have been heavily simplified.

Table 2-1 organizes the above discussed literature by the different modeling approaches used for network traffic signal control. Except RL and CA, most methods use networks with up to 100 control variables. For model inputs, link flows are widely used, as in ASC systems. System-level performance objectives include delay, throughput, travel time, and queue length.

Table 2-1 Literature on Network traffic signal control

Solution Method	Author	Network Size	Control Variables	# of Var (estimated)	Inputs	CSC /DSC	Objectives
GA	Hajbabaie, A. (2012)	4x4	Cycle, phase split, and offset	96	Upstream Volume from Gate Signals	CSC	Min Delay, Travel Time; Max Throughput, Trips
ACO	Putha et al. (2012)	4x5 one-way	Green time for throughput traffic	40	Link flows	CSC	Max Throughput
MILP	Beard et al. (2007)	1x2	Cycle, phase split, and offset	12	OD Demand	CSC	Min Total Travel Time
BP (CSC)	Gregoire et al. (2015)	8x8	Phase type	64	Queue Length	CSC	Min Total Queue Length
BP (DSC)	Chow, Sha, & Li (2020)	Bloomsbury network (15 intersections)	Green time for throughput traffic	30	Link flows	CSC & DSC	Min Total Delay
RL	Chen et al. (2020)	Manhattan Network (2510 intersections)	Phase type	2510	Link flows	DSC	Min Average Travel Time & Max Total Throughput
CA	Adacher & Tiriolo (2020)	Rome network (39 intersections)	Cycle, phase split, and offset	234	Path Demand	DSC	Min Total Delay

For CSC, despite excellent algorithms and machine learning methods, centralized signal control optimization problems remain NP-complete without simplifying traffic signal timing plans and their representation in models. Furthermore, both signal plan representation and model constraints must be kept relatively simple in order to apply algorithms such as BP and ACO. DSC appears to be a better approach for a network-level traffic signal control system, due to lower computational times achieved with more realistic signal control representations. However, how to best decompose a DSC network has not been well defined or investigated. In addition,

many previous works look at network control but still use flows or other local counts in a current technological paradigm.

As such, this research addresses the need to explore DSC systems for large-scale networks considering greater realism of traffic signal control, new forms of data inputs, and new network decomposition method.

3 BASIC ASSUMPTIONS & MODEL INPUTS

Chapter 3 gives basic assumptions and inputs of models used in this thesis. Firstly, the network traffic signal control system is based the new technical paradigm assumed as a MEC-enabled CV environment. Assumptions related to this environment is provided as three points. Secondly, this thesis using DTA model to predict traffic dynamics for network traffic signal optimization. Assumptions on DTA model include “No departure time choice”, “No reroute”, “FIFO”, and “Point Queue”. Thirdly, short-term demand is considered as inputs of models in this thesis. This chapter continues to discuss why short-term demand is used and define three types of short-term demands for this thesis.

3.1 CV Environment with Mobile Edge Computing (MEC)

Traditional signal control systems are under the control of a central Traffic Management Center (TMC). In Chapter 1, it was summarized that real-time traffic signal control for large-scale network was difficult to implement due to three issues: data quality, computational capacity, and control latency. The introduction of MEC enables an environment with distributed computation and low data transmission latency.

Low control latency is a key requirement for network-wide signal control to be implemented in real time. Control latency consists of computational latency and data transmission latency. For the whole process, data transmission latency includes vehicle-to-infrastructure (V2I), infrastructure-to-infrastructure (I2I), and infrastructure-to-vehicle (I2V). While more devices are connected, types of data with large size are supposed to be sent to the central TMC. There is a massive bandwidth demand to guarantee low data transmission latency. Computational latency is the time required for data analysis, which depends on the complexity of the problem. For real-time traffic control, the control system will provide signal timing plans for all the intersections in the network before new vehicles enter the control area. For example, if a vehicle sends their OD information 30 seconds before entering the network, it will require the system to generate data from all vehicles, predict traffic dynamics for the whole network, and then provide control plans within the 30 seconds. Otherwise, when the system changes signal controls after the traffic dynamics have changed, the control plan will become outdated. In

addition, low data transmission latency can facilitate the collection of data with high resolution and then increase the accuracy of traffic state estimation, which is also a benefit for real-time network signal control. Thus, low data transmission latency is very important for the real-time control system.

According to the introduction part of Chapter 1, data quality will not be a critical problem due to various data sources in the near future. Once data transmission latency is low enough, the only concern remaining is the computational latency arising from the data analysis required to solve the NP-complete CSC problem. Thus, this thesis is to use a distributed structure with MEC devices to increase computational efficiency. Data quality is not a problem due to various data sources. Once data transmission latency is low enough, the only control latency concerned is computational latency coming from data analysis for solving the NP-complete CSC problem. Thus, this thesis is to use a distributed structure with MEC devices in order to increase computational efficiency.

With the help of MEC, optimization calculations can be completed at the location where data is generated. This benefits the system via reduced data transmission latency, and the capabilities of distributed control. Similar to CSC system, in DSC system data transmission latency is assumed too low to be counted within the MEC-enabled CV environment.

Figure 3-1 compares network traffic signal control system with and without MEC.

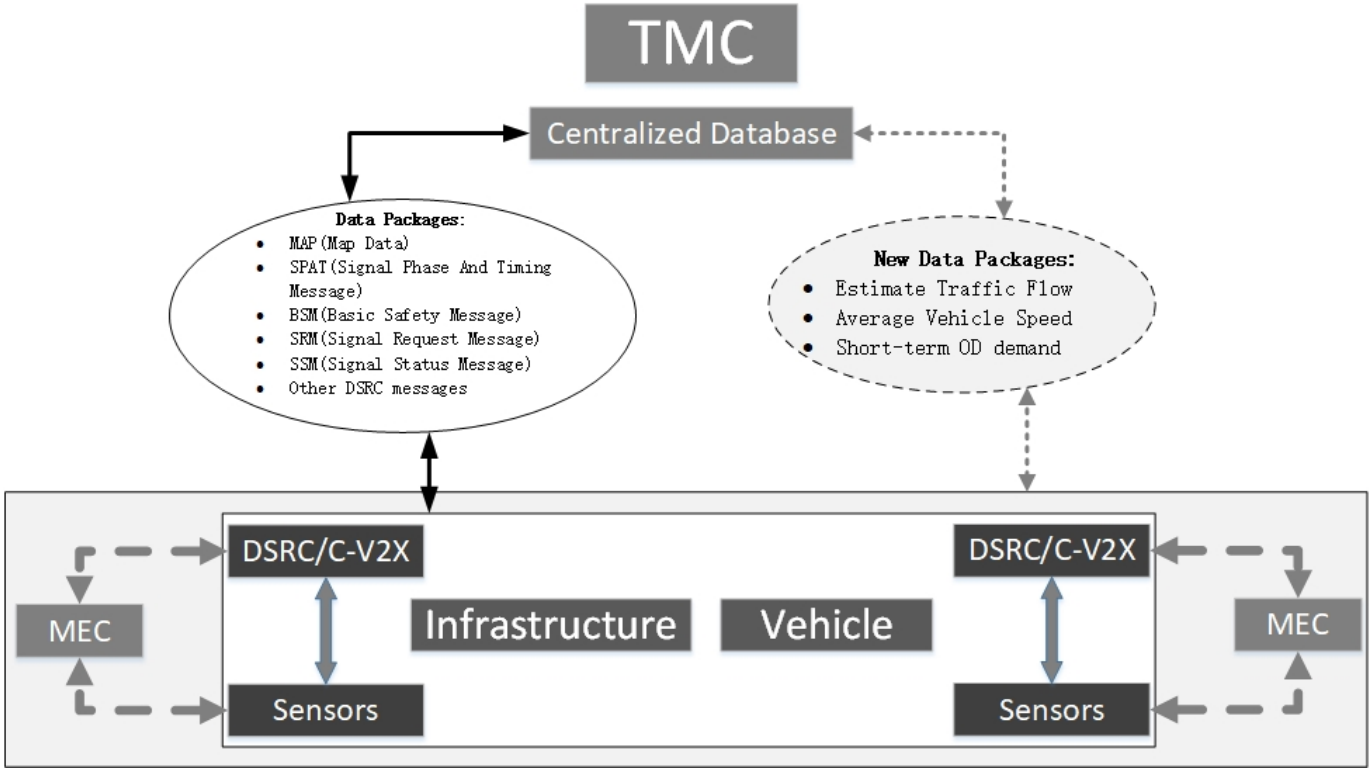


Figure 3-1 CV environment with/without MEC

The intercommunication in a CV environment, both with MEC and without MEC, includes three major parts: the TMC, infrastructure and vehicles. In traditional ITS architecture without MEC, all data is sent to and processed at a centralized TMC. However, the limited bandwidth of the backhaul connection between the centralized TMC (Central-TMC) and local infrastructure will be exhausted rapidly by the volume of data gathered by the increasing number of sensors in the ITS. Likewise, the latency will be insufficient for CV applications that require large amounts of CV data.

After introducing MEC into the framework, data can be processed and applications can be run on any local edge unit, regardless of whether it is an infrastructure edge or vehicle edge. Thus, the heavy computation workload for the central TMC can be evenly distributed to the lower-level mobile edge computers, increasing computation capacity and proximity for real-time data processing.

In summary, this work develops network traffic signal control system within the MEC-enabled CV environment. This environment is proposed for the testbed in South Campus,

Edmonton, Alberta, Canada. All the settings will be applied in field gradually within the next four years. MEC devices are located at each intersection of the network, collecting and delivering data, and controlling the traffic signal. These MEC devices at each intersection are called local MEC and include a computing server, roadside equipment (RSE), and a controller. Each vehicle has onboard equipment (OBE), and Central TMC has a cloud computing server. In addition, the whole network is decomposed into subnetworks with several intersections. Each subnetwork has a regional MEC controller. Each regional MEC controller is designed to receive messages from surrounding regional MEC controllers, make decisions for the traffic signal control of intersections inside the subnetwork and send messages to surrounding intersections.

There are three assumptions within the MEC-enabled CV environment:

1) *Short-term OD demands are available for the control needs.*

2) *Data transmission latency can be ignored because it is very low compared to the computational latency in this environment.*

3) *Computation of network traffic signal control can be distributed over MEC devices.*

Noted that, MEC-enabled CV environment assumptions are based on the fact that MEC technology has been and will be implemented in fields of transportation for both research and practice, see in Section 2.1. The framework of DSC physically consists of local MECs and regional MEC controllers, whose definitions coming from recent literatures as well.

3.2 Basic Assumptions on DTA

In this thesis, a queue-based DTA model is proposed to predict network traffic dynamics. Four basic assumptions are used:

1) *No Departure Time Choice:* Since traffic signal control requires a real time control strategy, travelers will not stop on their route to avoid high volume traffic. When travelers arrive at the network, they must move towards a path immediately.

2) No Reroute: When travelers enter the network, they will follow the same path until they exit the network, regardless of signal control plan changes.

3) *First In First Out (FIFO)*: The first vehicle that enters the link will also be the first to leave the link.

4) *Point Queue:* Upstream links are assumed to hold the queues at each signalized link. The link has infinite vehicle storage. Vehicles exit the queue with a constant discharge rate w.

3.3 Short-Term Demand

Traditional traffic control systems use highly localized counts collected by fixed location sensors as inputs to optimize traffic. With MEC and CV2X technologies, new data sources are available as inputs for network traffic signal control systems. There are some literatures highlighting the importance of adopting short-term origin-destination (OD) counts (as opposed to using these localized counts) for future traffic control (Ke et al., 2017)(Vlahogianni et al., 2014). This section discusses why short-term OD counts are superior to local counts for traffic management, and provides three definitions of short-term demand used for the rest of this thesis.

3.3.1 Limitation of Local Counts

Traffic data collected by fixed sensors such as loop detector and cameras include data types such as vehicle counts, mean vehicle speeds, density etc. These data types are used to describe traffic dynamics within several traffic models most notably METANET (Papageorgiou et al., 1989) and the Cell Transmission Model (CTM) (Daganzo, 1994). These two models have been used for and extended to many developments in traffic demand management over the last several decades. One of the basic assumptions in these two models is that flow dynamics follow the macroscopic fundamental diagram (MFD).

Figure 3-2 shows a typical MFD describing the relationship of space mean speed, traffic density, and flow. Before it has been demonstrated on empirical local counts (Geroliminis & Daganzo, 2008), the network flow model appeared early in 1969 (Godfrey, 1969). Because of its ease of use, MFD has been extensively applied to traffic studies, including the development of network-wide control strategies (Lele Zhang et al., 2020).

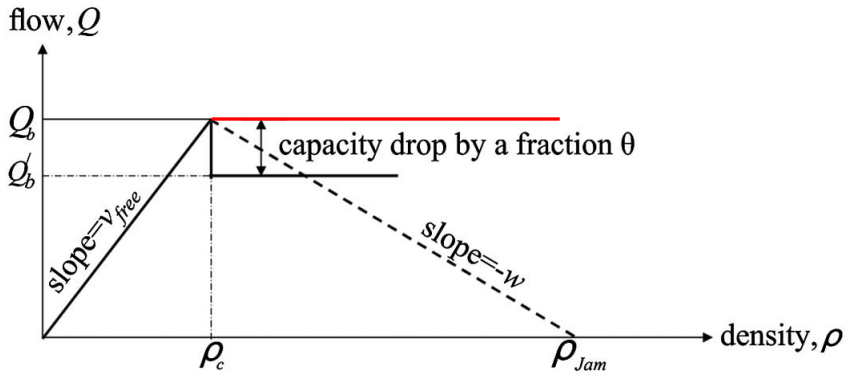


Figure 3-2 Macroscopic fundamental diagram.

However, in MFD, capacity dynamics and demand dynamics are both ignored. Flow equals demand when demand is less than capacity; however, when demand exceeds capacity, this is not the case any longer. In the fundamental diagram, shown in Figure 3-2, when density ρ is higher than the critical density ρ_c, capacity drops by a fraction θ, and demand should remain at least at the red line. And an investigation via simulation data also indicates that rapidly changing traffic demand affect MFD shape drastically(Ji et al., 2010). MFD cannot describe demand dynamics quite well, especially when traffic network is oversaturated. In addition, empirical evidence shows that freeway capacity is not constant even with the same condition for different days (Papageorgiou et al., 2008). MFD does present variations over time (Nguyen et al., 2016). This adds difficulty for MFD related models to capture capacity dynamics and demand dynamics in real-time control cases.

In addition, traffic state prediction from local counts is also short ranged from both space and time. Traffic flow collected by local counts on one link can affect travel cost of other links in the traffic network depending on ways it is distributed into the network. Because of its short range from space, traffic flow prediction in arterial area need to be updated frequently to maintain accuracy. Simulations have been made to show the poor performance in improving traffic efficiency of using local counts only for decentralized traffic control system (Tomforde et al., 2010). A survey on traffic state prediction shows the focus of research in this area is more on data with network attributes, such as short-term OD demand and vehicle trajectories (Yuan & Li, 2021).

3.3.2 Advantage of Short-Term Demand

Short-term demand forecasting has been the focus of countless papers over the decades (Vlahogianni et al., 2014). There are several challenges concluded by the author: complex arterial traffic environment, data resolution, excellent responsive algorithm, and travel time predict. In this thesis, an assumption is made that short-term OD demand is available in the MEC-enabled CV environment in Section 3.1.

Short-term demand is the general inputs for Dynamic Traffic Assignment (DTA) models. The literature review of Section 2.2 discussed that DTA was developed extensively over the past decades, proving an effective tool for determining traffic control strategies despite their complexity and computational burdens. And DTA model produces network traffic dynamic, captures the interactions between travelers and the whole transportation network, and thus becomes more suitable for network-level traffic signal control. The major advantage from using short-term OD demand is the ability to apply DTA model for network traffic signal control.

As the increasing usage of navigation systems (e.g., Google Maps), ride-sharing applications (e.g., Uber), and GPS-equipped vehicles, there will be also more chances to access short-term demand data which should include vehicle origin, destination, and start time. The foreseeable availability of short-term demand is the second advantage.

3.3.3 Definitions of Short-Term Demand

Three types of short-term demand – to be used in the remainder of this thesis – are described below:

1) Short-term demand for path

Short-term demand for path is defined as time-dependent flow entering a path, i.e., number of vehicles entering a path during one timestamp.

It is notated as $d_p(1, t)$, where p is the path id, t is timestamp and 1 means it is the first link of the path.

2) Short-term demand for links on the path

Suppose path p has l links, short-term demand for the k^{th} link on path p is defined as time-dependent flow on path p entering the k^{th} link, i.e., number of vehicles on path p enter k^{th} link during one timestamp.

It is notated as $d_p(k,t)$, where p is the path id, t is timestamp and k means it is the k^{th} link of the path.

Noted: One link can be shared by many paths, but short-term demand for links on path p only counts flow on p. Short-term demand for the same link on other paths may be different.

3) Short-term OD demand

Short-term OD demand is defined as time-dependent flow entering from origin node o and heading to destination node d, i.e., number of vehicles traveling from origin node o to destination node d during one timestamp.

It is notated as $d(od,t)$ where od is the OD pair and t is the timestamp.

Three types of short-term demand are differed by their objectives, such as paths, links, and OD pairs. Since short-term demand for path is equal to short-term demand to enter the first link on the path, it is a special case of short-term demand for links on path. Set of short-term demand for all paths is contained in set of short-term demand for links on all paths. In addition, short-term OD demand is the input for DTA model to predict short-term demand for links on all paths in traffic network. Thus, the relationship between the three short-term demand elements defined above is shown in Figure 3-3.

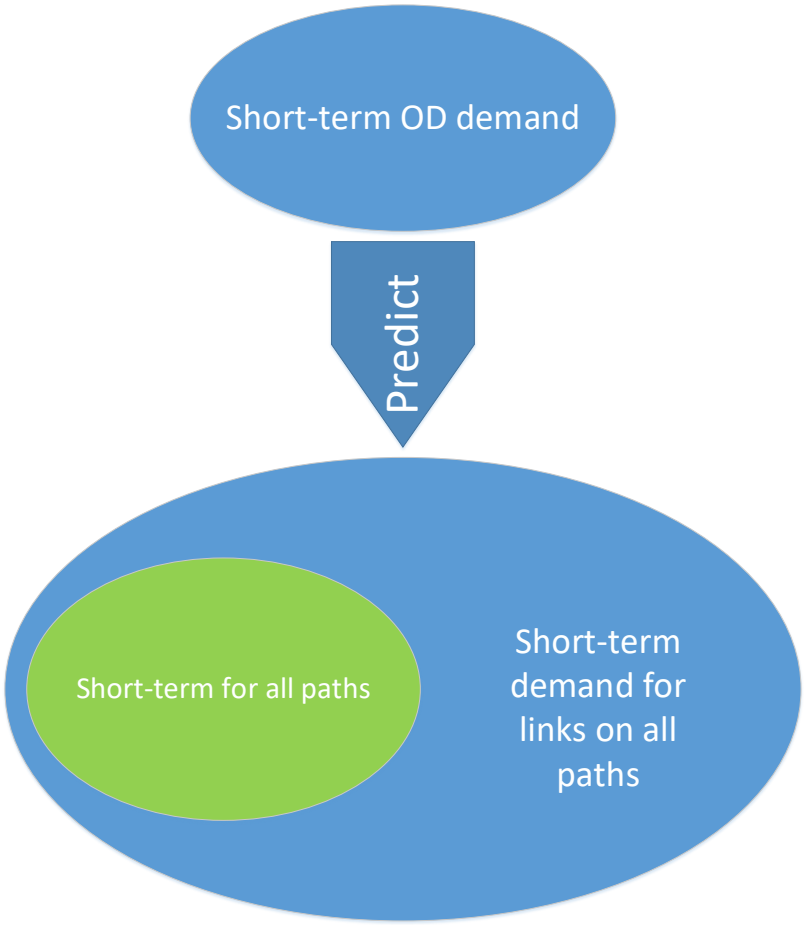


Figure 3-3 Relationship of three definitions of short-term demand.

4 A QUEUE-BASED DTA MODEL

This chapter a queue-based dynamic traffic assignment (DTA) model to predict traffic dynamics for network with real traffic signal timing plan. Traffic signal timing is formulated as a component of the link performance function with three control variables: cycle length, phase split, and offset. A G/G/n/FIFO queueing network is proposed to describe experienced travel times for links and paths. In addition, a numerical simulation is implemented in MATLAB to solve both user-optimal (UO) and system-optimal (SO) DTA problems of a corridor network among twelve cases with different demand inputs and traffic signal plans.

4.1 Notations

Table 4-1 shows the notations used in this chapter.

Table 4-1 Notations of Variables and Parameters

Symbol	Description
i, j, o, d	Index of node.
t	Index of timestamp, a unit of t does not stand for any second or minute.
a	Index of link. Let $a = (i, j)$, then the start node for link a is i and the end node for link a is j.
A	Set of all the nodes for the network.
p	Index of path. If link a is the k^{th} link on path p, $a = p(k)$.
$L(p)$	Index of length of path, i.e., number of links belonging to the path.
P	Set of all the paths for the network.
w_0	Maximum discharging flow rate for each lane of the link.
$f_{in}^{a,p}(t)$	Inflow rate for the lane on link a belongs to path p at time t.
$f_{out}^{a,p}(t)$	Outflow rate for the lane on link a belongs to path p at time t.
$f_{queue_{in}}^{a,p}(t)$	Inflow rate for the queue of lane on link a belongs to path p at time t.
$f_{queue_{out}}^{a,p}(t)$	Outflow rate for the queue of lane on link a belongs to path p at time t.
$T_{link}^{a,p}(t)$	Travel cost for the lane on link a belongs to path p at time t.
$T_{fft}^{a,p}(t)$	Free flow travel time for the lane on link a belongs to path p at time t.
$T_{queue}^{a,p}(t)$	Time spent to clear the queue of lane on link a belongs to path p at time t.
$T_{signal}^{a,p}(t)$	Signal waiting time for the lane on link a belongs to path p at time t.
$L_{queue}^{a,p}(t)$	Index of length of queue, i.e., number of vehicles in the queue of lane on link a belongs to path p at time t.
$X_{phase}^{a,p}$	Length of phase for path p at the intersection of link a.
$X_{offset}^{a,p}$	Offset for path p at the intersection of link a.
X_{cycle}^{a}	Cycle length of signal plan for the intersection of link a.
$L_{allred}^{a,p}$	Length for the red clearance phase for path p at the intersection of link a.
$T_{path}^{p}(t)$	Travel cost for path p at time t.
$\tau_k^p(t)$	Timestamp of vehicle entering the network at time t arrive at k^{th} link on path p
$\pi(t)$	Min travel cost of all the paths at time t.
$\widetilde{T_{link}}^{a,p}(t)$	Marginal travel cost for the lane on link a belongs to path p at time t.
$\widetilde{T_{path}}^{p}(t)$	Marginal travel cost for path p at time t.
$\tilde{\pi}(t)$	Min marginal travel cost of all the paths at time t.
$q^p(t)$	Demand for path p at time t.
$q(t)$	Demand for the single OD network at time t.

4.2 Introduction

Wait times at signalized intersections can contribute heavily to overall travel times. The development of Connected Vehicle to Everything (CV2X) technology can provide more detailed traffic signal timing information, more detailed traffic signal timing information can be accessed by travelers prior to their arrival at the intersection than years before, and travelers have opportunities to change routes when facing high signal wait times. I assume that all travelers have access to experienced travel times for each path immediately before they enter the network.

Since the seminal work of Merchant and Nemhauser (M-N model) in 1978, Dynamic Traffic Assignment (DTA) has been a central focus of transportation research (Merchant & Nemhauser, 1978a, 1978b). Given the time-dependent nature of demand and network characteristics, DTA models are used principally to predict dynamic traffic flow patterns over a network.

The traffic state prediction in DTA is based primarily on two principles in the traffic assignment literature. The first is User-Optimal DTA (UO-DTA). Ran et al. used the optimal control approach to formulate a convex model for instantaneous UO-DTA problem-solving that considers link flows as variables (Ran et al., 1993). If the actual travel times experienced by travelers departing at the same time are equal and minimal, then the dynamic flow over the network is in a travel time-based ideal dynamic user-optimal state for each OD pair at each interval of time. The second is System-Optimal DTA (SO-DTA). The objective of SO-DTA is to minimize the total travel time within the network, i.e., the travel time experienced by all users in the network. SO-DTA with full control can be formulated as a linear programing problem (Ziliaskopoulos, 2000).

As discussed in Section 2.4, although many researchers have considered traffic signal control in their DTA models, traffic signal timing has been simplified without considering more realistic phase sequencing and duration restrictions. In this chapter, signal wait time is included in the experienced path travel time, and the signal waiting time function is dependent on the control variables of cycle length, phase split, and offset.

Queueing theory is proposed to describe traffic dynamics. The earliest well defined queueing system was developed in 1953 (Kendall, 1953), with the notation written as X/Y/m/D/L, where X represents the distribution of intervals between arrivals, Y represents the distribution of service durations, m represents number of servers, D represents the queueing principle, and L represents the number of customers. Since the most popular distribution used in queueing system is exponential distribution, there are two major types of models according to the distribution type (X and/or Y), Markovian Queueing Systems and Non-Markovian Queueing Systems (Cruz et al., 2010). There are several principles of queue discipline, such as "First In First Out" (FIFO), "Last In First Out with Preemptive" (LIFO-PR), "Processor Sharing" (PS), and "Infinite Server" (IS). Moreover, according to the number of customers, there are two major types: closed queueing systems with a constant number of customers and open queueing systems with varying numbers of customers. These definitions cover most types of queueing systems (Karyotis & Khouzani, 2016).

There are also a number of queueing network model applications. Closed queuing network models are applied to the bike sharing problem (Q. L. Li et al., 2016). The same closed queueing network is also found in studies on the car rental system (George & Xia, 2011). In another application, a simulated queueing network with blocking created by traffic signal control achieves the system optimal (SO) assignment (Osorio & Bierlaire, 2013). Moreover, a multi-class queueing network is defined for air transportation in a study of transient congestion, solved by a decomposition algorithm (Peterson et al., 1995).

With this as our background, the main objective of this chapter is to formulate a queue-based DTA model to predict traffic dynamics in traffic network with real traffic signal timing plan.

4.3 Model Setup

Modeling work continues using the four basic assumptions for DTA presented in Section 3.2 (No Departure Time Choice, No Reroute, FIFO, and Point Queue), in addition to two more assumptions that describe travel behaviors:

1) ***UO-DTA***

For UO-DTA cases in this chapter, all travelers will choose the path with minimal experienced travel time.

2) ***SO-DTA***

For SO-DTA cases in this chapter, travelers make decisions to minimize the total travel time, i.e., the sum of travel times experienced by all travelers.

There are two theorems from the literature that rely on the assumptions above, that will be used in model development.

Theorem 1 *Marginal Cost for SO-DTA* (Ziliaskopoulos, 2000)

> A necessary and sufficient condition for SO-DTA in a single destination network is that all traveled paths from any cell, and departure time interval to the destination cell, have equal cost to the marginal cost of an additional unit of demand at that cell and time interval, while all untraveled paths have costs higher than or equal to the marginal cost.

Theorem 2 *Dynamic Process* (Friesz et al., 2013)

> The path delay operators usually do not take on any closed form. Instead, they can only be evaluated numerically through the dynamic network loading procedure.

There is an extended version of FIFO for network as **Lemma 1 & 2**.

Lemma 1 *FIFO for Paths*

The first vehicle that enters the path will also be the first to leave the path.

Proof According to ***No Reroute*** assumption in Section 3.2, vehicles will not change their paths. In addition, ***FIFO*** assumption in Section 3.2means that the first vehicle that enters the path will be the first to exit each link of the path, which proves **Lemma 1**. #

Lemma 2 *FIFO for single OD networks*

The first vehicle that enters the network will leave the network first in UO-DTA cases for the single OD network. The latter vehicle will never overtake any previous vehicle.

Proof Suppose vehicle A enters the network early at t_a and leaves the network at t_A, and there is another vehicle B that enters the network early at t_b and leaves the network at t_B. In addition, suppose $t_a < t_b$ and $t_A > t_B$, means that vehicle A enters the network earlier than vehicle B, but exits the network later.

According to ***UO-DTA***, let A choose path p_a at time t_a and let B choose path p_b at time t_b. Then path travel time $T_{path}^{p_a}(t_a) \leq T_{path}^{p_b}(t_a)$, since p_a is the path with minimal experienced travel cost at time t_a.

If A chooses p_b, According to ***FIFO*** and $t_a < t_b$, then A is in front of B on any link of p_b. Thus, the time A exits the network $t_A' < t_B$, and since $t_B < t_A$, $t_A' < t_A$.

Then $t_A' - t_a < t_A - t_a \implies T_{path}^{p_b}(t_a) < T_{path}^{p_a}(t_a)$.

If there is a contradiction with $T_{path}^{p_a}(t_a) \leq T_{path}^{p_b}(t_a)$, then the assumption is wrong. The lemma holds. #

Models of predicting traffic dynamics are built based on ***Theorem 1***, and ***Lemmas 1*** and ***2***. Marginal cost functions with traffic signal control will be formulated in Section 4.4, from which SO-DTA cases can be investigated according to ***Theorem 2***.

Note that ***Lemma 2*** only holds for UO-DTA cases. For SO-DTA cases, an upstream vehicle may overtake another previous vehicle, due to the fact that the path with minimal marginal experienced cost is not always the same as the path with minimal experienced travel cost. As explained shortly, the queueing process used in our simulation can avoid these conflicts.

4.4 Model Formulation

4.4.1 Link Dynamics

A general link $a = (i, j)$ is constructed as in Figure 4-1. The start point and end point of a link is denoted i, j. According to ***Point Queue***, there is a queue at the end of each lane of the link, which takes no space in the link. In Figure 4-1, there is one lane and one queue each for the left-turn, through, and right-turn movements. In addition, $(j, k_1), (j, k_2), (j, k_3)$ are links for signal phase left-turn, through, and right-turn respectively.

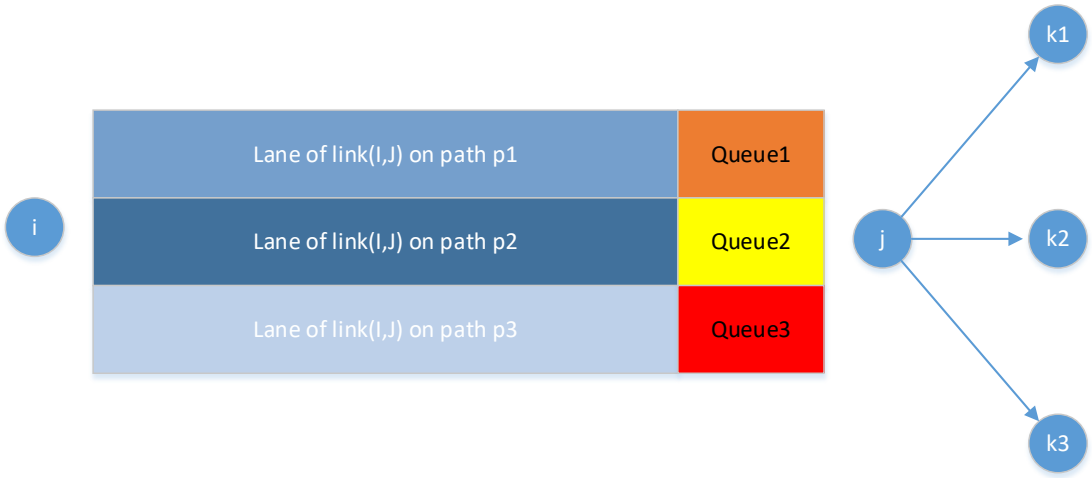


Figure 4-1 Sample of links.

Suppose link a is the k^{th} link on path p. Here Equations (4-1) - (4-3) describe conservations of flow and queue.

$$f_{in}^{a,p}\left(t - T_{fft}^{a,p}(t)\right) = f_{queue_{in}}^{a,p}(t) \tag{4-1}$$

$$f_{queue_{out}}^{a,p}(t) = f_{out}^{a,p}(t) \tag{4-2}$$

$$f_{in}^{p(k),p}(t) = f_{out}^{p(k-1),p}(t) \tag{4-3}$$

Equation (4-1) shows that the inflow rate of the queue for the lane of link a belonging to path p at time t equals the inflow rate of the lane at time $t - T_{fft}^{a,p}(t)$. That means when a vehicle travels to the end of the lane, it will join the queue directly. Equation (4-2) shows that the outflow queue rate will equal to the outflow of the link. Equation (4-3) shows that the inflow rate of the lane on path p equals the outflow rate of the lane of the previous lane on path p.

It is assumed that each traveler will know the experienced travel time for the path. For a single link, experienced travel time has three components, as in Equation (4-4). The first represents the time cost of travelers traveling from the start point of the link to the start point of the queue as $T_{fft}^{a,p}(t)$. Travelers arrive at the start point of the queue at time $t + T_{fft}^{a,p}(t)$. The time cost for travelers to get the front of queue is $T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right)$.

$$T_{link}^{a,p}(t) = T_{fft}^{a,p}(t) + T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right) + T_{signal}^{a,p}\left(t + T_{fft}^{a,p}(t) + T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right)\right) \quad (4\text{-}4)$$

When travelers reach front of queue at $t + T_{fft}^{a,p}(t) + T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right)$ and the signal is red, they will wait for it to turn green. The signal wait time is expressed as $T_{signal}^{a,p}\Big(t + T_{fft}^{a,p}(t) + T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right)\Big)$.

If $T_{fft}^{a,p}(t)$ is given, the second component $T_{queue}^{a,p}\left(t + T_{fft}^{a,p}(t)\right)$ indicates time spent to clear the queue, which is formulated as Equation (4-5). It is the ceiling for the ratio of queue length (number of vehicles in the queue) to maximum discharging flow rate w_0.

$$T_{queue}^{a,p}(t) = \left\lceil \frac{L_{queue}^{a,p}(t)}{w_0} \right\rceil \quad (4\text{-}5)$$

Figure 4-2 shows the relationship between queue length and time to clear queue. It is a step function that increases after the number of points that represent the maximum discharging rate w_0.

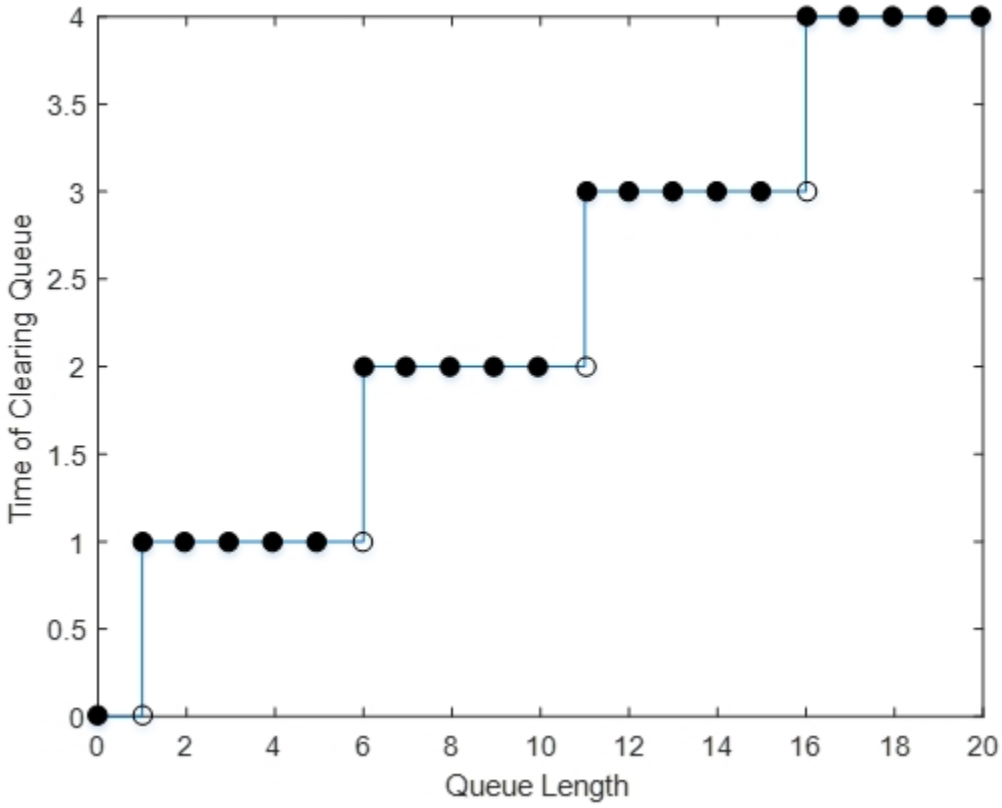


Figure 4-2 Relationship between queue length and time to clear queue.

Queue length function is calculated as Equation (4-6). It is the queue inflow rate minus the queue outflow rate, summed over previous period. When time index $t < T_{fft}^{a,p}(t)$, the initial queue length is 0. According to Equation (4-1) and (4-2), here gets Equation (4-7).

$$L_{queue}^{a,p}(t) = \sum_{\tau=T_{fft}^{a,p}(t)}^{t} \left(f_{queue_{in}}^{a,p}(\tau) - f_{queue_{out}}^{a,p}(\tau) \right) \quad (4\text{-}6)$$

$$L_{queue}^{a,p}(t) = \sum_{\tau=T_{fft}^{a,p}(t)}^{t} \left(f_{in}^{a,p}\left(\tau - T_{fft}^{a,p}(\tau)\right) - f_{out}^{a,p}(\tau) \right) \quad (4\text{-}7)$$

The outflow rate for the link is dependent on traffic signal control as Equation (4-8). When the traffic signal is red, i.e., signal wait time is positive, the outflow rate is 0. When the traffic signal is green, i.e., signal wait time is 0, the outflow rate is the minimum of queue length and maximum discharging flow rate w_0.

$$f_{out}^{a,p}(t) = \begin{cases} 0 & T_{signal}^{a,p}(t) > 0 \\ \min\left(L_{queue}^{a,p}(t), w_0\right) & T_{signal}^{a,p}(t) = 0 \end{cases} \quad (4\text{-}8)$$

The signal waiting time function is a cyclic function with a cycle length based on the cycle length of the traffic signal control in Figure 4-3.

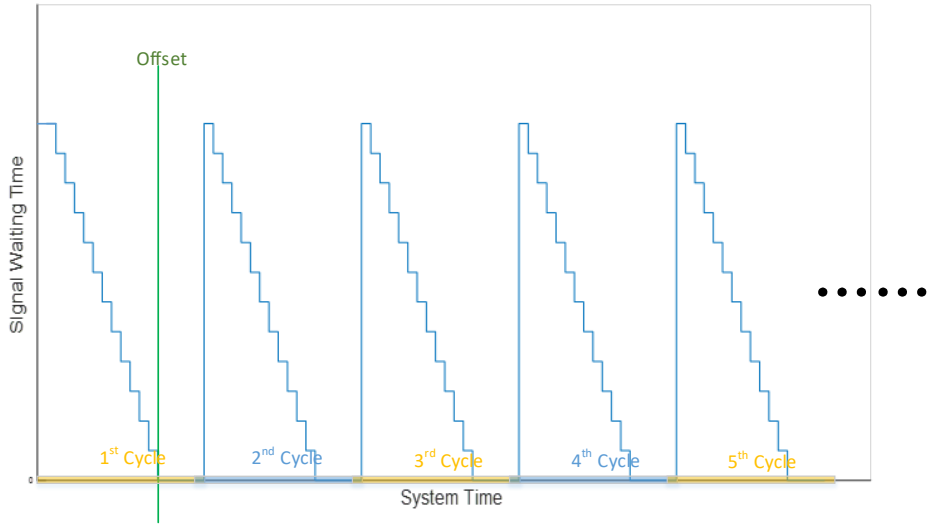


Figure 4-3 Signal waiting time function.

In addition, this part will formulate the signal waiting time function with cycle length X_{cycle}^{a}, phase split $X_{phase}^{a,p}$ (green time for the phase), and offset $X_{offset}^{a,p}$ as Equation (4-9). $\hat{t}$ is the reference time point in one cycle, which is equal to the remainder of $t + X_{offset}^{a,p}$ dividing X_{cycle}^{a} as per Equation (4-10). $X_{offset}^{a,p}$ is defined as the start point of green in the first cycle for the phase on link a belonging to path p. For example, offset is the green vertical line in Figure 4-3. Each phase will start with a red clearance as $L_{allred}^{a,p}$.

$$T_{signal}^{a,p}(t) = \max\left\{X_{cycle}^{a} - X_{phase}^{a,p} + L_{allred}^{a,p} - \hat{t}, 0\right\} \quad (4\text{-}9)$$

$$\hat{t} \equiv \left(t + X_{offset}^{a,p} + X_{phase}^{a,p} - L_{allred}^{a,p}\right) mod\ X_{cycle}^{a}, 0 \le \hat{t} < X_{cycle}^{a} \quad (4\text{-}10)$$

Note that the arrival rate and the service rate for the queue in each lane, before the traffic signals, are the multiplicative inverses of inflow rate $f_{queue_{in}}^{a,p}(t)$ and outflow rate $f_{queue_{out}}^{a,p}(\tau)$

respectively. Thus, the time-varying demand and traffic assignment principle (UO or SO) will determine the arrival rate for each queue as a specific distribution, and traffic signal control patterns (cycle length, phase split, offset) will determine the service rate as a periodic uniform distribution. In addition, the number of queues in the network is fixed. Travelers will follow the FIFO principle to clear the queue, and the total number of users for the network is not fixed. The network dynamics can be regarded as a G/G/n/FIFO open queueing network.

4.4.2 Path Travel Time

Path p is defined as a vector of ordered links$\{a_i\}$ that can be written as $p = \{a_i\}$.. If link a is the k^{th} link on path p, $a = p(k)$. $L(p)$ is length of path p, i.e., number of links in path p.

When a traveler on path p arrives at the k^{th} link $p(k)$, the time is recorded as $\tau_k^p(t)$. $\{\tau_k^p(t)\}_k$ is a sequence calculated via Equation (4-11) and (4-12).

$$\tau_1^p(t) = t \tag{4-11}$$

$$\tau_{k+1}^p(t) = \tau_k^p(t) + T_{link}^{p(k),p}\left(\tau_k^p(t)\right), 2 \le k < L(p) \tag{4-12}$$

Then the path travel time can be formulated as Equation (4-13).

$$T_{path}^p(t) = \sum_{k=1}^{L(p)} T_{link}^{p(k),p}\left(\tau_k^p(t)\right) \tag{4-13}$$

4.4.3 Marginal Travel Time for Link

The marginal cost of a link is the cost of an additional unit of demand added to a link. The calculation of differentials is based on this principle to compute the limit of $\frac{\Delta T}{\Delta f}$. A link's marginal cost can be formulated as Equation (4-14). Since $T_{link}^{a,p}(t)$ has three components as Equation (4-4) (free-flow travel time, queuing time, signal wait time), $\frac{\partial T_{link}^{a,p}(t)}{\partial f_{in}^{a,p}(t)}$ will be calculated as Equation (4-15).

$$\widetilde{T_{link}}^{a,p}(t) = T_{link}^{a,p}(t) + f_{in}^{a,p}(t) \times \frac{\partial T_{link}^{a,p}(t)}{\partial f_{in}^{a,p}(t)} \tag{4-14}$$

$$\frac{\partial T_{link}^{a,p}(t)}{\partial f_{in}^{a,p}(t)} = \frac{\partial T_{fft}^{a,p}(t)}{\partial f_{in}^{a,p}(t)} + \frac{\partial T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)}{\partial f_{in}^{a,p}(t)} + \frac{\partial T_{signal}^{a,p}\left(t+T_{fft}^{a,p}(t)+T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)\right)}{\partial f_{in}^{a,p}(t)} \quad (4\text{-}15)$$

Since $T_{fft}^{a,p}(t)$ is a constant function during the control period, here is

$$\frac{\partial T_{fft}^{a,p}(t)}{\partial f_{in}^{a,p}(t)} = 0 \quad (4\text{-}16)$$

According to Figure 4-2, Equation (4-1), and Equation (4-5) - (4-7), here is

$$\frac{\partial T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)}{\partial f_{in}^{a,p}(t)} = \frac{\partial T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)}{\partial f_{queue_{in}}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)} \begin{cases} 0 & L_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right) \ mod \ w_0 \not\equiv 0 \\ 1 & L_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right) \ mod \ w_0 \equiv 0 \end{cases} \quad (4\text{-}17)$$

According to Figure 4-3, and Equation (4-9) & (4-10), here is

$$\frac{\partial T_{signal}^{a,p}(t^*)}{\partial f_{in}^{a,p}(t)} = \frac{\partial T_{signal}^{a,p}(t^*)}{\partial t^*} \frac{\partial t^*}{\partial f_{in}^{a,p}(t)} \quad (4\text{-}18)$$

$$t^* = t + T_{fft}^{a,p}(t) + T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right) \quad (4\text{-}19)$$

$$\frac{\partial T_{signal}^{a,p}(t^*)}{\partial t^*} = \begin{cases} X_{cycle}^{a} - X_{phase}^{a,p} + L_{allred}^{a,p} & t^* \equiv X_{phase}^{a,p} + X_{offset}^{a,p} - L_{allred}^{a,p} mod \ X_{cycle}^{a} \\ 0 & otherwise \end{cases} \quad (4\text{-}20)$$

$$\frac{\partial t^*}{\partial f_{in}^{a,p}(t)} = \frac{\partial T_{queue}^{a,p}\left(t+T_{fft}^{a,p}(t)\right)}{\partial f_{in}^{a,p}(t)} \quad (4\text{-}21)$$

4.4.4 Marginal Travel Time for Path

The marginal cost of a path is the cost of one additional demand unit added into the path. According to Equation (4-13) & (4-15), here is

$$\widetilde{T_{path}^{p}}(t) = T_{path}^{p}(t) + f_{in}^{p(1),p}(t) \times \frac{\partial T_{path}^{p}(t)}{\partial f_{in}^{p(1),p}(t)} \quad (4\text{-}22)$$

The additional user entering the first link of path p at time t will arrive at the k^{th} link at time $\tau_k^p(t)$.

$$\frac{\partial T_{path}^{p}(t)}{\partial f_{in}^{p(1),p}(t)} = \sum_{k=1}^{L(p)} \frac{\partial T_{link}^{p(k),p}\left(\tau_k^p(t)\right)}{\partial f_{in}^{p(k),p}\left(\tau_k^p(t)\right)} \tag{4-23}$$

4.4.5 UO-DTA & SO-DTA

With the discussion above, the dynamic user optimal (DUO) conditions can be stated as Equations (4-24) - (4-27) according to ***FIFO***. And the dynamic system optimal (DSO) conditions can be stated as Equations (4-28) - (4-31) according to ***Theorem 1***. The UO-DTA and SO-DTA problems are to find feasible solutions for time-varying path demand $d^p(t)$ for all the paths constraint with DUO conditions and DSO conditions respectively (Chiu et al., 2010).

DUO conditions:

$$T_{path}^{p}(t) - \pi(t) \geq 0 \tag{4-24}$$

$$d^p(t)\left(T_{path}^{p}(t) - \pi(t)\right) = 0 \tag{4-25}$$

$$d^p(t) \geq 0 \tag{4-26}$$

$$\sum_{p \in P} d^p(t) = d(t) \tag{4-27}$$

DSO conditions:

$$\widetilde{T_{path}^{p}}(t) - \tilde{\pi}(t) \geq 0 \tag{4-28}$$

$$d^p(t)\left(\widetilde{T_{path}^{p}}(t) - \tilde{\pi}(t)\right) = 0 \tag{4-29}$$

$$d^p(t) \geq 0 \tag{4-30}$$

$$\sum_{p \in P} d^p(t) = d(t) \tag{4-31}$$

The main difference between UO-DTA and SO-DTA is that users choose paths with minimal experienced travel time in UO-DTA cases, while users are assigned to paths with minimal marginal experienced travel time in SO-DTA cases.

4.5 Numerical Example

The simulation uses a single OD network with seven paths. Both UO-DTA and SO-DTA cases are tested. Simulations are based on traffic dynamics discussed in previous Section 4.4.

4.5.1 Network Description

Figure 4-4 illustrates the single OD network with seven possible paths demarcated by colored arrows. The major corridor runs through the middle, labeled as path #5. The numbers on each link represents the free flow link travel time, while the numbers on the arrows at the bottom of the figure represent the free flow path travel time. There are two signalized intersections. The intersection with nodes 7,8,9,10 is noted as '***Intersection A***', and the other intersection is noted as '***Intersection B***'. For links inside each intersection, the left-turn lane costs 3 time units, the right-turn lane 2 time units, and the through lane 1 time unit. Each link is a one-way link with three lanes at most. Middle links $(1,10), (8,14)$ each have 3 lanes for left, right, and through movements each. For the SO-DTA case in this network, the traffic flow entering later has no impact on the previous flow. If they are on different paths sharing lanes as $(1,10), (10,8), or\ (8,14)$, they will abide by ***Lemma 1***; otherwise, they will not share lanes. The maximum discharge rate is set as $w_0 = 3$. The red clearance $L_{allred}^{a,p} = 2$ time units for each phase of the two intersections.

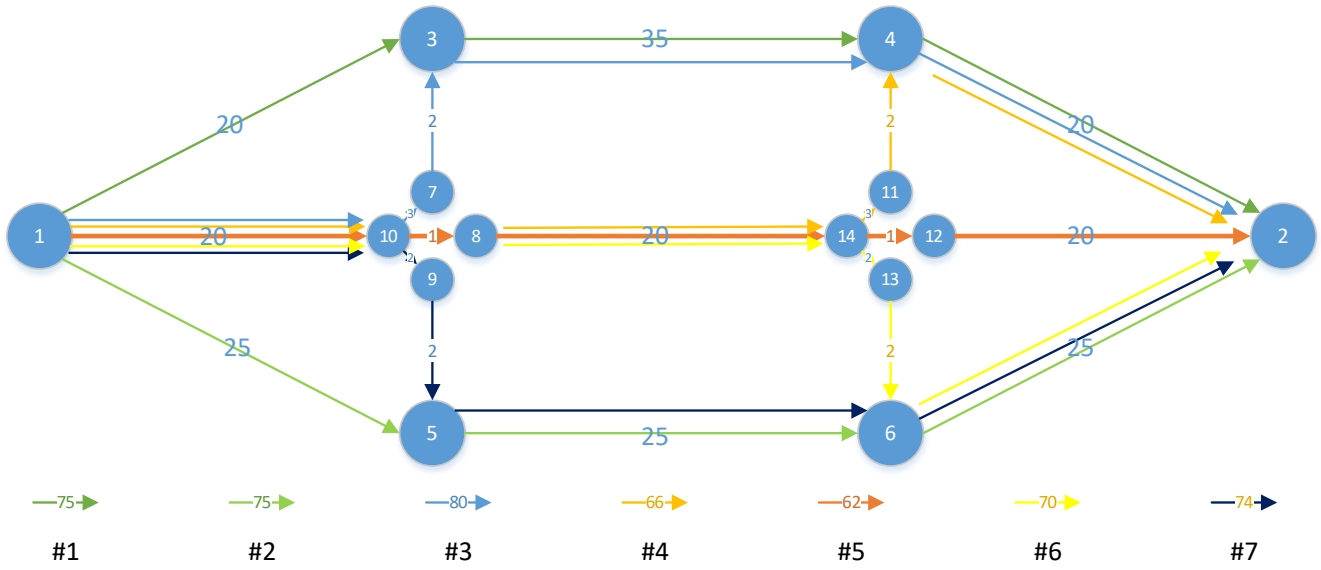


Figure 4-4 A Single OD Network.

The 3-phase signal plan for each intersection is shown as Figure 4-5. Offset is generally referred as the time relationship between coordinated phases-defined reference point and a defined master reference (master clock or sync pulse). Here offset for each intersection is defined as the same value as offset for the left-turn phase in Figure 4-5, which was mentioned before Equation (4-6).

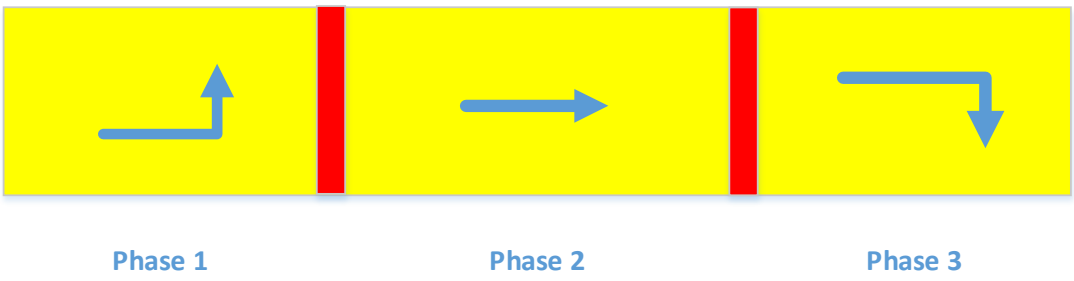


Figure 4-5 3-phase traffic signal plan.

4.5.2 Case Description

Each signalized intersection has a traffic signal timing plan for each group, as shown in Table 4-2. For example, green time of each phase for both intersections in ***Group 1*** is 10 time units. In addition, for each group, three different short-term demand profiles are tested. The time horizon is 100 time units for all demand profiles. Three demand profiles – Levels 1 through 3 – have 1, 3, and 5 vehs per time unit, respectively, for the network in Figure 4-4.

Table 4-2 Three Groups of Phase Plan for Two Intersections

Time	**Intersection A**				**Intersection B**			
Units	Phase 1	Phase 2	Phase 3	Offset	Phase 1	Phase 2	Phase 3	Offset
Group 1	10	10	10	0	10	10	10	0
Group 2	20	20	20	0	20	20	20	0
Group 3	10	40	10	0	10	40	10	0
Group 4	10	40	10	0	10	40	10	20

The phase split and offset are the same for ***Groups 1*** and ***2***; the impact of changing cycle length between these two groups will be investigated. The cycle length and offset are the same for ***Groups 2*** and ***3***, and the impact of changing phase split will be investigated. The cycle length and phase split are the same for ***Groups 3*** and ***4***, and the impact of changing offset between these two groups will be investigated as well.

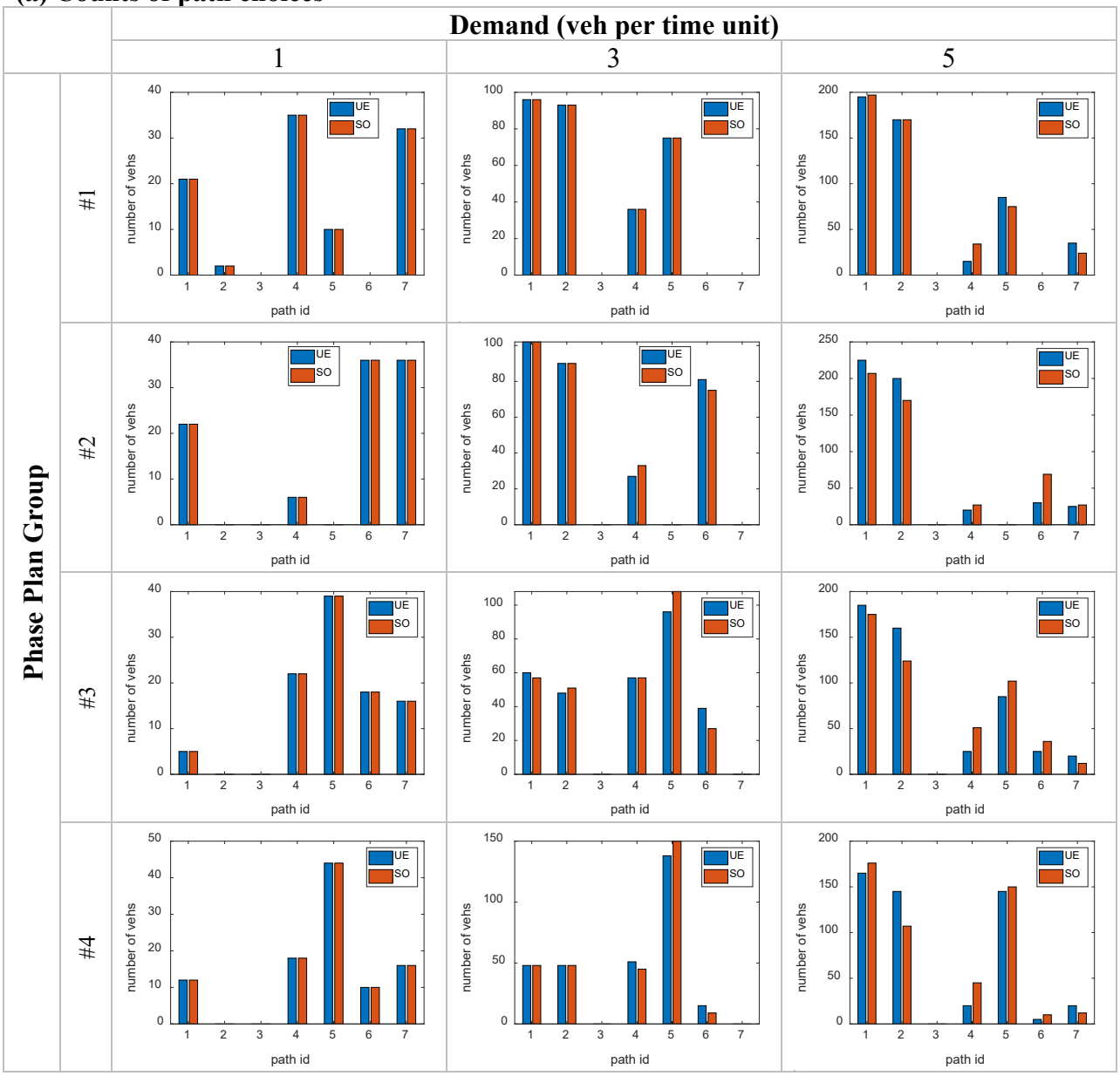
According to Section 4.3, travelers will choose paths with minimum experienced travel time in cases of UO-DTA, while they will choose paths with minimum marginal experienced travel time in cases of SO-DTA. Traffic assignment follows DUO and DSO conditions as per Equations (4-24) - (4-31).

4.5.3 Results

Results are show in Table 4-3.

Table 4-3 Results of 12 Scenarios

(a) Counts of path choices



(b) ATT for each user

UO(SO)		Demand Level		
		Level 1	Level 2	Level 3
Group of Phase Plan	Group 1	72.46(72.46)	74.18(74.18)	79.13(75.17)
	Group 2	72.76(72.76)	74.55(74.53)	79.70(75.16)
	Group 3	68.50(68.50)	73.02(72.84)	78.87(74.45)
	Group 4	67.50(67.50)	71.41(71.24)	78.03(74.51)

According to Tables 4-3(a) and 4-3(b), there are no differences between the UO-DTA cases and SO-DTA cases in either path sharing or ATT at ***Demand Level 1***. There are small differences at ***Demand Level 2***. However, there are significant differences at ***Demand Level 3***. As per Equations (4-17) and (4-20), the marginal experienced travel time increases when: 1) a vehicle joins the queue with more cost to clear the queue than the vehicle ahead of it, and/or 2) it meets red indication once clearing the queue. These two situations become more frequent at higher demands. As a result, SO-DTA cases performs better performance as much lower ATT than UO-DTA cases at ***Demand Level 3***. In addition, when demand increases from ***Level 1*** to ***Level 3***, ATTs for both UO-DTA and SO-DTA cases in the same group are increasing as well.

The path sharing rates differ among different groups (shown in columns of Table 4-3 (a)) because the signal wait times (and thus, queueing) are largely responsible for the path cost changes.

Group 2 has a phase plan with a higher cycle length than that of ***Group 1***. In most cases, ATT of cases in ***Group 2*** is higher than in ***Group 1***. However, when demand increases to 5 at ***Demand Level 3***, the SO-DTA case in ***Group 2*** performs the same as that of ***Group 1*** (See in Table 4-3 (b) – 75.16 vs 75.17). Given the stochastic simulation, no conclusion can be drawn that there is lower ATT at higher cycle lengths, and this should be further investigated in future work.

When increasing the green times for phase 2 in the signal plans of both intersections, there is a significant increase in volumes for path #5 in cases for ***Group 3***, from no volumes for this main path in cases of ***Group 2***, since path #5 benefits the most among all paths from phase 2 as shown in Figure 5.4 & 5.5. Increasing the green time for specific phases will absolutely change the path choices of travelers.

The offset of ***Intersection B*** is increased to 20 in ***Group 4*** from that in ***Group 3***. As shown in Figure 4-4**,** the free flow travel time for link (8,14) is 20 time units. This increase in offset makes it more likely that travelers on path #5 will meet a green indication when arriving at ***Intersection B***. Thus, improved offset makes most cases in ***Group 4*** with the best performance in ATT at the same demand level compared to other groups.

4.6 Summary

This chapter investigates the impacts of changing traffic signal control parameters – specifically, combinations of cycle length, phase spilt, and offset – on traffic assignment according to UO-DTA and SO-DTA.

To simulate traffic dynamics, this chapter first introduces a G/G/n/FIFO open queueing network. Equations are formulated for path travel time and marginal path travel time. In addition, DUO and DSO conditions help describe cases of UO-DTA and SO-DTA respectively.

12 cases are generated for simulation via three different demand profiles and four different traffic signal timing plans. ATT is used as the scale to evaluate all the cases. As expected, SO-DTA cases are shown to outperform UO-DTA cases in terms of less ATT. And SO-DTA cases perform more better than UO-DTA at higher demand level. In addition, when comparing the impact of traffic signal controls variables, such as cycle length, phase split, and offset, the results showed that all three variables have significant effects on path sharing and ATT.

The main contribution of this chapter was to integrate traffic assignment model with real traffic signal timing plan. The queue process described in this chapter also shows great potential as using for traffic dynamics prediction. Thus, this thesis continues to apply the queue-based process to optimize network traffic signal control in Chapter 5.

There are some limitations on the research presented in this chapter. Firstly, the model is only suitable for single-OD network. Travelers in arterial network with more ODs will influence the trajectory of each other. Thus, it is hard to formulate a close form of any link travel time function or path travel time function. To ignore this limitation and continue work on larger size network, instantaneous path travel time will be applied instead of experienced path travel time in Chapter 5. Secondly, more cases should be investigated to explore more solid results. Thirdly, this work only consider network with one-way links and three-phase traffic signal timing plan, further work can add more realism such as network with two-way links and standard four-phase traffic signal timing plan.

5. CENTRALIZED AND DECENTRALIZED SIGNAL CONTROL

This chapter develops and compares centralized and decentralized systems for optimal traffic signal control with short-term Origin-Destination (OD) demand as inputs. To add greater realism to the network traffic signal control optimization problem, a standard ring-and-barrier diagram with phase plans and all-red intervals are used for all intersections in test grid networks. The control variables include cycle length, green times of phases, and offset for each signalized intersection of the network. The optimization problem is formulated with a simulation-based Queueing Network Model (QNM), to minimize Average Travel Time (ATT) and Total Travel Delay (TTD). Computational latency is important for real-time applications. Thus, Computational Time (CT) is chosen to compare the performances of the Centralized Signal Control (CSC) and Decentralized Signal Control (DSC) systems as well. In addition, studies have shown network decomposition to have an important impact on system performance (Adacher & Tiriolo, 2020) (Chow et al., 2020). Therefore, the performance of CSC and DSC systems is investigated under different decomposition setups, using test scenarios constructed with varying demand profiles and grid networks.

From literature review in Chapter 2, concepts of centralized systems and decentralized systems can be defined in different ways from both hardware and software. Centralized signal control (CSC) system in this chapter uses central TMC to connect all the local devices and process all the data. The computation capacity of central TMC, therefore, is significantly high. The optimization algorithm runs inside the central TMC, which is simple and straightforward to optimize the traffic signals with all the data from the whole network.

Unlike CSC system, decentralized signal control (DSC) system in this part is a spatial interconnected distributed system. It is proposed to decompose the original network into a set of subnetworks. Each subnetwork has a regional MEC controller, which can communicate with its neighbors. Three functions as route guidance, traffic dynamics (traffic state prediction), and traffic signal optimization are embedded in the central TMC, the regional MEC controllers, and local MECs respectively, to fully use the computational resources and increase computational efficiency. Thus, from software side, programs of DSC are distributed over the whole network as well.

5.1 Notations

Table 5-1 contains notations and variables used in this chapter.

Table 5-1 Notation and Variables

Variables	Description
ij	Directed link from node i to node j; $ij \in A$
A	Set of links
$\{T_{ij}(t)\}_{ij\ in\ A}$	Travel cost matrix at timestamp t
$st(ij,t)$	Signal waiting time for signalized link ij at timestamp t
$gt(ij,t)$	Green time remaining for signalized link ij at timestamp t
$ft(ij)$	Free-flow travel time for link ij
FFT	Matrix of free-flow travel time (FFT)
M	Set of intersections
x_m	Control variable vector for intersection m, where $x_m = (cl_m, g_m^1, g_m^2, g_m^3, g_m^4, of_m)$
X	Set of control variables for all intersections
od	Original-Destination index, where o is origin node, and d is destination
OD	Set of od pairs
$d(od,t)$	Short-term demand for OD pair od at timestamp t
Δ	Set of short-term OD demand
T	Time horizon
P_{od}, P_s, P	Set of paths for od, for subnetwork s, and the whole network respectively
p	Path index, $p = \big(p(1), p(2), p(3), \dots, p(l)\big)$ is a vector with ordered nodes on the path
$l(p)$	Length of path p, i.e., the total number of nodes in path p

$tc(p,t)$	Instantaneous travel cost for path $p \in P$ at timestamp t
$d_p(k,t)$	Short-term demand at node $p(k)$ for path $p \in P$ at timestamp t
$d_p^s(k,t)$	Short-term demand at node $p(k)$ for path $p \in P_s$ in subnetwork s at timestamp t
w	Maximum discharging flow rate for each lane of the link
$L_{queue}(ij,t)$	Queue length for signalized link ij at timestamp t
$TA(\cdot,\cdot,\cdot)$	Function for the traffic assignment, inputs are short-term OD demand, traffic signal control variables, and FFT, output is short-term demand for the first link of all paths
$Q(\cdot,\cdot)$	Function for queueing process, inputs are short-term demand for the first link of all the path and traffic signal control variables, output is short-term demand for all paths.
s	Subnetwork index
S	Set of subnetwork indexes
$\Gamma_s(p,p_s)$	Path decomposition index

5.2 Introduction

The general goal to optimize signal timing control plans for an urban transportation network is to increase network traffic capacity, improve mobility (including reducing travel delays), and reduce traffic emissions. However, in many modern cities across the world today, adaptive traffic signal control systems such as SCATS (Lowrie, 1990) and SCOOT (Hunt et al., 1981) based on fixed sensors such as loop detectors are still widely used. These systems are poorly maintained and ineffective to deal with dynamic traffic demand. In addition, current computing capabilities of a centralized traffic management centre, coupled with the relatively high latency, low bandwidth, and reduced connectivity of communication between vehicles and infrastructures, as well as high data needs, means that network-wide traffic signal control optimization remains elusive.

To build the network traffic signal control system in this chapter, a MEC-enabled CV environment is proposed, within which vehicles and infrastructure can exchange data. MEC is assumed to handle data on the devices where it is generated, rather than data being sent to and received from the Traffic Management Center (TMC) cloud, which requires more time and therefore compromises real-time decision-making at the local level. In addition, MEC devices are located at each intersection of the network, collecting and delivering data, and controlling the traffic signal. These MEC devices at each intersection are called local MEC and include a computing server, roadside equipment (RSE), and a controller.

In this work, the framework of our DSC system is constructed within a Connected Vehicle (CV) environment using Mobile Edge Computing (MEC) to ensure a low impact of data transmission latency and enable distributed computing for network traffic signal control. Centralized systems are designed to solve the optimal control problem by processing all data inside a central Traffic Management Center (TMC). However, without simplifying variables or constraints or relaxing the control objective, the optimal control CT is high, causing high computational latency for real-time control. Data transmission latency worsens as more devices are connected to the central TMC. By adding MEC technology, data can be analyzed at the location where it is collected. Thus, traffic data can be transmitted through the network with low latency. In the DSC system, the network is divided into subnetworks, and each subnetwork has a

dedicated agent to control signals inside it. Thus, the DSC system decomposes the network traffic signal control problem in the CSC system, making it a distributed structure allowing reduced computational latency.

In addition, more kinds of data will soon be readily available in the fast-developing CV fields. As mentioned in the introduction, Origin-Destination (OD) demand data can be generated from a variety of sources such as navigation systems (e.g., Google Maps), ride-sharing applications (e.g., Uber), and GPS-equipped vehicles; these data can include vehicle origin, destination, and start time. However, in literature reviews, fixed sensor data, such as loop detector vehicle counts, is considered by many recent studies for network signal control problem. Traffic data collected by induction loop detectors and other standard sensors is short-ranged in both space and time scales. However, OD demand, a widely used input for network modeling, can help resolve these limitations.

Assumptions have been made in Section 3.1 that within the MEC enabled CV environment: **1)** Short-term OD demands are available for the control needs; **2)** Data transmission latency can be ignored, which is too low compared to computational latency in this environment. **3)** Computation of network traffic signal control can be distributed over MEC devices. Thus, short-term OD demand can be used as inputs for the control systems in this chapter as suggested from Section 3.3.

In addition to the work done in Chapter 4, this chapter will use a simulation-based Queueing Network Model (QNM) to predict traffic dynamics. Instead of UO-DTA assumption in previous chapter, traveler will choose the path with minimal instantaneous travel time.

This chapter is the main part for this thesis, work done in this section objects to present an analysis of centralized and decentralized signal control systems with short-term OD demand for network traffic, investigating the effect of different network decompositions on the performance of both systems.

5.3 Network Description

Since the assumption of traveler's behavior is different from Chapter 4 and the main purpose of this chapter is to solve traffic signal optimization problem instead of UO-DTA and SO-DTA problems. Network description will have some differences from previous chapter.

5.3.1 Travel Cost Matrix

The topology of the network is described using the travel cost matrix $\{T_{ij}(t)\}_{ij\in A}$, where ij is the link and t is a timestamp index. There are two types of links shown in Figure 5-1: signalized links (dotted) and general links (solid). All the links used in this work are two-way as per the arrow indications.

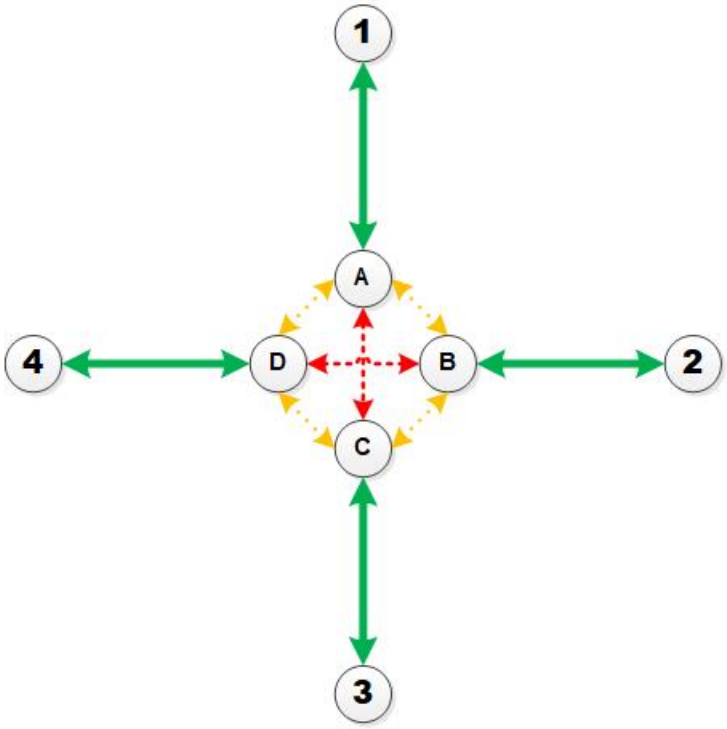


Figure 5-1 Sample of an intersection.

The travel time $T_{ij}(t)$ for a signalized link consists of signal wait time $st(ij,t)$ and turning time (also considered free-flow travel time $ft(ij)$ for signalized link), while that of general links consists of only free-flow travel time $ft(ij)$. $FFT = \{ft(ij)\}_{ij\in A}$ is the matrix of

free-flow travel times for all links. If node i is not directly connected to node j, then ij is not a link, and the travel cost is infinite. Equation (5-1) describes the travel cost matrix for all links and the network topology.

$$T_{ij}(t) = \begin{cases} 0 & i = j \\ \infty & ij\ is\ not\ a\ link \\ st(ij,t) + ft(ij) & ij\ is\ an\ intersection\ link \\ ft(ij) & otherwise \end{cases} \quad (5\text{-}1)$$

5.3.2 Control Variables

Figure 5-2 shows a standard ring-and-barrier diagram for a 4-leg intersection signal control strategy. our numerical simulations use this phase plan with control variables $\mathrm{X} = \{x_m = (cl_m, g_m^1, g_m^2, g_m^3, g_m^4, of_m)\}_{m \in M}$, where x_m is a vector consisting of cycle length, green times for the four phases and offset for signalized intersection m. Phase 1 is a protected left-turn phase for northbound and southbound movements; phase 2 is protected for northbound and southbound through and right movements; phases 3 and 4 repeat phases 1 and 2 for the eastbound and westbound directions. The bars between phases represent all-red intervals. Offset is between intersections on a corridor for the matching through-phase, based on a reference point or master reference.

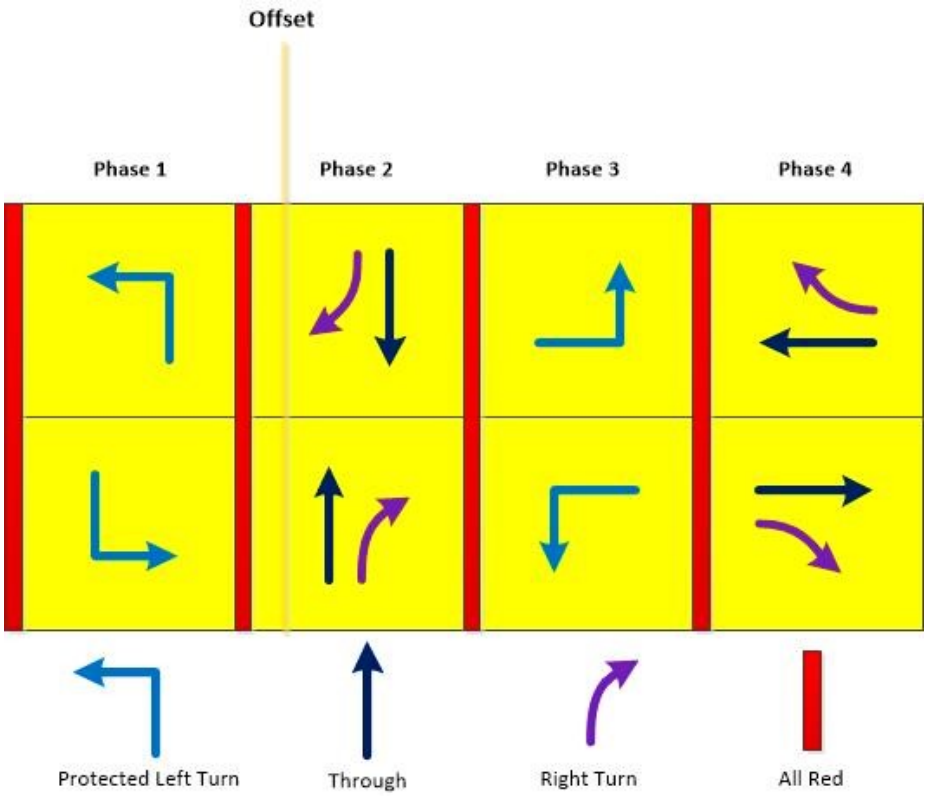


Figure 5-2 Standard Ring-and-Barrier Diagram.

ij is the signalized link for intersection m. Signal waiting time $st(ij,t)$in Equation (5-1) is computed according to the control variable x_m, which is a periodic function as per Figure 5-3. Similarly, the green time remaining function $gt(ij,t)$ can be inferred according to the control variable x_m.

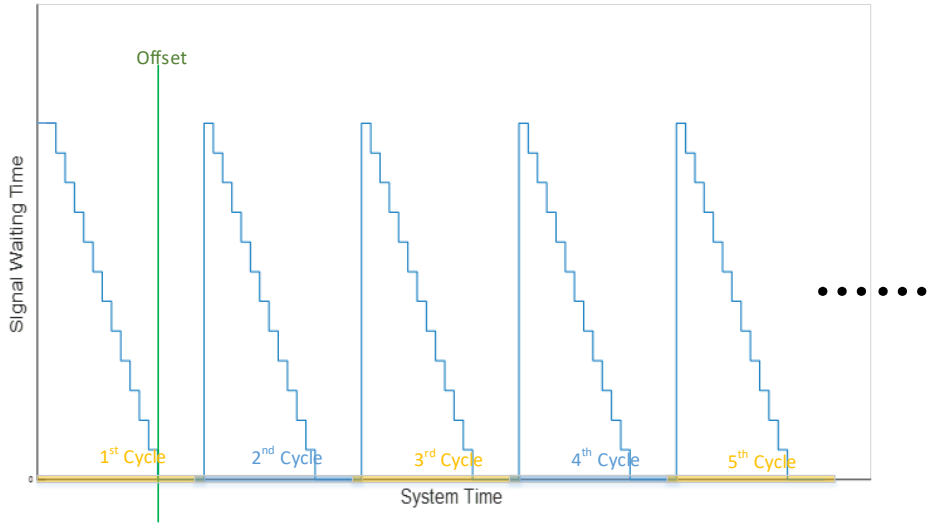


Figure 5-3 Sample of signal waiting time function.

5.3.3 OD demand and Traffic Assignment

Short-term OD demand represents the traffic volume between each origin and destination pair entering the network in each timestamp. Here uses $\Delta = \{d(od,t)\}_{od\in OD, t\in Z\ in\ [1,T]}$ to represent short-term OD demand for the network during the period $[0,T]$. T is the total time horizon.

'1-0' principle is applied to assign short-term OD demand, i.e., a traveler will choose the path with the minimum instantaneous travel time. It also assumes that travelers will not change routes.

A path is defined as a vector with ordered nodes along the path:

$$p = \big(p(1), p(2), p(3), \dots, p(l)\big) \tag{5-2}$$

The path length, $l(p)$, is quantified as the number of nodes on the path. The instantaneous travel time for path p at timestamp t is represented as follows:

$$tc(p,t) = \sum_{i=1}^{l(p)-1} T_{p(i)p(i+1)}(t) \tag{5-3}$$

The path with minimum instantaneous travel time for short-term demand $d(od,t)$ is expressed as:

$$p_{od}^{*} = \arg_{p \in P_{od}} \min tc(p,t) \coloneqq \{p_{od}^{*} \in P_{od} : tc(p_{od}^{*},t) \leq tc(p,t), \forall p \in P_{od}\} \quad (5\text{-}4)$$

Dijkstra's Shortest Path First algorithm (SPF algorithm) is used to search for Equation (5-4).

Here introduces demand for path $d_p(k,t)$ to describe the traffic flow dynamics. $d_p(k,t)$ represents traffic volume at node $p(k)$ for path p at timestamp t. And here assigns OD demand $d(od,t)$ to path p_{od}^{*} with minimal cost:

$$d_{p_{od}^{*}}(1,t) = d(od,t) \quad (5\text{-}5)$$

$TA(\Delta, \mathrm{X}, FFT)$ in Equation (5-6) is to describe the traffic assignment process that takes short-term OD demand, traffic signal control variables, and matrix of FFT as inputs to estimate short-term demand for the first link of all paths.

$$\{d_p(1,t)\}_{p \in P = \cup_{od \in OD} P_{od}, t \in Z \; in \; [1,T]} = TA(\Delta, \mathrm{X}, FFT) \quad (5\text{-}6)$$

5.3.4 Queueing Process

Three assumptions are used here as described in Section 3.2 for DTA.

The input for the queueing process is short-term demand for the first link of all paths $\{d_p(1,t)\}_{p \in P, t \in Z \; in \; [1,T]}$ plus traffic signal control variables X during the time horizon.

Figure 5-4 shows a sample link. The link has three lanes for left, right, and through movements, respectively. It is a general link followed by a signalized intersection. Here assumes that general links have three lanes (per direction) for each turning movement, while the downstream signalized links have one lane per direction.

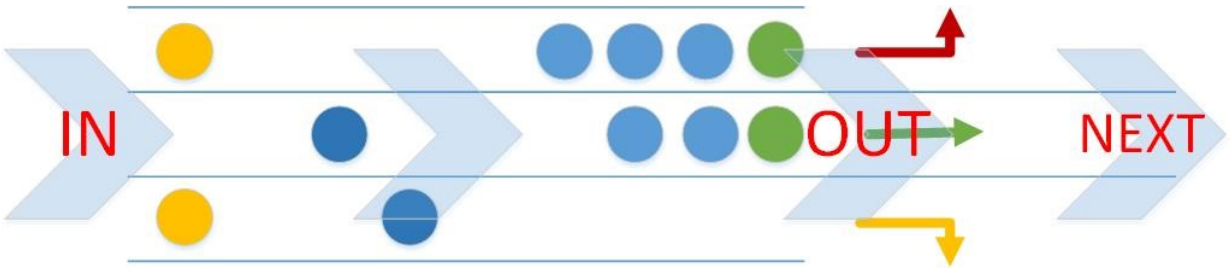


Figure 5-4 Sample link.

Suppose the queue length on a signalized link at timestamp t is $L_{queue}(ij,t)$. Considering the figure describes the movement for short-term OD demand $d_p(k,t)$, then node $n_{in} = p(k), n_{out} = p(k+1)$, and $n_{next} = p(k+2)$. In addition, after the time required to pass the link $ft(n_{in}n_{out})$, vehicles will join the queue such that $L_{queue}\left(n_{out}n_{next}, t + ft(n_{in}n_{out})\right)$. $n_{out}n_{next}$ represents the signalized link.

Suppose the signalized link $n_{out}n_{next}$ belongs to the h^{th} phase for intersection m. According to the control variable $x_m = (cl_m, g_m^1, g_m^2, g_m^3, g_m^4, of_m)$, additional information can be listed:

(1) Cycle length for the signalized link $n_{out}n_{next}$ is cl_m;

(2) Green time of the phase for the signalized link $n_{out}n_{next}$ is g_m^h;

(3) Signal waiting time for the signalized link $n_{out}n_{next}$ is $st(n_{out}n_{next}, t)$;

(4) Signal green time remaining for the signalized link $n_{out}n_{next}$ is $gt(n_{out}n_{next}, t)$

Traffic dynamics then can be predicted via a close loop:

Step 0 (Initialization): $t = 0, k = 1$, input demand profile for all paths

$$\{d_p(1,t)\}_{p \in P = \cup_{od \in OD} P_{od}, t \in Z \; in \; [1,T]}$$

Step 1 (Demand Update): $\forall p \in P = \cup_{od} P_{od}$,

$$n_{in} = p(k), n_{out} = p(k+1), n_{next} = p(k+2); \quad (5\text{-}7)$$

Find intersection m, which contains the signalized link $n_{out}n_{next}$.

if $\alpha = d_p(k,t) > 0$, then

$$L_{queue}(n_{out}n_{next}, t+\beta) = L_{queue}(n_{out}n_{next}, t+\beta-1) + \alpha; \quad (5\text{-}8)$$

$$\beta = ft(n_{in}n_{out}); \quad (5\text{-}9)$$

Equation (5-7) selects the k^{th} link on path p. Link layout is the same as Figure 5-4. Equation (5-8) updates the queue length for each signalized link. β in Equation (5-9) is the first part of total queueing time - the free flow travel time spent on link $n_{in}n_{out}$.

Calculate queueing time

Case 1: When signal is green, $gt(n_{out}n_{next}, t+\beta) > 0$

$$\Delta t = \beta + \left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta) - w \cdot gt(n_{out}n_{next}, t+\beta)}{w \cdot g_m^h} \right\rceil \times cl_m +$$

$$mod\left(\left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta)}{w} \right\rceil - 1, g_m^h\right) + 1 \quad (5\text{-}10a)$$

If a vehicle cannot pass the intersection during the green time in the current cycle of the signalized link, it will wait for another cycle for the next green phase. The time for the last vehicle in the queue to reach the front is $\left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta)}{w} \right\rceil$. The maximum discharging flow rate w describes the maximum number of vehicles that can traverse the signalized link in one timestamp.

$\lceil \cdot \rceil$ is an operator calculating the ceiling of the inside value, which is a minimal integer greater than $the\ value$. For example, if $L_{queue}(n_{out}n_{next}, t+\beta) = 3$ and $w = 2$, then $\left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta)}{w} \right\rceil = \left\lceil \frac{3}{2} \right\rceil = 2$. $mod(a, b)$ calculates the remainder of a divided by b. For example, $mod(7,4) = 3$.

Case 2: When signal is red, $gt(n_{out}n_{next}, t+\beta) = 0$

$$\Delta t = \beta + st(n_{out}n_{next}, t+\beta) + \left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta)}{w \cdot g_m^h} - 1 \right\rceil \times cl_m + mod\left(\left\lceil \frac{L_{queue}(n_{out}n_{next}, t+\beta)}{w} \right\rceil - 1, g_m^h \right) + 1 \quad (5\text{-}10b)$$

In Equation (5-10b), the additional term $st(n_{out}n_{next}, t+\beta)$ is the wait time (for the red signal turning green).

Updating Path Demand Function

$$d_p(k+1, t+\Delta t) = d_p(k+1, t+\Delta t) + \alpha \quad (5\text{-}11)$$

After exiting the k^{th} link on path p, traffic will enter the $(k+1)^{th}$ link on path p. Since constant maximum discharging rate w is used, $\alpha \leq w$ for each iteration.

Step 2 (Queue Length Update):

After demand for all the paths has been assigned, the queue length is updated.

$$L_{queue}(n_{out}n_{next}, t+\beta) = \begin{cases} L_{queue}(n_{out}n_{next}, t+\beta) & if\ T_{green}(n_{out}n_{next}, t+\beta) = 0 \\ L_{queue}(n_{out}n_{next}, t+\beta) - w & if\ T_{green}(n_{out}n_{next}, t+\beta) > 0 \end{cases} \quad (5\text{-}12)$$

From ***Step 1***, the length of the queue is increased for each positive demand unit. When the signal is green, queue length is reduced by the number of discharging vehicles; when the signal is red, the queue length will be the same as ***Step 1***.

If $k < l(p) - 1$, $k = k + 1$, go to ***Step 1***; otherwise, $k = 1$, go to ***Step 3***.

Step 3 (Stop Condition):

If $t < T + \delta$, $t = t + 1$, go to ***Step 1***; otherwise, Stop.

After all the path demand functions are updated for the current timestamp t, the next iteration will start for timestamp $t + 1$. δ is extra time for all the vehicles exiting the network.

In summary, Equations (5-7) – (5-12) describe the queueing process as $Q\left(\{d_p(1,t)\}_{p\in P,t\in Z\ in\ [1,T]}, \mathrm{X}\right)$ as shown in Equation (5-13), which takes short-term demand on the first link for all paths and traffic signal control variables as inputs, and short-term demand of all paths as outputs. The short-term demand for a given path consists of the demand for each link for the path and timestamp.

$$\{d_p(k_p,t)\}_{k_p\in Z\ in\ [1,l(p)],p\in P,t\in Z\ in\ [1,T]} = Q\left(\{d_p(1,t)\}_{p\in P,t\in Z\ in\ [1,T]}, \mathrm{X}\right) \quad (5\text{-}13)$$

Since general distributions are used for the demand (the arrival rate) and service rate, finite servers (signalized links), and FIFO assumption, network traffic dynamic model in this work can be regarded as a G/G/n/FIFO Queueing Network Model (QNM). The QNM is also open since the number of customers (vehicles) is not fixed.

In summary, network dynamics can be predicted using inputs including short-term OD demand, traffic signal control variables (cycle length, phase split, and offsets) for all intersections, and free flow travel time for each link. The data flow is summarized in Figure 5-5.

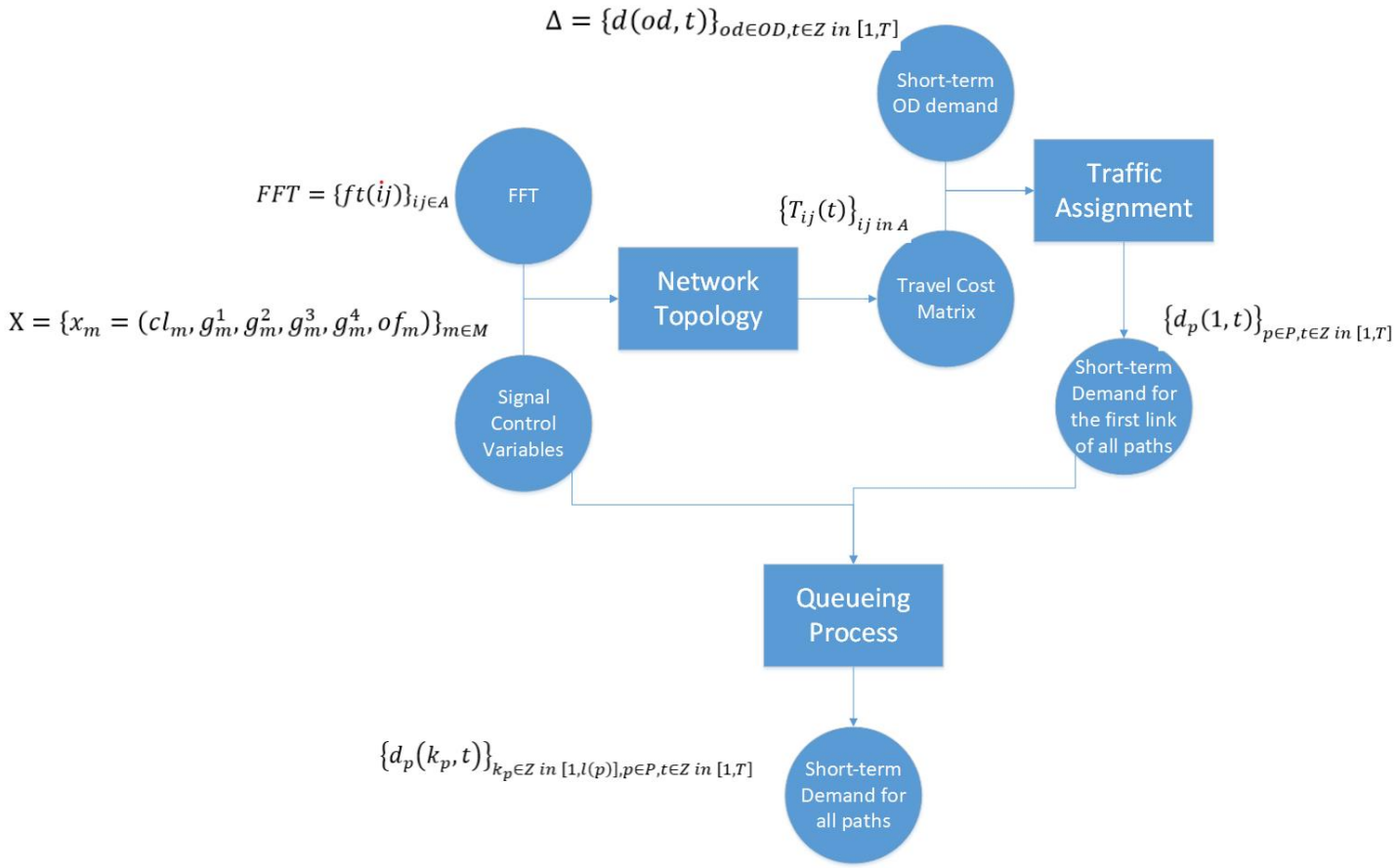


Figure 5-5 Data flow of network traffic dynamics prediction.

The travel cost matrix is estimated at the first step with inputs free-flow travel time (FFT) and signal timing. Traffic assignments are based on the '1-0' principle that assumes travelers will choose the minimal instantaneous path. The SPF algorithm is used to generate the shortest path. Thus, OD demand is transferred into volumes for the first link of all paths. Then, the traffic dynamic is estimated within QNM, combined with traffic signal control variables used in the initial step. This estimation is completed via simulation in MATLAB in the numerical simulation section.

5.4 Objectives of Traffic Signal Control

As one of the effective traffic control strategies, traffic signal control impacts traffic mobility, traffic safety, and environmental benefits. Input and output of traffic signal control models are categorized into these three subjects.

As mentioned in Chapter 2, as literature reviews on DTA-based network-level traffic signal control, the most common objective is maximum mobility of network traffic. However, maximum mobility can be translated into kinds of targets. In addition to minimizing Total Travel Time (TTT), Karoonsoontawong and Waller use the weighted sum of the expected total system travel time (TSTT) as an adopted robustness measurement in a Cell Transmission Model (CTM) (Karoonsoontawong & Waller, 2010). The input of this CTM is average time dependent on demand. Another well-known method for increasing traffic mobility is to minimize total delay (TD). Representatives of such traffic signal control systems are OPAC (Gartner, 1983), RHODES (Lucas et al., 2000), SCAT (A. Sims, 1979) and SCOOT (Hunt el al., 1982). The input of these systems is either time dependent link flows or inflow rates of traffic. Besides the direct objective, some researchers focus on increasing transit signal priority (TSP) (Christofa & Skabardonis, 2011) (Dion et al., 2004). Minimizing average delay is another measurement that can be used instead of minimizing TD (Mirchandani & Head, 2001), which uses sensors to detect a delay when a vehicle arrives at each intersection in the network.

When considering traffic safety, it seems that signal-warning flashers and speed controls are more effective for deterring accidents than intersection traffic signal control in terms of linear regression models (Wu el al., 2013). In addition, measurements for traffic safety usually rely on statistical models, such as before-and after studies based on historic accident/collision data, or prediction models, such as time-to-collision based on human factors and environmental parameters.

Environmental factors are often related to vehicle emissions. Researchers use measurements such as average/instantaneous speed, stops, acceleration, and deceleration in microscopic models (Rakha et al., 2000). However, although microscopic models are data intensive, calibration of them for a large-scale network is not usually easy work (Barceló & Casas, 2005). Some researchers, instead, tend to use macroscopic models to estimate emissions.

For example, CO2 (Carbon Dioxide) emissions are calculated using average speed and acceleration/deceleration in an AVENUE model (Advanced & Visual Evaluator for road Networks in Urban arEas) (Oda et al., 2004). However, macroscopic models lack accuracy, so in order to fill the gap between microscopic and macroscopic models, mesoscopic models are considered. For example, there are some researchers using functions of space mean speed to calculate Carbon monoxide (CO) emissions using a mesoscopic traffic flow model (Aziz & Ukkusuri, 2012).

According to the literature review above, each optimization objective requires specific data sets. This work only considers traffic dynamics without any safety or emission models. Thus, two macroscopic objectives are considered in this chapter: Average Travel Time (ATT) and Total Travel Delay (TTD).

5.5 Centralized Signal Control System

Section 6.3 outlines how to predict traffic dynamics using traffic signal control variables, short-term OD demand, and the FFT matrix within the G/G/n/FIFO open QNM. This work continues to develop a Centralized Signal Control (CSC) system to optimize network-wide signal control. Low data transmission latency is assumed within the MEC-enabled CV environment in both the CSC and DSC systems. Data transmission latency will be investigated in our network traffic signal control system in a future study after completing the device test in our Edmonton, Canada testbed. The objective functions are formulated using minimal Average Travel Time (ATT) and Total Travel Delay (TTD) and apply a three-step naïve method to optimize the network traffic.

5.5.1 Problem Formulation

Average Travel Time

ATT has been used as an objective or evaluation index in many network traffic signal control studies as shown in Table 2-1. ATT is the mean time spent by each vehicle traveling inside a network and is computed using the difference between a vehicle's network entry and exit times, as in Equation (5-14). However, However, assumption in Chapter 3 cannot generate trajectory data (time points of vehicles entering and exiting network) from short-term OD demand to calculate ATT.

$$ATT = mean(\Delta t_{veh}) = \frac{\sum_{i=1}^{n}\left(t_{veh_i}^{exit} - t_{veh_i}^{enter}\right)}{n} \quad (5\text{-}14)$$

Equation (5-15) simply rearranges the RHS of Equation (5-14) to represent all vehicles' mean exit and entrance time. Thus, ATT can be calculated by short-term demand of links for all paths as (5-16).

$$\frac{\sum_{i=1}^{n}\left(t_{veh_i}^{exit} - t_{veh_i}^{enter}\right)}{n} = \frac{\sum_{i=1}^{n} t_{veh_i}^{exit}}{n} - \frac{\sum_{i=1}^{n} t_{veh_i}^{enter}}{n} \quad (5\text{-}15)$$

$$ATT = \frac{\sum_{t=1}^{T+\delta}\sum_{p\in P} d_p(l(p),t)\times t}{n} - \frac{\sum_{t=1}^{T+\delta}\sum_{p\in P} d_p(1,t)\times t}{n} \quad (5\text{-}16)$$

δ is the additional time required for all demand in T to exit the network. Equation (5-17) is the total number of vehicles – the sum of short-term OD demand over the time horizon.

$$n = \sum_{t=1}^{T} d(od,t) \quad (5\text{-}17)$$

From Section 5.3, short-term demand for a path is generated from short-term OD demand Δ, control variable X, and FFT. Here can write one function – Equation (5-18) – to compute ATT from these three inputs.

$$ATT = \varphi(\Delta, \mathrm{X}, FFT) \quad (5\text{-}18)$$

Total Travel Delay

Total travel delay is the additional time all vehicles spend in the network due to signal waiting time, idling time in queues, and decelerating and accelerating. In this work, only signal waiting time and idling time in queue are considered due to the macroscopic representation of vehicle behavior.

TTD is calculated based on total queueing time with Equations (5-8) - (5-11), while total queueing time is calculated as Equation (5-19). Like Equation (5-18), here is one function to compute TTD from the original inputs as Equation (5-20).

$$TTD = \sum_{k\in Z\ in\ [1,l(p)]}\sum_{p\in P}\sum_{t\in Z\ in\ [1,T]} \Delta t \times \alpha \quad (5\text{-}19)$$

Where Δt and α were previously defined.

$$TTD = \phi(\Delta, \mathrm{X}, FFT) \tag{5-20}$$

5.5.2 Centralized Signal Control

Suppose the cycle length is the same for all intersections, while phases split and offset for different intersections can vary with control objectives. Assuming that an intersection signal's cycle length is the sum of green times, and offset is a positive integer less than the cycle length, the CSC problem can be formulated as follows:

$$\min_{\mathrm{X}\in\Omega} ATT \ \ or \ TTD \tag{A}$$

Subject to: $ATT = \varphi(\Delta, \mathrm{X}, FFT)$ or $TTD = \phi(\Delta, \mathrm{X}, FFT)$ (5-21)

$$\sum_{k=1}^{4} g_m^k \ = cl_m, \forall m \in M \tag{5-22}$$

$$cl_m = cl_n, \forall m, n \in M \tag{5-23}$$

$$0 \le of_m < cl_m, \forall m \in M \tag{5-24}$$

Where Ω is the feasible set for X,

$$\Omega = \left\{\mathrm{X}: cl_m = cl_n \in \Omega_{\text{cycle}}, (g_m^1, g_m^2, g_m^3, g_m^4) \in \Omega_{\text{phase}}, of_m \in \Omega_{\text{offset}}, \forall m, n \in M\right\}$$

$\Omega_{\text{cycle}} \subset Z_+$and $\Omega_{\text{offset}} \subset Z_+$ are finite subsets of Z_+. $\Omega_{\text{phase}} \subset Z_+^4$ is finite subset of Z_+^4.

Ω_{cycle}, Ω_{phase}, and Ω_{offset} are the feasible sets of cycle length, phase split, and offset for a signalized intersection. If the number of elements in each feasible set is n_{cycle}, n_{phase}, and n_{offset} respectively, then the size of the feasible set Ω is $|\Omega| = n_{cycle} \times n_{phase}^{4|M|} \times n_{offset}^{4|M|}$. The total number of intersections inside the network is $|M|$.

The feasible set is not convex and Constraint (5-21) is nonlinear. Network traffic signal control is known to be an NP-complete problem (Adacher et al., 2014) such that no method currently exists to solve the problem without simplifying variables or constraints, or relaxing the

control objective when the network size is relatively large. This work now attempts to address this issue.

5.5.3 Three-Step Naïve Method

As there is no suitable algorithm available to solve the optimization problem, a simple Three-Step Naïve Method is proposed.

After choosing ATT or TTD as objectives, this method decomposes the original ***Problem (A)*** into three sub-problems: ***Problem (A1)***, ***(A2)***, and ***(A3)***. The simple method is described as follows.

Three-Step Naïve Method

Step 1 Solve Problem (A1): Optimal cycle length: fix phase splits and offsets for all intersections, then find the optimal cycle length as per the optimization objective.

Step 2 Solve Problem (A2): Optimal phase split: based on the optimal cycle length in Step 1, fix the offset and generate all combinations of phase splits, then find the optimal phase split as per the optimization objective.

Step 3 Solve Problem (A3): Optimal offset: based on the first two steps, generate all the combinations of offsets for all intersections, then find the optimal offsets plan as per the optimization objective.

This idea to decompose original problem is coming from the traditional signal optimization steps, which has been implemented in the simulation software VISSIM as well. This naive method is not efficient as ACO, BP or RL methods used in the literature review. However, ACO, BP and RL methods have heavily simplified the assumption of signal timing plan, which are less realistic than assumptions used in the model of this chapter. In addition, although the Three-Step Naïve method is not advanced, it can give the global optimization solution for each subproblem, unlike some methods using algorithms such as ACO and RL only generate solutions less than the optimal.

This chapter consistently uses this method to solve network signal control in CSC system and in each subnetwork of DSC system. The main contribution of this work is to focus on the framework of DSC system, which is compared to the CSC system. Extended work of this thesis may improve the algorithm used for network signal control which have been discussed in the conclusion part of this thesis.

Problem (A1) - (A3) is listed below:

$$\min_{X\in\Omega_1} ATT \ \ or \ TTD \qquad \textbf{(A1)}$$

Subject to: $ATT = \varphi(\Delta, X, FFT)$ or $TTD = \phi(\Delta, X, FFT)$ (5-25)

$$g_m^k = \frac{cl_m}{4}, \forall m \in M, \ \forall k \in \{1,2,3,4\} \quad (5\text{-}26)$$

$$cl_m = cl_n, \forall m, n \in M \quad (5\text{-}27)$$

$$of_m = 0, \forall m \in M \quad (5\text{-}28)$$

Where $\Omega_1 = \left\{X: cl_m = cl_n \in \Omega_{\text{cycle}}, g_m^h = \frac{cl_m}{4}, of_m = 0, \forall m, n \in M, \forall h \in \{1,2,3,4\}\right\}$.

ATT or TTD is a convex curve with respect to cycle length, and ***Problem (A1)*** can be solved via the Bisection Algorithm (BA). If cl^* is the optimal cycle length from ***Problem (A1)***, then ***Problem (A2)*** is formulated as follows:

$$\min_{X\in\Omega_2} ATT \ \ or \ TTD \qquad \textbf{(A2)}$$

Subject to: $ATT = \varphi(\Delta, X, FFT)$ or $TTD = \phi(\Delta, X, FFT)$ (5-29)

$$\sum_{k=1}^{4} g_m^k = cl^*, \forall m \in M \quad (5\text{-}30)$$

$$of_m = 0, \forall m \in M \quad (5\text{-}31)$$

Where $\Omega_2 = \left\{X: cl_m = cl^*, (g_m^1, g_m^2, g_m^3, g_m^4) \in \Omega_{\text{phase}}, of_m = 0, \forall m \in M, \forall h \in \{1,2,3,4\}\right\}$.

Although ***Problem (A2)*** is simpler than ***Problem (A)***, the feasible set of the problem is still large and challenging to solve. Although some Genetic Algorithms (GA) may be applied to the problem, they cannot guarantee the solution is optimal, and therefore a global search is applied.

Suppose $\{(g_m^{1*}, g_m^{2*}, g_m^{3*}, g_m^{4*})\}_{m\in M}$ is the optimal phase split for all intersections from ***Problem (A2)***. Then ***Problem (A3)*** is formulated as follows:

$$\min_{\mathrm{X}\in\Omega_3} ATT \ \ or \ TTD \qquad \textbf{(A3)}$$

Subject to: $ATT = \varphi(\Delta, \mathrm{X}, FFT)$ or $TTD = \phi(\Delta, \mathrm{X}, FFT)$ (5-32)

$$0 \le of_m < cl^*, \forall m \in M \qquad (5\text{-}33)$$

Where $\Omega_3 = \{\mathrm{X}: cl_m = cl^*, (g_m^1, g_m^2, g_m^3, g_m^4) = (g_m^{1*}, g_m^{2*}, g_m^{3*}, g_m^{4*}), of_m \in \Omega_{\text{offset}}, \forall m \in M\}$.

The feasible set of this problem is convex, and thus a gradient search approach with a constant step size is used to update offset vector value.

If there are n intersections, the control variable can be written as a $n \times 6$ matrix $\tilde{X}$. Each row represents the traffic signal control variable $x_m, m = 1,2,3, \dots n$. The last column $\tilde{X}_{of}$ is a vector of offsets for all intersections, which is the control variable for ***Problem (A3)***.

The control variables are updated according to Equations (5-34) - (5-35).

$$\tilde{X}_{of}^{(i+1)} = \tilde{X}_{of}^{(i)} + \alpha d^{(i)} \qquad (5\text{-}34)$$

$$d^{(i)} = \arg_{d^{(i)}\in D} \min\left\{\tilde{X}_{of}^{(i)} + \alpha d^{(i)}\right\} \qquad (5\text{-}35)$$

Where $D = \{e_1, e_2, \dots, e_n\}$ is the standard basis for vector space R^n, $\alpha = \frac{cl^*}{\gamma}$, and γ is constant value which is a divisor of cl^*.

Finally, the solution for the network-wide traffic signal optimization can be achieved as

$\mathrm{X}^* = \{x_m^* = (cl^*, g_m^{1*}, g_m^{2*}, g_m^{3*}, g_m^{4*}, of_m^*)\}_{m\in M}$.

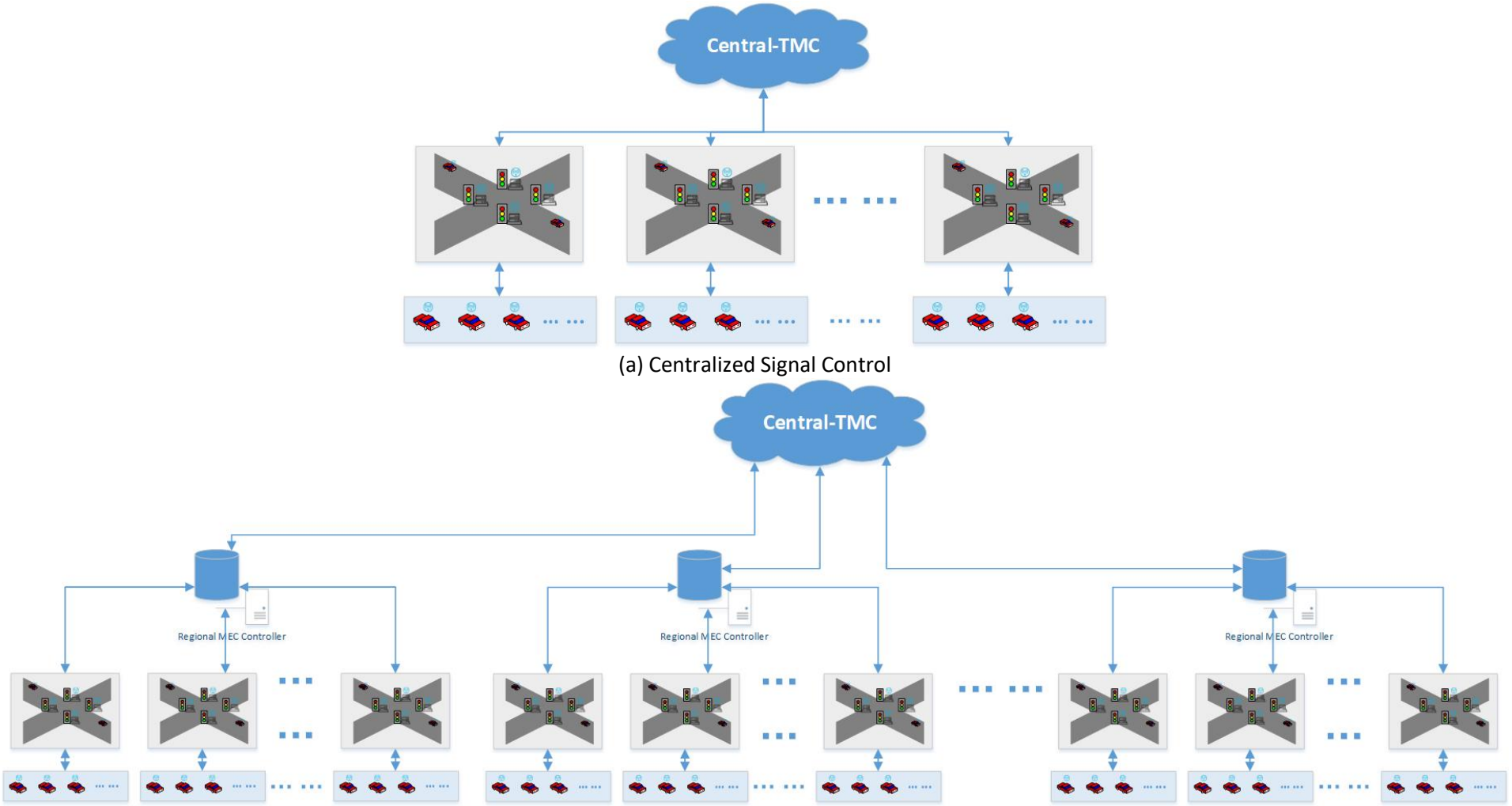
In summary, Section 5.5 formulated the optimization problem for centralized signal control (CSC) with short-term OD demand for a network. ATT and TTD are introduced as the control objectives with inputs as short-term demands for all paths. Because this optimization problem is NP-complete due to non-convex variables and nonlinear constraints, the problem is decomposed with a Three-Step Naïve Method whereby smaller sub-problems can be solved one by one. However, this problem remains costly to solve (which is demonstrated in Section 5.7) without developing a new algorithm for the problem.

5.6 Decentralized Signal Control System

Decentralized signal control (DSC) is also referred to as distributed signal control, and its development is motivated by the difficulty of solving CSC. Here, this chapter addresses this by developing a DSC system with short-term OD as inputs to control network traffic. Section 5.6.1 introduces the CV environment with Mobile Edge Computing (MEC). Section 5.6.2 proposes the decomposition method. Section 5.6.3 describes the two-layer process for our DSC system.

5.6.1 CV Environment with MEC

In addition to assumptions in Section 3.1, here introduces the CV environment by comparing the data flow of CSC and DSC.



(a) Centralized Signal Control

(b) Decentralized Signal Control

Figure 5-6 Data flow comparison between CSC and DSC.

In the CSC system (Figure 5-6(a)), each vehicle has on-board equipment (OBE) to communicate with each intersection. There is local MEC at each intersection to facilitate communicate between OBE and Central-TMC. There is a central cloud server for the Central-TMC to communicate with local MEC, store and analyze data, and optimize the traffic signal control.

OBE units collect the basic safety message (BSM) data from vehicles. BSM data include real-time speed and real-time location information and are sent to the local MEC at each intersection by the OBE units. Local MEC will collect and analyze data locally before send to the Central-TMC. According to the data received, the Central-TMC then analyses and optimizes for the signals. Finally, the Central-TMC sends the updated signal strategy to the local MEC at each intersection. The local MEC then changes the signal plan for the intersection and controls the traffic according to the updated signal strategy. In addition, local MEC sends the signal information to OBE units of nearby vehicles.

In the DSC system (Figure 5-6(b)), there is a virtual regional MEC controller for several intersections in a small region. Local MECs in the regional will follow the commands from the regional MEC controller. And the regional MEC controller will collect data from local MECs. In addition, the regional MEC controller will assign computation work to all the local MECs in order to compute parallelly. This additional capability facilitates the DSC system to divide the network into subnetworks, unlike the CSC system. A regional MEC controller is assigned to each subnetwork, that combines local MECs as a regional computing server, and is able to connect to the central cloud server of Central-TMC as well as to the local MECs and OBEs inside the subnetwork.

The vehicle's OBE collects and sends time-varying OD information and BSM data to the local MECs. All the data will be analyzed locally by the local MEC and send to the regional MEC controller. The regional MEC controller then analyzes the local data and sends it to the Central-TMC. The Central-TMC assigns the time-varying demand to paths, which is the route guidance information. This information is delivered to the regional MEC controller and then to each end of the traffic network. Meanwhile, the regional MEC controllers estimate the traffic dynamics for the whole network with local MECs parallelly. Finally, each regional MEC

controller obtains the time-varying OD demand for each subnetwork and optimizes the signals inside the subnetwork, respectively.

5.6.2 Network Decomposition

Network decomposition is highlighted in recent DSC research (Adacher & Tiriolo, 2020) as an important factor for optimizing DSC system. In my work, a grid network is decomposed into grid subnetworks. Besides spatial decomposition, this work also decomposes demand for paths of the whole network into demand for paths for each subnetwork. Thus, each subnetwork has an individual short-term demand profile to optimize network signals inside it as a small CSC system.

Figure 5-7 shows an example for a 6x6 network decomposed into 9 2x2 subnetworks.

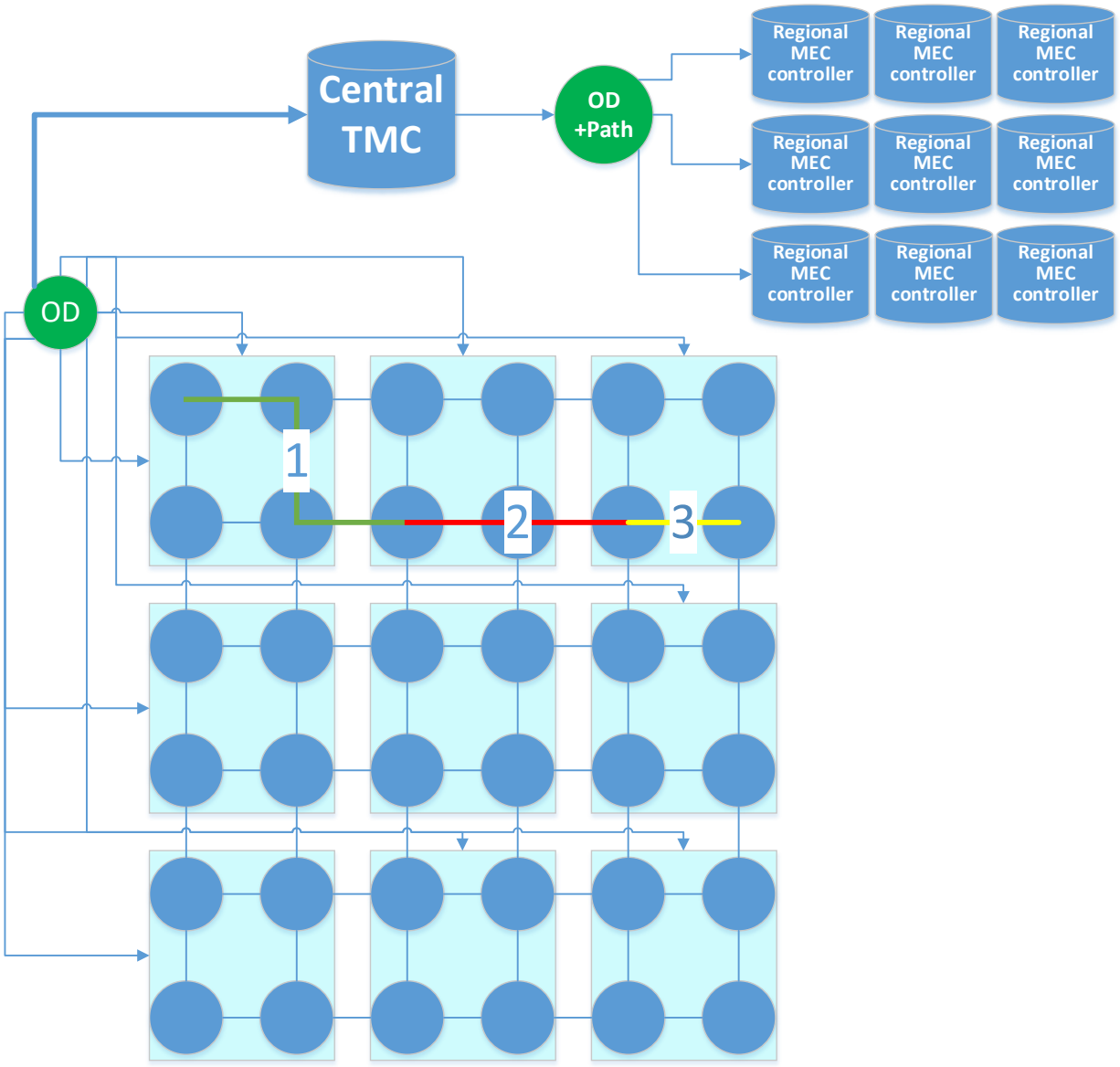


Figure 5-7 An example of network decomposition.

After short-term OD demand is collected at each end of the network, the Central TMC will distribute OD and path information to regional MECs in subnetworks. As '*No Reroute*' assumed in Section 3.2, travelers will not reroute on their way. The path from the node in the top left corner to the second-to-top right node, illustrated in the figure, is divided into three sub-paths through three subnetworks marked as 1, 2, and 3.

Path decomposition is achieved by the Path Decomposition Index (PDI) $\Gamma_s(p, p_s)$, where s is the index for the subnetwork, p is path for the whole network, and p_s is path for subnetwork s. $\Gamma_s(p, p_s) = m$ if the m^{th} node of path p is the first node of path p_s for subnetwork s. For example, suppose the index for the top middle subnetwork in Figure 5-7 is 2. For path p shown, since the 4^{th} node for the whole path p is the first node of the subpath p_2 inside the subnetwork, then $\Gamma_2(p, p_2) = 4$.

Say $d_p^s(k, t)$ is defined as short-term demand at node $p(k)$ for path $p \in P_s$ in subnetwork s at timestamp t. Suppose $\Gamma_s(p, p_s) = m$. Then,

$$d_{p_s}^s(1, t) = d_p(m, t) \tag{5-36}$$

5.6.3 Two-Layer Process

With the introduction of a CV environment with MEC and network decomposition with PDI, Figure 5-8 illustrated the two-layer process for DSC system with short-term OD demand used for this work.

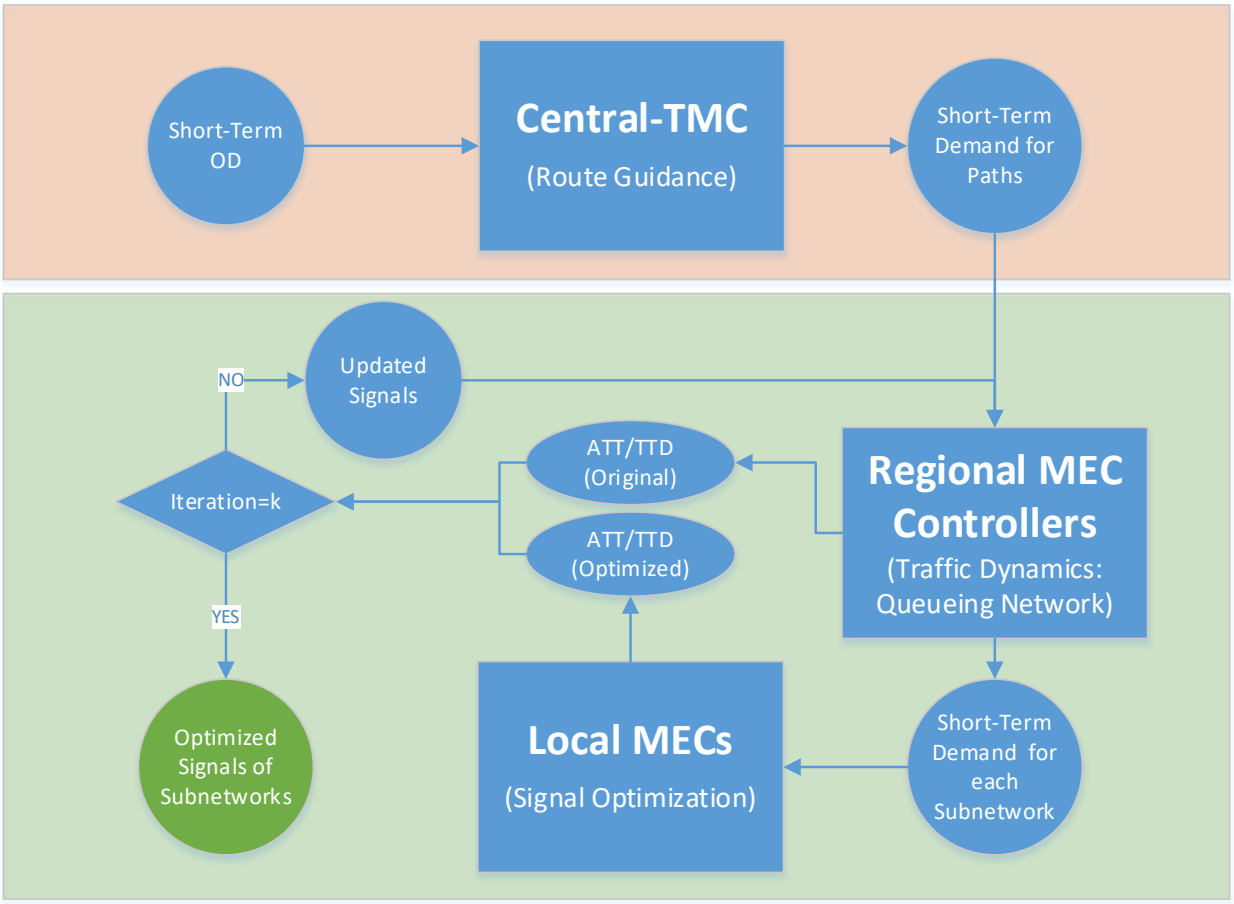


Figure 5-8 Two-layer process for decentralized signal control system.

In the first layer, the input is the short-term OD matrix. The Central-TMC will gather this information from the regional MEC controllers. The output from this first layer – the short-term demand for each path – is input to the second layer. The second layer has two functions: traffic dynamics (traffic state prediction) on regional MEC controllers and traffic signal optimization on local MECs inside each subnetwork.

Each subnetwork can be considered a small CSC system since the traffic signal optimization is achieved via the three-step naïve method, the same as for CSC systems. In addition, traffic dynamics are achieved using the same queueing process. Since the network traffic reacts to changes in traffic signal timing plans, traffic dynamics will be re-predicted after updating signals. This loop will run until the convergent or iteration number is up to a pre-set constant integer k.

In the CSC system, all three functions are run inside the cloud server of central TMC. However, in the DSC system, the sole function of the Central-TMC is to calculate short-term path demand to provide route guidance. The remaining optimization tasks are allocated to the regional MEC controller inside each subnetwork. In addition, local MECs in the subnetwork are involved in the computation of the traffic signal optimization function. All the computation resources of the DSC system are fully used.

In summary, Section 5.6 introduced the environment and decomposition method of DSC system in the first instance. The structure of DSC system was described as a two-layer process.

The DSC system has three main functions: route guidance in central TMC, traffic dynamics prediction in regional MEC controller, and signal optimizations with local MECs in each subnetwork. PDI plays a vital role in allocating sub-jobs to each regional MEC controller. Considering that each subnetwork functions as a small CSC system, the DSC system, as outlined here, is an extension of the CSC system that enables network expansion.

5.7 Numerical Simulation

This work uses analytic simulation implemented via MATLAB to demonstrate the proposed model, instead of an existing traffic simulation software such as SUMO, Aimsum or VISSIM. The main reason for choosing analytic simulation over these software programs is that the network signal control problem addressed in this thesis includes subproblems such as network decomposition and traffic assignment, which would require coding of external APIs in order to embed into the simulators. Mainly, however, this thesis focuses on macroscopic traffic optimization, which does not require the additional functionality and detail that the simulators named provide. As a result, an analytic simulation approach was deemed more suitable for this work.

The computation is finished in the Compute Canada server. The maximum number of cores is 40, and the speed for each core is 2.4 GHz. Both the CSC and DSC systems are demonstrated below in five scenarios. 1×1, 2×2, 3×3, 4×4, 6×6 grid networks are tested as illustrated in Figure 5-9.

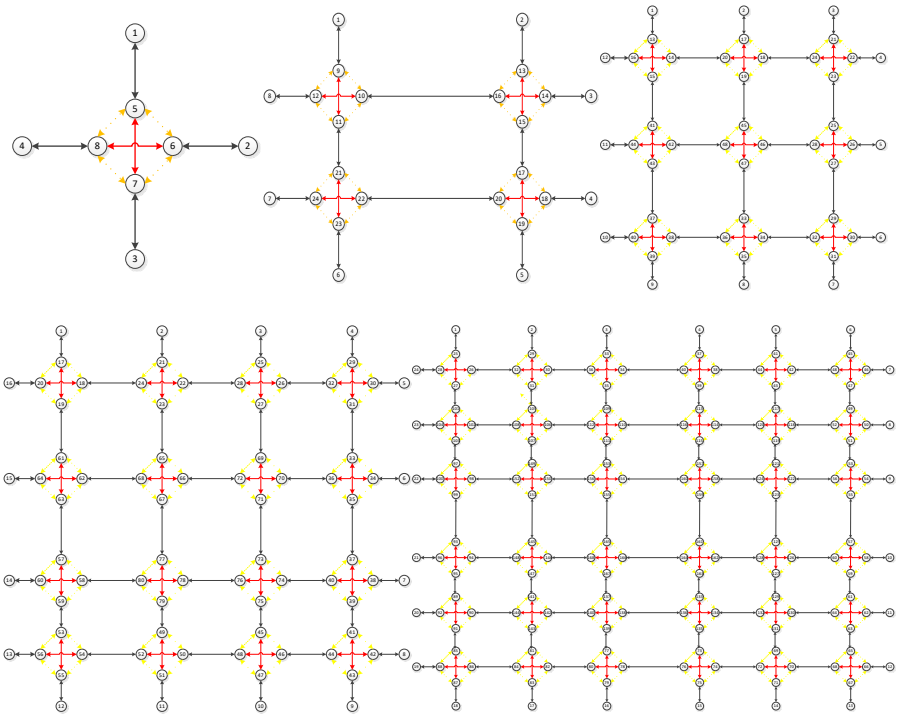


Figure 5-9 Grid networks.

The basic settings for all the cases are as follows:

(1) Time horizon: $T = 300$ time units, demand is randomly generated from a uniform distribution (variance equals mean value) for OD pairs within the time horizon;

(2) Free-flow travel time for all general links: 10 time units;

(3) Left turn, right turn, and through for signalized links cost 3, 2, and 1 time units, respectively. All-red time for signal control is 2 time units. The discharge rate for each signalized link is $w = 2$ vehicles per time unit;

(4) A standard ring-and-barrier diagram (Figure 5-2) is used for all intersection signal timings;

(5) Cycle length is the same for all intersections. Phase splits (i.e., green times) and offsets may vary among different intersections. The unit of control variables and two objectives are in time units.

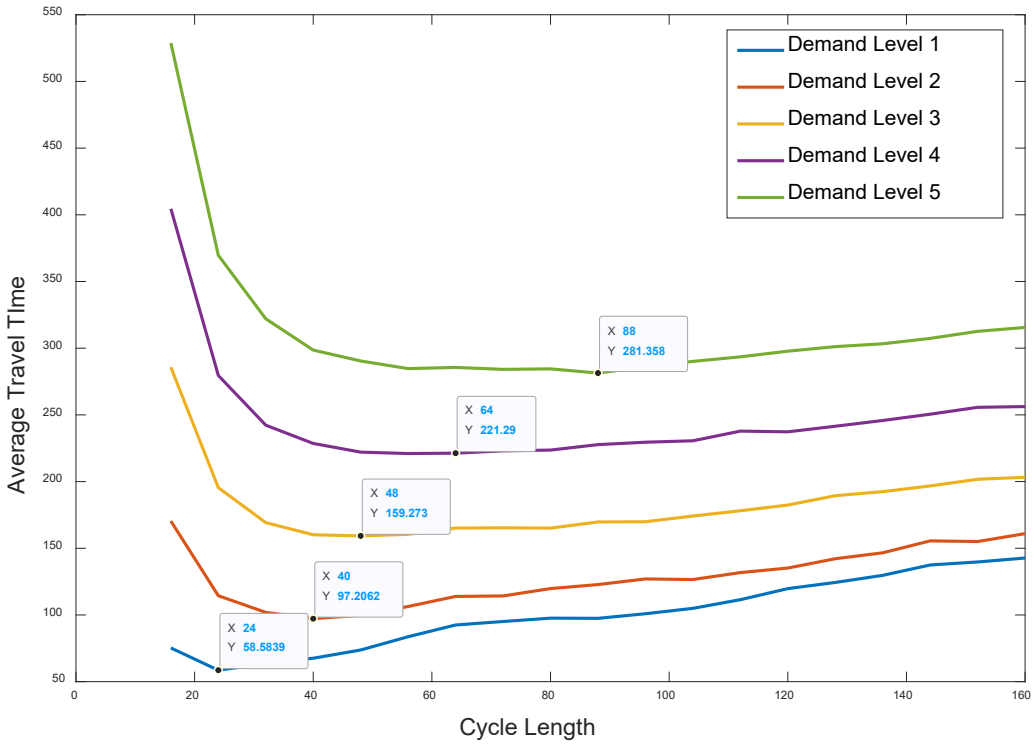
For **DSC-1** cases, the subnetwork size is $\mathbf{1 \times 1}$ with **one intersection**. For **DSC-2** cases, the subnetwork size is $\mathbf{2 \times 2}$ with **four intersections**. If **CSC** case is considered as the whole network decomposed into **one subnetwork**, then CSC, DSC-1, and DSC-2 are regarded as different network decompositions for the same network.

5.7.1 Scenario 1 Three-Step Naïve Method: CSC, $\mathbf{2 \times 2}$ Network

Scenario 1 consists of CSC for a 2×2 network. Results of the Three-Step Naïve Method is displayed step by step.

a) Cycle Length

The first step is to optimize the cycle. Five different demand levels (levels 1-5) are used: 560, 1120, 1680, 2240, 2800 total vehicles over the time horizon. Results are shown in Figure 5-10. Both ATT and TTD are convex, which provides evidence that Bisection Algorithm (BA) can be applied to search for the local optimal point of the cycle length. Points of minimal value show that when demand is increasing, the optimal cycle length is increasing as well. These results for cycle length have the same trend as the initial work done by Webster (1958) for a single intersection.



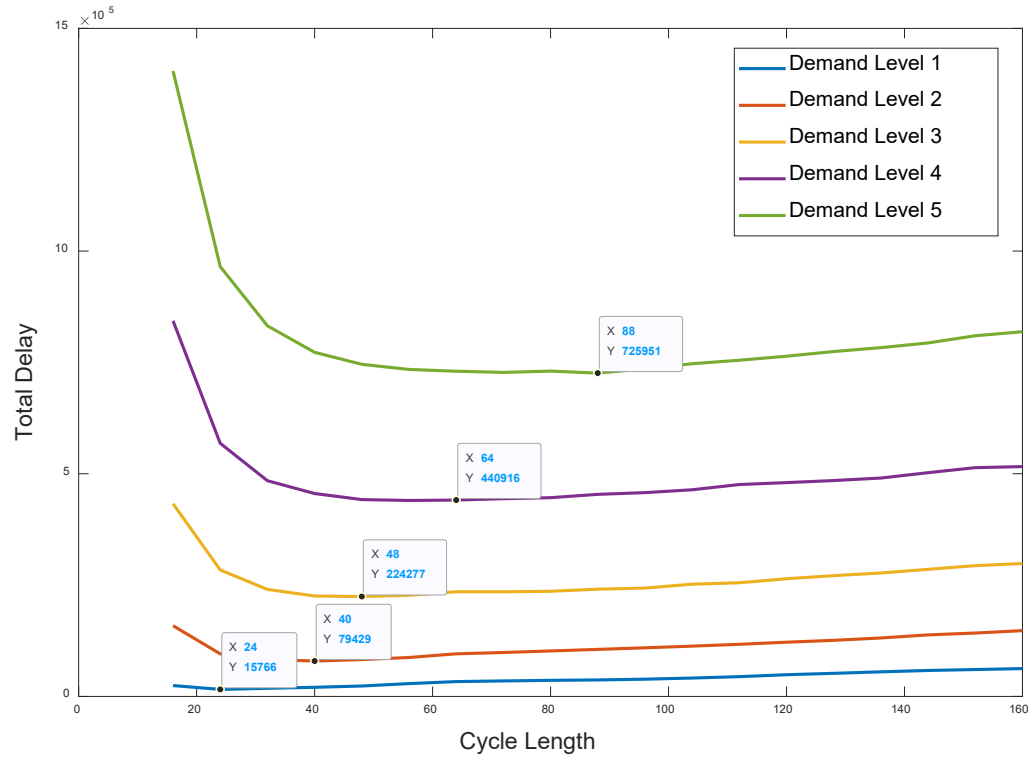


Figure 5-10 Curves of ATT and TTD under different demand levels.

b) Phase Split

Demand level 3 has an optimal cycle length $cl^* = 48$. The offset is fixed as the initial value 0, and here solves the problem via the global search method.

Set the feasible set of phase splits for ***Problem (A2)*** as

$$\Omega_{\text{phase}} = \{(12\ 12\ 12\ 12), (16\ 16\ 8\ 8), (16\ 8\ 16\ 8), (16\ 8\ 8\ 16), (8\ 16\ 16\ 8), (8\ 16\ 8\ 16), (8\ 8\ 16\ 16)\}$$

The optimal phase is $g_m^* = (12\ 12\ 12\ 12), m = 1,2,3,4$, which results from the uniformly distributed demand. The minimal $ATT = 159.27$ and $TTD = 2.24 \times 10^5$.

c) Offset

Continue with the values above for cl^*, g_m^*, and m. Set $\gamma = 6$, then the step size for the gradient search approach $\alpha = \frac{cl^*}{\gamma} = 8$. The resulting optimal offsets are $of_1 = 16$, $of_2 = 24$, $of_3 = 32$, $of_4 = 0$, final $ATT = 154.74$, and $TTD = 2.19 \times 10^5$. ATT is reduced by 2.84% and TTD by 2.33% from the previous step. $CT = 32.45$ seconds.

The results show that the Three-Step Naïve method can solve the CSC problem for a 2×2 network. In addition, the bisection, global search, and gradient search methods are applicable for step 1, step 2, and step 3, respectively. CT is acceptable for the time horizon $T = 300$ time units.

5.7.2 Scenario 2 Two-Layer Process: DSC-2, $\mathbf{4 \times 4}$ *Network*

Scenario 2 shows the two-layer process of the DSC system. The example used is a DSC-2 case for a 4×4 network. The stop iteration number is 5. The cycle length is fixed to 60. There are 351 vehicles over the time horizon $T = 300$. The number of CPU cores is 16. $\gamma = 6$ for the Three-step Naïve Method. Results for each iteration are shown in Table 5-2.

Table 5-2 Iterations of Scenario 2

# of Iteration	**1**	**2**	**3**	**4**	**5**
ATT	180.64	144.67	133.38	**133.35**	141.16
TTD	4.31E+04	3.04E+04	2.65E+04	**2.65E+04**	2.96E+04
Cycle Length	60	60	60	60	60
Phase Split & Offset in each iteration (Each row is green times of 4 phases, and offset for each intersection)	#1: 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 15 15 15 15 0 #3: 20 20 10 10 40 10 10 20 20 0 10 20 10 20 40 10 20 20 10 20 20 10 10 20 40 15 15 15 15 0 20 20 10 10 30 10 10 20 20 20 10 20 10 20 40 20 20 10 10 50 10 20 10 20 10 10 10 20 20 40 20 20 10 10 0 20 20 10 10 10 10 10 20 20 20 10 20 10 20 30	#2: 15 15 15 15 0 10 20 20 10 30 10 20 10 20 30 20 20 10 10 0 20 10 10 20 20 15 15 15 15 0 15 15 15 15 20 10 10 20 20 40 10 20 10 20 40 20 20 10 10 40 15 15 15 15 0 10 10 20 20 40 20 20 10 10 10 20 20 10 10 20 10 10 20 20 30 15 15 15 15 0 #4: 20 20 10 10 20 10 20 10 20 30 10 20 10 20 50 20 20 10 10 0 20 10 10 20 30 15 15 15 15 0 20 20 10 10 30 10 10 20 20 10 10 20 10 20 10 10 20 10 20 10 10 20 10 20 20 10 10 20 20 40 20 20 10 10 40 10 20 10 20 40 10 20 10 20 50 10 20 10 20 20			

	#5:	20	20	10	10	30
		10	20	20	10	40
		10	20	10	20	20
		10	20	10	20	0
		20	10	10	20	0
		15	15	15	15	0
		20	20	10	10	40
		20	20	10	10	0
		10	20	20	10	20
		10	20	10	20	0
		10	20	10	20	10
		15	15	15	15	40
		20	20	10	10	50
		10	20	20	10	0
		10	20	10	20	30
		15	15	15	15	0

The optimal solution results in $ATT = 133.35$ and $TTD = 2.65 \times 10^4$. In this scenario, $CT = 151.81$ secs. From the first iteration, ATT is reduced by 22.87%, while TTD is reduced by 38.52%.

Table 5-2 shows each iteration in the DSC case. The remaining scenarios measure the impact of CPU core numbers, stop iteration numbers, and different network decompositions based on the performance measures used (ATT, TTD, and CT).

5.7.3 Scenario 3 CPU Cores: CSC, 2×2

Core numbers are tested to observe the impact on CT. Here uses a CSC case with a 2×2 network, with 192 vehicles over the time horizon.

Figure 5-11 shows that when the number of CPUs increases, CT will decrease from over 250 seconds to around 10 seconds. However, the curve is convex, which means the marginal benefit decreases with respect to the number of CPU cores. This decrease demonstrates the computational limitations of a centralized system even with a powerful computing cloud server.

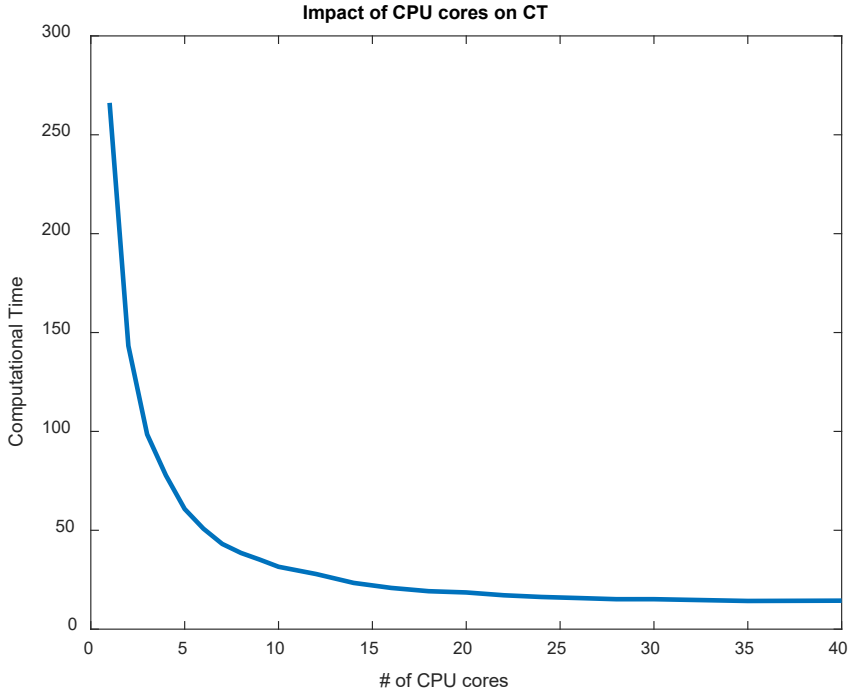


Figure 5-11 Impact of CPU cores on CT.

5.7.4 Scenario 4 Iteration Number: DSC-1, $\mathbf{2 \times 2}$

Here demonstrates how the number of iterations impacts DSC system performance, using a 2×2 network decomposed into four 1×1 subnetworks. There are 192 vehicles over the time horizon, the same demand as in Scenario 3.

The objective function converges as the number of iterations increase (Figure 5-12), evidence of the convergence for the two-layer process. In other words, there exists an equilibrium between demand and traffic signal optimization. The two objectives, ATT and TTD, correlate since they have similar trends when the iteration number increases.

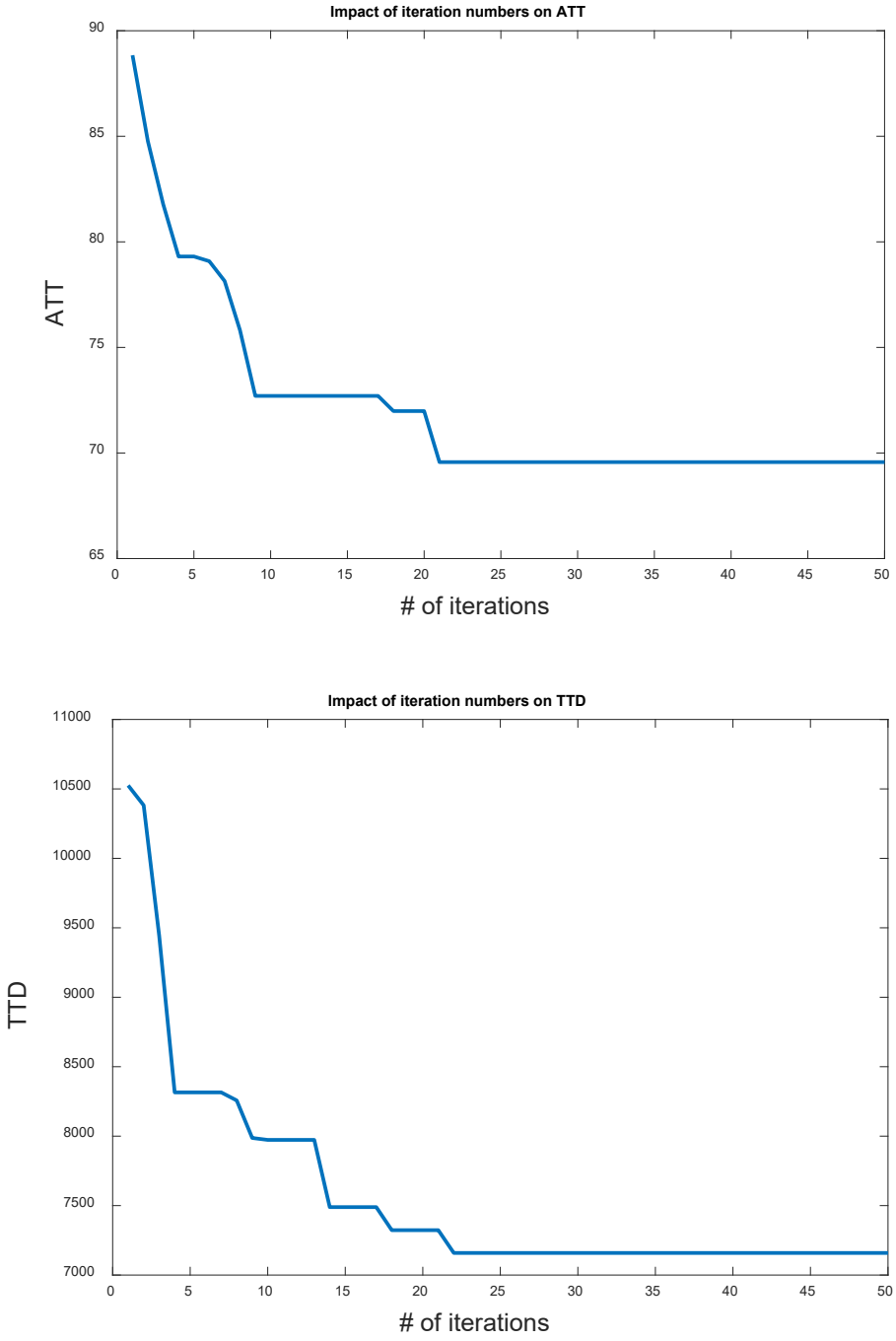


Figure 5-12 Number of iterations on performance of DSC system

5.7.5 Scenario 5 Network Decompositions

Scenario 5 is designed to compare different decompositions under different short-term OD demand profiles. The demand input is shown in Table 5-3. Demand is increased from level 1 to level 3. Short-term OD Demand is generated randomly following uniform distribution over OD pairs and the time horizon. The mean value and standard variance are $\tau, 2\tau, 3\tau$ /od for the three levels, respectively. τ is the adjust coefficient for different network sizes to balance saturation rate under the same demand level for all the networks. The number of control variables for one intersection is 6, so the maximum number of control variables in the simulation is 216 for the 6×6 network.

Table 5-3 Demand Profile of Scenario 5

Network Size	Level 1 (τ/od)	Level 2 (2τ/od)	Level 3 (3τ/od)	τ	# of intersections	# of control variables
$\mathbf{1 \times 1}$	96	168	228	12	1	6
$\mathbf{2 \times 2}$	192	360	552	6	4	24
$\mathbf{3 \times 3}$	284	520	816	4	9	54
$\mathbf{4 \times 4}$	351	753	1089	3	16	96
$\mathbf{6 \times 6}$	574	1082	1205	2	36	216

For DSC-1 cases, subnetwork size is 1×1 with one intersection. For DSC-2 cases, subnetwork size is 2×2 with four intersections. Considering CSC case as decomposing the whole network into one subnetwork, CSC, DSC-1, and DSC-2 are regarded different network decompositions for the same network.

Since the primary calculation of the system is in the local layer of the two-layer process, and the local MEC has a computing unit located at each intersection, the total number of cores

used for the DSC system is equal to the number of intersections inside the network. Limited to the test environment, here will use clusters with the same number of CPU cores to intersections instead. The numbers of cores used for DSC cases are 4, 9, 16, 36 for 1×1, 2×2, 3×3, 4×4, 6×6 networks, respectively. Since CSC system computation is completed in the cloud server, 40 cores are used.

Table 5-4 outlines all the results for Scenario 5. Each case is the result of 10 iterations.

Table 5-4 Test Results of Scenario 5

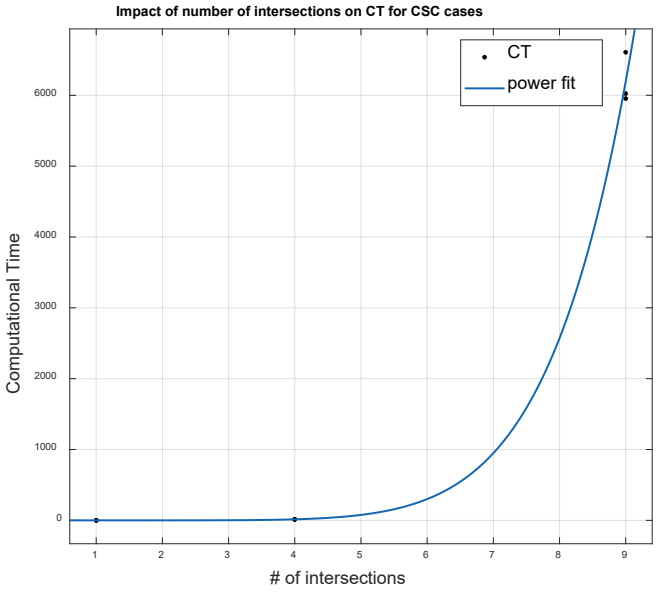
Network Size	Demand Level	DSC-1(subnetwork is 1×1)			DSC-2(subnetwork is 2×2)			CSC		
		Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
1×1	ATT							26.76	24.42	19.68
	TTD							2.04E+03	3.54E+03	4.46E+03
	CT (sec)							0.55	0.64	0.65
2×2	ATT	71.98	74.69	71.73				64.89	68.39	64.84
	TTD	7.44E+03	1.63E+04	2.46E+04				7.32E+03	1.43E+04	2.29E+04
	CT (sec)	6.28	5.51	5.52				14.36	14.76	14.51
3×3	ATT	105.38	110.81	117.91				100.55	108.08	112.86
	TTD	1.70E+04	3.39E+04	6.03E+04				1.17E+04	2.18E+04	3.80E+04
	CT (sec)	16.44	16.01	16.13				6023.88	5954.00	6607.22
4×4	ATT	135.25	176.37	170.77	129.46	164.24	169.21			
	TTD	2.74E+04	8.98E+04	1.26E+05	2.51E+04	8.13E+04	1.26E+05			
	CT (sec)	45.25	44.17	44.81	299.63	303.61	299.01			
6×6	ATT	234.17	299.44	285.00	210.02	281.35	245.14			
	TTD	8.86E+04	2.34E+05	2.50E+05	7.47E+04	2.00E+05	2.15E+05			
	CT (sec)	287.56	286.19	283.41	772.50	769.49	521.23			

* Value of ATT and TTD are in time unit.

Table 5-4 outlines the case for CSC only since the 1×1 network only has one intersection. In addition, 2×2 and 3×3 networks cannot be divided into more 2×2 subnetworks. The DSC-2 case is not possible for either network. Since CT increases exponentially as the CSC case network size increases, there are no CSC cases for 4×4 and 6×6 networks. CT for the CSC case of the 3×3 network is ~1h 30min, which is high for any real-time application, thereby making it problematic for a larger-sized network in these conditions.

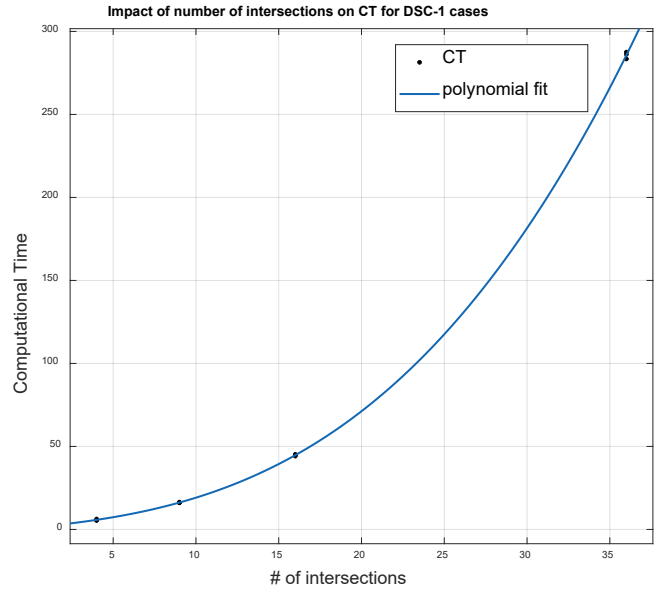
In Table 5-4, the best value among different demand levels for cases of the same network and decomposition (DSC-1, DSC-2, or CSC) is shaded in grey. For TTD, the best case is always the one with the lowest demand. Since demand reflects the total number of vehicles in the network, the resulting TTD is correlated. However, for ATT, the best case is not consistent with demand level. For example, the best case is demand level 3 for network 1×1 in CSC cases and network 2×2 in both the DSC-1 and CSC cases due to the varying distribution of demand profiles. For larger networks, cases with lowest demand have the best ATT and TTD because larger networks seem to be oversaturated in the high demand level as compared to smaller networks. For example, ATTs in both DSC-1 and DSC-2 cases of network 6×6 increase by 30% approximately from demand levels 1 to 2. In summary, TTD is positively correlated to demand while ATT will depend on demand distribution and signal time, especially in cases of small networks.

Figure 5-13 illustrates the results of comparing CT of cases in the same network decomposition.



General model Power1:
f(x) = a*x^b
Coefficients (with 95% confidence bounds):
a = 0.0004653 (-0.02222, 0.02315)
b = 7.466 (-14.73, 29.66)

Goodness of fit:
SSE: 2.573e+05
R-square: 0.9967
Adjusted R-square: 0.9962
RMSE: 191.7



Linear model Poly3:
f(x) = p1*x^3 + p2*x^2 + p3*x + p4
Coefficients (with 95% confidence bounds):
p1 = 0.004032 (0.002062, 0.006003)
p2 = 0.04922 (-0.06002, 0.1585)
p3 = 0.9085 (-0.6358, 2.453)
p4 = 1.09 (-4.451, 6.632)

Goodness of fit:
SSE: 10.02
R-square: 0.9999
Adjusted R-square: 0.9999
RMSE: 1.119

Figure 5-13 Impact of number of intersections on CT, CSC and DSC-1 cases

Trend of CTs of CSC and DSC-1 cases are investigated by increasing number of signalized intersections. Here, the goodness of fit for CSC cases is measured by power function. Since the power value is over 7, this indicates that the CT of CSC increases exponentially when number of intersections increase. Given that the relationship is convex, the goodness of fit for DSC-1 is measured using 3-degree polynomial function. However, the coefficients for x^2 and x^3 are relatively small, indicating that the relationship between CT and number of intersections in

the DSC-1 cases is nearly linear. R-square values for both cases are over 99%. Fits for the CSC cases using polynomial functions are low on goodness of fit and thus, not shown here. The coefficients shown here just demonstrate the trend of CTs in both CSC and DSC-1 cases. The number may differ in different simulation environment. When network size increases, based on results comparison from CT, DSC-1 system is more computational efficient, which is more suitable for real-time applications of larger size networks than the CSC system.

For cases of same network size, this work further compares the performance of different decompositions, as per Table 5-5. Since the intersection of CSC cases and DSC-2 cases are empty, the performance ratios over DSC-1 cases is used for other cases with same demand level and network size.

Table 5-5 Comparison Between Different Decompositions

Network		Demand Level 1	Demand Level 2	Demand Level 3
CSC/DSC-1				
2 × 2	**ATT**	0.90	0.92	0.90
	TTD	0.98	0.88	0.93
	CT	2.29	2.68	2.63
3 × 3	**ATT**	0.95	0.98	0.96
	TTD	0.69	0.65	0.63
	CT	366.42	371.80	409.72
DSC-2/DSC-1				
4 × 4	**ATT**	0.96	0.93	0.99
	TTD	0.92	0.91	1.01
	CT	6.62	6.87	6.67
6 × 6	**ATT**	0.90	0.94	0.86
	TTD	0.84	0.92	0.80
	CT	2.69	2.69	2.76

From the ratio of CSC versus DSC-1 cases in the upper half of Table 5-5, evidence can be found that DSC-1 cases emerge as the most computationally efficient. Ratios for CT in all cases are greater than one. In addition, CT is reduced by around 55% and 99.7% as compared to CSC cases for 2×2 and 3×3 networks, respectively. However, ATT and TTD for DSC-1 cases are inferior. Ratios for ATT and TTD for all cases are smaller than one. In general, DSC-1 cases result in much better CT than CSC cases, but demonstrate poorer performance in ATT and TTD.

When comparing CSC with DSC-1 cases, the improvement of ATT is around 10% for 2×2 network, while the improvement is less than 5% for 3×3 network for the CSC cases. However, there is significant improvement of TTD for CSC cases in 3×3 network compared to 2×2, for which ratios are less than 0.7 when compared to those of 2×2 networks (over 0.88). The same trend happens when comparing DSC-2 with DSC-1 cases. Cases of 6×6 network are better in terms of TTD as compared to 4×4 network cases. Suitable objective functions should be considered for optimizing cases in different scenarios.

In addition to comparing DSC-2 and DSC-1 cases, CT ratios of 6×6 networks are approximately 2.7, smaller than the ratios of the 4×4 networks. These ratios demonstrate that expanding DSC-2 to a more extensive network does not sacrifice much computational efficiency. However, the ratio of CT jumps from less than 3 to over 300 with increases in network size when looking at CSC cases. This change highlights the limitation of CSC cases when expanding network size.

As a final note, results of DSC-2 and CSC cases cannot be directly compared in Table 5-5. However, feasible solutions from the DSC system are a theoretical subset of feasible solutions from CSC systems. Thus, the CSC system could achieve better ATT and TTD than all DSC systems for the same network. The CT ratio for DSC-2 over DSC-1 in a larger network is superior; the CT ratio for CSC over DSC-1 in a larger network is inferior. As a result, it can be expected that DSC-2 cases are more computationally efficient than CSC cases in the same-sized network.

In summary, Section 5.7 has developed five scenarios to test CSC and DSC systems. Scenarios 1 and 2 demonstrate the Three-step Naïve Method in a CSC system of a 2×2 network and two-layer process to optimize signals in a 4×4 network by decomposing it into four subnetworks, respectively. Scenario 3 shows that an increasing number of cores for the computation can increase computational efficiency while marginal revenue decreases. Scenario 4 shows a convergence of the DSC system when iteration numbers increase. Scenario 5 compares cases with different network decompositions and demand levels. Results demonstrate that DSC cases have significantly better performance on CT and weaker performance on ATT and TTD. Network decomposition also has an impact on three scales. Increases in the subnetwork size may see reductions in ATT and TTD and loss of computational efficiency for the cases tested.

5.8 Summary

This chapter aims to develop and compare CSC and DSC systems using short-term OD demand as inputs in an MEC-enabled CV environment and to investigate the impact of network decomposition on performances for both systems. Control variables are considered cycle length, phase split (green times for all phases), and offset for each intersection inside the network. In addition, two control objectives are formulated in the optimization model: minimal Average Travel Time (ATT) and Total Travel Delay (TD).

Network dynamics are based on a G/G/n/FIFO open queuing network model, solved by simulation in MATLAB. Signal timing is proposed using a standard ring-and-barrier diagram with four phases and all-red intervals. Considering each phase as a virtual link, it is assumed that travelers will choose the paths with minimal instantaneous travel time. The optimal control problem for network signals is formulated with OD demand and free-flow travel time as inputs. However, the original problem is NP-complete due to non-convex variables and nonlinear constraints. Therefore, a Three-step Naïve Method is applied to decompose it into three subproblems and develop a DSC system within a CV environment with MEC. Thus, the DSC system decomposes the network into subnetworks, with each subnetwork controlled by an individual agent (regional MEC controller). Agents can exchange information in real-time. Finally, a two-layer process is proposed to solve the DSC system.

Both structures of CSC and DSC systems are constructed in the MEC-enabled CV environment. And this environment guarantees low data transmission latency. However, data analysis in CSC is processed by the cloud server in the central TMC, while data analysis in DSC is divided into three major functions run on the cloud server and MEC devices (local MECs and regional MEC controllers), which fully uses the computational resources in the network.

Numerical simulations were performed via five scenarios where demand was randomly generated following a uniform distribution. Scenarios 1 and 2 helped identify the basic settings and results of the CSC and DSC systems. Scenario 3 demonstrated the limitations of solving CSC cases against increasing the number of CPU cores. Scenario 4 gave an example for the convergence of the loop in the two-layer process for DSC cases. Scenario 5 investigated both systems in cases with different network decompositions and demand levels.

The results show that network decomposition with smaller subnetworks results in better computational efficiency but reduced performance on ATT and TTD. For example, the CSC case for a 3×3 network has a CT of 90 minutes, while the DSC-1 case with the same demand takes 16 seconds – a reduction between 55% - 99.7%, which is more than the 40% reduction of a recent study (Chow et al., 2020). In addition, the DSC system can be regarded as the physical combination of several small CSC systems if considering each end of the traffic network as connected to the central TMC with fiber. The Three-step Naïve Method used in CSC systems to solve traffic signal optimization is the same method used to solve optimal traffic signal timing plans in each subnetwork of the DSC systems. In this case, the method – a combination of the bisection method, global search, and gradient search approaches – is highly inefficient. Thus, the main difference between the systems is structure. The results show notable improvement in computational efficiency with some, but not significant, loss of ATT and TTD compared to the CSC system, thus demonstrating the value of the DSC system structure.

The results also revealed the ability of the DSC structure, with the scalable decomposition method, to apply to more extensive networks, which suffer acceptable losses of computational efficiency compared to the CSC system. Although the performances of ATT and TD are positively correlated, some cases revealed that suitable objectives should be chosen for the different cases.

This work used a simulation-based method to gather more realism to the framework of network signal optimization. Limitations of this work should be highlighted as future research to pursue. First, the network traffic signal control system needs to be tested in real-world traffic networks in a CV environment with MEC. The simulation in a cloud server with parallel computing is not equal to a DSC system within a real-world CV environment with MEC, constituting the need for field tests. Second, algorithms for traffic signal optimization are simple and time-consuming in the CSC system, and more efficient algorithms must be developed for real-world applications. Finally, the route guidance function/traffic assignment model in Central TMC is under-investigated in our study, requiring further research into its limitations.

6 CONCLUSIONS

This chapter provides a summary of the research, key findings, contributions, and limitations and future work.

6.1 Research Overview

This thesis develops a traffic signal control system for large-scale networks. The broad goals in optimizing the signal timing control plan for an urban traffic network include more efficient traffic capacity utilization, improved vehicle mobility. However, network-level traffic signal optimization is a Nondeterministic Polynomial Time (NP)-complete problem, and remains unsolved with respect to optimizing both objectives and computation resources. It is also difficult to implement due to its requirement for high-quality data, low data transmission latency, and wide bandwidth communication network.

The development of new technologies, such as Connected Vehicle (CV) and Mobile Edge Computing (MEC), will provide high-quality data and guarantee low data transmission. In fact, 5G and MEC projects with TELUS are in planning phases in a testbed in Edmonton, AB, Canada. Thus, the network traffic signal control system developed in this thesis is assumed under the MEC-enabled CV environment.

Chapter 2 provides a literature review in four key areas related to this research: mobile edge computing (MEC), dynamic traffic assignment (DTA), network traffic signal control, DTA-based network traffic signal control. The review reveals a need to explore DSC systems for large-scale networks considering greater realism of traffic signal control, new forms of data inputs, and new network decomposition methods.

Chapter 3 introduces basic assumptions and inputs of models used in this work. The network traffic signal control system in this thesis is assumed to be constructed in the MEC-enabled CV environment. This guarantees the availability of short-term OD demand, low data transmission latency, and ability of distributed computation. In addition, assumptions on DTA model are made such as “No Departure Time Choice”, “First-In-First-Out (FIFO)”, “No Reroute”, and “Point Queue” to help predict traffic dynamics within the network traffic signal

control systems. At last, this chapter also describes limitation of local counts and advantage of short-term demand. Three types of short-term demand used in this thesis are defined.

Chapter 4 develops a queue-based DTA model with traffic signal timing plan with cycle length, phase split, offset, and all-red phases. Combining DTA and traffic signal control is not an innovation. But most of the literature only consider simplified traffic signal timing plan. Within the model, travelers are assumed to choose the minimal experienced travel time and marginal experienced travel time for UO-DTA and SO-DTA cases respectively. Case study is proposed in a single OD corridor network. By solving both UO-DTA and SO-DTA.

Chapter 5 develops and analyzes a simulation-based framework of decentralized signal control system in this research. To extend the arterial network study of Chapter 4, a simulation-based G/G/n/FIFO open Queueing Network Model (QNM) is proposed to predict traffic dynamics. Traffic signal timing plans are represented using a standard ring-and-barrier diagram with four phases and all-red intervals. Considering each phase of traffic signal control as a virtual link with cost as waiting time, it is assumed that travelers will choose paths with minimal instantaneous travel time. The optimal control problem for network signals is formulated with short-term OD demand and free flow travel time as inputs. However, the original problem is NP-complete due to non-convex variables and non-linear constraints. Therefore, a Three-step Naïve Method is applied to decompose it into three sub-problems, and a DSC system is further developed within the MEC-enabled CV environment. The DSC system decomposes the network into subnetworks, with each subnetwork controlled by its own agent (regional MEC controller). Agents can exchange information in real-time. Lastly, a two-layer process is proposed to solve the DSC system. The proposed control systems are applied to a set of test scenarios constructed using different demand levels in different grid networks. This chapter also investigates the impact of network decomposition strategy on performance of the network traffic signal control system. Performance is evaluated via Computational Time (CT), Average Travel Time (ATT), and Total Travel Delay (TTD).

6.2 Research Findings

Research findings are presented in the order of the research questions of Section 1.2.

1. Coordinated network traffic signal optimization is based on prediction of traffic dynamics. How do we predict traffic dynamics in a large-size traffic network with signalized intersections?

Network traffic signal control is the management of network traffic demands by adjusting the traffic signal timing plan for each intersection. Section 3.3 suggests applying short-term OD demand as inputs to predict traffic dynamics, since traditional traffic dynamics prediction based on local counts is shorted ranged in both time and space. However, DTA model, which take short-term OD demand as inputs, shows its advantage to capture the interactions between travelers and the whole traffic network. Additional discussion is made about the foreseeable availability of short-term demand with new technologies.

Chapter 4 continues to explore this question by developing a queue-based DTA model to predict traffic dynamics, using short-term OD demand as inputs in a corridor network with two signalized intersections. 12 cases based on three demand profiles and four different signal phase plans are tested. Results show the ability of the queue-based DTA model to solve UO-DTA and SO-DTA problems, which are the fundamental problems of network traffic assignment. ATT is used to evaluate all the cases. As expected, SO-DTA cases are shown to be superior to UO-DTA cases in terms of ATT, with this difference increasing with higher demand. In addition, when comparing the impact of variables for traffic signal controls as cycle length, phase split, and offset, all three variables have significant effects on both path sharing and ATT.

2. In the MEC-enabled CV environment, how do we design the structure of a decentralized signal control (DSC) system? And how do we evaluate the performance of these systems?

This question is answered in Chapter 5. In the MEC-enabled CV environment described in Chapter 3, each intersection has a local MEC and each subnetwork has a MEC controller. Based on that, this thesis develops a two-layer process of DSC system as show

in Figure 5-8. The Central-TMC will gather this OD information from the regional MEC controllers. The output from this first layer – the short-term demand for each path – is input to the second layer. The second layer has two functions: traffic dynamics (traffic state estimation) on regional MEC controllers and traffic signal optimization on local MECs inside each subnetwork. In the CSC system, all three functions are run inside the cloud server. However, in the DSC system, the sole function of the Central-TMC is to calculate short-term path demand to provide route guidance. Remaining optimization tasks are allocated to the regional MEC controller inside each subnetwork. In addition, local MECs in the subnetwork are involved in the computation of the traffic signal optimization function. All the computation resources of the DSC system are fully used.

In Chapter 5, overall network performance is evaluated using measures of Computational Time (CT), Average Travel Time (ATT), and Total Travel Delay (TTD). Five scenarios are designed to understand the new structure of DSC system. Findings from scenarios 1-4 are listed below:

- The Three-Step Naïve method helps to optimize traffic signals in CSC cases and traffic signals in subnetworks of DSC cases, which improves ATT and TTD for all systems.
- The two-layer process can produce convergent solutions for DSC cases. The results of ATT or TTD for DSC system will remain the same even with iteration number large enough.
- The optimization of traffic signals in CSC cases is still an NP-complete problem, which limits CSC system for any application even with powerful computing cloud server. CT of CSC system will increase exponentially when network size increases.

3. There are many characteristics which impact DSC system performance, such as network size, traffic demands, and how a network is decomposed with respect to the traffic signal control system. How does each of these affect traffic network performance?

This question is answered by results from Scenario 5 (See Tables 5-4 & 5-5). Network size is tested via grid networks varying from 1×1 to 6×6. Three demand levels are tested for each network size. Considering CSC case as a specific network decomposition method, there are three types of network decompositions are investigated: CSC, DSC-1, DSC-2. For CSC, the network is not divided. For DSC-1, the network is divided into subnetworks with size 1×1. And for DSC-2, the network size is divided into subnetworks with size 2×2.

The key findings are as follows.

- When network size increases, ATT, TTD and CT all increases for cases with the same demand profile and network decomposition. CT is increasing exponentially in CSC cases and increasing almost linearly in DSC-1 cases when network size increases.
- When demand increases, ATT, TTD increases for cases with the same network size and network decomposition. However, CT remain almost the same, i.e., it is not sensitive to demand.
- ATT and TTD are positively correlated among cases. However, there are also results from a few cases which show that suitable objectives should be chosen for different scenarios.
- Network decomposition with smaller subnetworks results in better computational efficiency (less CT), but reduced performance on ATT and TTD.
- The scalable decomposition method in DSC systems is extendable to larger networks, which suffer acceptable losses of computational efficiency as compared to the CSC system. Since the curve of CT on number of intersections for DSC-1 cases is close to linear, and the CT ratio between DSC-2 and DSC-1 cases is reduced when size of network increases.
- The Three-Step Naïve Method used for both systems is highly inefficient resulting from CT. However, this highlights the value of the DSC system structure since the results show significant improvement in computational efficiency with some, but not significant, loss of ATT and TTD as compared to CSC system.

6.3 Research Contribution

This thesis contributes to both academic research and practice in decentralized signal control system for large-scale network.

Academic Contributions

1) A queue-based DTA model for traffic network with real traffic signal timing plan

Although DTA has been an effective tool to describe traffic dynamics for traffic optimization, and many researchers have considered traffic signal control in their models, signal timing representations have been simplified without considering more realistic and critical phase sequence and duration restrictions. This work formulates traffic signal timing as a component of the link performance function with three control variables: cycle length, phase split, and offset. In addition, this work also solves UO-DTA and SO-DTA in a corridor network. In addition, the queue-based DTA model in Chapter 4 is extended to a simulation-based QNM in Chapter 5 to predict traffic dynamics in larger networks.

2) A simulation-based framework of DSC system within the MEC-enabled CV environment

Although many researchers have developed network traffic signal control systems within current technological paradigm, this work is proposed within a future technological paradigm, with the assumption that the DSC system is based on a MEC-enabled CV environment. A two-layer process is proposed, in which distributes three functions route guidance, traffic dynamics prediction, and traffic signal optimization by decompose the whole network into subnetworks. Each subnetwork has a MEC controllers that optimize signals within the subnetwork with local MECs at each intersection. This framework comes with new data inputs, solving algorithm, and decomposition method compared to previous research. In addition, the computational structure of DSC system gains significant reduction comparing to CSC system, which is superior to some existing work.

3) A scalable and extendable decomposition method for a DSC system

The decomposition method used for this study is based on size of subnetwork. The size of subnetwork can be 1×1, 2×2, or any other grid size. A Path Decomposition Index (PDI) is proposed to explain how the decomposition method not only decomposes the geographic of the network, but also decomposes the path demands. The results also explain the ability of the scalable decomposition method to apply to larger networks, which suffer acceptable losses of computational efficiency. Compared to the previous work only consider intersection-based decomposition or subnetwork-based decomposition only for specific networks. This decomposition method is more scalable and extendable.

Practical Contributions

1) Guidance to transportation engineers and planners about how to optimize decentralized signal control system via network decomposition

Chapter 5 results show that network decomposition with smaller subnetworks results in better computational efficiency (less CT), but reduced performance on ATT and TTD. In other words, if higher performance on ATT, TTD, or any other objective is required, network decomposition with larger subnetworks should be applied. Meanwhile, increasing subnetwork size in DSC system for a large-size network will require more computational capacity. As such, constrained to existing computational capacity, maximizing subnetwork size is a general way to get the best performance of the whole traffic network in terms of the control objective.

2) Significant reduction in computational time for DSC system compared to CSC system

Among the three performance metrics ATT, TTD, and CT, the most critical to enable a real-time control application is low computational time (CT). Significant reduction of CT in DSC system compared to CSC system demonstrates the importance of a DSC system in realizing real-time network traffic signal control. For example, the CSC case for a 3×3 network has a CT of 90 minutes while the DSC-1 case with the same demand profile takes 16 seconds, demonstrating a reduction of over 99%. In addition, there remains significant room to improve the DSC system documented in this thesis. For example, the Three-Step Naïve Method used for traffic signal optimization is highly computational inefficient, replacing it with excellent algorithm will make CT much lower.

6.4 Research Limitations and Future Work

Details on research limitations and potential methods of improvement are described as follows.

1. The DSC system is not tested in any real-world traffic network and the CV environment with MEC is described in a high level. And simulation in cloud server with parallel computing is not equal to a DSC system within a real-world CV environment with MEC. However, ongoing work is taking place in the testbed of our research group in South Campus. 5G and MEC devices will be fully installed in the testbed in the next four years as in the project plan with Telus. After devices are ready, there will be chances to test the DSC system in real world.

2. Algorithms for traffic signal optimization are simple and time-consuming in the CSC system. The three-step naïve method is inefficient. When network size increases to 3×3, CT increases to more than 1.5 hours. This leads the further work to develop innovative algorithms to optimize traffic signal controls in CSC system. In addition, each subnetwork in DSC system in this work can be regarded as a small CSC system since it runs the same algorithm with CSC. Future work will also investigate the potential to develop an algorithm specifically for traffic signal control optimization problem in DSC system.

3. Although the focus of this work is network traffic signal control, it requires use and assumptions of short-term demand prediction and dynamic traffic assignment. Part of the

work on short-term demand prediction and DTA modeling can be improved in the future work.

4. As discussed in Chapter 2, optimization on network decomposition is one key to optimizing DSC system. However, this thesis has not investigated how to optimize this decomposition itself. In future work, since the network decomposition method is scalable in this research, there is great potential to develop models and algorithms for optimizing network decomposition. And this improvement will also give decision-makers more directed guidance on choosing locations of MEC devices.

REFERENCES

Adacher, L., Gemma, A., & Oliva, G. (2014). Decentralized spatial decomposition for traffic signal synchronization. *Transportation Research Procedia*, *3*(July), 992–1001. https://doi.org/10.1016/j.trpro.2014.10.079

Adacher, L., & Tiriolo, M. (2017). Distributed urban traffic signal optimization based on macroscopic model. *2016 6th International Conference on Innovative Computing Technology, INTECH 2016*, 605–610. https://doi.org/10.1109/INTECH.2016.7845123

Adacher, L., & Tiriolo, M. (2020). Performance analysis of decentralized vs centralized control for the traffic signal synchronization problem. *Journal of Advanced Transportation*. https://doi.org/10.1155/2020/8873962

Allsop, R. E., & Charlesworth, J. A. (1977). TRAFFIC IN A SIGNAL-CONTROLLED ROAD NETWORK: AN EXAMPLE OF DIFFERENT SIGNAL TIMINGS INDUCING DIFFERENT ROUTEINGS. *Traffic Engineering and Control*.

Aziz, H. M. A., & Ukkusuri, S. V. (2012). Integration of Environmental Objectives in a System Optimal Dynamic Traffic Assignment Model. *Computer-Aided Civil and Infrastructure Engineering*, *27*(7), 494–511. https://doi.org/10.1111/j.1467-8667.2012.00756.x

Barceló, J., & Casas, J. (2005). Methodological notes on combining macro, meso and micro models for transportation analysis. *... on Traffic Modeling " ...*, *1*, 1–40. http://www.aimsun.com/papers/Sedona-Modeling_Workshop_2005.pdf%5Cnhttp://www.researchgate.net/publication/242154775_METHODOLOGICAL_NOTES_ON_COMBINING_MACRO_MESO_AND_MICRO_MODELS_FOR_TRANSPORTATION_ANALYSIS/file/72e7e52cd298803454.pdf

Beard, C., & Ziliaskopoulos, A. (2007). System Optimal Signal Optimization Formulation. *Transportation Research Record: Journal of the Transportation Research Board*, *1978*, 102–112. https://doi.org/10.3141/1978-15

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H., & Mishalani, R. (1998). *Massachusetts Institute of Technology Intelligent Transportation Systems Program DynaMIT: a simulation-based system for traffic prediction*. 1–12. https://pdfs.semanticscholar.org/8a22/ecb3650f9dae2fd1a644a460f52824f1040c.pdf

Ben-Akiva, M., Cuneo, D., Hasan, M., Jha, M., & Yang, Q. (2003). Evaluation of freeway control using a microscopic simulation laboratory. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/S0968-090X(02)00020-7

Cai, C., Wong, C. K., & Heydecker, B. G. (2009). Adaptive traffic signal control using approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, *17*(5), 456–474. https://doi.org/10.1016/j.trc.2009.04.005

Carey, M. (1987). Optimal Time-Varying Flows on Congested Networks. *Operations Research*, *35*(1), 58–69. https://doi.org/10.1287/opre.35.1.58

Carey, M., & Subrahmanian, E. (2000). An approach to modelling time-varying flows on congested networks. *Transportation Research Part B: Methodological*, *34*(3), 157–183. https://doi.org/10.1016/S0191-2615(99)00019-3

Carey, M., & Watling, D. (2012). Dynamic traffic assignment approximating the kinematic wave model: System optimum, marginal costs, externalities and tolls. *Transportation Research Part B: Methodological*, *46*(5), 634–648. https://doi.org/10.1016/j.trb.2012.01.008

Chen, C., Wei, H., Xu, N., Zheng, G., Yang, M., Xiong, Y., Xu, K., & Li, Z. (2020). Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *Proceedings of the AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i04.5744

Chen, H. K., & Hsueh, C. F. (1998). A model and an algorithm for the dynamic user-optimal route choice problem. *Transportation Research Part B: Methodological*, *32*(3), 219–234. https://doi.org/10.1016/S0191-2615(97)00026-X

Chiou, S. W. (2007). Reserve capacity of signal-controlled road network. *Applied Mathematics and Computation*, *190*(2), 1602–1611. https://doi.org/10.1016/j.amc.2007.02.041

Chiu, Y., Bottom, J., Mahut, M., & Paz, a. (2010). A Primer for Dynamic Traffic Assignment. *Transportation Research …*. https://doi.org/10.3141/1994-15

Chow, A. H. F., Sha, R., & Li, S. (2020). Centralised and decentralised signal timing optimisation approaches for network traffic control. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2019.05.007

Christofa, E., & Skabardonis, A. (2011). Traffic signal optimization with application of transit signal priority to an isolated intersection. *Transportation Research Record*, *2259*, 192–201. https://doi.org/10.3141/2259-18

Cruz, F. R. B., van Woensel, T., MacGregor Smith, J., & Lieckens, K. (2010). On the system optimum of traffic assignment in M / G / c / c state-dependent queueing networks. *European Journal of Operational Research*, *201*(1), 183–193. https://doi.org/10.1016/j.ejor.2009.03.006

D'Andrea, R., & Dullerud, G. E. (2003). Distributed Control Design for Spatially Interconnected Systems. *IEEE Transactions on Automatic Control*. https://doi.org/10.1109/TAC.2003.816954

Dafermos, S. (2008). Traffic Equilibrium and Variational Inequalities. *Transportation Science*, *14*(1), 42–54. https://doi.org/10.1287/trsc.14.1.42

Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, *28*(4), 269–287. https://doi.org/10.1016/0191-2615(94)90002-7

Dion, F., Rakha, H., & Zhang, Y. (2004). Evaluation of potential transit signal priority benefits

along a fixed-time signalized arterial. *Journal of Transportation Engineering*, *130*(3), 294–303. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(294)

Ferdowsi, A., Challita, U., & Saad, W. (2017). *Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems*. http://arxiv.org/abs/1712.04135

Filali, A., Abouaomar, A., Cherkaoui, S., Kobbane, A., & Guizani, M. (2020). Multi-access edge computing: A survey. In *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3034136

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., & Wie, B. W. (1993). A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, *41*(1), 179–191. https://doi.org/10.1287/opre.41.1.179

Friesz, T. L., Han, K., Neto, P. A., Meimand, A., & Yao, T. (2013). *Dynamic user equilibrium based on a hydrodynamic model q*. *47*, 102–126.

Friesz, T. L., & Tobin, R. L. (1989). *Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem Author (s): Terry L . Friesz , Javier Luque , Roger L . Tobin and Byung-Wook Wie Published by : INFORMS Stable URL : https://www.jstor.org/stable/171471 REFERENCE*. *37*(6), 893–901.

Gao, J., Shen, Y., Liu, J., Ito, M., & Shiratori, N. (2017). Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network. In *arXiv*.

Gartner, N. H. (1983). OPAC: A demand-responsive strategy for traffic signal control. *Transportation Research Record*, *1983*(906), 75–81.

Ge, X., Li, Z., & Li, S. (2017). 5G software defined vehicular networks. *IEEE Communications Magazine*. https://doi.org/10.1109/MCOM.2017.1601144

George, D. K., & Xia, C. H. (2011). Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2010.12.015

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*. https://doi.org/10.1016/j.trb.2008.02.002

Godfrey, J. W. (1969). The mechanism of a road network. *Traffic Engineering and Control*.

Gokulan, B. P., & Srinivasan, D. (2010). Distributed geometric fuzzy multiagent urban traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, *11*(3), 714–727. https://doi.org/10.1109/TITS.2010.2050688

Gregoire, J., Qian, X., Frazzoli, E., De La Fortelle, A., & Wongpiromsarn, T. (2015). Capacity-aware backpressure traffic signal control. *IEEE Transactions on Control of Network Systems*, *2*(2), 164–173. https://doi.org/10.1109/TCNS.2014.2378871

Hajbabaie, A. (2012). Intelligent dynamic signal timing optimization program. *ProQuest Dissertations and Theses*, 216. https://search.proquest.com/docview/1416413244?accountid=26759%0Ahttp://www.yidu.edu.cn/educhina/educhina.do?artifact=&svalue=Intelligent+dynamic+signal+timing+optimization+program&stype=2&s=on%0Ahttp://159.226.100.141/Reader/union_result.jsp?title=1&word

Hu, Y. C., Patel, M., Sabella, D., Sprecher, N., & Young, V. (2015). Edge Computing: a key technology towards 5G. *ETSI White Paper No. 11 Mobile*, *11*, 1–16. http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf

Hunt, P. B., Robertson, D. I., & Bretherton, R. D. (1982). The SCOOT on-line traffic signal optimisation technique (Glasgow). *Traffic Engineering & Control*, *23*(4), 190–192.

Hunt, P. B., Robertson, D. I., Bretherton, R. D., & Winton, R. I. (1981). SCOOT - A TRAFFIC RESPONSIVE METHOD OF COORDINATING SIGNALS. *TRRL Laboratory Report (Transport and Road Research Laboratory, Great Britain)*.

Jayakrishnan, R., Mahmassani, H. S., & Hu, T. Y. (1994). An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research Part C*, *2*(3), 129–147. https://doi.org/10.1016/0968-090X(94)90005-1

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., & Qian, X. (2010). Investigating the shape of the macroscopic fundamental diagram using simulation data. *Transportation Research Record*. https://doi.org/10.3141/2161-05

Karoonsoontawong, A., & Waller, S. T. (2010). Integrated Network Capacity Expansion and Traffic Signal Optimization Problem: Robust Bi-level Dynamic Formulation. *Networks and Spatial Economics*, *10*(4), 525–550. https://doi.org/10.1007/s11067-008-9071-x

Karyotis, V., & Khouzani, M. H. R. (2016). Elements of queuing theory and queuing networks. *Malware Diffusion Models for Wireless Complex Networks*, 235–254. https://doi.org/10.1016/b978-0-12-802714-1.00023-2

Ke, J., Zheng, H., Yang, H., & Chen, X. (Michael). (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2017.10.016

Kendall, D. G. (1953). Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*. https://doi.org/10.1214/aoms/1177728975

Lee, J., Park, B. (Brian), & Yun, I. (2013). Cumulative Travel-Time Responsive Real-Time Intersection Control Algorithm in the Connected Vehicle Environment. *Journal of Transportation Engineering*, *139*(10), 1020–1029. https://doi.org/10.1061/(asce)te.1943-5436.0000587

Li, P., Mirchandani, P., & Zhou, X. (2015). Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork. *Transportation Research Part B: Methodological*, *81*(P1), 103–130. https://doi.org/10.1016/j.trb.2015.08.011

Li, Q. L., Fan, R. N., & Ma, J. Y. (2016). *A unified framework for analyzing closed queueing networks in bike sharing systems*. *638*, 177–191. https://doi.org/10.1007/978-3-319-44615-8_16

Lin, J., Yu, W., Yang, X., Zhao, P., Zhang, H., & Zhao, W. (2020). An Edge Computing Based Public Vehicle System for Smart Transportation. *IEEE Transactions on Vehicular Technology*. https://doi.org/10.1109/TVT.2020.3028497

Liu, S., Liu, L., Tang, J., Yu, B., Wang, Y., & Shi, W. (2019). Edge Computing for Autonomous Driving: Opportunities and Challenges. *Proceedings of the IEEE*. https://doi.org/10.1109/JPROC.2019.2915983

Lowrie, P. R. (1990). SCATS—A Traffic Responsive Method of Controlling Urban Traffic. In *Sales information brochure published by Roads and Traffic Authority*.

Lucas, D. E., Mirchandani, P. B., & Head, K. L. (2000). Remote simulation to evaluate real-time traffic control strategies. *Transportation Research Record*, *1727*, 95–100. https://doi.org/10.3141/1727-12

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D., & Mirchandani, P. (2003). ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems. *Transportation Research Record*. https://doi.org/10.3141/1856-19

Manolis, D., Papamichail, I., Kosmatopoulos, E. B., & Papageorgiou, M. (2016). Automated tuning of ITS management and control systems: Results from real-life experiments. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2015.08.020

McKenney, D., & White, T. (2013). Distributed and adaptive traffic signal control within a realistic traffic simulation. *Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2012.04.008

Mehrabipour, M., & Hajbabaie, A. (2017). A Cell-Based Distributed-Coordinated Approach for Network-Level Signal Timing Optimization. *Computer-Aided Civil and Infrastructure Engineering*, *32*(7), 599–616. https://doi.org/10.1111/mice.12272

Merchant, D. K., & Nemhauser, G. L. (1978a). A Model and an Algorithm for the Dynamic Traffic Assignment Problems. *Transportation Science*, *12*(3), 183–199. https://doi.org/10.1287/trsc.12.3.183

Merchant, D. K., & Nemhauser, G. L. (1978b). Optimality Conditions for a Dynamic Traffic Assignment Model. *Transportation Science*, *12*(3), 200–207.

https://doi.org/10.1287/trsc.12.3.200

Mirchandani, P., & Head, L. (2001). A real-time traffic signal control system: Architecture, algorithms, and analysis. *Transportation Research Part C: Emerging Technologies*, *9*(6), 415–432. https://doi.org/10.1016/S0968-090X(00)00047-4

Nagurney, A. (2003). Network economics: a variational inequality approach. *European Journal of Operational Research*. https://doi.org/10.1016/0377-2217(94)90433-2

Nguyen, H. N., Fishbain, B., Bitar, E., Mahalel, D., & Gutman, P. O. (2016). Dynamic Model for estimating the Macroscopic Fundamental Diagram. *IFAC-PapersOnLine*. https://doi.org/10.1016/j.ifacol.2016.07.050

Nishi, T., Otaki, K., Hayakawa, K., & Yoshimura, T. (2018). Traffic Signal Control Based on Reinforcement Learning with Graph Convolutional Neural Nets. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC.2018.8569301

Nowé, A., Vrancx, P., & De Hauwere, Y. M. (2012). Game theory and multi-agent reinforcement learning. In *Adaptation, Learning, and Optimization*. https://doi.org/10.1007/978-3-642-27645-3_14

Oda, T., Kuwahara, M., & Niikura, S. (2004). *Traffic Signal Control for Reducing Vehicle Carbon Dioxide Emissions on an Urban Road Network*. *1*, 1–8. http://www.plan.civil.tohoku.ac.jp/kuwahara/publications/2004-013.pdf

Osorio, C., & Bierlaire, M. (2013). A simulation-based optimization framework for urban transportation problems. *Operations Research*, *61*(6), 1333–1345. https://doi.org/10.1287/opre.2013.1226

Papageorgiou, M., Blosseville, J.-M., & Hadj-Salem, H. (1989). Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transportation Research Part B: Methodological*, *23*(1), 29–47. https://doi.org/10.1016/0191-2615(89)90021-0

Papageorgiou, M., Kosmatopoulos, E., & Papamichail, I. (2008). Effects of Variable Speed Limits on Motorway Traffic Flow. *Transportation Research Record: Journal of the Transportation Research Board*, *2047*(1), 37–48. https://doi.org/10.3141/2047-05

Peeta SrinivasandZiliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, *1*(August 2001), 233–265. https://link.springer.com/content/pdf/10.1023%2FA%3A1012827724856.pdf

Perry Sr, F. (2017). *UFTI DSRC and Other Communication Options for Transportation Connectivity Workshop Overview of DSRC Messages and Performance Requirements*. https://www.transportation.institute.ufl.edu/wp-content/uploads/2017/04/HNTB-SAE-Standards.pdf

Peterson, M. D., Bertsimas, D. J., & Odoni, A. R. (1995). Decomposition Algorithms for

Analyzing Transient Phenomena in Multiclass Queueing Networks in Air Transportation. *Operations Research*. https://doi.org/10.1287/opre.43.6.995

Pol, E. van der, & Oliehoek, F. A. (2016). Coordinated deep reinforcement learners for traffic light control. *30th Conference on Neural Information Processing Systems (NIPS)*.

Prashanth, L. A., & Bhatnagar, S. (2011). Reinforcement learning with average cost for adaptive control of traffic lights at intersections. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC.2011.6082823

Putha, R., Quadrifoglio, L., & Zechman, E. (2012). Comparing Ant Colony Optimization and Genetic Algorithm Approaches for Solving Traffic Signal Coordination under Oversaturation Conditions. *Computer-Aided Civil and Infrastructure Engineering*, *27*(1), 14–28. https://doi.org/10.1111/j.1467-8667.2010.00715.x

Rakha, H., Van Aerde, M., Ahn, K., & Trani, A. A. (2000). Requirements for evaluating traffic signal control impacts on energy and emissions based on instantaneous speed and acceleration measurements. *Transportation Research Record*, *1738*, 56–67. https://doi.org/10.3141/1738-07

Ran, B., Boyce, D. E., & LeBlanc, L. J. (1993). A New Class of Instantaneous Dynamic User-Optimal Traffic Assignment Models. *Operations Research*, *41*(1), 192–202. https://doi.org/10.1287/opre.41.1.192

Sims, A. (1979). the Sydney Co-Ordinated Adaptive Traffic (Scat) System. *Publication of: Australian Road Research Board*, *V*(May), 130–137. http://trid.trb.org/view.aspx?id=147832

Sims, A. G., & Dobinson, K. W. (1980). The Sydney Coordinated Adaptive Traffic (SCAT) System Philosophy and Benefits. *IEEE Transactions on Vehicular Technology*. https://doi.org/10.1109/T-VT.1980.23833

Sun, D., Benekohal, R. F., & Waller, S. T. (2006). Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization. *Computer-Aided Civil and Infrastructure Engineering*, *21*(5), 321–333. https://doi.org/10.1111/j.1467-8667.2006.00439.x

Tomforde, S., Prothmann, H., Branke, J., Hähner, J., Müller-Schloer, C., & Schmeck, H. (2010). Possibilities and limitations of decentralised traffic control systems. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2010.5596886

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2014.01.005

Wang, G., & Xu, F. (2020). Regional intelligent resource allocation in mobile edge computing based vehicular network. *IEEE Access*, *8*, 7173–7182.

https://doi.org/10.1109/ACCESS.2020.2964018

Wang, S. H., & Davison, E. J. (1973). On the Stabilization of Decentralized Control Systems. *IEEE Transactions on Automatic Control*. https://doi.org/10.1109/TAC.1973.1100362

Webster, F. V. (1958). Traffic signal settings. In *Road Research*.

Wei, H., Xu, N., Zhang, H., Zheng, G., Zang, X., Chen, C., Zhang, W., Zhu, Y., Xu, K., & Li, Z. (2019). Colight: Learning network-level cooperation for traffic signal control. *International Conference on Information and Knowledge Management, Proceedings*. https://doi.org/10.1145/3357384.3357902

WEI, H., ZHENG, G., GAYAH, V., & LI, Z. (2019). A survey on traffic signal control methods. In *arXiv*.

Withanawasam, J., & Karunananda, A. (2018). Multi-agent based road traffic control optimization. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-March*, 977–981. https://doi.org/10.1109/ITSC.2017.8317896

Wu, Y. T., & Ho, C. H. (2009). The development of Taiwan arterial traffic-adaptive signal control system and its field test: A Taiwan experience. *Journal of Advanced Transportation*. https://doi.org/10.1002/atr.5670430404

Wu, Z., Sharma, A., Mannering, F. L., & Wang, S. (2013). Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. *Accident Analysis and Prevention*, *54*, 90–98. https://doi.org/10.1016/j.aap.2013.01.016

Yang, H., & Yagar, S. (1995). Traffic assignment and signal control in saturated road networks. *Transportation Research Part A*, *29*(2), 125–139. https://doi.org/10.1016/0965-8564(94)E0007-V

Yuan, H., & Li, G. (2021). A Survey of Traffic Prediction: from Spatio-Temporal Data to Intelligent Transportation. *Data Science and Engineering*, *6*(1), 63–85. https://doi.org/10.1007/s41019-020-00151-z

Zhang, Lei, Lee, M., Chang, H., & Carrion, C. (2017). An approach to grouping traffic signals for coordination using clustering methods. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings*, 792–797. https://doi.org/10.1109/MTITS.2017.8005620

Zhang, Lele, Yuan, Z., Yang, L., & Liu, Z. (2020). Recent developments in traffic flow modelling using macroscopic fundamental diagram. *Transport Reviews*. https://doi.org/10.1080/01441647.2020.1738588

Ziliaskopoulos, A. K. (2000). Linear programming model for the single destination System Optimum Dynamic Traffic Assignment problem. *Transportation Science*, *34*(1), 37–49. https://doi.org/10.1287/trsc.34.1.37.12281