NITC
NATIONAL INSTITUTE for TRANSPORTATION and COMMUNITIES

Final Report 1235
April 2020

Connected Vehicle System Design for Signalized Arterials

Xianfeng Terry Yang, Ph.D.
Mingyue Ji, Ph.D.
Qinzheng Wang

THE UNIVERSITY OF UTAH

Connected Vehicle System Design for Signalized Arterials

Final Report

NITC-RR-1235

by

Xianfeng Terry Yang, Mingyue Ji, and Qinzheng Wang
University of Utah

for

National Institute for Transportation and Communities (NITC)
P.O. Box 751
Portland, OR 97207

April 2020

Technical Report Documentation Page

1. Report No. NITC-RR-1235	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Connected Vehicle System Design for Signalized Arterials		5. Report Date April 2020
		6. Performing Organization Code
7. Author(s) Xianfeng Terry Yang, Mingyue Ji, and Qinzheng Wang		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Utah		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Institute for Transportation and Communities (NITC) P.O. Box 751 Portland, OR 97207		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

It can be expected that connected vehicles (CVs) systems will soon go beyond testbed and appear in real-world applications. To accommodate a large number of connected vehicles on the roads, traffic signal control systems on signalized arterials would require supports of various components such as roadside infrastructure, vehicle on-board devices, an effective communication network, and optimal control algorithms. In this project, we aim to establish a real-time and adaptive system for supporting the operations of CV-based traffic signal control functions. The proposed system will prioritize the communication needs of different types of CVs and best utilize the capacity of the communication channels. The CV data sensing and acquisition protocol, built on a newly developed concept of Age of Information (AoI), will support the feedback control loop to adjust signal timing plans. Our multidisciplinary research team, including researchers from transportation engineering and electrical engineering, will carry out the project tasks along four directions that capitalized on the PIs' expertise: (i) Data collection and communication, in which the proposed system will be based on the AoI, prioritize the data needs of different types of CVs, and optimize the communication network; (ii) Dynamic traffic signal coordination, which will concurrently facilitate the progression of traffic flows along multiple critical paths; (iii) Smart traffic signal control, where both operational efficiency and safety improvement are accounted for at signalized intersections; and (iv) Multimodal system design, which will integrate transit signal priority (TSP) and suppression controls for accommodating connected buses. This project addresses the urgent needs in CV system designs and offers control foundations to support the operations of urban signalized arterial in a CV environment.

17. Key Words Connected vehicles, traffic signal control, communication optimization	18. Distribution Statement No restrictions. Copies available from NITC: www.nitc-utc.net

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 80	22. Price

ACKNOWLEDGEMENTS

This project was funded by the National Institute for Transportation and Communities (NITC) under grant number (NITC-RR-1235). The PI Xianfeng Terry Yang would also like to acknowledge the data support from the Utah Department of Transportation (UDOT).

DISCLAIMER

RECOMMENDED CITATION

Yang, Xianfeng, Ji, Mingyue, & Wang, Qinzheng. *Connected Vehicle System Design for Signalized Arterials*. NITC-RR-1235. Portland, OR: Transportation Research and Education Center (TREC), 2020.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

It can be expected that connected vehicles (CVs) will soon go beyond testbed and appear on real-world road networks. Through vehicle-to-infrastructure communication channels, much more enriched data (e.g., the trajectories of CVs) will be available to operate more efficient and responsive traffic control modules. Many existing studies also showed the effectiveness of CV-based transportation systems could be improved with the increase of CV penetration rates.

Hence, to accommodate a large number of connected vehicles on the roads, traffic signal control on signalized arterials would require supports of various components such as roadside CV infrastructures, vehicle on-board devices, an effective communication network, and optimal control algorithms. In this project, we aim to establish a real-time and adaptive system for supporting the operations of CV-based traffic signal control functions. The proposed system will best utilize the capacity of the communication channels with optimization functions. The CV data sensing and acquisition protocol, built on Age of Information (AoI) as a function of the system overall latency and communication load, will support the feedback control loop to adjust signal timing plans.

Our multidisciplinary research team, including researchers from transportation engineering and electrical engineering, carry out the project tasks along four directions that capitalized on the PIs' expertise: (i) Data collection and communication, in which the proposed system is based on the AoI and optimize the communication network; (ii) Dynamic traffic signal coordination, which will concurrently facilitate the progression of traffic flows along multiple critical paths; (iii) Smart traffic signal control, where both operational efficiency and safety improvement are accounted for at signalized intersections; and (iv) Multimodal system design, which will integrate transit signal priority (TSP) and suppression controls for accommodating connected buses. This project addresses the urgent needs in CV system designs and offers control foundations to support the operations of urban signalized arterials in a CV environment.

1.0 INTRODUCTION

1.1 RESEARCH BACKGROUND

Car ownership has increased tremendously all over the world in recent years, resulting in an increase in travel time and aggregated traffic congestion. A recent technical report showed that 5.5 billion hours and 2.9 billion gallon of fuel were spent by American drivers each year due to traffic congestions (Schrank et al., 2012). The severity of those problems rises significantly at intersections because of the conflict of the right-of-way. Statistics results of the U.S. Department of Transportation (USDOT) and National Highway Traffic Safety Administration (NHTSA) further revealed that over half of all traffic congestion is caused at intersections and by traffic incidents (Florin and Olariu, 2015).

In the meantime, with the rapid development of communication technologies connected vehicles (CVs) are considered as the next frontier for automotive revolution. According to Foley & Lardner LLP's Survey (2017), it is predicted that more than 20% of vehicles in transportation networks will be CVs by 2025. By exchanging real-time information between vehicles and infrastructure, CV applications have demonstrated promising benefits in improving safety and mobility of transportation and reducing emissions. CV technology has also shown great potential in mitigating traffic congestion and improving the efficiency of transportation systems (Lu et al., 2014; Diakaki et al., 2015). The CVs equipped with onboard units can communicate via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) infrastructure in real time. V2V technology allows CVs to exchange critical vehicle status data, such as vehicle speeds, location, acceleration, etc. The V2I platform supports vehicles' communication with infrastructure (e.g., signal phase and timing data from the signal controller). For exploring potential benefits from CV technology, many researchers have investigated its applications in various traffic control domains (Wang et al., 2016; Ge and Orosz, 2014; Rao et al., 1994; Grumert and Tapani, 2017), which helps traffic agencies deploy CV applications in the real world.

Hence, there is an urgent need to develop reliable and an efficient communication network (e.g., V2V and V2I) (Lu et al., 2014), and real-time functions for supporting CV-based traffic signal control (Zhu and Ukkusuri, 2015). Those works will become even more challenging when the system has to accommodate a large number of CVs on the road network. In a review of the literature, although great research efforts have been made to study connected and/or autonomous vehicle technology, most of them didn't distinguish vehicles' need in connectivity and automation very well (Talebpour and Mahmassani, 2016), which leaves many unanswered questions in control function designs. To support the CV corridor operations, our research team at the University of Utah aims to develop a systematic protocol that leverages the interdependencies between fundamental theories of wireless communication and the next generation of intelligent transportation systems (ITS). More specifically, the developed system includes four key modules such as communication network optimization, dynamic traffic signal

coordination, smart intersection signal timing control, and multimodal traffic signal optimization (transit signal priority and suppression controls).

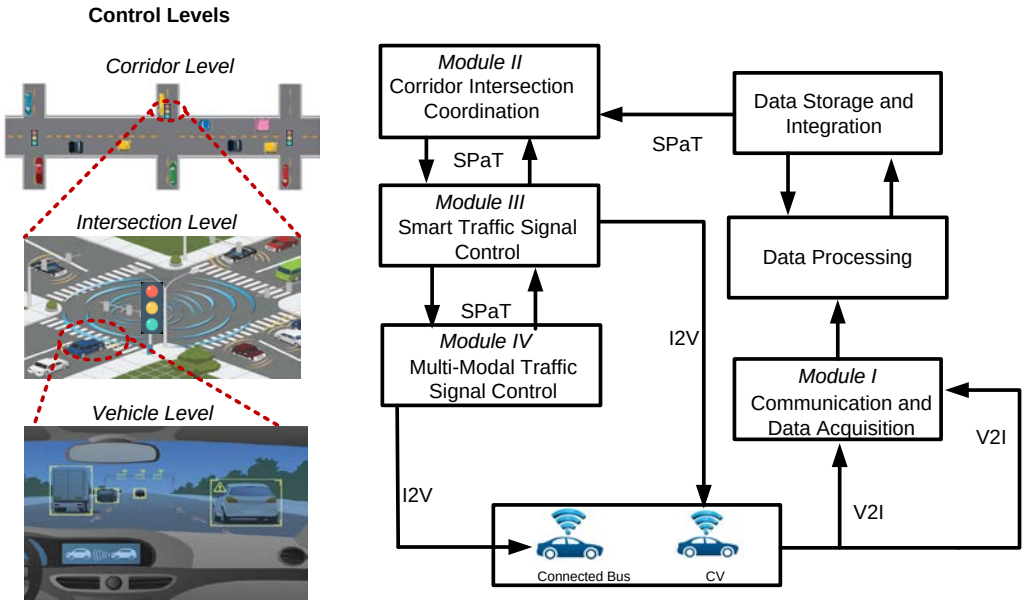


Figure 1.1: Overview of the key control modules of the system

As shown in Figure 1.1, Module-I presents a communication and data acquisition protocol built off the concept of Age of Information (AoI) as a function of overall system latency and communication load, which measures the "freshness" of information, introduced in Kaul et al. (2012). It has been well recognized that "freshness" is essential for sensor systems, especially for the vehicular applications (Kaul et al., 2011; Sun et al., 2017). The module is designed in such a way to minimize AoI under some system models (single vs. multihop transmissions) for all CVs to ensure optimal control of status updates received at roadside units (RSU). Then under the CV environment, Module-II proposes a methodology to adaptively change the signal offsets to provide dynamic progression to multiple critical path-flows. The critical paths, determined by CV trajectories, refer to the routes that connect those OD pairs with high traffic volumes. Following the multipath progression principle, a real-time optimization model is constructed to design a signal coordination plan with a control objective of maximizing green bandwidth along the determined critical paths. To solve the optimization problem, a solution algorithm based on dynamic programming is then proposed.

At the intersection level, Module-III further develops a control logic that can integrate CV data and traffic sensor information to concurrently address the needs of improving urban arterial safety and mobility. Under the mixed traffic pattern of CVs and human-driven vehicles (HVs), the system aims to achieve three primary objectives: proactively preventing rear-end collisions, reactively protecting side-street traffic from red-light-running vehicles, and effectively facilitating speed harmonization along local arterials. The embedded safety function will integrate CV and roadside sensor data to compute the distribution of dilemma zones for vehicles of different approaching speeds in real time. Such data fusion will enable the proposed system to offer the advice of either "stop" or "go" to both CVs and HVs so as to prevent rear-end collisions

and side-angle crashes. Given the locations and speeds of CVs, and the number of vehicles monitored by sensors, the proposed system can further compute the time-varying intersection queue length. Then the embedded mobility function will optimize the arterial signal plan in real time based on the output of Module-II, and produce the speed advisory to approaching vehicles to facilitate their progression through intersections.

To extend the application of the developed system under a multimodal environment, Module-IV further introduces a control function to accommodate connected buses. Particularly, the function aims to address one of the most difficult cases that median-island stations are located beside intersections. Due to the limited storage space in the stations, high bus volumes may cause queue spillbacks and consequently block the nearby intersections. Traditional transit signal priority (TSP) control systems may deteriorate such situations when signal priorities are granted to approaching buses. Hence, at intersections with nearby stations, a new operational logic is required to ensure the operational efficiency and safety of traffic signals. To satisfy such a need, this module presents a real-time signal control system, using the CV technology, to integrate both priority control strategies (i.e., green extension and red truncation) and suppression control strategies (i.e., green truncation and red extension). Particularly, the suppression strategies are only implemented to the far-side directional traffic so as to prevent station queue overflow.

1.2 PROJECT TASKS

This project includes the following five key tasks:

- **Literature Review**, where our project team conducts comprehensive reviews on existing research on various domains including sensing, data acquisition, and CV communications; device-to-device (e.g., V2V) coded caching with minimal (AoI), urban corridor adaptive control, CV and CAV trajectory control, and multimodal traffic signal control.
- **Optimal Dynamic Control for V2V and V2I Downlink Communications**, where an optimization model is developed to best utilize the capability of the communication networks so as to accommodate the data exchanging needs of CVs.
- **Dynamic Signal Progression Control Based on CV Technology**, where a real-time traffic signal coordination model is formulated to facilitate the movement of CV and HV flows along different critical paths.
- **SMART Control for Improving both Safety and Operational Benefits**, where an innovative control logic that can integrate CV data and traffic sensor information is developed to concurrently address the needs of improving urban arterial safety and mobility.
- **Multimodal Traffic Signal Control for Connected Buses**, where both priority control strategies (i.e., green extension and red truncation) and suppression control strategies (i.e., green truncation and red extension) are provided to connected buses based on the queue length in roadside stations and their lateness conditions.

2.0 LITERATURE REVIEW

2.1 SENSING, DATA ACQUISITION, AND CV COMMUNICATIONS

Vehicular network combines wireless communication with sensing devices installed on vehicles. With in-vehicle sensing technologies, we are capable of collecting abundant driving data, such as speed and engine parameters, from a large number of vehicles. Such data are characterized as large volume, multifrequency and multisource, which largely reflect the vehicle and the road traffic status and thereby are widely used to adjust the traffic signal control and evaluate human driving behaviors. When infrastructure senses vehicle status via V2I, the key for the sensing system is to optimize the "freshness" of data instead of minimizing delay. The concept of AoI is therefore uniquely suited for our application, which focuses on status update instead of data communications. AoI was introduced in Kaul et al. (2012) to quantify the "freshness" of knowledge about the status of remote systems. More specifically, it is the time elapsed since the generation of the last successfully received source message containing the updated information about its source system. Utilizing different communication models, in a series of papers (Kaul et al, 2012; Yates et al., 2012; Costa et al., 2014; Yates, 2015; Pappas et al., 2015; Kam et al., 2016; Kadota et al., 2016; Klei et al., 2017; Sun et al., 2017; Wu et al., 2018), AoI was characterized analytically. The concept has evolved and expanded to be more tractable or more suitable such as the Peak Age of Information (PAoI) (Costa et al., 2016) and average age penalty, opening even more research opportunities. In order to guarantee high freshness of information, keeping the AoI low is of high interest when AoI is being treated as a tool to facilitate the timely update of information that will eventually improve performance metrics in different contexts. Facilitated by the sensing system, data acquisition in vehicular networks is an important and challenging task due to the various requirements, complex and dynamic network environments. The efficient use of the wireless communication medium is one of the basic issues (Ashraf et al., 2016, 2017). Based on wireless communication technology, sensed data including localizations, speeds, directions, accelerations, etc. can be gathered at RSUs, which can determine accurately the traffic flow characteristics based on these data. The application of vehicular networks is capable of monitoring every single vehicle dynamically.

Reliable and seamless V2V and V2I data communication is the critical component of CV technology applications (Dey et al., 2016). Various wireless technologies have been used to support the data transfer requirement of diverse ITS applications, such as Bluetooth, ZigBee, Passive RFID, Ultra Wide Bandwidth (UWB) and mmWave communications (Lu et al., 2014). The selection of a wireless communication option relies on the accessibility and feasibility of wired and wireless communication options and data transfer requirements of particular applications. While existing ITS applications are infrastructure-based (i.e., installed at the roadside locations) (Silva et al., 2017), the next major deployment of wireless technologies within the transportation grid is the high-speed wireless communication between moving vehicles and transportation infrastructure. To meet the requirements of CV applications (e.g., fast acquisition, low latency with high reliability, highest security and privacy standards), the Federal Communications Commission had previously allocated the 5.9 GHz band (5.85GHz to 5.925GHz) for V2V/V2I communications, also known as dedicated short-rangecommunications (DSRC) (Bai et al., 2010; Wang and Hassan, 2008; Ma et al., 2009; Hafeez et al., 2013). These

communications include that of safety messaging, geographic locations, sensor data (e.g., tire pressure), which involves periodic broadcast of time-sensitive status information by vehicles. We focus on the DSRC technology as our baseline communication schemes in this project.

2.2 DEVICE-TO-DEVICE CODED CACHING WITH MINIMAL AGE-OF-INFORMATION (AOI)

The V2V or CV communication is one example of the so-called device-to-device (D2D) communication network in the area of wireless communications and networking. In the remainder of this section, we will refer to the vehicles as devices. In this research, based on a practical communication protocol such as DSRC, for device i under the assumption that the inter-updated time Z_i among two packets are independently and identically distributed (i.i.d.) following an exponential distribution, and the transmission slot is a constant time slots for each packet, in this case, the AoI is given by $c_1 \frac{\mathbb{E}[Z_i^2]}{\mathbb{E}[Z_i]} + c_2$, where c_1 and c_2 are some constants. It can be shown that the AoI $c_1 \frac{\mathbb{E}[Z_i^2]}{\mathbb{E}[Z_i]} + c_2$ is a monotonic function of the system overall communication load in the D2D networks. In order to minimize the AoI, we can simply minimize the communication load. In this work, we will exploit a newly developed concept of coded caching in order to minimize the overall communication load in the D2D networks. Coded caching has been shown to be an efficient approach to handle dramatically increased traffic in the current internet. Maddah-Ali and Niesen (MAN, 2014) introduced a centralized shared-link caching network model, where a central controller serves *K* users, each of which is equipped with a cache of size *M* files from a library of *N* files, via an errorless broadcast link (shared link). Here, a file could simply be a packet and the library could just mean all the sensing packet in some duration of time. In order to achieve the optimal worst-case rate (i.e., communication load) under uncoded cache placement, a cache placement and a coded delivery schemes were proposed and required to partition each file into $\binom{K}{t}$ packets where $t = K\mu$, $\mu = M/N$ (normalized cache memory size). It can be easily seen that for a fixed μ, $\binom{K}{K\mu}$ scales exponentially with *K*, which makes the implementation of the coded caching in wireless networks less practical due to unacceptably high subpaketization complexity. Later, it was shown that that this file subpacketization level is necessary to achieve the optimal rate under a so-called placement delivery array (PDA) design based on uncoded cache placement (Yan et al., 2017). In order to reduce the subpacketization levels, various combinatorial design-based schemes (Shanmugam et al., 2016; Shanmugam et al., 2017; Shangguan et al., 2018; Tang et al., 2018; Jin et al., 2019) have been proposed and showed that the subpacketization can be reduced at a cost of a higher transmission rate (or higher traffic load) for the shared-link network model. Ji, Molisch and Caire extended the shared link caching model to D2D coded caching networks, where no central controller is present and all users serve each other via individual shared links (Ji et al., 2016). It can be seen that the V2V network is a special case of the D2D wireless network. Under uncoded cache placement, Ji et al. proposed a caching scheme referred to as the *JCM scheme* that achieves the optimal worst-case rate of $R(M) = \frac{N}{M}\left(1 - \frac{M}{N}\right)$ when $N \geq K$. In this case, $R(M)$ is surprisingly not a function of *K* and hence it is scalable. In order to achieve this rate, the required number of subpackets (*subpacketization level*) is equal to $F^{JCM} = t\binom{K}{t}$, which can be impractical for large *K*. Efforts have been made in reducing the subpacketization levels

for D2D coded caching problem (Woolsey et al., 2017; Wang et al., 2017; Woolsey et al., 2018). For example, a design approach named *D2D placement delivery array* (DPDA) was introduced by Wang et al. (2017), which designed new DPDA schemes when $t = 2, t = K-2$, for which the JCM scheme is actually not optimal in terms of subpacketization although it achieves the optimal rate. Though some evidence of potential sunpacketization reduction opportunities was shown, the scheme therein is far from a general design since it essentially works only for two points of caching parameter *t*. In this work, we aim to minimize the overall system AoI and also aim to make the proposed system more practical by reducing the subpacketization level. Hence, we focus on the subpacketization level reduction problem of D2D coded caching under the assumption that the optimal communication load (AoI) must be preserved. Our new design result showed that order or constant reduction of subpacketization complexity of D2D coded caching can be achieved, which is a huge step in making coded caching techniques more practical.

2.3 URBAN CORRIDOR ADAPTIVE CONTROL

As one of the most commonly deployed traffic control systems, actuated signal control can accommodate variable phase sequences (e.g., optional protected left-turn phase), variable green time, and variable signal cycle length (Boillot et al., 1992). In both semi-actuated control and fully actuated control systems, there are three key parameters that need to be determined for operations: 1) the minimum green time (Kell, 1998); 2) the passage time (TRB, 2015); and 3) the maximum green time (Zhang and Wang, 2011). Using the real-time detected data, adaptive signal control is to dynamically adjust the signal plans and consequently improve the operational efficiency at local intersections. A lot of well-developed systems have been promoted by the transportation researchers, such as Spite Cycle and Offset Optimization Technique (SCOOT) (Day et al., 1998); Sydney Coordinated Adaptive Control System (SCATS) (Lowrie, 1990); Optimized Policies for Adaptive Control (OPAC) (Gartner, 2001); Real-time Hierarchical Optimizing Distributed Effective System (RHODES) (Mirchandani and Head, 2001); PRODYN (Khoudour et al., 1991); and UTOPIA (Mauro and Taranto, 1990). Grounded on the logic of adaptive signal control, more recent studies have focused on developing real-time signal control models with mixed CAVs and regular vehicles in the traffic (Feng et al., 2015; Hu et al., 2014, 2015). Since CV trajectory data are mainly used to estimate traffic flows within a centralized computational environment, those models often require a relatively high CV penetration rate (Flint et al., 2002; Zhou et al., 2017; Ma et al., 2017; Bang and Ahn, 2017; Munigety et al., 2016). CAV technology has the potential to enable precise control of individual vehicle trajectories (Ahn et al.,2013; Wang et al., 2014). With such trajectory control, vehicles can either adjust their driving behavior (e.g., speed, deceleration, acceleration) based on the approaching traffic signal timing plan (Kamalanathsharma et al., 2013) or coordinate with other vehicles to pass through an intersection during green (Dresner and Stone, 2008; Lee and Park, 2012). However, the human-in-the-loop feature, explicated by how other drivers would interact with CVs in the traffic stream, is not fully addressed.

2.4 CV AND CAV TRAJECTORY CONTROL

CAV technology has the potential to enable precise control of individual vehicle trajectories (Ahn et al., 2013; Wang et al., 2014a; Wang et al., 2014b). With such trajectory control, vehicles can either adjust their driving behavior (e.g., speed, deceleration, acceleration) based on the

approaching traffic signal timing plan (Kamalanathsharma et al., 2013) or coordinate with other vehicles to pass through an intersection during green (Dresner and Stone, 2008; Lee and Park, 2012).

Quite a few existing studies focused on addressing the individual vehicle trajectory control (He et al., 2015; Wu et al., 2015; Schouwenaars et al., 2001). Dynamic programming (DP)-based algorithm lays the theoretical foundation for single-vehicle trajectory optimization. However, DP-based trajectory optimization is oftentimes inefficient for real-world application due to its computational complexity, let alone applying to the scenarios involving multiple-vehicle trajectories control. To fill this research gap, Flint et al. (2002) proposed an approximate DP algorithm for multiple AVs to move cooperatively. McNaughton (2011) developed a five-dimensional search space formulation for AV motion planning. More recently, Zhou et al. (2017) presented a parsimonious shooting heuristic (SH) algorithm that can effectively smooth trajectories of a stream of CAVs approaching a signalized intersection. A time geography-oriented approach is combined with Newell's simplified car-following model to control detailed acceleration profile. Other studies also incorporated Newell's simplified car-following model for CAV platooning strategy (Bang and Ahn, 2017; Munigety et al., 2016). Following their preceding study (Zhou et al., 2017), Ma et al. (2017) discussed the SH algorithm on optimality and computational complexity. They proposed a framework for CAVs under centralized control for multitrajectory optimization with only a few control variables. Wei et al. (2017) developed mathematically rigorous optimization models and computationally tractable algorithms to model the dynamic process of tight platoon formation and system-level control for AVs. The vehicle trajectories are optimized subject to minimal safe driving distance between cars, as well as different entrance and exit boundary conditions.

2.5 MULTIMODAL TRAFFIC SIGNAL CONTROL

Developing public transportation systems has long been recognized as the most promising strategy to reduce the use of passenger cars and mitigate traffic congestion in urban networks (Al-Deek et al., 2017; Truong et al., 2017). Depending on the available information of operating buses, the functions of signal controllers, and the geometric layout of bus stations, a variety of transit signal priority (TSP) control systems have been implemented in practice since Wilbur et al. (1968) first conducted bus preemption experiments to reduce bus travel time. However, it remains a challenging research topic on how to implement TSP at those intersections with median stations which are often close to intersections. When the number of dwelling buses researched the station storage capacity, a granted TSP may cause queue overflows and consequently block the nearby intersection. Therefore, it is essential to develop a real-time control system which can concurrently reduce unnecessary time waste and prevent blocking the intersection.

In reviews of literature for TSP control, many control methods have been reported to reduce bus travel time at signalized intersections. Yagar and Han (1994) proposed some unconditional priority strategies which grant green extension or red truncations when approaching buses were detected. Ling et al. (2004) and Hounsell et al. (2008) later developed conditional priority strategies to improve bus punctuality based on the priority rules and the target bus's performance with respect to its schedule adherence. Besides the rule-based control logic, other research was conducted to minimize the total delay of detected buses (Head et al., 2006; Ma et al., 2010; Ma et al., 2013), the total vehicle delay of buses and passenger cars (Mirchandani et al., 2001; Li et al., 2011; Christofa and Skabardonis, 2011), the total person delay of buses and passenger cars

(Chang et al., 1996; Wu et al., 2012; Song et al., 2018), and the person delay of in-bus passenger delay and passenger waiting delay at next bus stop (Lin et al., 2013; Ye and Xu, 2017). Different from the previous methods considering delay as control objective, headway-based TSP control logic is to grant priority to a bus if it can improve bus regularity by considering the adjacent three buses belonging to the same route (Lin et al., 2017; Hu et al., 2015).

With the recent advancement of wireless communication technology, the connected vehicle (CV)-based system has attracted a lot of research efforts recently. Transit systems, managed by public transportation agencies, have been treated as a feasible testbed of CV technology. For example, the Utah Department of Transportation (UDOT) launched a project in 2017 to install DSRC devices on 30 intersections of Redwood Road in Salt Lake City to support the CV-based TSP control. On the research side, a next-generation TSP strategy was proposed (He et al., 2014; Hu et al., 2014) to simultaneously implement coordinated-actuated signal control in CV systems. Later on, Hu et al. (2015) presented a person-delay-based optimization algorithm by splitting the green time of bus approach under the CV environment, and then extended it for multiple conflicting priority requests (Hu et al., 2016). In addition, other researchers demonstrated that the conditional signal priority can be used to improve bus reliability under different operation modes via a mathematical model based on Brownian motion again (Anderson, 2018).

3.0 OPTIMAL DYNAMIC CONTROL FOR V2V AND V2I DOWNLINK COMMUNICATIONS

3.1 INTRODUCTION

The vehicle-to-vehicle (V2V) or connected vehicle (CV) communication is often referred to as a device-to-device (D2D) communication network in the area of wireless communications and networking. Based on a practical communication protocol for V2V and V2I communications such as DSRC, for device i under the assumption that the inter-updated time Z_i among two packets are independently and identically distributed (i.i.d.) following an exponential distribution, and the transmission slot for each device is a constant number of time slots for each packet. In this case, the AoI is given by $c_1 \frac{\mathbb{E}[Z_i^2]}{\mathbb{E}[Z_i]} + c_2$, where c_1 and c_2 are some constants. It can be shown that the AoI given by $c_1 \frac{\mathbb{E}[Z_i^2]}{\mathbb{E}[Z_i]} + c_2$ is a monotonic function of the system overall communication load in the communication networks (e.g., V2V or V2I). In order to minimize the AoI, we can simply minimize the communication load. In this report, we specifically focus on the V2V networks. However, our proposed scheme can be applied to the V2I infrastructure as well.

Coded caching is a recently proposed technique which has the potential to significantly reduce the network peak communication load buy utilizing the device cache memory in an efficient manner. More specifically, coded caching exploits the coordinated or random cache placement to create simultaneous broadcasting opportunities, improving the transmission efficiency and thus reducing delay and AoI. Note that broadcasting opportunities, meaning that one packet could be successfully received by more than one device in a wireless D2D (V2V) network, is automatically provided in wireless networks. Coded caching is first proposed for the shared-link network (e.g., downlink for a V2I network) where a central server serves a set of devices (users), each equipped with a cache memory. The achievable scheme therein (i.e., coded caching) is able to improve the data transmission efficiency (inverse of delay, or square root of AoI) by a multiplicative factor which is proportional to the aggregate cache size of all the users in the network (See Figure 3.1). This new factor is referred to as *global caching gain*, in contrast to the conventional local caching gain. The achievable communication load of coded caching is shown to be exactly optimal under the assumption of uncoded cache placement and worst-case load. Later, the idea of coded caching is extended to many different network settings, among which the D2D extension is probably the most important one. While enjoying the global caching gain, D2D coded caching also faces the problem of exponentially high subpacketization level (i.e., the number of packets per file), which makes the coded caching gain less promising when applying to real-world systems.

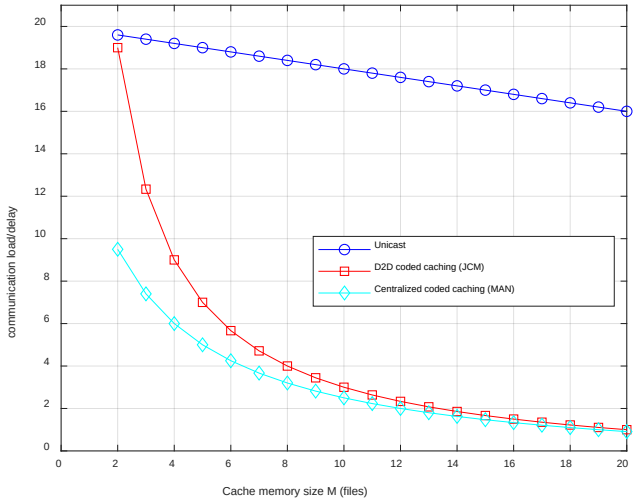


Figure 3.1: Comparison of communication load/delay versus cache memory size for traditional unicast, D2D coded caching and centralized coded caching for a system with $N = 20$ files and $K = 20$ devices.

In this project, we aim to minimize the overall system AoI. Meanwhile, we also guarantee that the proposed system is more practical by reducing the subpacketization level significantly in coded caching. In other words, we revisit the D2D coded caching subpacketization reduction problem and propose new schemes which work for general system parameters in a general D2D caching networks. We show that subpacketization reduction can be obtained for a large range of systems parameters while preserving the same (optimal) rate as the JCM scheme, revealing there is a fundamental difference between the share-link coded caching and the D2D coded caching in terms of subpacketization. More specifically, we proposed a new design framework called *Packet Type-based (PTB) design* tailored for D2D coded caching with reduced subpacketization level but optimal rate compared to the JCM scheme. In particular, in the PTB design, D2D users (or nodes) are first partitioned into multiple groups, each of which may or may not contain the same number of nodes. Then the corresponding packet types and multicasting group types are specified based on the prescribed node partition (referred to as node grouping). We show that certain packet types can be excluded under a given node grouping and we refer to this reduction gain as *raw packet saving gain*. Moreover, based on a careful selection of the transmitters within each multicasting group of different types a so-called *further splitting ratio gain* can also be obtained. While preserving the optimal rate, the joint effect of the *raw packet saving gain* and the *further splitting ratio gain* can lead to an order-wise or constant subpacketization reduction compared to the JCM scheme, where none of these gains is available in the JCM scheme. In fact, the PTB design problem can be cast into an integer optimization problem subject to node cache constraints and the design variables are the choices of possible transmitters within each multicasting group. Moreover, according to our knowledge, it is the first time in the literature showing that the employment of subpackets with heterogeneous lengths is a key to achieve subpacketization reduction for some system parameters.

In order to achieve the optimal rate, the JCM scheme (Ji at al., 2016) proposed a direct translation from MAN scheme (Maddah-Ali and Niesen, 2014) by splitting each packet further into t subpackets. It turns out that when the cache placement is uncoded and the delivery scheme is one-shot, this translation holds in general and it seems that the design procedure for the D2D

coded caching scheme is 1) design a shared link coded caching scheme and 2) translate it into D2D coded caching scheme. As a byproduct of the PTB design, we show that the above design methodology is not optimal in terms of subpacketization in general. Hence, in order to achieve good subpackeitzations in D2D coded caching networks, a new design framework is indeed needed.

The contribution of this study includes:

1) A Fundamental Difference Between Shared-link (e.g., downlink a V2I network) and D2D (e.g., a V2V network) Coded Caching: We showed that D2D coded caching schemes with optimal rates but strictly lower subpacketization levels exist for general system parameters, implying that direct translation of shared-link schemes to D2D setting is not optimal, at least in the sense of subpacketization. Therefore, a new design methodology is indeed needed for the design of D2D coded caching schemes.

2) A Unified Design Framework for Minimum AoI: We proposed the packet type-based (PTB) design paradigm which includes the following standard design steps: (*1) Node Grouping*: D2D nodes are first partitioned into multiple groups where each group may or may not contain an identical number of nodes (equal and unequal groupings). (*2) Classification of Packet and Multicasting Group Types*: Based on the prescribed node grouping in Step (1), packets and multicasting groups are specified and classified into multiple types according to their geometric structures; The rationale behind this classification is that different packet types can be treated differently, resulting in heterogeneous subpacketization for different packet types. (*3) Determination of Local Further Splitting Ratios (FSR)*: For each multicasting group type, assuming that a certain subset of nodes is selected as transmitters and all other nodes only receive but do not transmit, the local FSR for each involved packet type can be correspondingly determined. (*4) Determination of Global FSR*: After obtaining the local FSR vectors for all packet types within each multicasting group type in Step (3), the least common multiple (LCM) vector of these local FSR vectors is derived, which specifies the ultimate further splitting ratio of all possible packet types. (*5) Memory Constraint Check*: Previous steps must be conducted such that the node cache memory constraints are satisfied regardless of the homogeneous or the heterogeneous subpacket length design.

In the PTB design framework, subpacketization reduction compared to the JCM scheme consists of two parts (i.e., the raw packet saving gain and further splitting ratio gain). In Step (2) of the above design process, certain packet types can be potentially excluded (i.e., raw packets of certain types will not appear in the coded caching scheme) which leads to the raw packet saving gain. Note that in the JCM scheme, all possible packet types are employed and, hence, raw packet saving gain is not available. Through a careful selection of transmitters within each multicasting group in Step (3), local FSR vectors are obtained such that the overall FSR vector (LCM vector) results in a low subpacketization level. This subpacketization reduction due to smaller further splitting ratios of raw packets is called further splitting ratio gain. The subpacketization reduction of the coded caching schemes designed under the PTB framework is a result of the joint effect of the two kinds of gain mentioned above. Moreover, under the PTB design framework, coded caching schemes employing subpackets with heterogeneous lengths can also be designed, which gives more flexibility in satisfying the node cache memory size constraints. This is the very first proposal to use subpackets with heterogeneous lengths in the

design of coded caching schemes, both for the shared-link and D2D settings. In summary, the PTB framework presents a new perspective to understand the subpacketization problem of coded caching and provides important clues in designing low subpacketization-level schemes with optimal rate, which we believe is one significant step in addressing the fundamental limits of subpacketization complexity of coded caching.

3) Construction of Multiple Classes of Rate-optimal Schemes with Low Subpacketization Levels for D2D (e.g, V2V) Networks: Following the design methodology of the PTB framework, several classes of D2D coded caching schemes were constructed with either order or constant subpacketzation reduction while preserving the optimal rate. More specifically, we first constructed a class of schemes with order reduction compared to the JCM scheme when the number of D2D nodes, K, and the caching parameter $t = KM/N$, are both even in the regime of large cache memory size. Nodes are divided into groups, each of which contains exactly two nodes, and through the transmitter selection process. A constant FSR vector is obtained which does not depend on K, implying that as K goes larger, order reduction is achievable. Second, when K is odd, we constructed a scheme with two node groups containing $(K+1)/2$ and $(K-1)/2$ nodes, respectively. Constant reduction can be obtained in this case. We also showed that the grouping $((K+1)/2, (K-1)/2)$ actually achieves the minimal subpacketization among a class of schemes with two groups in the node grouping. Moreover, we also discovered a new class of coded caching schemes with heterogeneous subpacket lengths and constant subpacketization reduction can be achieved.

3.2 MODEL DEVELOPMENT

Consider a D2D caching network with a user set U where $|U| = K$. Each file from a library of N files is partitioned into F subpackets, which may or may not have equal length. The system operates in two separate phases (i.e., the cache placement phase and delivery phase) as described in [8]. In the cache placement phase, each user k stores up to MF packets from the file library. This phase is done without the knowledge of the users' requests. In the delivery phase, each user k reveals its request for a specific file $W_{d_k}, d_k \in [N]$ to other users. Let $\boldsymbol{d} = (d_1, d_2, \dots, d_K)$ denote the user demand vector. Since users have already cached part of the files, the task in the delivery phase is to design a corresponding transmission scheme for each user based on the cache placement and the user demand vector so that the users' demands can be satisfied. In this paper, our goal is to propose a new design framework based on combinatorial optimization such that the subpacketization level of each file is significantly reduced while preserving the optimal rate, defined as the minimum total communication load in terms of file transmissions. In the rest of this paper, we use $F^{JCM} = \frac{KM}{N}\binom{K}{KM/N}$ to represent the subpacketization level of the JCM scheme.

In order to achieve the above goal, we propose a new D2D coded caching design framework called packet type-based (PTB) design, which classifies packets and multicasting groups into multiple types. We will present the PTB design framework by decomposing it into the concepts including *Node Grouping*, *Packet Type*, *Multicasting Group Type*, *Further Splitting Ratio (FSR)*, *Further Splitting Ratio Table (FSRT)*, *Memory Constraint Table (MCT)* and *PTB Design as an Integer Optimization Problem*. Overall, the design of D2D coded caching schemes under the PTB framework consists of three different phases: the *PTB design phase*, *prefetching phase* and

delivery phase. All the aforementioned seven key concepts regarding the PTB design framework are considered and conducted in the design phase, in which the file splitting and the schedule of the coded delivery are determined. The key idea to improve the subpacketization comes from the special way to do file splitting instead of simply employing all $\binom{K}{t}$ packets and then splitting each packet further into $t = KM/N$ subpackets. According to the determined file splitting strategy in the PTB design phase, the subpackets are then assigned and stored by the D2D nodes in the prefetching phase. In the delivery phase, D2D nodes form multicasting groups of certain types determined in the PTB design phase and the selected transmitters then send coded multicast messages to serve the demands of each other. For each part in the design phase, we will present a corresponding example to better illustrate these definitions and concepts.

1) *Node Grouping:* The user set U is partitioned into m non-empty groups denoted by $Q_1, Q_2, \dots, Q_m$ where the i-th group contains $|Q_i| = q_i$ nodes. We use a *partition vector* $\boldsymbol{q} = (q_1, q_2, \dots, q_m, ,0,0,..,0)$ to represent such a node grouping, satisfying $q_1 \geq q_2 \geq \cdots \geq q_m \geq 0$ and $\sum_{i=1}^{m} q_i = K$. For a specific partition, there are multiple ways to assign the set of nodes to different groups, but they are all considered the same partition. The number of groups m and the number of nodes contained in each group are parameters to be designed. Let N_d denote the number of distinct elements in $\boldsymbol{q}$. We define a *unique set* as the union of non-empty groups containing the same number of nodes in $\boldsymbol{q}$. The i-th unique set, denoted by U_i, contains ψ_i groups and each group contains β_i nodes. A node partition is called *equal grouping* is all the groups contains the same number of nodes. Otherwise, it is called *unequal grouping*.

For example, consider the set of nodes $U = [7]$. Then $\boldsymbol{q} = (3, 2, 1, 1, 0, 0)$ is an unequal grouping representing a partition of the seven nodes into four groups Q_1, Q_2, Q_3 and Q_4 satisfying $|Q_1| = 3$, $|Q_2| = 2$ and $|Q_3| = |Q_4| = 1$. Consider two possible assignments of the seven nodes to the four groups according to $\boldsymbol{q}$, with the first assignment being $Q_1' = \{1,2,3\}, Q_2' = \{4,5\}$, $Q_3' = \{6\}$ and $Q_4' = \{7\}$ and the second being $Q_1'' = \{4,3,5\}, Q_2'' = \{1,6\}$, $Q_3'' = \{2\}$ and $Q_4'' = \{7\}$. These two assignments, though different, have identical grouping structures.

2) *Packet Type*: A packet type refers to a partition of $t = KM/N$ nodes and is represented by a partition vector $\boldsymbol{v} = (v_1, v_2, \dots, v_t)$, satisfying $v_1 \geq v_2 \geq \cdots \geq v_t \geq 0$ and $\sum_{i=1}^{m} v_i = t$. Different partitions of t correspond to different packet types. A raw packet $W_{n,T}$ for some $T \subset U$, $|T| = t$ refers to a packet that is cached exclusively in a set of nodes in T. Note that the definition of raw packets is conceptual, and further splitting of these raw packets into smaller subpackets might be necessary to accommodate the coded delivery phase. Each packet type may contain multiple raw packets. Since not all packets types can appear under a given node grouping we can exclude some invalid packet types, meaning that these packet types will not be used in the PTB design. This corresponds to the raw packet saving gain. In the delivery phase, raw packets might be further split into multiple subpackets, i.e., $W_{n,T} = \{W_{n,T}^{(i)} : i = 1,2, \dots, \alpha(\boldsymbol{v})\}$ where $\boldsymbol{v}$ is the packet type and $\alpha(\boldsymbol{v})$ is called *further splitting ratio (FSR).* Raw packets of the same type must have the same further splitting ratio. Note that all the raw packets of any packet type $\boldsymbol{v}$ have the same further splitting ratio $\alpha^{JCM} = t$ in the JCM scheme. The following example illustrates the concept of packet types and raw packets.

Example 3.1: (Packet Type) For $(K, t) = (6, 3)$ and $U = [K]$, consider the (equal) node grouping $\boldsymbol{q} = (3, 3)$ with a specific node assignment $Q_1 = \{1, 2, 3\}, Q_2 = \{4, 5, 6\}$, which is shown in Fig.

1. There are two different types of (raw) packets, i.e., $\boldsymbol{v}_1 = (3,0)$, meaning picking three nodes from either one of the two groups, and $\boldsymbol{v}_2 = (2,1)$, meaning picking two nodes from one group and one node from the other group. For example, the packet $W_{n,\{4,5,6\}}$ which is cached by nodes 4, 5, 6 is a type-$\boldsymbol{v}_1$ packet. The packet $W_{n,\{3,5,6\}}$ } which is cached by nodes 3, 5, 6 is a type-$\boldsymbol{v}_2$ packet. It can be seen that there are in total $F(\boldsymbol{v}_1)$ =2 type-$\boldsymbol{v}_1$ packets and $F(\boldsymbol{v}_2)$ =18 type-$\boldsymbol{v}_2$ packets. These two packets are called raw packets since they have not been further split into subpackets for the purpose of code multicasting in the delivery phase.

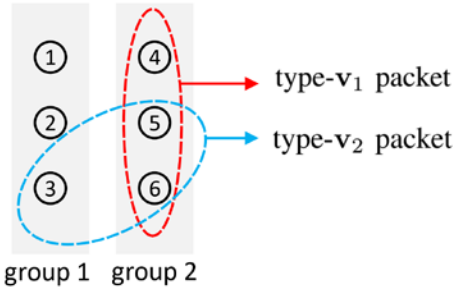


Figure 3.2: An illustration of packet types under node grouping $q = (3,3)$ with node assignment $Q_1 = \{1,2,3\}, Q_2 = \{4,5,6\}$. The red dashed circle represents a type $v_1 = (3,0)$, raw packet and the blue circle represents a type $v_2 = (2,1)$ raw packet.

3) *Multicasting Group Type:* A multicasting group is a set of $t+1$ node among which each node of a subset broadcasts some packets needed by the remaining t nodes. A *multicasting group type refers* to a specific partition of $t+1$ which is represented by a partition vector $\boldsymbol{s} = (s_1, s_2, \dots, s_{t+1})$where all the elements sum up to $t+1$ and are not increasing with the subscript index. Different partitions of $t+1$ correspond to different multicasting group types. A *unique set* $\widehat{U}_i, i \in [N_d]$ within $\boldsymbol{s}$ represents the set of groups containing the same number of nodes. N_d denotes the number of different unique sets in $\boldsymbol{s}$.

4) *Further Splitting Ratio (FSR):* The further splitting ratio of a packet type $\boldsymbol{v}$, denoted by $\alpha(\boldsymbol{v})$, means that all the raw packets of type- $\boldsymbol{v}$ need to be split into $\alpha(\boldsymbol{v})$ subpackets during the delivery phase. For a multicasting group S of type $\boldsymbol{s}$ containing N_d different unique sets, a set of nodes $Tx \subseteq S$ is selected to serve as transmitters for the coded multicasting transmissions in S. We can select Tx in such a way that it can be expressed as a union of $|D_T|$ different unique sets where, in the multicasting group S of type $\boldsymbol{s}$, $D_T \subseteq [N_d]$ is defined as the set of the indices of the unique sets which are selected as transmitters, i.e., $Tx = \cup_{i\in[D_T]} \widehat{U}_i$. Denote $g_i = |\widehat{U}_i|$ as the number of nodes contained in the unique set $\widehat{U}_i$ then we have $|Tx| = \sum_{i\in[D_T]} g_i$. Under such a selection of transmitters, the further splitting ratios for the involved packet types is $\alpha(\boldsymbol{v}_i) = \sum_{j\in[D_T]} g_j - 1$ if $i \in [D_T]$ and $\alpha(\boldsymbol{v}_i) = \sum_{j\in[D_T]} g_j$ if $i \notin [D_T]$.

Consider the multicasting group $S = \{3,4,5,6\}$ in the above example which is composed of two unique sets $\widehat{U}_1 = \{4,5,6\}, \widehat{U}_2 = \{3\}$. Each $D_T \subseteq [2]$ and $D_T \neq \emptyset$ corresponds to a specific choice of transmitter selections. Hence, there are three different choices which are $D_T = \{1\}, \{2\}$ and $D_T = \{1,2\}$, respectively.

Choice 1: $D_T = \{1\}$. This means that the nodes in the first unique set (nodes 4, 5 and 6) are selected as transmitters (i.e., Tx) $= \{4, 5, 6\}$ and node 3 will not transmit anything but only receive in the delivery phase. In this case, the type-$\boldsymbol{v}_1$ raw packet $W_{d_3,\{4,5,6\}}$ needed by node 3 is transmitted by nodes 4, 5 and 6. To obtain the optimal rate (or equivalently, t nodes are served simultaneously by one coded transmission), the packet $W_{d_3,\{4,5,6\}}$ needs to be further split into $\alpha(\boldsymbol{v}_1) = 3$ subpackets each of which is delivered by a node $\{4, 5, 6\}$. On the other hand, the type-$\mathbf{v}_2$ raw packet $W_{d_4,\{3,5,6\}}$needed by node 4 is transmitted by node 5 and 6 (node 3 does not transmit) and, hence, needs to be further split into $\alpha(\mathbf{v}_2)$ = two subpackets, each of which is delivered by node 5 or 6. Similarly, the other two type-$\mathbf{v}_2$ packets $W_{d_5,\{3,4,6\}}$, $W_{d_6,\{3,4,5\}}$ also need to be split into two subpackets. Since all the raw packets in the JCM scheme are further split into $\alpha^{JCM} =$ three subpackets and here we have $\alpha(\mathbf{v}_2) = 2 < \alpha^{JCM} = 3$, less number of further splitting is required, which corresponds to the further splitting ratio gain mentioned previously. It can be seen that only the further splitting ratio gain can be obtained under this choice.

Choice 2: $D_T = \{2\}$. This means that the second unique set (node 3) is selected as the only transmitter (i.e., Tx) $= \{3\}$, then nodes 4, 5 and 6 will only receive but not transmit subpackets in the delivery phase. In this case, the type-$\mathbf{v}_2$ raw packet $W_{d_4,\{435,6\}}$ needed by node 4 is only transmitted by node 3. Hence, there is no need to further split it into multiple subpackets. Similarly, there is no need to split the other two type-$\mathbf{v}_2$ raw packets $W_{d_5,\{3,4,6\}}$, $W_{d_6,\{3,4,5\}}$. We use $\alpha(\mathbf{v}_2) = 1$ to indicate that no further splitting is needed. However, the type-$\mathbf{v}_1$ raw packet $W_{d_3,\{4,5,6\}}$ needed by node 3 is not transmitted by any other node. We can simply exclude this raw packet without sacrificing the completeness of the caching scheme because node 3's need of $W_{d_3,\{4,5,6\}}$ is eliminated though no nodes are transmitting this packet. We use $\alpha(\mathbf{v}_1) = 0$ to indicate that type-$\mathbf{v}_1$ packets are excluded. Actually, all type-$\mathbf{v}_1$ raw packets can be excluded here, which corresponds to the raw packet saving gain mentioned previously. In this case, since $\alpha(\mathbf{v}_2) = 1 < \alpha^{JCM} = 3$, both the raw packet saving gain and the further splitting ratio gain are available.

Choice 3 (JCM scheme): $D_T = \{1, 2\}$. This means that both unique sets, U_1 and U_2 (i.e., all four nodes in S_1) are selected as transmitters. In this case, any raw packet $W_{d_i,S\backslash\{i\}}$, $i \in S$, either type $\mathbf{v}_1$ or type-$\mathbf{v}_2$, needed by node i is transmitted by the other three nodes in S. To preserve the optimal rate, all these four packets should be further split into $\alpha(\mathbf{v}_1) = \alpha(\mathbf{v}_2)$ = three subpackets. We can see that neither the raw packet saving gain nor the further splitting ratio gain are obtainable here. This actually corresponds to the design of the JCM scheme: For any multicasting group S, all the $t + 1$ nodes in S are selected as transmitters and a further splitting ratio of $\alpha(\mathbf{v}) = |S| - 1 = t$ is required for any packet type $\mathbf{v}$, which leads to a subpacketization level of $F^{JCM} = t\binom{K}{t}$. However, as discussed above, it is not always necessary to select all $t + 1$ within S to serve as transmitters. A selection of one or more (not all) unique sets within S as transmitters will lead to smaller further splitting ratios of the raw packets, providing an opportunity to reduce the subpacketization. The overall subpacketization reduction of the proposed design framework is the result of both the raw packet saving gain and the further splitting ratio gain.

5) *Further Splitting Ratio Table (FSRT):* Given a node grouping $\boldsymbol{q}$, let V, S denotes the total number of different valid packets types and multicasting group types, respectively. A *further splitting ratio table* is a matrix $\boldsymbol{\Lambda} = [\alpha_{ij}]_{S\times V}$ which specifies the local further splitting ratios of

the involved packet type set derived from all the different S multicasting types. More specifically, the i-th, $i \in [S]$ row of the FSRT, which is referred to as the *local further splitting ratio vector (local FSR vector)*, denoted by $\boldsymbol{\alpha}_i$, consists of further splitting ratios $\alpha(\boldsymbol{v}_j)$ for all packet types $\boldsymbol{v}_j$ involved in multicasting group type $\boldsymbol{s}_i$ and all the other entries are left empty. Note that a further splitting ratio of $\alpha = 0$ is not the same as an empty entry. To determine the overall further splitting ratio for all the V types of packets, we need to derive the *Least Common Multiple (LCM) vector* $\boldsymbol{\alpha}^{\text{LCM}}$ (defined below) of the S different local further splitting ratio vectors.

Definition 3.1: (*Least Common Multiple (LCM) Vector*) For a set of n row vectors $A = \{a_i\}_{i\in[n]}$ in which $|a_i| = V$ and a_i may contain 'empty' entries. The LCM vector of A, denoted by $\alpha^{LCM} = \text{LCM}(A)$, is defined as: $\exists\ z_1, z_2, \cdots, z_n$ which are all positive integers, such that

$$\begin{aligned} z_1 a_1 &= z_2 a_2 = ... = z_n a_n \\ a^{LCM} &= combine\{zi * a_i\}_{i\in[n]} \end{aligned} \tag{3.1}$$

Where $\{z_i*\}_{i\in[n]} = \arg\min \| combine\{z_i * a_i\}_{i\in[n]} \|$ are the optimal combing coefficients. The combined operation returns a vector a^{LCM} of length V whose j-th entry takes the value of the non-zero and non-empty value among the set of n j-th ($j \in [V]$) entries of all the n vectors $z_i a_i, i \in [n]$. We assume that 1) the product of any integer and an empty entry is still an empty entry; 2) entry '0' is equal to any other entries, including non-zero entries and empty entries; and 3) empty entry is equal to any other zero/non-zero entries.

For example, let us consider $A = \{a_1, a_2, a_3\}$ where $a_1 = (1, 2, 3, 0)$, $a_2 = (\bullet, 4, \bullet, 3)$ and $a_3 = (\bullet, 0, 2, 1)$. It can be easily seen that the optimal coefficients are $z_1^* = 2, z_2^* = 1, z_3^* = 3$ and $a^{LCM} = \text{LCM}(A) = \text{combine}(\{2a_1, a_2, 3a_3\}) = (2,4,6,3)$.

The LCM vector may not always exist. If it exists, it must be unique. In the PTB design, the global splitting ratio vector, denoted by a^{LCM}, is obtained by deriving the LCM vector of the set of local splitting ratio vectors $\{a_i\}_{i\in[S]}$, i.e., $a^{LCM} = \text{LCM}(\{a_i\}_{i\in[S]})$. The LCM operation of ensures the coded transmissions within multicasting groups of different types agree with each other. That is, for example, if a packet type v is involved in two different multicasting group type s_1 and s_2 which yields two distinct local FSR for $\boldsymbol{v}$, i.e., $a'(\boldsymbol{v}) = 1$ for s_1 and $a''(\boldsymbol{v}) = 2$ for s_2, then the LCM operations gives a global FSR of $a(\boldsymbol{v}) = LCM\big(a'(\boldsymbol{v}), a''(\boldsymbol{v})\big) = 2$, implying that type-**v** packets should be split into two subpackets in the final PTB design. This also means that each desired message of the nodes within type-s_1 multicasting groups should be a concatenation of two type-v subpackets while in type-s_2 multicasting groups the desired message is only one individual type-**v** subpacket.

Continue with Example 3.1. There are $S = 2$ multicasting group types $\boldsymbol{s}_1 = (3, 1)$ and $\boldsymbol{s}_2 = (2, 2)$. There are also $V = 2$ packets types $\boldsymbol{v}_1 = (3, 0)$ and $\boldsymbol{v}_2 = (2, 1)$. The corresponding involved packet type sets are $\boldsymbol{\rho}_1 = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\boldsymbol{\rho}_2 = \{\mathbf{v}_2\}$. The selection of transmitters is as follows. For $\boldsymbol{s}_1$, we choose the second unique set as transmitters. We use the superscript * to mark the transmitters within a multicasting group type, *i.e.*, $\boldsymbol{s}_1 = (3, 1^*)$. For example, in aspecific type-$\boldsymbol{s}_1$

multicasting group $S_1 = \{3, 4, 5, 6\}$, the transmitters is node 3. This selection will result in a local further splitting ratio vector $\boldsymbol{\alpha}_1 = (\alpha(\mathbf{v}_1), \alpha(\mathbf{v}_2)) = (0, 1)$. For $\boldsymbol{s}_2$, the only choice is to select all $t + 1 = 4$ nodes as transmitters since there is only $N_d = 1$ unique set which has to be selected. This results in a further splitting ratio vector $\boldsymbol{\alpha}_2 = (\bullet, \alpha(\mathbf{v}_2)) = (\bullet, 3)$ in which the symbol • denotes an empty entry since type-$\mathbf{v}_1$ packets do not appear in type-$\mathbf{s}_2$ multicasting groups. As a result, the FSRT is

Table 3.1: The FSRT of Example 3.1

FSRT	$\mathbf{v}_1$	$\mathbf{v}_2$
$\boldsymbol{\alpha}_1$	0	1
$\boldsymbol{\alpha}_2$	•	3

from which we can easily obtain $\boldsymbol{\alpha}^{LCM} = \text{LCM}(\alpha_1, \alpha_2) = (0, 3)$, implying that in the PTB design, type-$\mathbf{v}_1$ raw packets are excluded while each type-$\mathbf{v}_2$ raw packet is further split into three subpackets. It can be verified that this PTB design is complete and correct, the rate is optimal and the required subpacketization level is

$$F = \alpha(v_1)F(v_1) + \alpha(v_2)F(v_2) = 54$$

while $F^{JCM} = 60$ and the reduction of the six subpackets is due to the exclusion of type-$\mathbf{v}_1$ packets.

6) *Memory Constraint Table (MCT*): Given a node grouping $\boldsymbol{q}$ containing N_d different unique sets, a memory constraint table is a matrix $\Omega = [\omega_{ij}]$ with $\omega_{ij} = F_i(\boldsymbol{v}_j)$ where $F_i(\boldsymbol{v}_j)$ denotes the number of type $\boldsymbol{v}_j$ raw packets cached by a node in the *i*-th unique set. Denote $F_i = [F_i(\boldsymbol{v}_1), F_i(\boldsymbol{v}_2), \dots, F_i(\boldsymbol{v}_V)]$ as the *i*-th row of Ω. Also denote the raw packet number vector as $F_i = [F(\boldsymbol{v}_1), F(\boldsymbol{v}_2), \dots, F(\boldsymbol{v}_V)]$ where $F(\boldsymbol{v}_j)$ represents the total number of raw packets of type $\boldsymbol{v}_j$ in a PTB design. Furthermore, we define the *node cache difference vector* as $\Delta F_i = [\Delta_{i1}, \Delta_{i2}, \dots, \Delta_{iV}]$ in which $\Delta_{ij} = F_{i+1}(\boldsymbol{v}_j) - F_i(\boldsymbol{v}_j)$ is the difference of the number of type-$\boldsymbol{v}_j$ raw packets cached by nodes in the *i*-th and (*i+1*)-th unique sets. Assuming all the subpackets have the same size, the memory constraint can be expressed as

$$\alpha^{LCM}\Delta F_i^T = \alpha^{LCM}(F_{i+1} - F_i)^T = 0, \forall i \in [N_d - 1] \quad (3.2)$$

Since all the nodes have identical cache memory size, caching the same number of subpackets of equal length satisfies the memory constraint. Note that for equal node grouping, where there is only one unique set, the memory constraint is automatically satisfied. The exact length of the subpackets can be determined by the fact that each node has a cache memory size of *M* files.

7) *PTB Design as An Integer Optimization:* With all the above definitions, under the condition of equal-length subpacketizations (all subpackets have identical length), we introduce the following integer optimization problem that determines the optimal LCM vector which results in the minimum F.

$$\begin{aligned} &min\ F = \alpha^{LCM} F^{T} \\ &s.t.\ \ \alpha^{LCM} \in \Phi \\ &\qquad \alpha^{LCM} \Delta F_i^{T} = 0, \forall i \in [N_d - 1] \end{aligned} \tag{3.3}$$

in which Φ represents the set of all possible LCM vectors derived from the *S* local further splitting vectors based on the set of all possible node grouping $\boldsymbol{q}$ and the set of all possible selections of transmitters within each multicasting group type under each $\boldsymbol{q}$. Any feasible solution of the above optimization problem will give a valid D2D coded caching scheme and the optimal solution yields a scheme with minimal subapcketization level. Moreover, we can extend this optimization problem to the case of heterogeneous subpacket length setting by a minor modification to the above optimization problem formulation.

3.3 CASE STUDY

In this section we present our main result, which consists of three theorems, on subpacketization level reduction problem for D2D coded caching using the PTB design framework. For each theorem, one or more corresponding examples are provided to illustrate the achievability of our results. We will use the notation $\bar{t} \triangleq K - t = K(1-\mu)$ in the following.

Theorem 1: (*Order Reduction on F*) For even $\bar{t} \coloneqq K - t$, where $K = 2m$, using the PTB design framework, the optimal rate of D2D caching networks is achievable and

$$\frac{F}{F^{JCM}} = \Theta\left(\frac{f(\bar{t})}{K - \bar{t}}\right) \tag{3.4}$$

where $f(\bar{t}) \coloneqq \prod_{i=1}^{\bar{t}/2}(2i-1)$ is a function which depends only on $\bar{t}$. Moreover, for all $K \geq 2\bar{t}$ and $\bar{t} = O(loglogK)$, F/F^{JCM} vanishes as K goes to infinity.

Table 3.2 shows a comparison of the subpacketization level between the proposed PTB design for systems with $\bar{t} = K - t = 2$ and even $K \geq 4$, and the JCM scheme. This comparison shows that the PTB design of Theorem 1 can significantly reduce the subpacketization level.

Table 3.2: Comparison of subpacketization level of the PTB design in Theorem 1

K	10	20	40	80	160
F	40	180	760	3120	12640
F^{JCM}	360	3420	29640	246480	2009760

In the following, we highlight the implications of Theorem 1 and its connections to some prior works.

1) *Order reduction in large cache memory regime:* From Theorem 1, it can be seen that when K ≤ 2t is even and *t* is large enough (i.e., *t* = *K* − O(*loglog K*)), an order gain in terms of subpacketization can be obtained using the PTB design compared to the JCM scheme while preserving the optimal rate. More specifically, for a fixed $\bar{t}$, it can be seen that F/F JCM =

$\Theta(1/K)$, implying the order gain. Note that when $\bar{t}$ is kept fixed, the normalized cache size $\mu = t/K = (K - \bar{t})/K = 1 - \Theta(1/K)$ =and hence $\lim_{K\to\infty} \mu = 1$, which essentially means that the order gain is only obtainable in a larger cache memory regime. In fact, this order reduction is achieved by a specific equal node grouping method with $m = K/2$ groups each containing exactly $q = 2$ nodes, i.e., $q = (2, 2, \cdots, 2)$. Order reduction can also be obtained using other equal grouping methods, as we will show later.

2) *Asymmetric multicast transmission matters:* In the achievable scheme of Theorem 1, both the raw packet saving gain and further splitting ratio gain are available. The key idea behind the PTB design is that the enforcement of asymmetry in the coded is multicasting transmission (i.e., only a subset of nodes within each multicasting group are selected as transmitters and all other remaining nodes will only receive instead of transmitting. To preserve the optimal rate, (i.e., each coded multicast message should be simultaneously useful for t different nodes) further splitting of the raw packets is necessary. The number of subpackets that each raw packet needs to be split into is equal to the number of transmitters which transmit the stored raw packet within a particular multicasting group. Since in any multicasting group of $t + 1$ nodes, every t nodes share a raw packet which is desired by the other node (assuming no raw packet is excluded), selecting t nodes to transmit will result in a further splitting ratio of $\alpha = t$. However, asymmetry can be exploited, where we can choose less than t nodes to serve as transmitters, resulting in a smaller further splitting ratio.

3) *Connection to JCM scheme:* In the PTB design framework, if the node grouping contains only one unique set, (i.e., $\boldsymbol{q}' = (K, \mathbf{0}^{(K-1)})$ or $\mathbf{q}'' = (\mathbf{1}^{(K)})$) then there will be V = 1packet type and S = 1 multicasting group type, which are $\mathbf{v} = (t, \mathbf{0}^{(t-1)})$, $\mathbf{s} = (t + 1, \mathbf{0}^{(t)})$ for $\boldsymbol{q}'$ and $\mathbf{v} = (\mathbf{1}^{(t)})$, $\mathbf{s} = (\mathbf{1}^{(t+1)})$ for $\boldsymbol{q}''$. In both cases, there is only one unique set within the multicasting group type. Hence, the only unique set has to be chosen as the transmitters. The only type of raw packets has to be split into $\alpha = t$ subpackets. No packet types can be excluded and a further splitting ratio of t has to be enforced. This exactly corresponds to the design of the JCM scheme.

4) *Comparison to DPDA* (Wang et al., 2017)*:* The DPDA proposed several D2D coded caching schemes achieving a lower subpacketization level than the JCM scheme while preserving the optimal communication rate, and showed the subpacketization optimality via the PDA argument. However, DPDA design only treats four points of the caching parameter (i.e., $t = 1,2, K - 1, K - 2$), which is far from a general design. Our PTB design matches these cases of DPDA and we can show the subpacketization optimality from the perspective of the PTB design. Therefore, PTB is more general than the DPDA design. More specifically, it is shown (Wang et al., 2017) that 1) For $t = 1, F \geq K$ (matching the JCM scheme) ; 2) For $t = K - 1, F \geq K(K - 1)$ (matching the JCM scheme); 3) For $t = 2, F \geq K^2/4$ (lower than the JCM scheme, tight when K is even); 4) For $t = K - 2, F \geq K(K - 2)$ for odd K and $F \geq K(K - 2)/2$ for even K (lower than the JCM scheme). These results, including the achievability and converse, can all be easily obtained under the PTB design framework. For example, the subpacketization $F = K^2/4$ when $t = 2$ can be achieved by equal node grouping $\boldsymbol{q} = (K/2, K/2)$. More specifically, there are two possible packet types, $\boldsymbol{v}' = (2,0), \boldsymbol{v}'' = (1,1)$ where $\boldsymbol{v}' = (2,0)$ is excluded and $\boldsymbol{v}''$does not need to be further split due to the selection of transmitters within the only multicasting group type $\boldsymbol{s} = (2, 1^*)$. Since the number of type-$\boldsymbol{v}''$ packets is equal to $(K/2)^2 = K^2/4$, the subpacketization is equal to $F = K^2/4$. The converse bound on the subpacketization can be

shown via the PTB framework from the memory constraint satisfaction argument and the existence of least common multiple vector of local further splitting ratio vectors.

5) *Connection to Coded Distributed Computing (CDC)* (Li et al., 2017)*:* It is shown that for single-round (s =1) MapReduce distributed computing, the proposed CDC scheme (communication- computation tradeoff) in [14] is equivalent to the JCM scheme (rate-memory tradeoff) for D2D coded caching. More specifically, the number of raw packets and subpackets in D2D coded caching correspond to the number of input data files (sub-tasks) and the number of intermediate values per task in CDC, respectively. As a result, the PTB design framework also works for the distributed computing case and can reduce the subpacketization complexity (number of sub-tasks per task, and the number of intermediate values per sub-task) without sacrificing the optimal communication load.

Theorem 2: (*Order Reduction on F*) Let $\zeta(t) \triangleq \lim_{K\to\infty} F/F^{JCM}, 2 \le t \le K$. For $(K,t) = (2q, 2r)$ with $q \ge t+1, r \ge 1$, using the PTB design framework with the two-group equal grouping (i.e., $\boldsymbol{q} = (K/2, K/2)$), the optimal rate of D2D caching networks is achievable by the further splitting ratio vector $\alpha^{LCM} = (0,1,2,\dots,r)$. When $r \ge 2$, we have

$$\zeta(t) < \frac{1}{2}\left(1-\frac{1}{2^{t-1}}\right) \tag{3.5}$$

We next highlight the implications of Theorem 2 as follows.

1) *Constant reduction*: Under the corresponding PTB design, a more-than-half reduction on F can be achieved compared to the JCM scheme. This design applies to any integer value of t as long as it is an even number in the range $2 \le t \le K-2$. Note that we have proved that for $t = 1, K-1$, the JCM scheme is actually optimal in terms of F in the discussions following Theorem 1. This result also demonstrates that for even caching parameter t, the JCM scheme is not optimal in subpacketization in general. The achievable design employs a simple two-group equal node grouping $\boldsymbol{q} = (K/2, K/2)$ and the corresponding FSR vector of $\alpha^{LCM} = (0,1,2,\cdots,r)$ suffices to achieve subpacketization level $F < F^{JCM}/2$.

2) *Connection to Theorem 1:* For some system parameters K and t, there can be two PTB designs with different F, according to Theorem 1 and Theorem 2. In fact, the PTB designs are not unique (i.e., for the same set of system parameters), multiple rate-optimal schemes can be designed via the PTB approach with different subpacketization levels since different node grouping might be used, different raw packet types can be excluded and different further splitting ratios can be chosen. The key difference is that Theorem 1 indicates an order gain in the large cache regime while Theorem 2 only achieves constant gain, implying that the design of Theorem 1 is better. However, in other cache regimes, this may not be true. We will simply choose the one that yields a smaller F.

Theorem 3: (*Heterogeneous Subpacketization*) Let $\zeta(t) \triangleq \lim_{K\to\infty} F/F^{JCM}, 2 \le t \le K$. For $(K,t) = (2q+1, 2r)$ with $q \ge 2r+1, r \ge 1$, using the PTB design framework with the two-group equal grouping (i.e., $\boldsymbol{q} = ((K+1)/2, (K-1)/2)$) the optimal rate of D2D caching

networks is achievable by the further splitting ratio vector $\alpha^{LCM} = (0,2,4,\ldots,t-2,t,t,\ldots,t)$. Moreover, when $t \geq 4$, we have

$$\zeta(t)=\frac{1}{t}\frac{\sum_{i=2}^{r}\frac{2}{(i-2)!(t-i+1)!}+r(r!)^{-2}}{\sum_{i=1}^{r}\frac{2}{(i-1)!(t-i+1)!}+(r!)^{-2}}<\frac{1}{t}\left(\binom{t}{r}/2^{t}-1\right)+1 \qquad (3.6)$$

Theorem 3 states that a constant reduction on *F* can be achieved when the number of nodes is added and the caching parameter *t* is even. The novelty of Theorem 3 lies in that subpackets with heterogeneous size are employed in the PTB design, which according to our knowledge is the very first approach in the literature to try heterogeneous subpacketization. Note that optimal rate is still preserved when using the heterogeneous subpacketization design. Since $\frac{1}{t}\left(\frac{\binom{t}{r}}{2^{r}}-1\right)<0$ always holds under the condition of Theorem 3, the right-hand side of Equ. (3.6) is always less than 1, indicating a constant factor reduction. When $K \to \infty$, the expression of $\varsigma(t)$ is derived under the condition that $t = KM/N$ is fixed. This means that the cached fraction of the file library is small (i.e., $M/N \to 0$) as *K* approaches infinity. Hence, Theorem 3 basically says that there is constant reduction gain in asymptotically small cache fraction in contrast to Theorem 1, which characterizes an order-wise reduction in the asymptotically large fraction region.

In the following, we provide several examples corresponding to the above theorems to show the supremacy of the PTB design over existing schemes.

Table 3.3: The FSRT of Example 3.2

FSRT	$\boldsymbol{v}_1$	$\boldsymbol{v}_2$	$\boldsymbol{v}_3$
α_1	•	1	0
α_2	4	3	•
α^{LCM}	4	3	0

Example 3.2: (*Order Reduction under Equal Node Grouping*) Consider the system parameters $(K, \bar{t}) = (3m,3)$ where $m \geq \bar{t} = 3$ ($K \geq 9$) and equal node grouping $\boldsymbol{q} = (3, 3, \cdots, 3)$. In this case, $(S, V) = (2, 3)$. The packet types, multicasting group types and involved packet type sets are (the unique sets which are selected as transmitters within each multicasting group type are marked with the superscript *):

$$v_1 = (3^{(m-3)}, 2^{(3)}), \quad v_2 = (3^{(m-2)}, 2^{(1)}, 1^{(1)}), \quad v_3 = (3^{(m-1)}, 0^{(1)})$$
$$s_1 = (3^{(m-1)}, 1^{(1)*}), \quad \rho_1 = \{v_2, v_3\}$$
$$s_2 = (3^{(m-2)}, 2^{(2)*}), \quad \rho_2 = \{v_1, v_2\}$$

and the FSRT is from which we obtain α^{LCM} = (4, 3, 0), implying that type $\mathbf{v}_3$ is excluded (raw packet saving gain) and $\mathbf{v}_1$, $\mathbf{v}_2$ have further splitting ratios of 4, 3 < $\alpha^{\text{JCM}} = t$ ($t = K - \bar{t} = K - 3 \geq 6$ since $K \geq 9$) respectively (*further splitting ratio gain*). Hence, the number of subpackets is equal to $F = (4,3,0)[F(v_1), F(v_2), F(v_3)]^T = K(K-3)(2K-3)/3 = \Theta(K^3/3)$ in which $F(v_1) = K(K-3)(K-6)/6$, $F(v_2) = K(K-3)$ and $F(v_3) = K/3$. On the other hand, we have $F^{JCM} = K(K-1)(K-2)(K-3)/6 = \Theta(K^4/6)$. Therefore, we have

$$\frac{F}{F^{JCM}} = \frac{2(2K-3)}{(K-1)(K-2)} = \Theta\left(\frac{4}{K}\right)$$

implying an order gain on F of the PTB design compared to the JCM scheme. Table 3.4 shows the comparison of subpacketizaiton level of Example 3.3 to that of the JCM scheme.

Table 3.4: Comparison of subpacketization level of the PTB design in Example 3.2

K	9	18	36	72	144
F	162	2970	27324	233496	1928880
F^{JCM}	3024	12240	235620	8230320	68715504

Next we consider the detailed delivery procedure for a system with K = 9, N = 3, M = 2 and t = 6. Consider a specific equal-grouping assignment Q_1={1, 2, 3}, Q_2 ={4, 5, 6} and Q_3 ={7, 8, 9}. Since α^{LCM} = (4, 3, 0), type- $\boldsymbol{v}_3$ packets are excluded and type- $\boldsymbol{v}_1$ and type- $\boldsymbol{v}_2$ packets need to be further split into four and three subpackets, respectively. In this case F = 270 while $F^{JCM} = 504$.

The cache placement is that node k stores any subpacket $W_{n,T}^{(j)}, k \in T$. Note that in this example the subpacketization reduction gain compared to the JCM scheme consists of two parts: *1)* raw packet saving gain: $tF(\boldsymbol{v}_3)$ = 18 subpackets and *2)* further splitting ratio gain: $(t-4)F(\boldsymbol{v}_1) + (t-3)F(\boldsymbol{v}_2)$ = 216 subpackets. We can see that the reduction is mainly due to smaller further splitting ratios of type-$\boldsymbol{v}_1$ and type-$\boldsymbol{v}_2$ packets in this example.

For a type-s_1 multicasting group $S_1 = [7]$, node 7 is the only transmitter and it transmits three coded multicast messages: $\oplus_{k\in[6]} W_{d_k,S_1\backslash\{k\}}^{(j)}, j = 1,2,3$ to all the other nodes in S_1. Each node k recovers its desired subpackets $\{W_{d_k,S_1\backslash\{k\}}^{(j)}, j = 1,2,3\}$ with the help of their cached contents while node 7 itself only transmits but receives nothing. For a type-s_2 multicasting group $S_2 = [9]\backslash\{6,9\}$, the set of type-$\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ packets involved are $\{W_{d_k,S_2\backslash\{k\}}^{(j)}, j = 1,2,3,4, k = 1,2,3\}$ and $\left\{W_{d_k,S_2\backslash\{k\}}^{(j)}, j = 1,2,3, k \in S_2\backslash Q_1\right\}$,respectively. Denote $W^{(j)} = \oplus_{k\in[3]} W_{d_k,S_2\backslash\{k\}}^{(j)}, j = 1,2,3,4$. Nodes 4, 5, 7, 8 each sends a coded multicast message as follows:

$$W_4 = W^{(1)} \oplus W^{(1)}_{d_5, S_2\setminus\{5\}} \oplus W^{(1)}_{d_7, S_2\setminus\{7\}} \oplus W^{(1)}_{d_8, S_2\setminus\{8\}}$$
$$W_5 = W^{(2)} \oplus W^{(1)}_{d_4, S_2\setminus\{4\}} \oplus W^{(2)}_{d_7, S_2\setminus\{7\}} \oplus W^{(2)}_{d_8, S_2\setminus\{8\}}$$
$$W_7 = W^{(3)} \oplus W^{(2)}_{d_4, S_2\setminus\{4\}} \oplus W^{(2)}_{d_5, S_2\setminus\{5\}} \oplus W^{(3)}_{d_8, S_2\setminus\{8\}}$$
$$W_8 = W^{(4)} \oplus W^{(3)}_{d_4, S_2\setminus\{4\}} \oplus W^{(3)}_{d_5, S_2\setminus\{5\}} \oplus W^{(3)}_{d_7, S_2\setminus\{7\}}$$

from which we can see that all nodes can recover their desired subpackets. Since each coded message is simultaneously useful for $t = 6$ nodes, the rate is optimal. The transmission procedure for other multicasting groups is similar.

We will illustrate the two-group equal grouping scheme in Theorem 2 using the following two examples.

Example 3.3: Consider $K = N = 6$, $M = 2$, $t = KM/N = 2$ and the node grouping $\boldsymbol{q} = (3, 3)$ with a specific node assignment$Q_1 = \{1, 2, 3\}$, $Q_2 = \{4, 5, 6\}$. We have $V = S = t/2 + 1 = 2$. The packet and multicasting types are

$$v_1 = (2,0), \quad v_2 = (1,1)$$
$$s_1 = (3,0), \quad \rho_1 = (v_1)$$
$$s_2 = (2,1^*), \ \rho_2 = (v_1, v_2)$$

There are $F(\boldsymbol{v}_1) = 6$ type-$\boldsymbol{v}_1$ raw packets (per file) which are

$$\left\{W_{n,T} : T \subset Q_1 \text{ or } T \subset Q_2, |T| = 2\right\}, \ \forall n \in [N]$$

and there are $F(\boldsymbol{v}_2) = 9$ type-$\boldsymbol{v}_2$ raw packets (per file) which are

$$\left\{W_{n,T} : T = \{k', k''\}, k' \in Q_1, k'' \in Q_2\right\}, \ \forall n \in [N]$$

Therefore, there are $F(\boldsymbol{v}_1) + F(\boldsymbol{v}_2) = 15$ raw packets per file, which corresponds to the number of packets in the JCM scheme. There are $F(\boldsymbol{s}_1) = 2$ type-$\boldsymbol{s}_1$ multicasting groups $S_1 = Q_1, S_2 = Q_2$ There are are $F(\boldsymbol{s}_2) = 18$ type-$\boldsymbol{s}_2$ multicasting groups each of which contains one node from $Q_1 (Q_2)$ and two nodes from $Q_2 (Q_1)$. Then the global FSR is $\alpha^{LCM} = (0, 1)$, implying that type-$\boldsymbol{v}_1$raw packets are excluded. Since type-$\boldsymbol{s}_1$ multicasting groups only contain $\boldsymbol{v}_1$ packets and all type-$\boldsymbol{v}_1$ packets are excluded, the transmissions within each type-$\boldsymbol{s}_1$ multicasting groups become unnecessary. This is to say, type-s$\boldsymbol{s}_1$ multicasting groups are also excluded. Type-$\boldsymbol{v}_2$ packets have a FSR of 1, which means that these packets do not need to be further split to accommodate the coded delivery phase. Therefore, the total number of subpackets in the PTB design is equal to

$$F = (0,1)\left[F(v_1) + F(v_2)\right]^T = 9$$

Then the prefetching and delivery phases are described as follows.

Prefetching Phase: Each node $k \in [6]$ caches all the (type-$\boldsymbol{v}_2$) subpackets

$$Z_k = \left\{ W_{n,T} : k \in T, \forall n \in [6] \right\}$$

It can be seen that each node in total stores 18 subpackets, satisfying the memory constraint.

Delivery Phase: For each type-$\boldsymbol{s}_2$ multicasting group $S = \{k'\} \cup \{k'', k'''\}$ where node k' belongs to one of the groups Q_1, Q_2 and nodes k'', k''' belong to the other group, node k' sends a coded multicast message

$$\bigoplus_{k \in S \setminus \{k'\}} W_{n,d_k,S\setminus\{k\}}$$

which is useful to both nodes k'' and k'''. Note that in S, nodes k'', k''' do not send anything. Hence, node k' receives nothing. It can be seen that the desired subpacket can be successfully decoded. Each node *k* is involved in six multicasting groups, in each of which node *k* receives a different desired subpacket. Hence, each node receives six subpackets in the delivery phase of its desired file. Combined with the three subpackets in the cache, all nodes can recover their desired files. Since each coded message serves *t* = 2 nodes simultaneously, the rate is optimal. We have *F* = 9 while $F^{JCM} = 30$. The reduction of 21 subpackets is due to the joint effect of raw packet exclusion and a smaller FSR of type-$\boldsymbol{v}_2$ packets.

This design can be generalized fully to the case $\boldsymbol{q}$ = (*K*/2, *K*/2) for arbitrary even $K \geq 4$ and the [1] required subpacketization is $F = (0,1)\left[F(v_1) + F(v_2)\right]^T = K^2/4$. Then

$$\frac{F}{F^{JCM}} = \frac{1}{4}\frac{K}{K-1} \leq \frac{1}{3}$$

Since $K \geq 4$, implying a constant reduction.

Example 3.4: Consider *t* = 4 with $\boldsymbol{q}$ = (*K*/2, *K*/2). We have *V* = *S* = 3 and the packet and multicasting groups are

$$\begin{aligned} &v_1 = (4,0), \quad v_2 = (3,1), \quad v_3 = (2,2) \\ &s_1 = (5,0), \quad \rho_1 = (v_1) \\ &s_2 = (4,1^*), \ \rho_2 = (v_1, v_2) \\ &s_3 = (3,2^*), \ \rho_3 = (v_2, v_3) \end{aligned}$$

The global FSR vector is $\alpha^{LCM} = (0,1,2)$. As a result, we have

$$F = (0,1,2)\left[F(v_1) + F(v_2) + F(v_3)\right]^T = \frac{K^2(K-2)(5K-14)}{96}$$

On the other hand, $F^{JCM} = K(K-1)(K-2)(K-3)/6$. Since we require that $K \geq t + 2 = 6$, the following bound holds

$$\frac{F}{F^{JCM}} = \frac{0.3125K(K-2.8)}{(K-1)(K-3)} \leq 0.4, \quad \forall K \geq 6$$

and $\lim_{K\to\infty} \frac{F}{F^{JCM}} = 0.3125$, which is less than one-third. Table 3.5 shows the comparison of subpacketizaiton level of Example 3.4 to that of the JCM scheme.

Table 3.5: Comparison of subpacketization level of the PTB design in Example 3.4

K	6	12	24	48	96
F	24	690	13992	249504	4205184
F^{JCM}	60	1980	42504	778320	13287840

In the following, we will illustrate the general achievable scheme for Theorem 1 and some examples to show that order reduction of the subpacketization level can be achieved while preserving the same (optimal) rate as the JCM scheme. We assume that K is sufficiently large while $\bar{t} = K - t$ is kept fixed. We will show later that t can be extended to a higher order of $\bar{t} = O(log_2 log_2 K)$.

For $(K, \bar{t}) = (2m, 2r)$ with $m \geq \bar{t} + 1$, $r \geq 1$, consider the node grouping $\mathbf{q} = (\mathbf{2}^{(m)})$ with $m = K/2$ groups each containing exactly two nodes. In this case, the number of multicasting types and packet types are $(S, V) = (r, r+1)$. More specifically, the i-th ($i \in [r]$) multicasting type $\mathbf{s}_i$ and the j-th ($j \in [r+1]$) packet type $\mathbf{v}_j$ are (transmitters are marked by the superscript*).

$$\mathbf{s}_i = \left(\mathbf{2}^{(m-(r+i)+1)}, \mathbf{1}^{(2i-1)\star}, \mathbf{0}^{(r-i)}\right)$$

$$\mathbf{v}_j = \left(\mathbf{2}^{(m-(r+j)+1)}, \mathbf{1}^{(2(j-1))}, \mathbf{0}^{(r-j+1)}\right)$$

The i-th involved packet type set is $\rho_i = \{\boldsymbol{v_{i+1}}, \boldsymbol{v_i}\}$ and the corresoding local FSR vector is $a_i^{local} = [a(\boldsymbol{v}_i), a(\boldsymbol{v}_{i+1})] = [2(i-1), 2i-1]$. As a result, the global FSR vector will be $a^{LCM} = (a_1, a_2, \dots, a_{r+1})$ in which the entries are $a_1 = 0$ (type $\boldsymbol{v}_1$ is excluded), $a_2 = 2^{r+1}\prod_{i=1}^{r-1} i$, $a_r = 2(r-1)\prod_{i=1}^{r-1}(2i-1)$, $a_{r+1} = \prod_{i=1}^{r-1}(2i-1)$ and $a_j = \prod_{i=1}^{r-1}(2i-1)\left(\prod_{i=j-1}^{r-1} 2i\right), \forall j \in [3, r-1]$. The number of subpackets required is therefore equal to $F = \boldsymbol{a}^{LCM}\boldsymbol{F}^T = \sum_{j=1}^{r+1} \boldsymbol{a}_j F(\boldsymbol{v}_j)$.

The entries of $\boldsymbol{a}^{LCM}$ is actually strictly increasing as shown in the following lemma.

Lemma 1: For the derived LCM vector $\boldsymbol{a}^{LCM}$, the entry sequence $\{a_j\}_{j=1,2,\dots,r+1}$ is strictly increasing.

Proof of Lemma 1: for $3 \leq j \leq r-2$, we have $a_{j+1} < a_j$ due to

$$\frac{\alpha_{j+1}}{\alpha_j} = \frac{\prod_{i=1}^{(j+1)-1}(2i-1)\left(\prod_{i=(j+1)-1}^{r-1}2i\right)}{\prod_{i=1}^{j-1}(2i-1)\left(\prod_{i=j-1}^{r-1}2i\right)} = \frac{2j+1}{2(j-1)} > 1$$

Also, $a_3 > a_2$ due to

$$\frac{\alpha_3}{\alpha_2} = \frac{\prod_{i=1}^{3-1}(2i-1)\left(\prod_{i=3-1}^{r-1}2i\right)}{2^{r-1}\prod_{i=1}^{r-1}i} = 3/2 > 1$$

It can be seen that $a_r > a_{r-1}$ due to

$$\frac{\alpha_r}{\alpha_{r-1}} = \frac{2(r-1)\prod_{i=1}^{r-1}(2i-1)}{\prod_{i=1}^{(r-1)-1}(2i-1)\left(\prod_{i=(r-1)-1}^{r-1}2i\right)} = \frac{2r-3}{2r-4} > 1$$

Therefore, we have $a_{j+1} < a_j, \forall j \in [r]$. Hence, we finish the proof of Lemma 1.

To derive an upper bound on F/F^{JCM}, we have the following definition and lemma.

Definition: (Dominant Packet Type) Under a specific node grouping, the dominant packet type, denoted by $\boldsymbol{v}_{dom}$, is defined as the packet type containing the largest number of raw packets for sufficiently large K.

Lemma 2: For a specific packet type v, which is a portion of t, and a given equal node grouping $\boldsymbol{q} = (\boldsymbol{q}^{(m)})$ in which each group contains $q = \frac{M}{m} \in \mathbb{Z}^+(m \geq \bar{t}+1)$ nodes, the dominant packet type $\boldsymbol{v}_{dom} = ((\boldsymbol{q}-\boldsymbol{1})^{(\bar{t})}, \boldsymbol{0}^{(\boldsymbol{m}-\bar{t})})$.

Proof of Lemma 2: For the node grouping $\boldsymbol{q}$, a packet type which is a partition of the integer t, can be equivalently represented by a partition of the integer $\bar{t} = K - t$, meaning that excluding the $\bar{t}$ nodes from $\boldsymbol{q}$ and the remaindering t nodes form a packet type. Let the partition

$$(\boldsymbol{\beta}_1^{(\psi_1)}, \boldsymbol{\beta}_2^{(\psi_2)}, \cdots, \boldsymbol{\beta}_{N_d}^{(\psi_{N_d})}, \boldsymbol{0}^{(m-\bar{t})})$$

satisfying

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_{N_d} > 0$$

and

$$\sum_{i=1}^{N_d}\beta_i\psi_i = \bar{t}, \sum_{i=1}^{N_d}\psi_i = N_p$$

(N_p denotes the number of parts) represent a partition of $\bar{t}$ (corresponding to the packet type). Then the number of type-v packets can be calculated as

$$
\begin{aligned}
F(\mathbf{v}) &= \prod_{i=1}^{N_d} \binom{m-d_i}{\psi_i} \binom{q}{\beta_i}^{\psi_i} \\
&= \prod_{i=1}^{N_d} \prod_{j=0}^{\psi_i - 1} (m - d_i - j) f(\beta_i, \psi_i) \\
&\overset{(a)}{\approx} \prod_{i=1}^{N_d} m^{\psi_i} f(\beta_i, \psi_i) \\
&= m^{\sum_{i=1}^{N_d} \psi_i} \prod_{i=1}^{N_d} f(\beta_i, \psi_i) \\
&= m^{N_p} \prod_{i=1}^{N_d} f(\beta_i, \psi_i) \\
&= K^{N_p} \frac{\prod_{i=1}^{N_d} f(\beta_i, \psi_i)}{q^{N_p}} \\
&= C K^{N_p}
\end{aligned}
$$

where in (a) we let K be sufficiently large and keep q fixed. The notations are as follows: $\psi_0 = 0$ and $d_i = \sum_{j=1}^{i} \psi_{j-1}$. The function f is defined as

$$f(\beta_i, \psi_i) \triangleq \binom{q}{\beta_i}^{\psi_i} / \psi_i!$$

and C is given by

$$\prod_{i=1}^{N_d} f(\beta_i, \psi_i) / q^{N_p}$$

Which is a constant which does not depend on K.

From the above equation, we see that $F(\boldsymbol{v}) = \Theta(K^{N_p})$. Hence, the dominant type will be the type with maximum.

Thus, we finish the proof of Lemma 2.

Next, we derive a stick upper bound the ratio F/F^{JCM}. Using lemma 1, we have $a_j \leq a_{r+1}, \forall j \in [r+1]$. As a result, for any $K \geq 2\bar{t}$, we have

$$\frac{F}{F^{\mathrm{JCM}}} = \frac{\sum_{j=1}^{r+1} \alpha_j F(\mathbf{v}_j)}{t \sum_{j=1}^{r+1} F(\mathbf{v}_j)} < \frac{\alpha_{r+1}}{t} \frac{\sum_{j=1}^{r+1} F(\mathbf{v}_j)}{\sum_{j=1}^{r+1} F(\mathbf{v}_j)} = \frac{\alpha_{r+1}}{K - \bar{t}} = \frac{\prod_{i=1}^{\bar{t}/2} (2i-1)}{K - \bar{t}}$$

It can be seen that the upper bound $\frac{\prod_{i=1}^{\bar{t}/2}(2i-1)}{K-\bar{t}} = \Theta(1/K)$ for a fixed $\bar{t}$, implying an order reduction of F compared to the JCM scheme.

The next lemma shows that the upper bound in above equation vanishes as K goes to infinity if $\bar{t} = O(loglogK)$. A special case is that $\bar{t}$ is kept fixed as a constant when K grows.

Lemma 3: For $\bar{t} = \mathrm{K} - \mathrm{t} = O(loglogK)$, the upper bound on F/F^{JCM} vanishes as K goes to infinity.

Proof of Lemma 3: Without loss of generality, assume for some constant c and $K = 2^{2^{\kappa}}$. Then

$$\begin{aligned}
\lim_{K\to\infty} \frac{F}{F^{\mathrm{JCM}}} &= \lim_{K\to\infty} \frac{\prod_{i=1}^{c\log\log K}(2i-1)}{K - 2c\log\log K} \\
&= \lim_{K\to\infty} \frac{\prod_{i=1}^{\kappa c}(2i-1)}{2^{2^{\kappa}} - 2\kappa c} \\
&= \lim_{K\to\infty} \frac{\prod_{i=1}^{\kappa c}(2i-1)}{2^{2^{\kappa}}} \\
&= \lim_{K\to\infty} \prod_{i=1}^{\kappa c} \frac{2i-1}{2^{\frac{2^{\kappa}}{\kappa c}}}
\end{aligned}$$

Since each individual term f_i satisfies

$$f_i \triangleq \frac{2i-1}{2^{\frac{2^{\kappa}}{\kappa c}}} \leq \frac{2\kappa c - 1}{2^{\frac{2^{\kappa}}{\kappa c}}} \overset{K\to\infty}{=} 0$$

As a result, we obtain

$$\lim_{K\to\infty} \frac{F}{F^{\mathrm{JCM}}} = \lim_{K\to\infty} \prod_{i=1}^{\kappa c} f_i = \prod_{i=1}^{\kappa c} \lim_{K\to\infty} f_i = 0$$

Hence, we finish the proof of Lemma 3.

From Lemmas 1-3, we finish the proof Theorem 1. Theorems 2 and 3 can be proved using similar approaches and thus we omit the proof of these theorems.

4.0 DYNAMIC SIGNAL PROGRESSION CONTROL BASED ON CV TECHNOLOGY

4.1 INTRODUCTION

With the communication optimization model introduced in Chapter 3.0, CVs can exchange real-time information with roadside infrastructures and provide much enriched data to support various traffic operation and control tasks. To mitigate urban traffic congestion, optimal traffic signal control is one of the most promising methods by allocating green times to various vehicle movements and to regulate traffic flows. Currently, the most commonly used signal control strategies are fixed-time control, actuated control, and adaptive control. Fixed-time control schedules signal timing plans for different time of day based on historical data. It assumes that the traffic demand does not vary significantly within the entire control period, but may fall short of efficiency when traffic demand fluctuates quickly in practice (Feng et al., 2015). Both actuated control and adaptive control utilize real-time data collected from roadside sensors (e.g., detectors and radars) to control signals. For actuated control, several logics including green extension, gap out, and max out are applied to adjust signal timing. Adaptive control employs the collected data to predict the incoming traffic situations and make control decisions to optimize some objectives (e.g., minimizing delay). Although actuated and adaptive control indicates better effectiveness compared to fixed-time control, two limitations are revealed with them. First, the collected data only show instantaneous information when vehicles pass over. Vehicle states such as heading and acceleration cannot be obtained. Besides, the sensors can only provide traffic information which is measured at discrete spatial points. Second, the cost of installing and maintaining sensors is relatively high. If a sensor malfunction occurred, the performance of those two control strategies will degrade significantly (Feng et al., 2015).

Recent years have witnessed a rapid development of CV technology. CV wireless communication could be able to provide enriched vehicle data including location, speed, acceleration, stop time, queue length, etc. With the high-resolution CV data, signal controllers can respond to the fluctuation of traffic condition more efficiently and adjust the signal timing plan accordingly. Moreover, the obtained CV trajectories, through V2I communications, allow the control system to better understand the OD patterns of the arterial network. Such information can greatly support the design of signal coordination plans to facilitate the movement of the most critical OD flows (or path-flows).

4.2 SYSTEM OVERVIEW

With the emerging and rapid development of CV technology, on-board units (OBU) are able to transmit messages to or receive from roadside units (RSU) through dedicated short-range communication (DSRC). The broadcasted messages are defined by Society of Automotive Engineers (SAE) *(*Draf and Jam, 2006*)*, including a) intersection geometry message (MAP) providing road geometry; b) signal phase and timing (SPaT) message reporting real-time signal information; c) basic safety message (BSM) offering real-time vehicle trajectories; d) signal request message (SRM) indicating requested information; and e) signal status message (SSM) recording priority status information. Those new high-resolution data enable traffic signals to be

responsive to real-time changes in traffic demand patterns. Thus, traffic signals could perform more effectively.

Based on SPaT and BSM, this study proposes a methodology of dynamic signal progression control for arterial coordination. Figure 4.1 shows the key control structure of this system, which includes two core functions: V2I communication and signal plan (offset) adjustment. It is expected that CVs will co-exist with HVs for a long time period and the market penetration of CVs will be low in the near future. Some researchers have been conducted to study the effects of low CVs penetration rates on mobility performance and test the applications with a wide range of penetration rates.

Table 4.1 summarizes the minimum penetration rates of CVs for several mobility applications. Depending on the results shown in Table 4.1, this study assumes the CV market penetration as 30% for providing vehicle trajectory data. As shown in Figure 4.1, the vehicle trajectory is firstly decoded based on the BSM to determine the critical paths. Second, the signal progression control algorithm functions to optimize the traffic signal timing plan based on the decomposed CVs data. In this study, the traffic signal timing plan will be optimized by adjusting offsets of each intersection along an arterial after each control period (e.g., five minutes). Notably, the length of the control period shall be set properly to ensure the pattern of critical paths remain unchanged between at least two adjacent control periods. Then the signal progression control algorithm is utilized to optimize the offsets for the next period based on the determined critical paths (using CV data) in the last control period. This study will test the proposed in Vissim, and BSM is generated in the driver model interface in this simulation environment.

Table 4.1: CVs applications and corresponding minimum penetration rate

Applications	Minimum CV penetration rate	Reference
Arterial performance measurement	10%-50%	Li et al., 2008; Argote et al., 2015
Traffic signal control	20%-30%	Feng et al., 2015; He et al., 2012; Priemer et al., 2009
Speed estimation	20%	Rim et al., 2011; Goodall et al., 2014
Queue estimation	30%	Tiaprasert et al., 2015; Ban et al., 2011

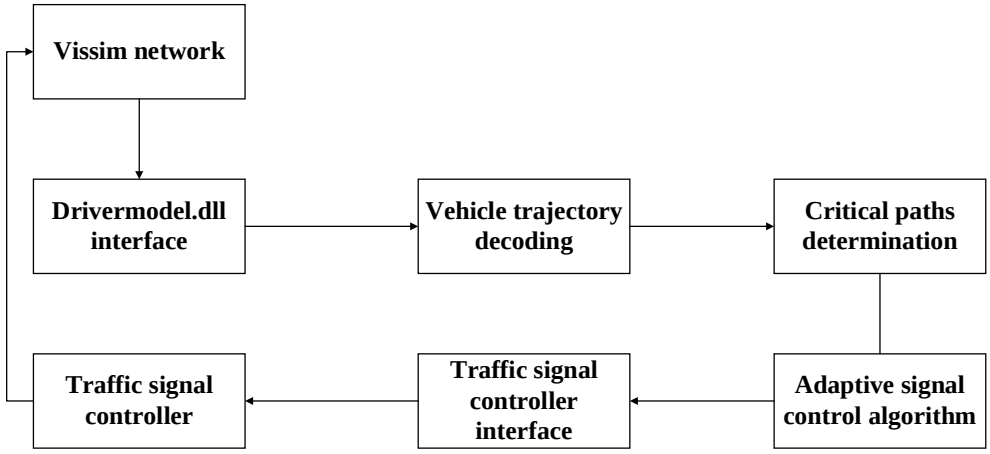


Figure 4.1: Flow chart of signal progression control system

Note that one important part of this progression control system is to determine critical paths. The definition of critical paths in this paper is illustrated by Yang et al. (2015), which refers to the routes connecting to origin-destination (OD) pairs with the highest volumes. Taking the road network in Chupei, Taiwan, as an example, as shown in Figure 4.2. The flow from node 6 to node 1 shows the highest volume with 702 vph. And other paths (path2-path5) also exhibit high volumes. Hence, the five paths are defined as critical paths in this example network.

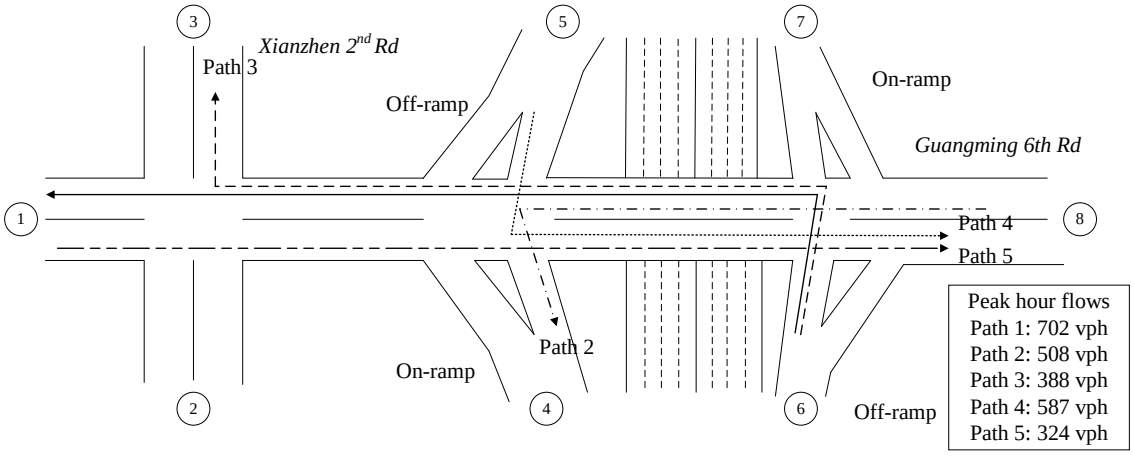


Figure 4.2: Critical paths of the road network in Chupei, Taiwan (Yang et al., 2015)

4.3 DYNAMIC SIGNAL PROGRESSION CONTROL

4.3.1 Model Development

This study adopts the bandwidth-based method to design signal coordination plans for vehicles traveling along an arterial. More specifically, the offsets of all intersections after each control period (e.g., every five minutes) to maximize the green bandwidth of all critical path-flows along both outbound and inbound directions of the arterial, which is denoted by Eq. (4.1).

$$\max(\sum_i \sum_p \ \omega_p\ (j) w_{p,i}(j) + \sum_i \sum_p \ \bar{\omega}_p\ (j) \bar{w}_{p,i}(j)) \tag{4.1}$$

where j denotes the control period, represented by index 1, 2, 3, etc. (e.g., number 1 represents the first control period); $w_{p,i}(j)$ and $\overline{w}_{p,i}(j)$ denote the green bandwidth of critical path p for outbound and inbound between intersection i and $i-1$, respectively, in seconds; $\omega_p(j)$ and $\overline{\omega}_p(j)$ denote the weighting factor for path p. It should be noted that those factors are identified by traffic demands along various paths.

The green bandwidth of path p for outbound and inbound directions between two adjacent intersections are computed by Eq. (4.2) and Eq. (4.3).

$$w_{p,i}(j) = max\big(w_{r,p,i}(j) - w_{l,p,i}(j), 0\big) \tag{4.2}$$

$$\overline{w}_{p,i}(j) = max\big(\overline{w}_{r,p,i}(j) - \overline{w}_{l,p,i}(j), 0\big) \tag{4.3}$$

where $w_{r,p,i}(j)$ and $w_{l,p,i}(j)$ ($\overline{w}_{r,p,i}(j)$ and $\overline{w}_{l,p,i}(j)$) denote the right bound and left bound of the green band of path p for outbound (inbound), which are calculated by Eq. (4.4)-Eq. (4.7).

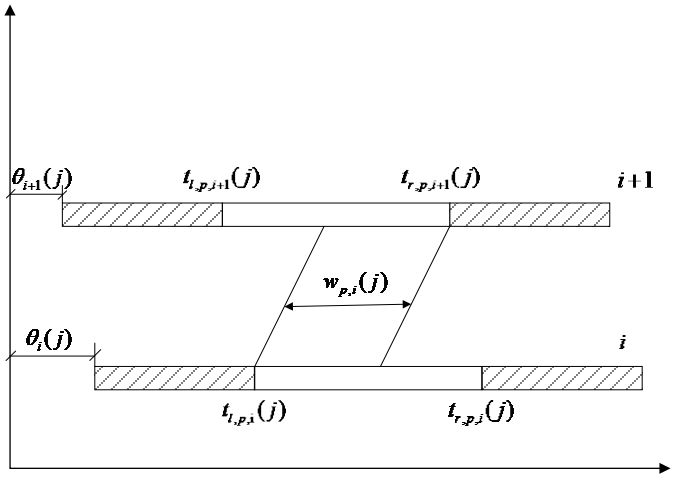
$$w_{r,p,i}(j) = min(t_{r,p,i}(j) + t_{i,i+1}(j), t_{r,p,i+1}(j)) \tag{4.4}$$

$$w_{l,p,i}(j) = max(t_{l,p,i}(j) + t_{i,i+1}(j), t_{l,p,i+1}(j)) \tag{4.5}$$

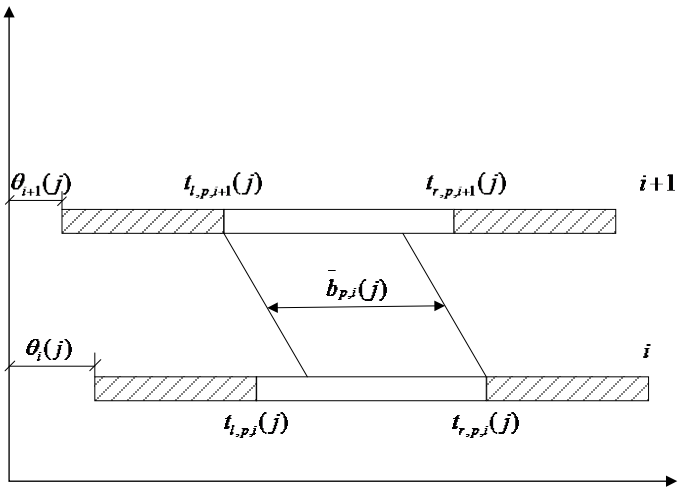
$$\overline{w}_{r,p,i}(j) = min(t_{r,p,i+1}(j) + t_{i,i+1}(j), t_{r,p,i}(j)) \tag{4.6}$$

$$\overline{w}_{l,p,i}(j) = max(t_{l,p,i+1}(j) + t_{i,i+1}(j), t_{l,p,i}(j)) \tag{4.7}$$

where $t_{l,p,i}(j)$ and $t_{r,p,i}(j)$ denotes the start and end of the green band for critical path p at intersection i, which are illustrated in Figure 4.3; $t_{i,i+1}(j)$ denotes the travel time from intersection i to intersection $i+1$.



(a) Illustration of green band for an outbound path-flow



(b) Illustration of green band for an inbound path-flow

Figure 4.3: Illustration of green band for both outbound and inbound directions

Figure 4.3 shows that the start and end of the green band for each path can be computed depending on the signal phase plan, offsets and green timings of intersections, as shown in Eq. (4.8) and Eq. (4.9).

$$t_{l,p,i}(j) = \sum_m \sum_n \beta_{m,p,i} * \varphi_{m,n} * g_{i,m}(j) + \theta_i(j) \tag{4.8}$$

$$t_{r,p,i}(j) = \sum_m \sum_n \beta_{m,p,i} * \varphi_{m,n} * g_{i,m}(j) + \sum_m \beta_{m,p,i} * g_{i,m}(j) + \theta_i(j) \tag{4.9}$$

where, $\beta_{m,p,i}$ is a binary variable to indicate the phase allocated to critical path p at intersection i (it equals to 1 if path p receives green time in phase m at intersection i, and 0 otherwise); $\varphi_{m,n}$ is a binary variable to determine the sequence of phases (it equals to 1 if phase m is before phase n, and 0 otherwise); $g_{i,m}(j)$ indicates the green time allocated to phase m for intersection i; and $\theta_i(j)$ denotes the offset of intersection i at jth control period, in seconds.

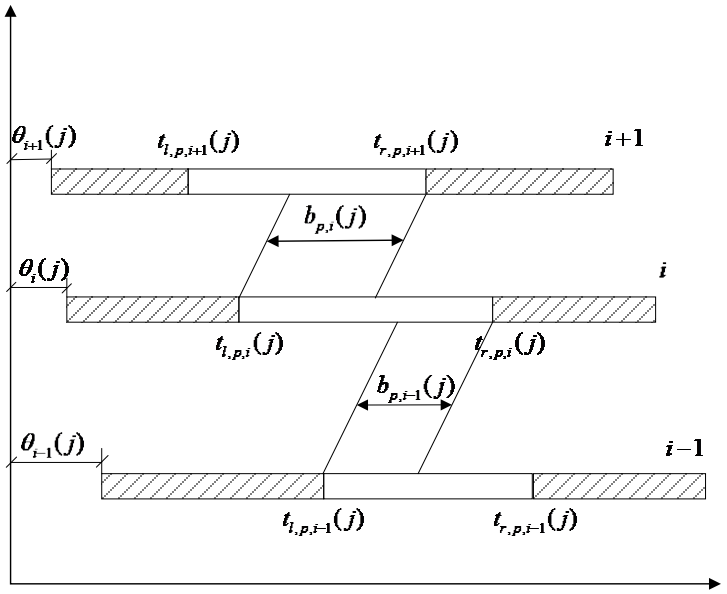
Note that if the green band for a path is not continuous between intersections along an arterial, as shown in Figure 4.4, vehicles may need to stop several times when traveling along this path. It will exert negative impacts on the effectiveness of the coordination system. To ensure the continuity of the green band for a path along multiple intersections. The following constraints need to be satisfied:

$$b_{l,p,i}(j) < b_{r,p,i+1}(j) - t_{i,i+1}(j) \tag{4.10}$$

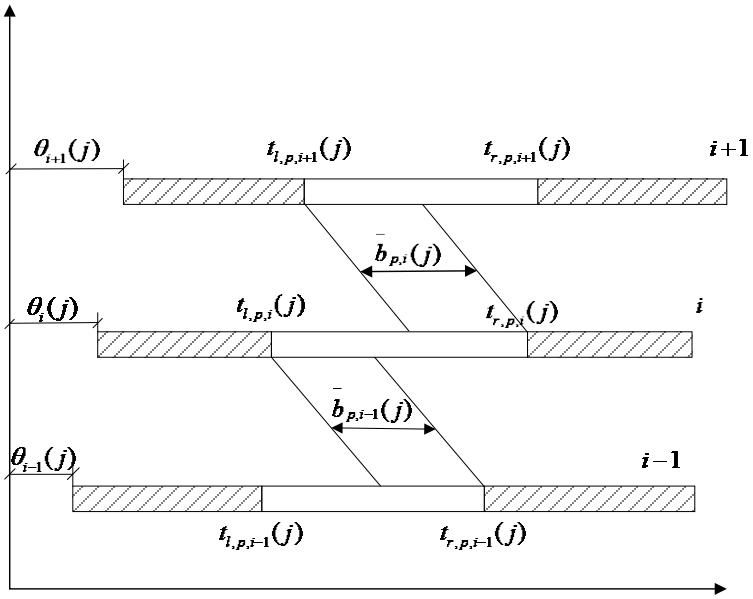
$$b_{r,p,i}(j) > b_{l,p,i+1}(j) - t_{i,i+1}(j) \tag{4.11}$$

$$\bar{b}_{l,p,i+1}(j) < \bar{b}_{r,p,i}(j) - t_{i,i+1}(j) \tag{4.12}$$

$$\bar{b}_{r,p,i+1}(j) > \bar{b}_{l,p,i}(j) - t_{i,i+1}(j) \tag{4.13}$$



(a) Non-continuous green band of an outbound critical path for three adjacent intersections



(b) Non-continuous green band of an inbound critical path for three adjacent intersections

Figure 4.4: Non-continuous green band of a critical path for three adjacent intersections

Eq. (4.14) is to ensure the offset change within a range at each control period.

$$\theta_{i-1}(j) - \Delta\theta_i \leq \theta_i(j) \leq \theta_{i-1}(j) + \Delta\theta_i \tag{4.14}$$

where $\Delta\theta$ denotes the maximum offset difference of two consecutive control periods for intersection i, in seconds.

4.3.2 Solution Algorithm

In this study, a solution algorithm depending on dynamic programming (DP) is developed to solve the proposed model shown in Eqs. (4.1)-(4.14). The main principle of the DP algorithm is that the decision problem is broken into a set of manageable decision stages and optimal decisions are made in a recursive manner. The basic elements for dynamic programming are stages, state variables, decision variables, and value functions. In this study, the stage is defined as the index of intersections, represented by $\{1,2,3,\dots,N_i\}$. The state variable is defined as the feasible new offset of each control period at each intersection, the feasible solution is denoted as follows:

$$S_i(j) = \{\theta_i(j-1) - \Delta\theta_i, \theta_i(j-1) - \Delta\theta_i + 1, \dots, \theta_i(j-1) + \Delta\theta_i\} \quad (4.15)$$

The value function, $V_i(\theta_i)$, can be calculated by Eq. (4.16)

$$V_i(\theta_i) = V_{i-1}(\theta_{i-1}^*) + B_i(\theta_i) \quad (4.16)$$

where, θ_{i-1}^* indicates the optimal offset determined at stage $i-1$; $\mathrm{B}(\theta_i)$ denotes the total green bandwidth of all critical paths at stage i when the offset involves θ_i (i.e. the total green bandwidth between intersection $i-1$ and intersection i), which could be calculated by Eq. (4.17).

$$B_i(\theta_i) = \sum_p \delta_{p,i-1,,m}\, \delta_{i,p,m}\omega_p(j) b_{p,i}(j) + \sum_p \delta_{p,i-1,m}\, \delta_{p,i,m}\bar{\omega}_p(j)\bar{b}_{p,i}(j) \quad (4.17)$$

where $\delta_{p,i,m}$ is a binary variable to identify the green time for paths at each intersection (it equals to "1" if green time is allocated to path p in phase m at intersection i, and "0" otherwise).

Based on the defined elements for dynamic programming, the solution algorithm is summarized as follows:

Step 1: define $i = 1$, $\theta_1(j) = 0$, and $V_i(0) = 0$;

Step 2: $i = i + 1$; update value function with Eq. (16) and determine the optimal value function

$V_i\left(\theta_i^*(j)\right) = min_{\theta_i(k)}\{V_{i-1}\left(\theta_{i-1}^*(j)\right) + B_i(\theta_i(j)) | \theta_i(j) \in S_i(j)\}$; Find the optimal solution at this stage, denoted as $\theta_i^*(j)$

Step 3: if $i < N_i$

go to step 2;

else

trace back to find the optimal solution for each stage.

4.4 NUMERICAL EXAMPLES

4.4.1 Experimental Design

To illustrate the applicability and evaluate the effectiveness of the proposed progression control system, an arterial simulation model with three intersections is established in VISSIM to perform experimental tests. The basic layout and signal plan of these three intersections are shown in Figure 4.5.

Key parameters of the proposed coordination control system are defined as follows:

1) The three intersections have the same cycle length which is 130 seconds;

2) The original offsets for those three intersections are 0 seconds, 6 seconds and 14 seconds, respectively;

3) The maximum offset difference between two consecutive control period is 6 seconds;

4) The average vehicle speed traveling along this arterial is 55 km/h;

5) The progression weighting factors for the critical paths at each control period is determined by the traffic flow along each path. More specifically, the weighting factor of one critical path is calculated by dividing the traffic flow along that path by the traffic demand of all determined critical paths.

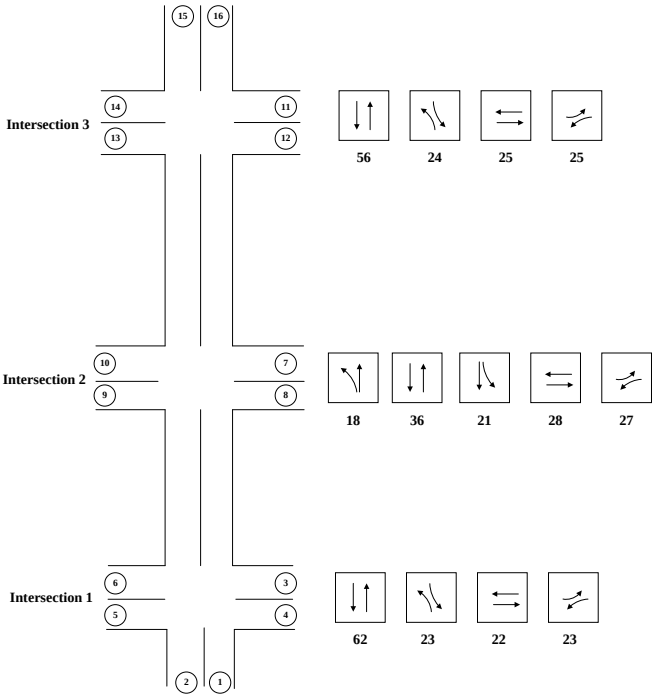


Figure 4.5: Illustration of the test arterial

The simulation runs 3,600 seconds and the CV penetration rate is assumed as 30% in VISSIM. In this simulation scenario, the length of control periods is set as 10 minutes, which means that the offsets of those three intersections change every 10 minutes based on the determined critical paths utilizing the CV trajectory data. The critical paths of each control period during the simulation are shown in Table 4.2.

Table 4.2: Critical path of each time period determined by CV trajectory data

Time (Sec)	0-600	600-1,200	1,200-1,800	1,800-2,400	2,400-3,000	3,000-3,600
Critical path	Node 1 to10	Node 1 to 10	Node 1 to 10	Node 1 to 10	Node 1 to 10	Node 1 to 10
	Node 1 to 14	Node 1 to 14	Node 1 to 14	Node 1 to 14	Node 1 to 14	Node 1 to 14
	Node 1 to 16	Node 1 to 16	Node 1 to 16	Node 1 to 16	Node 1 to 16	Node 1 to 16
	Node 15 to 3	Node 15 to 3	Node 15 to 3	Node 11 to 2	Node 11 to 2	Node 11 to 2
	Node 15 to 2	Node 15 to 2	Node 15 to 2	Node 15 to 2	Node 15 to 2	Node 15 to 2

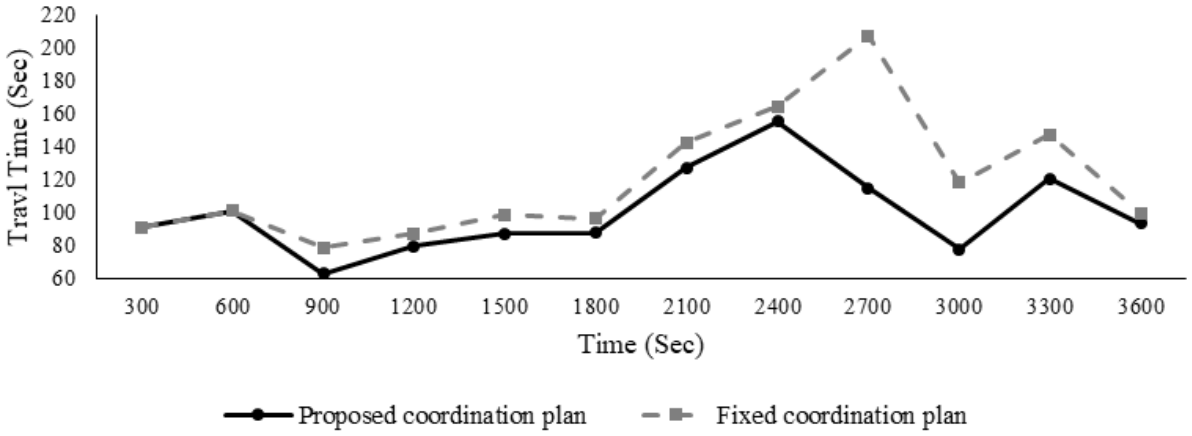
Table 4.2 shows that there are six most common critical paths during the simulation period. For simplicity of analysis, they are defined as path 1 to path 6 in this study, as shown in Table 4.3.

Table 4.3: Illustration of critical paths during the simulation period

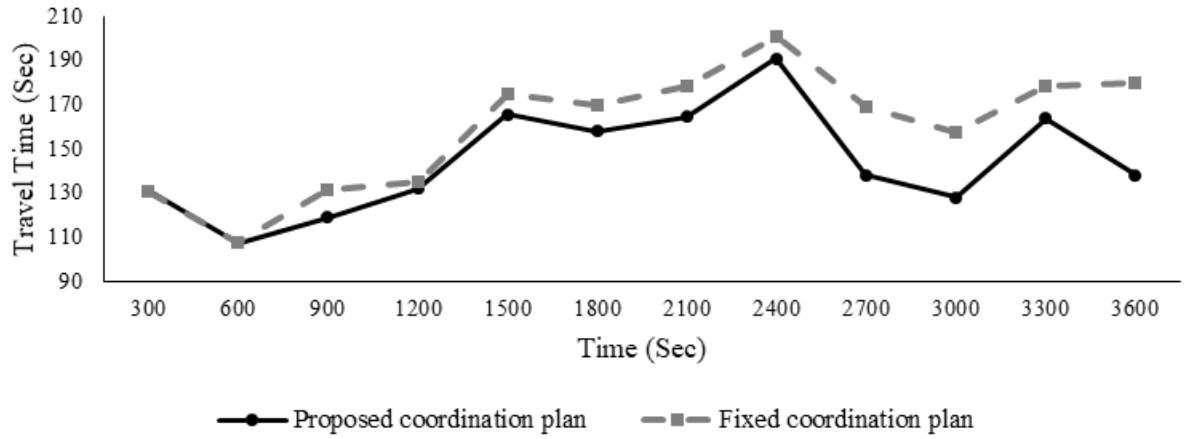
Critical path number	Critical path illustration
Path 1	Node 1 to Node 16
Path 2	Node 1 to Node 14
Path 3	Node 1 to Node 10
Path 4	Node 11 to Node 2
Path 5	Node 15 to Node 2
Path 6	Node 16 to Node 3

4.4.2 Results Analysis

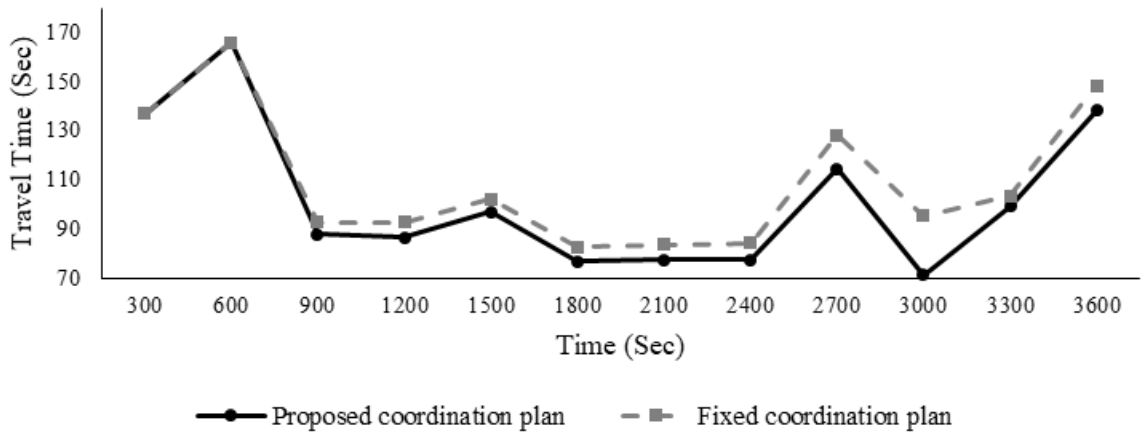
Figure 4.6 (a)-(f) shows the travel time of vehicles along the six critical paths under two scenarios. The first one is the benchmark in which the offsets of the three intersections are pre-set within a coordinated actuated signal system and the other is that the offsets of the three intersections are optimized depending on the proposed progression signal control algorithm.



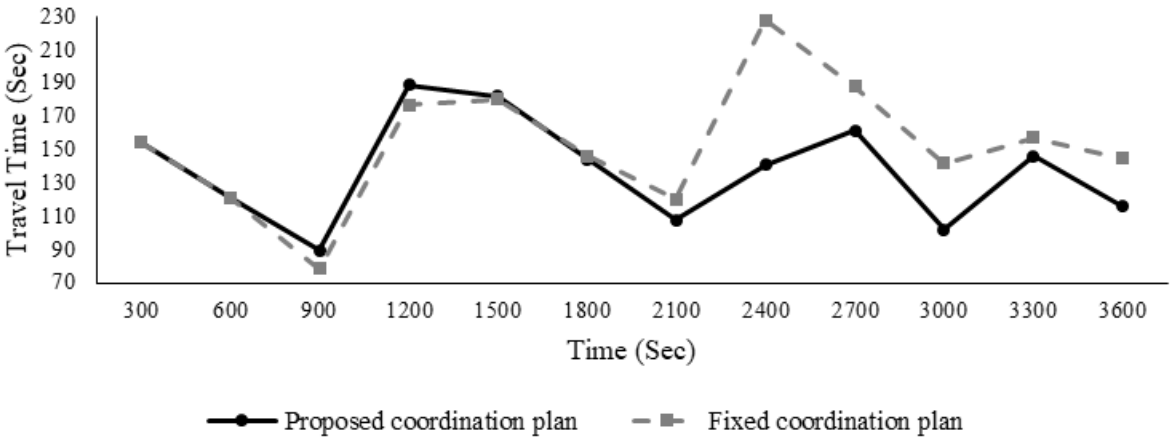
(a) The travel time of Path 1 depending on simulation time



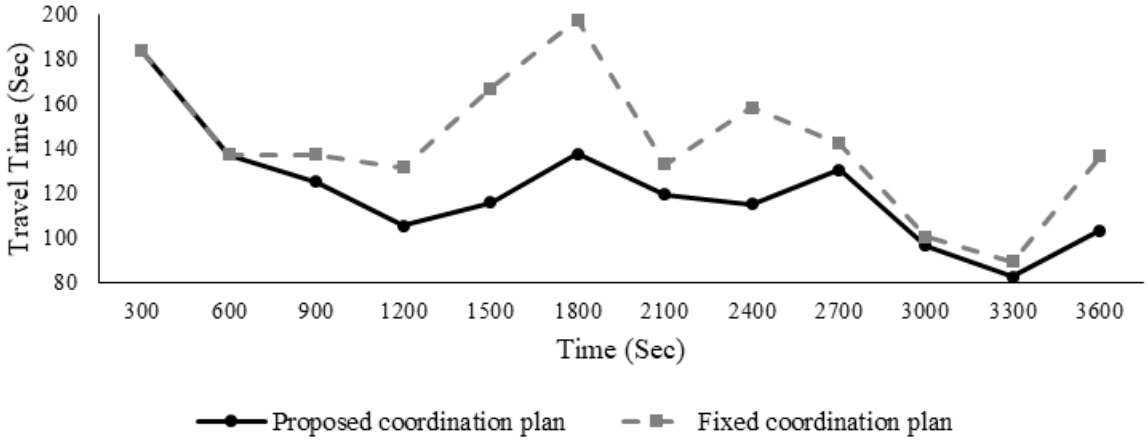
(b) The travel time of Path 2 depending on simulation time



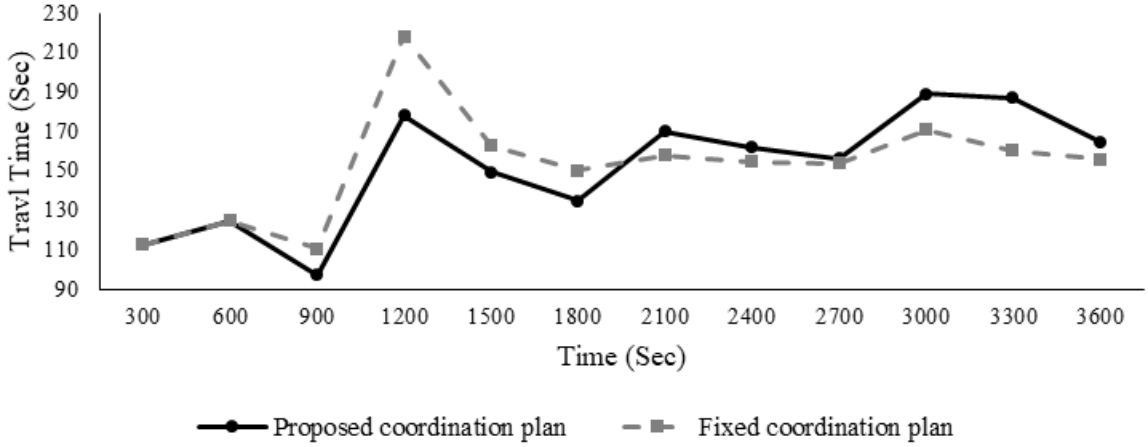
(c) The travel time of path 3 depending on simulation time



(d) The travel time of path 4 depending on simulation time



(e) The travel time of path 5 depending on simulation time



(f) The travel time of path 6 depending on simulation time

Figure 4.6: The travel time of critical paths depending on simulation time

It is observed that the proposed model can produce much lower travel time along path 1 compared to the fixed signal coordination plan, as evidenced by results in Figure 4.6 (a). This is because path 1 is always a critical path during the entire simulation period. The dynamic signal progression function can effectively adjust the offsets to provide progression priority for the movement along this path. Similar observations could be found for path 2 and path 3, as shown in Figure 4.6 (b) and (c), since the two paths are also the critical paths during the simulation period and they share most green phases with path 1.

Figure 4.6 (d) shows that the travel time difference between the two signal control systems is not significant before 1,800 seconds. The travel time along this path based on the proposed control system is even higher than that with the fixed coordination system. Such a phenomenon can be attributed to the fact that path 4 is not the critical path during the period before 1,800 seconds. Although the traffic demand of this path in this period is low, the progression offered to other critical paths may cause higher travel time of it. However, progression is provided to this path due to high traffic demand after 1,800 seconds. Thus, the travel time is lower compared to the fixed signal coordination plan. Path 5 has a similar fluctuation pattern; while it is not the critical path before 600 seconds, but the progression is provided to this path after 600 seconds.

As shown in Figure 4.6 (e), the travel time along path 6 with the proposed system is lower than that with the fixed coordination system before 1.800 seconds because the traffic demand along this path is high and thus the movement along this path receives the progression. But after that time no progression is provided to this path due to lower traffic demand. Therefore, the travel time is higher compared with the fixed coordination system.

In conclusion, the proposed coordination control system exhibits a better performance in reducing vehicle travel times along those critical paths. To further evaluate the proposed system, other operational performances with different measurement of evaluations (MOEs) are tested. Table 4.4 summarizes the average delay and the average number of stops for both critical paths and the entire network. It shows that the proposed system, as expected, can reduce the average delay and the average number of stops. In detail, the average delay and the average number of stops reduced by about 13% and 10%, respectively. It also shows that the proposed system can reduce the average delay and number of stops for the entire network. This may due to the fact that the traffic volume along those critical paths is relatively high, reducing the delay of those critical paths could result in delay reduction for the entire network significantly.

Table 4.4: Arterial and network performance with various control plan

MOEs	Fixed coordination plan	Proposed coordination plan
Average critical path delay (Sec)	84.03	73.84
Average number of stops for critical paths	1.63	1.47
Average network delay (Sec)	54.06	50.10
Average number of stops for network	1.00	0.97

5.0 SMART CONTROL FOR IMPROVING BOTH SAFETY AND OPERATIONAL BENEFITS

5.1 INTRODUCTION

Besides the mobility benefit brought by the dynamic signal progression, as introduced in Chapter 4.0, how to effectively design traffic signal control systems to improve the safety of urban arterials at the same time has long been recognized as a vital issue by the traffic community. With the recent advancements in wireless communications and computing techniques, CV technology has reached a level of maturity and further sheds new light on real-time signal control. V2V and V2I communication platforms allow CVs and roadside infrastructure to exchange real-time traffic data (Cho et al., 2009). Deployments of such technology have shown great promise in crash avoidance, injury prevention, and congestion relief at intersections. With a high penetration rate of CVs within the traffic, it is possible to remove traffic signals completely at intersections and only utilize CV data as input of signal control. However, it can be expected that both CVs and HVs will co-exist on the road in a long time period (Huang et al., 2018). Hence, how to deal with such mixed traffic patterns, with fusion of CV information and traffic sensor data, in both safety and mobility control functions become an urgent task at the current stage.

In daily operations, signalized intersections may experience two types of crashes: side-angle crash and rear-end crash. As reported in the literature (Gazis et al., 1960), trapping vehicles within the dilemma zone where vehicles can neither stop before or pass the stop line safely, is one of the most common causes that lead to side-angle crashes. By extending the yellow time, the dilemma zone can be eliminated. However, a longer yellow time may create an “indecision” zone where drivers are not able to make the right pass/stop decisions. For protecting drivers from becoming trapped in dilemma/indecision zones, existing efforts include both proactive and reactive control methods. The core logic of proactive control is to either provide advanced warning message to drivers to reduce their speed (Zimmerman et al., 2012; McCoy and Pesti, 2003) or/and adjust signal green times (early termination or extension) before max-out to prevent trapping vehicles into dilemma zones (Bonneson et al., 1977; Zegeer, 1977). Different from proactive control, reactive protection strategies aim to prevent side-angle crashes when vehicles fail to stop safely before a signal alters to the following phase. All-red extension is a commonly used reactive protection function that offers extended all-red intervals to accommodate red-light-running vehicles (Zhang et al., 2012). Based on field collected or simulated data, existing studies have reported the effectiveness of both proactive and reactive dilemma zone projection strategies in reducing red-light-running vehicles (Jahangiri et al., 2016; Gates, 2007) and number of side-angle crashes at intersections (Hurwitz et al., 2016; Chang et al., 2012). However, preventing rear-end crashes with real-time signal control, in contrast, remains a challenging issue in the literature (Park et al., 2016). Although field experiments (Hurwitz et al., 2016) showed signal green-extension has some potentials in reducing rear-end crash rate, a more effective strategy on affecting driver behaviors has not been fully studied yet.

In regard to improving arterial mobility performance, existing operational measures can be divided into two categories: real-time signal control and green light optimized speed advisory (GLOSA). In the first category, the main principle is to make real-time adjustments to intersection signal

timing according to collected traffic data. Actuated and adaptive control models (15) have been widely studied and implemented in practice. Under a CV environment, enriched real-time data from CV trajectories can also enhance the prediction/estimation of traffic flow so as to improve systems' mobility performance (Feng et al., 2015; Beak et al., 2017; He et al., 2012; Priemer et al., 2009; Lee et al., 2013; Goodall et al., 2013; Guler et al., 2014; Yang et al., 2016). The second category, GLOSA, is to suggest advisory speeds for approaching vehicles to stay in the progression band, where the arterial can be running with either a pre-timed signal coordination plan or coordinated-actuated signal with green extensions. Notably, the advisory speeds can be passed to vehicles by V2I and variable speed limit (VSL) technologies under a mixed CV and HV traffic pattern. One such application named Application for the Environment: Real-Time Information Synthesis Program was initiated by USDOT (24). With CV technology, each vehicle can receive Signal Timing and Phase (SPaT) data from roadside unite (RSU) and revise its own speed profile within the control boundary (Xia et al., 2013). Also, such speed harmonization can help reduce fuel consumptions by 7% to 13% (Katsaros et al., 2011). Unlike the CV technologies, which targets for individual drivers, VSL signs can broadcast advisory speeds to all incoming vehicles.

In summary, despite promising efforts that have been placed on improving either intersection safety or arterial mobility, existing control models for these two purposes are often carried out with separate devices and sensors. Integration of one set of equipment to concurrently satisfy both safety and mobility needs has not been well addressed yet. In recent years, Park et al. (2017) proposed a system that integrates the dilemma zone protection with a VSL-based speed harmonization function. Their simulation experiments confirmed the effectiveness of the proposed system on offering protection to red-light-running vehicles, and on improving traffic mobility with respect to fewer stops, reduced stop delays, and less fuel consumption. The shared utilization of deployed hardware devices also allows responsible agencies to best use available real-time signal operational resources. This study will follow the same line and extend the system's capability into the CV environment. The contributions of the new system are three-fold: 1) it fuses real-time data from both roadside microwave sensors and CVs to estimate queue evolutions at intersections; 2) it integrates efficiency improvement function into the dilemma zone protection system by optimizing a signal coordination plan and vehicle advisory speeds; and 3) it offers solutions to prevent potential rear-end crashes at intersections.

5.2 SYSTEM ARCHITECTURE

Figure 5.1 shows the overview of the proposed system architecture that contains five key components: 1) *long-range microwave detector* for tracking the speeds and locations of all vehicles within the detection zone; 2) *RSU* for supporting V2I communications; 3) *VSL signs* for providing advisory speed for incoming HVs to ensure safe stops or smooth progressions along the arterial; 4) *in-cabinet computer* for processing collected data, operating embedded algorithms, and making control decisions; and 5) *signal controller* for providing current SPaT and receiving instructions for an all-red extension or offset adjustment from an in-cabinet computer.

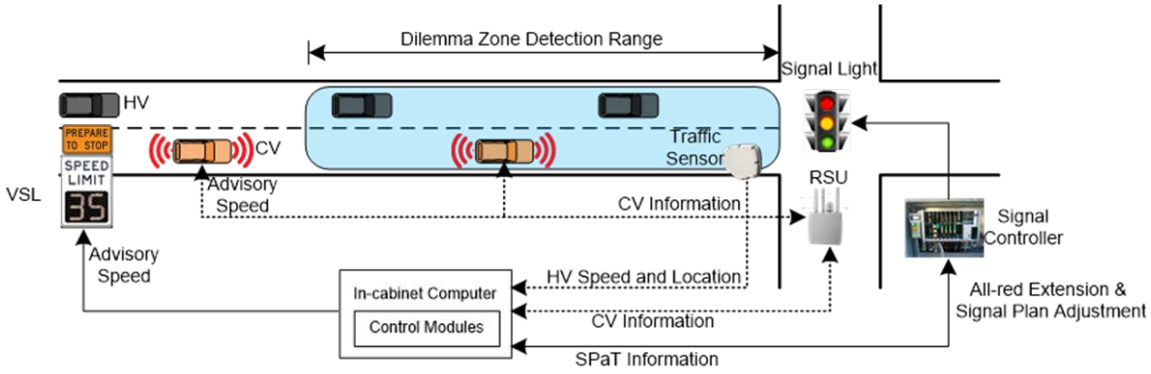


Figure 5.1: Overview of the proposed integrated signal control system

Figure 5.2 illustrates the data flowchart of the proposed system. The long-range microwave sensor can detect the speeds and locations (i.e., distances to the stop line) of both HVs and CVs within its detection zone. However, lane-based information (how vehicles distributes among different lanes) is usually not obtainable. RSU will collects CV trajectories in real time and send advisory speeds back to CVs when necessary. The signal controller would provide SPaT information to determine when and how to activate control modules including dilemma zone protection, real-time queue evolution estimation, real-time signal control, and rear-end crash prevention. Detailed information regarding the design of each module will be introduced in the following section.

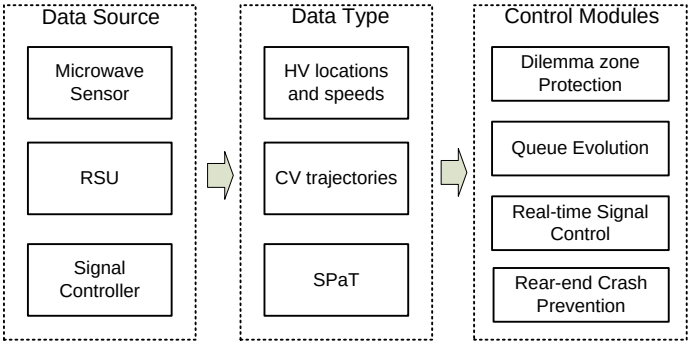


Figure 5.2: Data flow of the proposed system

5.3 CONTROL MODULES AND MODEL DEVELOPMENT

Module 1: Dilemma Zone Protection

This module is used to monitor all vehicles within the detection zone and provide an all-red extension to those trapped in the dilemma zone. In a previous study, Park et al. (2017) utilized the following logistic regression expression to predict each vehicle's passing probability at the onset of a yellow interval:

$$P_{pass}(i,t) = \frac{1}{1-e^{-\beta_0-\beta_1 v_i(t)-\beta_2 d_i(t)}} \tag{5.1}$$

where $v_i(t)$ and $d_i(t)$ are the speed and location (distance to the stop line) of vehicle i at time t; β_0, β_1, and β_2 are parameters calibrated with field data. When $P_{pass}(i,t) \geq 0.5$, vehicle i is identified as passing vehicle; otherwise it is identified as stopping vehicle.

However, in real-world applications, observations showed that some vehicles may change their pass/stop decision during the yellow interval. Hence, prediction executed at the onset of a yellow interval may fall short of accounting for such behavior. To address this issue, this study aims to predict vehicles' passing probability at ε seconds before the end of a yellow interval, where ε indicates the time needed for data transition and all-red extension activation. An extension of Eq. (5.1) is formulated as follows:

$$P_{pass}(i,t_\varepsilon) = Max(\frac{1}{1-e^{-\beta_0-\beta_1 v_i(t_\varepsilon)-\beta_2 d_i(t_\varepsilon)}}, \delta_i(t_\varepsilon)) \tag{5.2}$$

where $\delta_i(t_\varepsilon)$ is a binary variable which indicates whether vehicle i intends to accelerate:

$$\delta_i(t_\varepsilon) = \begin{cases} 1 & if\ v_i(t_\varepsilon) \geq v_i(t_\varepsilon - 1) \\ 0 & o.w. \end{cases} \tag{5.3}$$

By introducing $\delta_i(t)$, the system can monitor whether a vehicle changes its initial stopping decision to passing during a yellow interval. Given the vehicles' passing probabilities, speeds and locations, the system can estimate their required passing time by $d_i(t_\varepsilon) / v_i(t_\varepsilon)$. Then the required all-red extension time, ARE, can be calculated by:

$$ARE = \max_i \{\frac{d_i(t)}{v_i(t)} - \varepsilon - AR + \sigma\} \tag{5.4}$$

where AR is the pre-set all-red time and σ is additional all-red protection time to overcome potential estimation error of vehicle passing time.

Module 2: Queue Length Estimation

As accumulated queueing vehicles during a red interval can greatly affect the effectiveness of signal progression and increase the potential for rear-end collisions with short sight distance, this module aims to predict the lane-based queue evolution at the onset of the green signal and estimate the required clearance time. Utilizing the microwave sensor and V2I technology, Figure 5.3 shows available real-time information within the detection range (e.g., 900 feet), including trajectories of CVs and HVs. Notably, microwave sensors will detect the speeds and locations (distances to the stop line) of HVs but cannot identify their lane distribution.

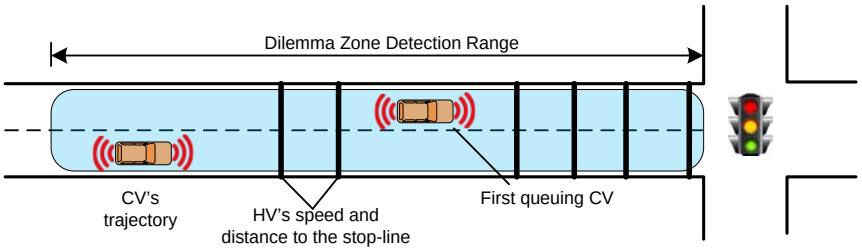


Figure 5.3: Vehicle trajectory information within the detection zone

This study divides the detection range into queueing zone and arriving zone. The length of a queueing zone can be estimated by identifying the location of queueing vehicles which have zero speeds. However, due to the lack of HV lane distribution data, the first step to estimate lane-based queue evolution is to assign queueing HVs to different lanes. As the locations of CVs provide direct observations of queueing vehicles, this model divides the queuing zone into a set of cells by treating locations of CVs and stop lines as boundaries. Also defining h as the average vehicle gap in the queue, which is calibrated by field data, HVs can be distributed in each cell by the following optimization model based on their locations, $d_i(t)$.

$$\begin{aligned} & Min \sum_i (g_i - h)^2 \\ & s.t.\ g_i \le d_k(t)\varphi_{kj} - d_i(t)\varphi_{ij} \quad \forall i\,;\, \forall k > i \\ & \sum_j \varphi_{ij} = 1 \qquad\qquad\qquad\quad \forall i \end{aligned} \tag{5.6}$$

where g_i denotes the distance between HV i and its following vehicle; φ_{ij} equals "1" if HV i is assigned to lane j. Notably, the first set of constraints of the optimization model (5.6) determines the actual gaps of vehicles on each lane, the second set of constraints is to ensure each vehicle can only be assigned to one lane, and the objective function is to minimize total square of differences between actual gap and average gap. Then the queue length on each lane, q_j, can be obtained. If there is no CV in the queuing zone, the lane assignment of HVs becomes an open problem which has no unique solutions. In such case, for safety concerns, we assume the queue lengths of all lanes equal $\max(d_i(t)|\ \forall\ i)$.

As the physical queue length (distance from the end of queue to stop line) will still increase for a few seconds after the onset of a green signal, the required queue clearance time, τ_j, on each lane j can be estimated by solving the following equation:

$$\upsilon_j \tau_j = q_j + \lambda_j \tau_j \tag{5.7}$$

where υ_j and λ_j are the discharging shockwave speed and arrival rate of lane j calculated from previous signal cycles.

Module 3: Signal Coordination and Speed Harmonization

This module aims to improve the operational efficiency of the arterial by better coordinating intersections and harmonizing vehicles' speeds. Given the estimated queue clearance time, the

control objective of this module is to minimize the stopped delay of upstream arriving vehicles due to initial queues at each intersection, along both outbound and inbound direction. For convenience of discussion, offset of intersection *i*, denoted as θ_i, is defined as its green onset time difference compared with its downstream intersection along the outbound direction. Then the stopped delay of the first arriving vehicle at intersection *i*, along outbound direction, $\zeta_{out,i}$, can be estimated by:

$$\zeta_{out,i} = \max(\theta_i + \max_{j \in \Psi_{out}(i)} \{\tau_j\} - \frac{L_{out,i} - \max_{j \in \Psi_{out}(i)} \{q_j + \lambda_j \tau_j\}}{v_{out,i}}, 0) \tag{5.8}$$

where $v_{out,i}$, and $L_{out,i}$ are the advisory speed and link length between intersection *i* and its downstream intersection along the outbound direction, and $\Psi_{out}(i)$, denotes the lane group along the outbound direction at intersection *i*. Similarly, the stopped delay of the first arriving vehicle along the inbound direction, $\zeta_{in,i}$, can be estimated by:

$$\zeta_{in,i} = \max(-\theta_i + \max_{j \in \Psi_{out}(i)} \{\tau_j\} - \frac{L_i - \max_{j \in \Psi_{out}(i)} \{q_j + \lambda_j \tau_j\}}{v_{in,i}}, 0) \tag{5.9}$$

Then the real-time optimization model for Module 3 can be formulated as follows:

$$\begin{aligned} & Min \sum_i (\zeta_{out,i} + \zeta_{in,i}) \\ s.t.\ & \theta_i^- - \Delta\theta \le \theta_i \le \theta_i^- + \Delta\theta \qquad \forall i \\ & v_{in,i}^- - \Delta v \le v_{in,i} \le v_{in,i}^- + \Delta v \qquad \forall i \\ & v_{out,i}^- - \Delta v \le v_{out,i} \le v_{out,i}^- + \Delta v \qquad \forall i \end{aligned} \tag{5.10}$$

where decision variables θ_i^-, $v_{out,i}^-$, and $v_{in,i}^-$ denote the offset, outbound direction advisory speed, and inbound advisory speed in the last signal cycle; $\Delta\theta$ and Δv denote the maximal allowable differences of offset and advisory speed, respectively, between consecutive cycles. Notably, due to the limited searching space of the optimal solution, the above optimization model can be easily solved in real time.

Module 4: Rear-end Crash Prevention

For preventing potential rear-end crashes at intersections, this module aims to address the following three cases:

Submodule 1 – vehicles are arriving with insufficient sight distance while the intersection has an uncleared initial queue after the onset of green.

To deal with this case, the proposed system will utilize the VSL sign to offer the advisory speed for HVs to prepare to stop. Given the lane-based speed evolutions estimated by Module 2 and the location of VSL sign, d_{VSL}, the advisory speed can be determined by the following expression:

$$vsl = \frac{d_{VSL} - \max_j \{\upsilon_j \tau_j\}}{\tau_j} \tag{5.11}$$

For CVs, the advisory speed will be sent via V2I communication channels.

Submodule 2 – vehicles are arriving with insufficient sight distance while intersection has uncleared initial queue after onset of red.

To deal with this case, the proposed system will utilize the VSL sign to offer advisory speed for HVs to prepare to stop. The advisory speed will be changed over time and is calculated based on the current queue length $q(t)$:

$$d_{VSL} - q(t) = vsl(t) * p_t + \frac{vsl(t)^2}{2a} \tag{5.12}$$

Notably, the current queue length $q(t)$ equals the last stopped vehicle's distance to the stop line. For CVs, the advisory speed will be sent via V2I communication channels.

Submodule 3 – some vehicles within the detection zone are predicted to be stopping during yellow and all-red time.

Recall that Eq. (5.1) to predict the passing/stopping behavior of each detected vehicle at the onset of a yellow signal. By identifying the location, denoted as d_s, and speed, denoted as v_s, of the last stopping vehicles, the advisory speed for ensuring a safe stop can be estimated by solving the following equation:

$$d_{VSL} - d_s = vsl * p_t + \frac{vsl^2 - v_s^2}{2a} \tag{5.13}$$

where a is the field-observed deceleration rate and p_t is the drivers' perception time

5.4 SYSTEM CONTROL LOGIC AND ACTIONS

To integrate those four control modules for concurrently improving efficiency and safety performance of arterial intersections, Figure 5.4 illustrates the system control logic by choosing different control objectives during yellow, red and green time. By defining a set of control scenarios the system may encounter in practice, this section will discuss the control actions in response to each scenario.

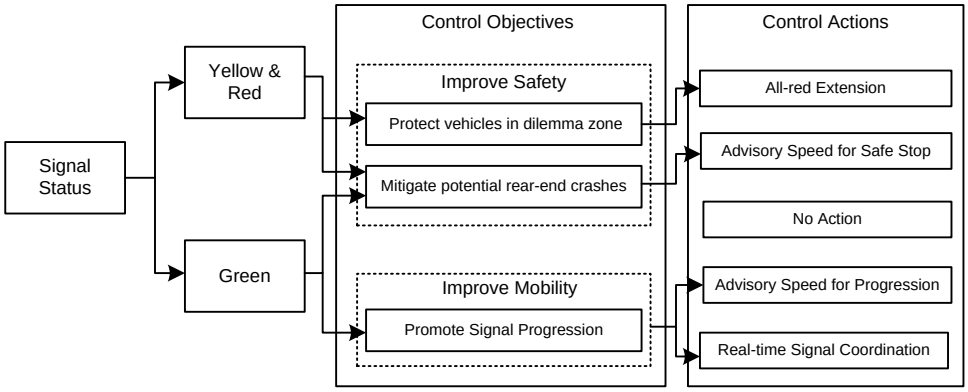


Figure 5.4: System control actions based on traffic signal status

Actions During Yellow and Red Intervals

When vehicles are arriving during yellow and red intervals, the proposed system may encounter the following two scenarios.

Scenario 1: vehicles are arriving during yellow and all-red time

Vehicles arrived during yellow and all-red time may neither stop before or pass the intersected area safely due to two reasons: 1) the yellow settings are too short and the vehicles are trapped in a dilemma zone; and 2) the drivers are in an indecision zone and make wrong "pass" decisions. Hence, the proposed system will take the control actions of "All-red Extension" and "Advisory Speed for Safe Stop" following several key steps:

Step 1: collect the speed and location of all vehicles within the dilemma zone detection range at ε seconds before the end of yellow signal;

Step 2: calculate the vehicles' passing probability using Eq. (5.2) and activate the submodule 3 of module 3. If all probabilities are below 0.5, stop; otherwise move to Step 3;

Step 3: use Eq. (5.4) to estimate the required all-red extension time by calculating the largest passing time of vehicles. If the obtained *ARE* is zero, stop; otherwise, extend the all-red time by *ARE* seconds.

Scenario 2: vehicles are arriving during the red interval

Vehicles arriving during the red signal interval must join the end of stopping queues at intersections. Due to improper alignment design of intersections or dark/light conditions, the approaching vehicles may not have sufficient sight distance to observe the stopping queue ahead. Hence, if the vehicles are travelling at a high speed, it can lead to potential rear-end collisions. To reduce such risk, the proposed system will take the control action of "Advisory Speed for Safe Stop" with the following steps:

Step 1: activate module 2 to estimate the queue length at the beginning of the red interval; keep updating the queue length based on the information of arriving vehicles;

Step 2: activate the submodule 2 of module 3.

Actions During Green Interval

When vehicles are arriving during yellow and red intervals, the proposed system may encounter the following three scenarios:

Scenario 4: vehicles are arriving at the beginning of the green interval

Vehicles arriving at the beginning of the green interval may be delayed since the intersection needs to take a few seconds to discharge vehicles queued during the red interval. Without sufficient sight distance, vehicles traveling at high speeds may crash with the vehicles stopping ahead. Under such a scenario, the proposed system will activate submodule 1 of module 3 for providing advisory speed to the approaching vehicles.

Scenario 5: vehicles are arriving during the green interval

During the signal green time, the main control objective is to facilitate signal progression along the arterial. For such a need, the proposed system will take the actions of "Advisory Speed of Progression" and "Real-time Signal Coordination" with the key steps described as follows:

Step 1: If the last time of adjusting signal offsets happened within five minutes ago, set $\Delta\theta$ as zero; otherwise; set $\Delta\theta$ as five seconds;

Step 2: estimate the queue clearance time using Eq. (5.7);

Step 2: optimize the offsets and vehicles' advisory speeds using the optimization model listed in Eq. (10);

Step 3: provide advisory speed to HVs and CVs by VSl and V2I, respectively.

Scenario 6: vehicles are arriving at the end of the green interval

Vehicles arriving at the end of the green interval may encounter a yellow signal when they get close to the intersection. Hence, the following steps of action shall be taken:

Step 1: collect the vehicles' speeds and locations;

Step 2: estimate the vehicles' arrival time at the signal stop line. If the arrival time is within a yellow and all-red interval, take the same actions as Scenario 1; otherwise, take the same actions as stated in Scenario 5.

5.5 NUMERICAL EXAMPLES

Simulation Platform Set-up

For evaluating the proposed system's capability on improving both arterial safety and mobility, this study selects a segment on Redwood Road in Salt Lake City. As shown in Figure 5.5, the arterial segment includes five intersections and it is a part of the CV corridor operated by UDOT.

All intersections are installed with DSRC RSUs for supporting V2I communications. The prevailing speed is set as 45mph and yellow timing is three seconds.

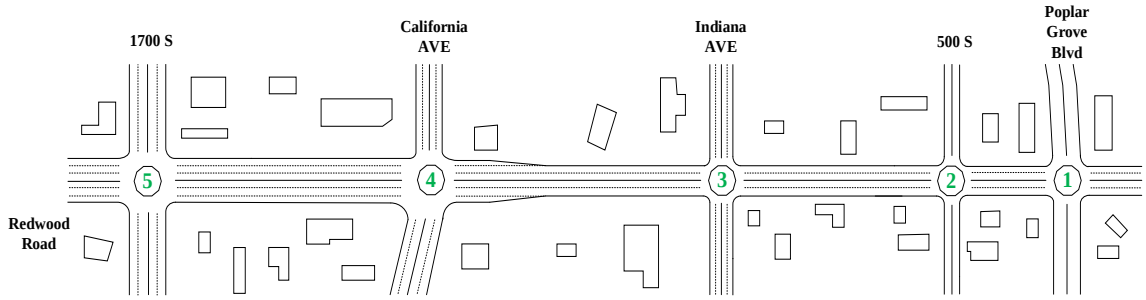


Figure 5.5: Overview of the study site

Table 5.1 summarizes the intersection turning volumes collected at the arterial on Feb. 15, 2018. The collected data are further used for simulation calibrations. In this study, our research team employed VISSIM as the unbiased simulation tool for system evaluations. Through a VB.NET developed VISSIM-COM interface, Figure 5.6 shows the architecture of the VISSIM platform. By defining two vehicle groups, one represents HVs and the other represents CVs, the program detects and records the real-time locations and speeds of vehicles within the detection range. The obtained trajectory data, along with Signal Phase and Timing (SPaT) information, will be sent to the computational program by an interval of one second. Depending on the current signal status (i.e., red, green, or yellow and all-red), the embedded modules will take proper actions and provide feedback control to the signal controllers and vehicles via VISSIM-COM.

Table 5.1: Summary of the collected intersection volumes (veh/hr)

Intersection	EB			WB			NB			SB		
	L	**T**	**R**	**L**	**T**	**R**	**L**	**T**	**R**	**L**	**T**	**R**
	116	36	152	212	40	24	164	508	104	28	664	128
1	304			276			776			820		
	32	24	44	84	4	148	52	700	8	36	632	56
2	100			236			760			724		
	188	132	112	128	168	164	32	704	68	176	700	112
3	432			460			804			988		
	140	316	100	192	184	80	164	716	208	192	748	140
4	556			456			1088			1080		
	200	124	76	68	232	268	280	672	116	116	844	60
5	400			568			1068			1020		

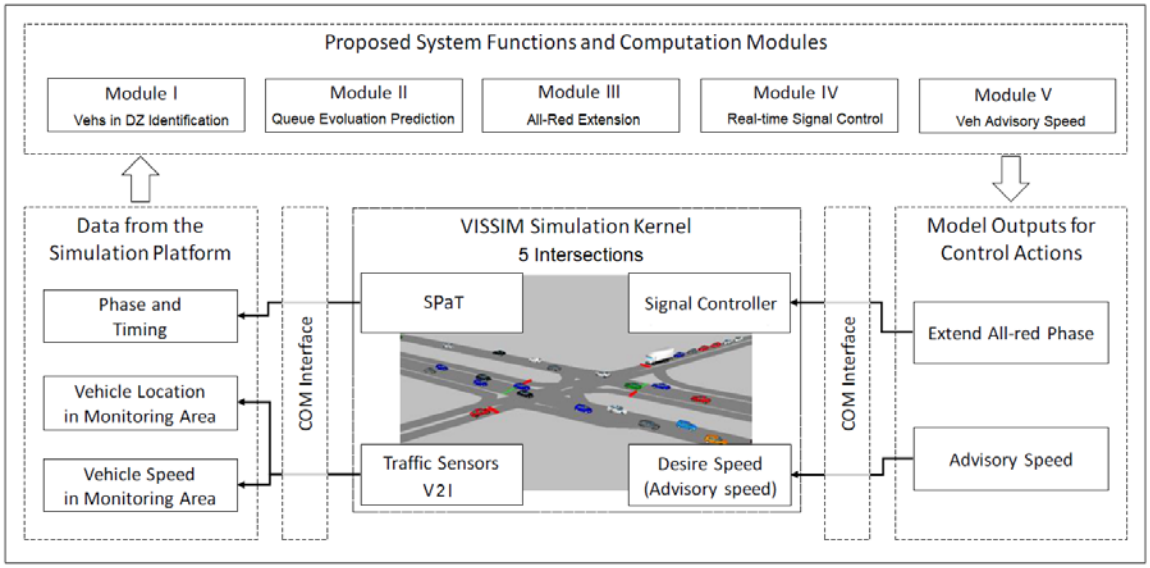


Figure 5.6: Overview of the VISSIM simulation platform

Notably, one of the key features of the proposed system is providing advisory speeds to both CVs and HVs through V2I communication and VSL, respectively. In this study, we assume that HVs may or may not follow the VSL, represented by a compliance rate, but CVs will follow the advisory speed they received. In the simulation platform, such actions are replicated by the following steps:

Step 1: break the simulation process;

Step 2: identify the type of a coming vehicle. If it is a CV, change its speed according the output of control modules; otherwise move to step 2;

Step 3: use a random number generator to determine whether the HV will follow the instruction of VSL;

Step 4: change the HV's freeway flow speed if it is a complying vehicle; and

Step 5: continue the simulation.

Measures of Effectiveness (MOEs)

In the simulated scenarios, 10% of vehicles are assumed to be CVs and the rest of them are HVs. Also, HVs' compliance rate to the VSL sign is assumed be 40%. Since the proposed system is unique in adopting one set of hardware to support both safety and mobility control functions, the measures of effectiveness shall cover both aspects.

Safety MOEs:

- Average number of vehicles trapped in the dilemma zone per signal cycle;
- Average number of potential side-angle crashes per signal cycle measured by vehicle trajectories;
- Average number of potential real-end crashes per signal cycle measured by the number of hard-braking vehicles (deceleration rate > 10ft/s^2); and
- Average number of red-light-running vehicles per signal cycle.

Mobility MOEs:

- Average number of stops;

- Average of vehicle delay where delay is defined as the difference between actual travel time and free flow travel time (using the original desired speed).

Based on the MOEs defined above, this study tests the following scenarios for comparisons:

- *Base Scenario*: the arterial is under the control of pre-timed traffic control system;
- *Scenario 1:* the arterial intersections are equipped with a dilemma zone protection system (DZPS); and
- *Scenario 2:* the arterial is under the control of the proposed system.

Results Analysis

By simulating the arterial network over a two-hour period, Table 5.2 summarizes the resulting MOEs under different scenarios. By comparing the safety MOEs between Base Scenario and Scenario 1, it can be observed that implementing DZPS can greatly reduce the average number of potential side-angle crashes (-84.7%). However, the performance differences between these two scenarios, in terms of average number of vehicles in the dilemma zone, average number of potential rear-end crashes, and average number of red-light-running vehicles, are not significant. This is due to DZPS activating the all-red extension function once some vehicles cannot safely pass the intersection. Hence, taking the all-red extension action can greatly reduce the potential for side-angle crashes. However, since the approaching vehicles will not receive advance notice regarding the signal status in both scenarios, the performance on the other safety MOEs are quite similar. In contrast, Scenario 2 with the proposed system can greatly outperform the Base Scenario in reducing vehicles in the dilemma zone (-22.2%), number of side-angle crashes (-84.7%), number of rear-end crashes (-55.5%), and number of red-light-running vehicles (-20.4%). Such comparisons can prove the effectiveness of the proposed functions on providing advisory speed to stop.

Regarding the mobility MOEs, this study examines both average number of stops and average vehicle delay over the entire arterial. The comparison between Base Scenario and Scenario 1 reveals that implementing DZPS will slightly deteriorate the traffic congestion, evidenced by the increased number of stops (+2.3%) and delay (+9.4). This is caused by the intersection capacity reduction due to granted all-red extensions. However, with the proposed signal coordination and speed harmonization functions, Scenario 2 can greatly improve the arterial's mobility performance.

Table 5.2: The system's safety and mobility performance under different scenarios

MOEs		**Base Scenario**	**Scenario 1**	**Scenario 2**
Safety MOEs	Ave # of vehicles in the dilemma zone	1.13	1.12 (-0.6%)	0.88 (-22.2%)
	Ave # of potential side-angle crashes	0.85	0.13 (-84.7%)	0.13 (-84.7%)
	Ave # of potential rear-end crashes	3.21	3.19 (-0.6%)	1.43 (-55.5%)
	Ave # of red-light-running vehicles	0.49	0.49 (-0.0%)	0.39 (-20.4%)
Mobility MOEs	Ave # of stops along the studied arterial	25.8	26.4 (+2.3%)	16.7 (-35.3%)

	Ave vehicle delay (seconds) over the entire arterial	135.7	148.9 (+9.4%)	112.7 (-17.0%)

Sensitivity Analysis on CV Penetration Rate

As CV penetration rate also plays a key role in affecting the proposed system's performance, this study further conducts a sensitivity analysis to analyze its impact on reducing the probability of occurring rear-end crashes. As shown in Figure 5.7, when the CV rate is below 20%, the resulting average number of potential rear-end crashes is around three vehicles per signal cycle. However, after the rate reached 20%, the corresponding number has been greatly reduced to 1.9 vehicles per signal cycle. By further examining the simulation animations, it has been found that the speed control of CVs can concurrently affect the speed of HVs as they are sharing the roadway. Hence, one can treat 20% as a critical CV rate to maximize the system's benefit in the studied case.

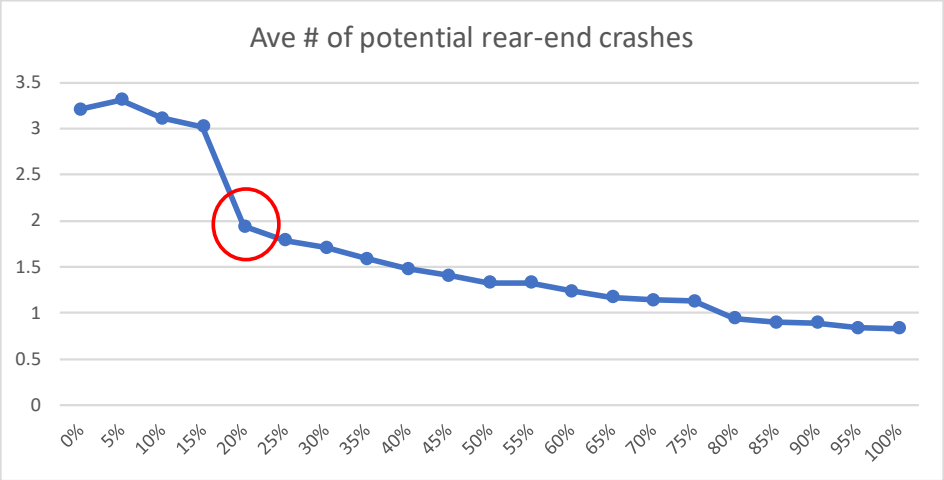


Figure 5.7: Average number of potential rear-end crashes with different CV rates

Intersections with short sight distance

To further evaluate the capability of Module 4 in reducing the potential of occurring rear-end crashes at intersections with short sight distance, this study further simulates the following three scenarios and evaluates the corresponding safety performance of the arterial:

- *Scenario 1*: all intersections are with sufficient sight distance;
- *Scenario 2:* all intersections have a sight distance of 250 feet; and
- *Scenario 3:* all intersections have a sight distance of 100 feet.

Based on the results from Table 5.3, it can be observed that shorter sight distance can greatly increase the risk of occurring rear-end crashes, evidenced by the largest number of hard-braking vehicles in Scenario 3. The comparison between "No protection" and "With proposed system" reveals that the proposed queue estimation module and rear-end prevention module can effectively provide optimal advisory speeds and make approaching vehicles be well-prepared to stop before getting close to the intersection.

Table 5.3: The system's safety performance under different scenarios

Scenarios	**Total # of hard-braking vehicles**		
	No Protection	With proposed system	Improvement
Scenario 1	304	41	-86.5%
Scenario 2	431	93	-78.4%
Scenario 3	751	105	-86.0%

6.0 MULTIMODAL TRAFFIC SIGNAL CONTROL FOR CONNECTED BUSES

6.1 INTRODUCTION

Besides the CV-based system introduced in Chapter 4.0 and 5.0, transit systems, managed by public transportation agencies, have been treated as a feasible testbed of CV technology. For example, UDOT launched a project in 2017 to install DSRC devices on 30 intersections of Redwood Road in Salt Lake City to support the CV-based TSP control. On the research side, a next-generation TSP strategy was proposed (He et al., 2014; Hu et al., 2014) to simultaneously implement coordinated-actuated signal control in CV systems. Later on, Hu et al. (2015) presented a person-delay-based optimization algorithm by splitting the green time of bus approach under the CV environment, and then extended it for multiple conflicting priority requests (Hu et al., 2016). In addition, other researchers demonstrated that the conditional signal priority can be used to improve bus reliability under different operation modes via a mathematical model based on Brownian motion again (Anderson and Daganzo, 2018).

Despite great efforts that have been devoted to TSP controls, those aforementioned methods are not suitable to be applied at intersections with median stations. In practice, many critical issues associated with the real-time signal control for connected buses remain to be addressed. For instance, the signal controllers may receive multiple priority requests from both directions within one signal cycle. Since those buses locate at different links, how to grant the signal priority in response to multiple bus requests, considering the potential impacts on both travel time saving and remaining capacity of nearside stations, will thus be a complex issue. The traditional TSP strategy may reduce the bus delay in one direction but at the cost of increasing risks to traffic safety and security in the other direction. To extend a new control system for this case, this study intends to address the scenario where the control system needs to determine how to dynamically adjust signal timing for controlling the connected buses which will enter (from the outbound direction) or leave (from the inbound direction) the nearside station. Key issues to be covered in this study include: (1) accounting for the uncertainty of bus dwelling time at nearside stations when granting signal priority to the buses from the inbound direction; (2) suppressing the arriving buses from the outbound direction by signals if queue spillover at downstream stations is detected; and (3) concurrently optimizing the total intersection performance considering both buses and passenger cars.

6.2 PROBLEM NATURE

Figure 6.1 shows an example at a TSP-controlled intersection with a nearby median station. When detecting the arrival of buses from both directions, the control system makes a priority decision and grants extra green time to both through directions as they share the same phase. Hence, although the near-side directional bus could pass the intersection without stopping by a red signal, the far-side bus will join the end of queue and wait for the clearance of the far-side station. Consequently, it can result in queue spillover and block the intersection.

To deal with the potential bus queue spillover, this study presents a novel priority suppression strategy of including green truncation and red extension. As shown in Figure 6.2, when detecting the approaching buses to the fully occupied far-side station, the system would activate the suppression control by truncating green time. Notably, when the buses approaching the far-side station are stopped by a red signal from green truncation, the bus flow in the opposite direction would experience additional delays. Hence, to coordinate all approaching buses from these two directions, this study aims to develop a novel control method that can concurrently avoid the spillback and reduce bus delay at the intersections.

Recognizing those vital research issues aforementioned at those intersections, the research objectives of this study are to prevent the presence of queue spillback and reduce the travel delay of CVs by proposing an effective signal priority strategy. Based on the real-time bus arrival information, the proposed strategy will dynamically adjust the current signal timings of the two directions, respectively. The detailed control logic along with the operational strategies will be provided in the following sections.

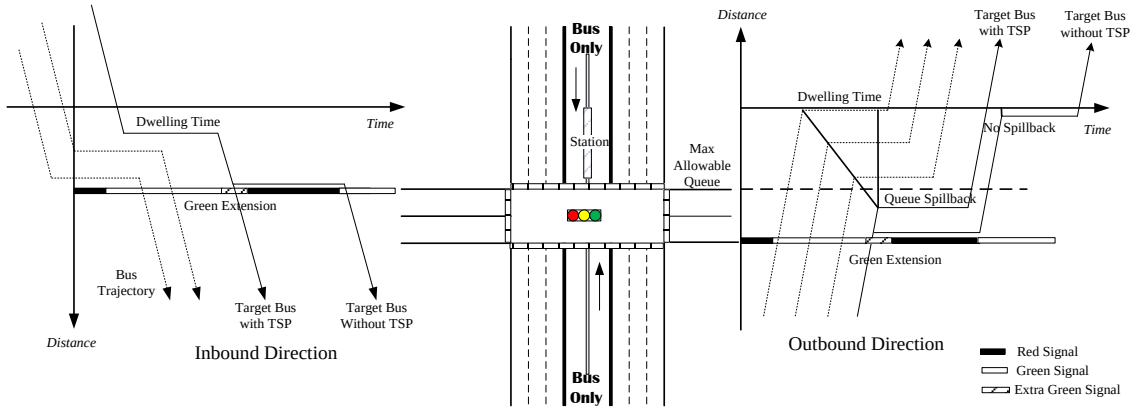


Figure 6.1: Bus operations under traditional TSP control with green extension

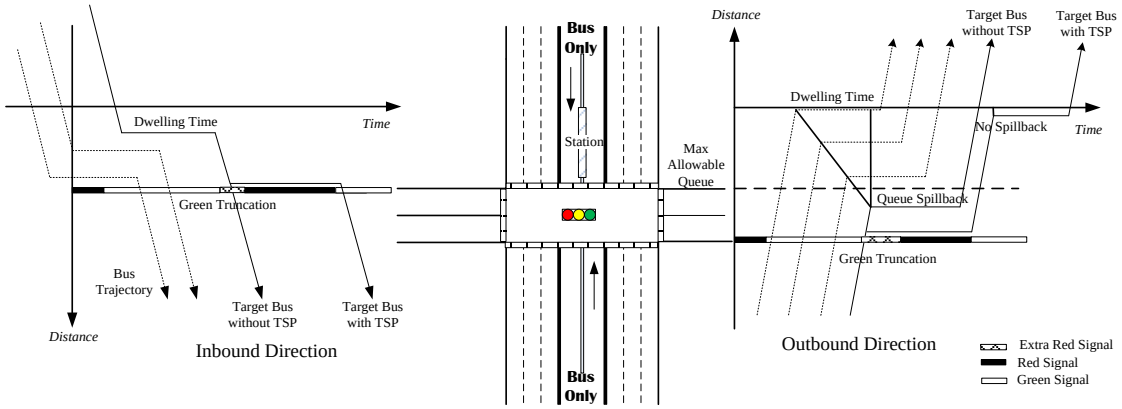


Figure 6.2: Bus operations under novel signal suppression control with green truncation

6.3 MODEL DEVELOPMENT

System Framework

This study proposed a novel real-time signal control system to account for the unique operational characteristics of intersections with median bus stations. Figure 6.3 illustrates the key control structure of the proposed system, which includes two core functions: vehicle-to-infrastructure (V2I) communication and signal timing adjustment. The model assumes that the locations of all buses are equipped with onboard units (OBUs) for providing their trajectories, and that the signal controllers are able to provide varying priority or suppression times. Based on the bus trajectories and their operation schedules, the control module simultaneously makes the control decisions for two directions. Before the end of green time, two possible control strategies may be implemented: green extension (grant priority to all buses) and green truncation (suppress the outbound approaching buses). Similarly, two other strategies, red truncation (grant priority to all buses) and red extension (suppress the outbound approaching buses), might be selected before the end of the red phase. Detailed control logic with respect to various traffic conditions will be discussed in the follows sub-sections.

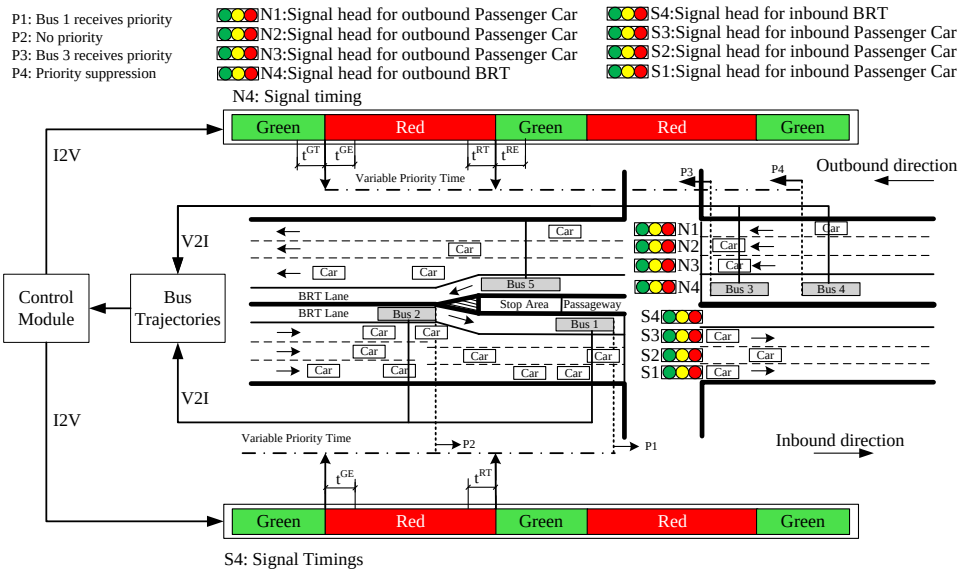


Figure 6.3: The demonstration of system control logic

Signal Control Strategies

Based on the detected buses locations, the signal control module is designed to make the priority decisions. Two control strategies could be found in the traditional TSP systems, green extension and red truncation, which are designed to provide extra green time to those approaching buses and progress buses to discharge the intersection without stops. In the proposed real-time multimodal signal control systems, two additional novel suppression control strategies are presented to prevent the queue spillback due to the overcapacity of near-side stations: green truncation and red extension. In summary, the potential control strategies with respect to different approaching directions are listed in Table 6.1.

Table 6.1: The control strategies and the corresponding control objectives

Strategies	Approaching directions	Objectives
Green extension	Inbound	Minimize bus delay through the station and intersection
	Outbound	
Red truncation	Inbound	
	Outbound	
Green truncation	Outbound	Prevent the queue spillback caused by the overcapacity at the station
Red extension	Outbound	Prevent the queue spillback caused by the overcapacity at the station

Note that the most essential control objective of the proposed system in this study is to implement the suppression control strategies (green truncation and red extension) to avoid the overflow at bus stations and the occurrence of queue spillback. Hence, the priority control strategies are activated only if the potential queue spillback is not predicted at the station. For the case that the bus exclusive lane and other through lanes use separate signal lights, the suppression control will be only activated on the bus signal light at the outbound direction. Figure 6.4 shows an illustrative control status of signals during the suppression time.

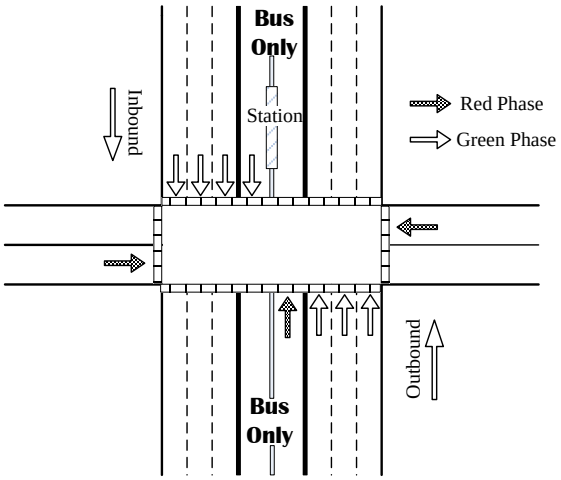


Figure 6.4: The demonstration of signal control status during suppression time

Signal Control Logics

To simplify the discussion, one can assume there are two signal statuses: green and red for both near-side and far-side directions. The entire decision-making process for TSP in response to multiple requests includes the following steps.

Step 1: Collect connected bus location and speed data via V2I communications and record the current signal timing plan at the target intersection.

Step 2: Estimate the maximal allowable duration for priority controls (green extension or red truncation) and suppress controls (green truncation or red extension) in terms of the traffic condition at the major and minor roads.

To limit the impact of priority/suppress control on other passenger-cars, the maximum allowable priority or suppress time is determined by the current v/c ratio on the crossing street and the major road:

$$\frac{(v_2 / c_2) \cdot t_{2,n}^g}{(t_{2,n}^g - t_n^{GE})} \leq \beta \tag{6.1-a}$$

$$\frac{(v_2 / c_2) \cdot t_{2,n}^g}{(t_{2,n}^g - t_n^{RT})} \leq \beta \tag{6.1-b}$$

$$\frac{(v_1 / c_1) \cdot t_{1,n}^g}{(t_{1,n}^g - t_n^{RE})} \leq \beta \tag{6.2-a}$$

$$\frac{(v_1 / c_1) \cdot t_{1,n}^g}{(t_{1,n}^g - t_n^{GT})} \leq \beta \tag{6.2-b}$$

where, v_1 and c_1 represent the bus volume and capacity of the bus exclusive lane in the outbound direction of the major road, respectively; v_2 and c_2 denote the passenger car volume and capacity of the crossing street, respectively; $t_{1,n}^g$ and $t_{2,n}^g$ are the green duration of the major road/crossing street phase in cycle n without the priority, respectively; $t_n^{GE}, t_n^{RT}, t_n^{RE}$, and t_n^{GT} are the duration of signal adjustment in green extension/red truncation/red extension/green truncation in the two intersection major approaches; and β is the maximum allowed (v/c) ratio in the signalized intersection after priority. Additionally, the minimum green time of each phase needs to be set such as to ensure the pedestrian can go through the intersection.

Step 3: Detect the approaching and dwelling buses around bus stops and calculate their current locations a few seconds (the sum of communication time and reaction time plus the maximal allowable duration) prior to the end of a green phase or a red phase.

Step 4: Estimate the potential benefits if granting the priority or suppression to a different number of detected buses.

The purpose of this step is to estimate the resulting benefits if a priority or suppression control for a specified length is granted to the major road. Depending on selected criterions, one can estimate the delay from the perspective of the bus passengers, and possibility of queue spillback caused by bus overflow at the station. However, this study does not consider the passenger car's delays due to difficulty of data collection in the current CV control framework. Since the objectives of the proposed control are to prevent the overflow at bus stations and reduce bus travel delay, the first step to develop the control logic is to compute the resulting travel time of each connected bus, given the current bus locations, if granted a priority or suppression.

Through V2I communication, the system can record the number of approaching buses and estimate their arrival time based on their locations and speeds. For outbound direction, one can describe bus arrivals at the stop line as follows:

$$T_A(t_S, t_E) = \{t_{a,1}, t_{a,2}, ..., t_{a,m}\} \tag{6.3}$$

where, $T_A(t_S, t_E)$ represents the set of bus arrival times during the time interval $[t_S, t_E]$; and $t_{a,i}$ denotes the arrival time of the *i*th bus; m is the total number of detected buses.

Then the system would calculate the occupancy of detectors at the inbound/outbound direction before the switch of signal phases and estimate the duration of priority control at the major road in Figure 6.3.

Step 4.1: If $t = nT + t_{1,n}^g$, based on the detected bus arrival information, one can calculate the required green extension time for the inbound direction buses as follows:

$$t_{n1}^{ge} = \begin{cases} t + t_n^{GE} - (t_1^* + t_{M1} + \Delta t_{10}) & f_A(t) = 1 \text{ and } t_{10}^* + t_{M1} + \Delta t_1 \in [t, t + t_n^{GE}] \\ t + t_n^{GE} - (t_{11}^* + \Delta t_1) & t_{11}^* + \Delta t_1 \in [t, t + t_n^{GE}] \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

$$t_{10}^* = \max(\{x \mid f_A(x) = 1, f_A(x-1) = 0, x \le t\}), t_{11}^* = \max(\{x \mid f_A(x) = 0, f_A(x-1) = 1, x \le t\}) \tag{6.5}$$

where, *t* denotes the current time; *T* is the length of signal cycle; t_{M1} is the average dwelling time of the inbound buses which might be estimated by the historical data; Δt_1 is the average travel time from the station to the stop line of the nearby intersection. In Equations (6.5-6.6), one can assume that the average dwelling time is more than the duration of maximum allowed green extension.

In contrast with the inbound direction, the required duration of green extension for the outbound direction buses can be determined by:

$$t_{n2}^{ge} = \begin{cases} \min(T_A(t, t + t_n^{GE})) - t & m(T_A(t, t + t_n^{GE})) = 1 \text{ and } f_B(t) = 0 \\ \min(T_A(t, t + t_n^{GE})) - t & m(T_A(t, t + t_n^{GE})) \ge 2 \text{ and } f_B(t) = 0 \text{ and } t_{20}^* + t_{M2} > \min(T_A(t, t + t_n^{GE})) + \Delta t_2 \\ t_{21}^* - t + \sigma & m(T_A(t, t + t_n^{GE})) \ge 2 \text{ and } f_B(t) = 0 \text{ and } t_{20}^* + t_{M2} \le t_{21}^* + \Delta t_2 \\ \min(T_A(t, t + t_n^{GE})) - t & f_B(t) m(T_A(t, t + t_n^{GE})) > 0 \text{ and } t_{22}^* + t_{M2} \le \min(T_A(t, t + t_n^{GE})) + \Delta t_2 \\ 0 & \text{otherwise} \end{cases} \tag{6.6-a}$$

$$t_{20}^* = \max(\{x \mid f_B(x) = 0, f_B(x-1) = 1, x \le t\}), t_{21}^* = \min(T_A(\min(T_A(t, t + t_n^{GE})) + 1, t + t_n^{GE})),$$
$$t_{22}^* = \max(\{x \mid f_B(x) = 1, f_B(x-1) = 0, x \le t\}) \tag{6.6-b}$$

where, Δt_2 denotes the average travel time from the V2I communication boundary and stop line; t_{M2} is the average dwelling time of the outbound buses; and t_{21}^* is the second minimum element in dataset $T_A(t, t+t_n^{GE})$.

Step 4.2: If $t \in [(n+1)T - t_n^{RT}, (n+1)T]$, based on the detected bus arrival information, one can estimate the duration of red truncation in the inbound direction as follows:

$$t_{n1}^{rt} = \begin{cases} \min((n+1)T - (t_{11}^* + \Delta t_1), t_n^{RT}) & t_{11}^* + \Delta t_1 > nT + t_{1,n}^g \\ 0 & \text{otherwise} \end{cases} \tag{6.7}$$

Similarly, the duration of outbound red truncation can be expressed by using the following equation:

$$t_{n2}^{rt} = \begin{cases} \min((n+1)T - \min(T_A(nT + t_{1,n}^g, t)), t_n^{RT}) & f_B(t) = 0 \text{ and } m(T_A(nT + t_{1,n}^g, t)) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.8}$$

Step 4.3: If $t = (n+1)T$, the duration of red extension is obtained by:

$$t_n^{re} = \begin{cases} \min((n+1)T - \min(T_A(nT + t_{1,n}^g, t)), t_n^{RE}) & f_B(t) = 1 \text{ and } m(T_A(nT + t_{1,n}^g, t)) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

Step 4.4: If $t \in [nT + t_{1,n}^g - t_n^{GT}, nT + t_{1,n}^g]$, based on the detected bus locations, compute the required duration of the green truncations strategy in the outbound direction using the following equation:

$$t_{n2}^{gt} = \begin{cases} nT + t_{1,n}^g - t_{30}^* & f_B(t) = 0 \text{ and } \mathrm{m}(T_A(t, nT + t_{1,n}^g)) \geq 2 \text{ and } t_{20}^* + t_{M2} \leq t_{30}^* + \Delta t_2 \\ nT + t_{1,n}^g - \min(T_A(t, nT + t_{1,n}^g)) & f_B(t) = 0 \text{ and } \mathrm{m}(T_A(t, nT + t_{1,n}^g)) \geq 2 \text{ and } t_{20}^* + t_{M2} > t_{30}^* + \Delta t_2 \\ nT + t_{1,n}^g - \min(T_A(t, nT + t_{1,n}^g)) & f_B(t) = 1 \text{ and } \mathrm{m}(T_A(t, nT + t_{1,n}^g)) \geq 1 \text{ and } \\ & t_{21}^* + t_{M2} \leq \min(T_A(t, nT + t_{1,n}^g)) + \Delta t_2 \\ 0 & \text{otherwise} \end{cases} \tag{6.10-a}$$

$$t_{30}^* = \min(T_A(\min(T_A(t, nT + t_{1,n}^g)) + 1, nT + t_{1,n}^g)) \tag{6.10-b}$$

where, t_{30}^* is the second minimum element in dataset $T_A(t, nT + t_{1,n}^g)$.

Note that the suppression control strategies will not be activated at the inbound direction.

Step 5: Execute the TSP control strategy.

Note that Step 4 will be executed multiple times within one signal cycle. Hence, the previous four strategies may be provided in one signal cycle for two directions. Hence, additional constraints will be considered after the second control strategy is provided during the identical signal cycle:

- To limit the traffic disruption to the crossing street, red truncation and green extension cannot be activated concurrently in one signal cycle for each direction;

- To reduce bus travel time and queue spillback as much as possible, the inbound and outbound directions are separately controlled; and

- To maintain the consistence of signal phases, using both green truncation and green extension (red truncation and red extension) for the outbound direction in one signal cycle is also prohibited.

Then the final optimal decision would be sent to the local signal controllers or control center to execute the dynamic control, including either the extended green duration for the priority phase and the green time reduction for the non-priority approach.

6.4 NUMERICAL EXAMPLES

Experimental Design

To illustrate the applicability and efficiency of the proposed real-time multimodal signal control strategy, this study has employed VISSIM as an unbiased tool for performance evaluation. Using the VISSIM-COM interface, this study developed an integrated program to simulate bus operations and actual signal controllers by VB.NET and MYSQL database. During the simulation, the program detects and records the real-time bus locations, automatically estimates the arrival time of each connected bus, and in real time adjusts the signal timings in terms of detected buses arrivals. Figure 6.5 shows the flow chart of the entire simulator and dynamic signal adjustments for the developed TSP operations in this study.

To evaluate the effectiveness of the proposed real-time signal control system for bus operations, this study has selected an intersection with two bus exclusive lanes constructed in the middle of the arterial as a test case. The key traffic and geometric parameters of this study site are listed below:

1) The target median station is located beside the intersection;

2) One road (Road 1) has four two-way passenger-car lanes, and two two-way bus exclusive lanes;

3) The other road (Road 2) has four two-way passenger-car lanes;

4) The target intersection is selected to offer real-time TSP controls based on the proposed model;

5) There are two bus routes in two directions;

6) Average bus dwelling time at the station is set to be 30s;

7) The mean headway of bus routes is about three minutes, and the observed variance on the target arterial ranges from about 2.5 minutes;

8) The average travel speed of buses is set to be 40km/h; and

9) The duration of the simulation time is 4,500 seconds.

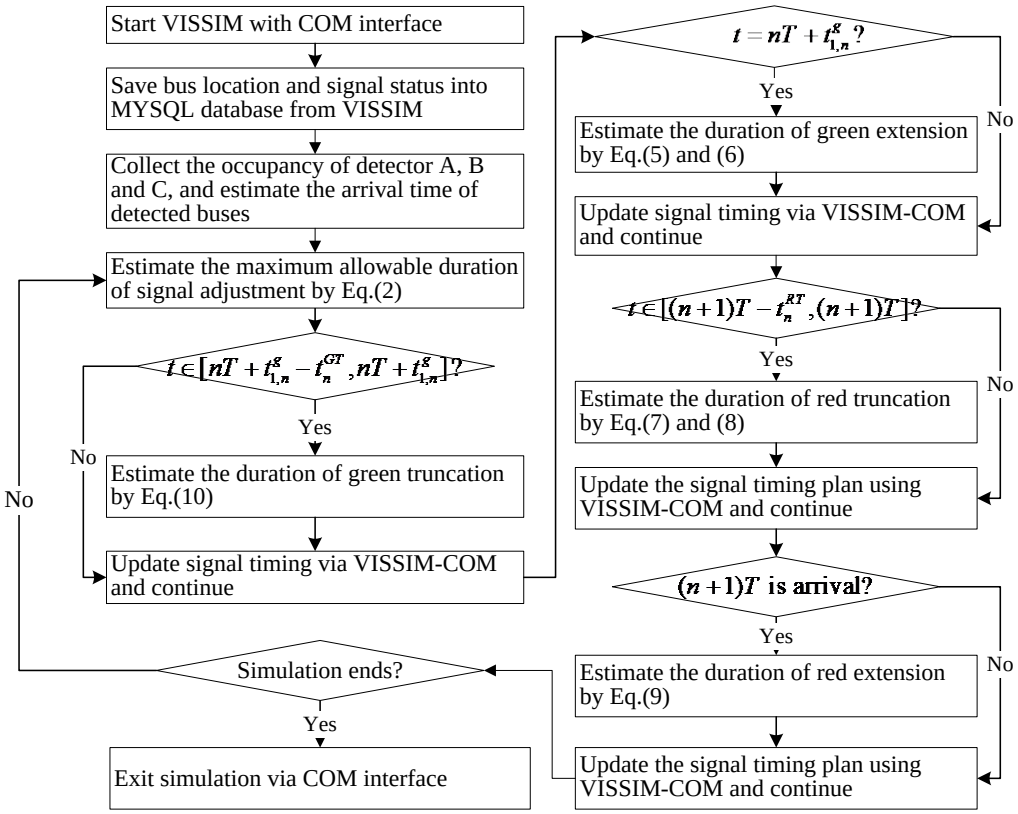


Figure 6.5: Flow chart of simulation evaluation

Experimental Results

For control efficiency and convenience of case illustration, this study evaluates the proposed strategy with the following three scenarios:

- Scenario 1: no TSP control (No TSP);
- Scenario 2: unconditional TSP control to all requested bus vehicles with 10s for green extension (Unconditional TSP);
- Scenario 3: proposed TSP control with maximum 10s for green extension and green truncation (Dynamic TSP).

Several MOEs are selected for model evaluation: bus queue length, the number of vehicle stops, average vehicle delay, and vehicle stops delay. Figure 6.5 has compared the results among different controls. Some key findings from simulation results are summarized below:

1) The proposed control can outperform the other two methods in terms of reduction in the maximal bus queue length at the outbound bus station every five-minute interval, due to priority strategies restriction in Figures 6.6 - 6.7. Moreover, the frequency of bus overflow at the outbound

bus station under three controls is 7, 8, and 5, respectively, when the outbound bus queue length equals or is greater than a threshold of three buses. Also, the possibility of bus overflow under unconditional TSP control is greater than others because it offers much more green time to discharge buses into the outbound station. One can obtain the same findings based on reduction in the number of bus stops in Fig. 8. The probability of having bus overflow under the scenario of real-time TSP control is 37.5% less than the one with the unconditional control.

2) Because the developed real-time multimodal signal control method concurrently considers the impacts of signal priority and suppression on buses and passenger cars, it also achieves a better performance than the other two methods with regard to reduction in bus passenger delay and passenger-car user delay (See Figure 6.9). This proposed control can significantly reduce the average stop delay and stops because in the priority-suppression control cases overflowed buses block other directional traffic flows and waste some green signal time. Particularly, the unconditional TSP case provides extra green time to bus vehicles without considering the capacity of the downstream station, which it has inevitably produced much more possibility of queue spillback (See Figure 6.7) and thus increased the potential occurrence of traffic accidents. Hence, one should recognize the limitation of traditional TSP control strategy in such cases.

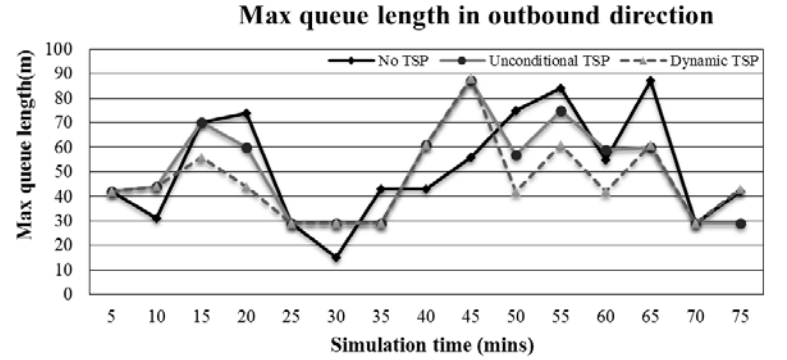


Figure 6.6: Max queue length at the outbound bus station

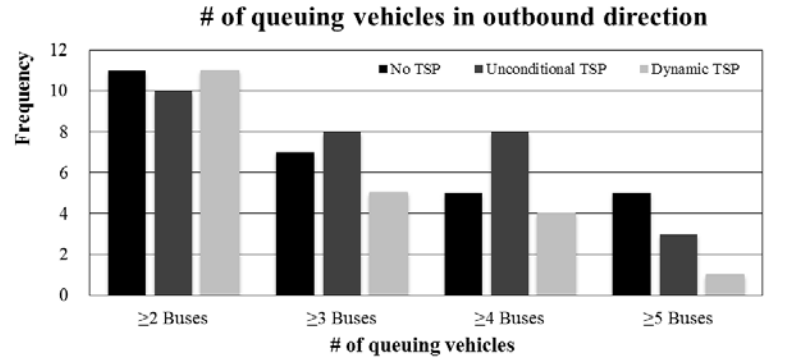


Figure 6.7: The number of queuing vehicle at the outbound bus station

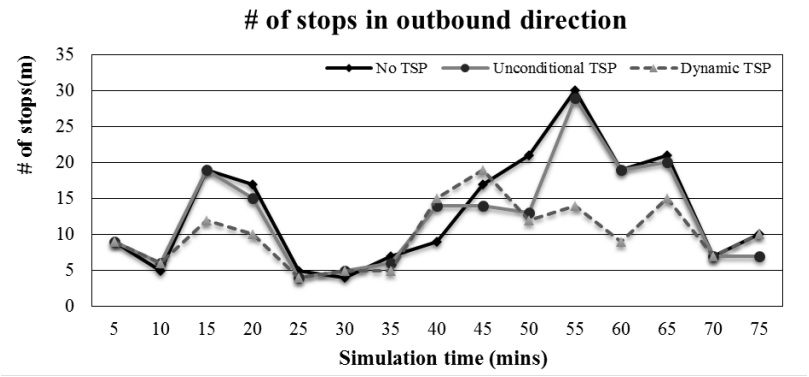
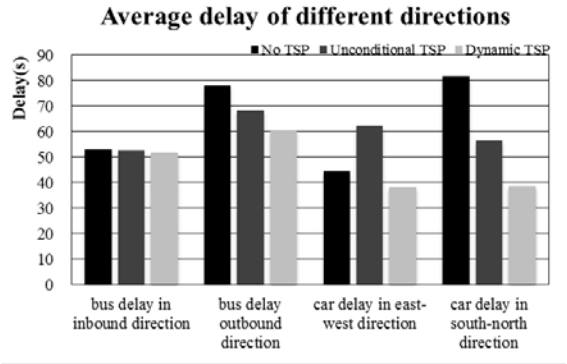
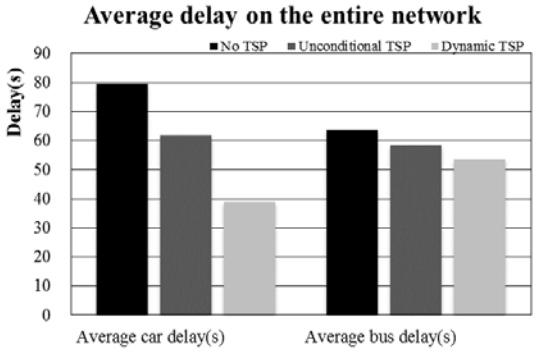


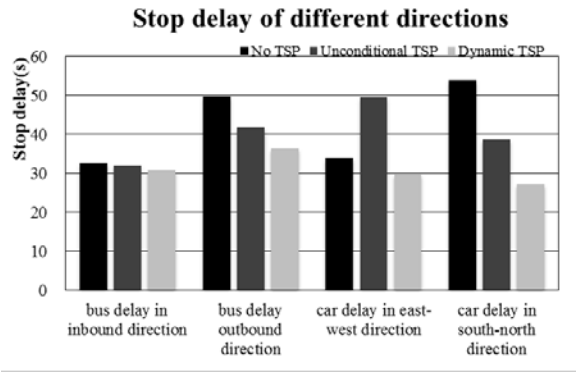
Figure 6.8: The number of stops at the outbound bus station



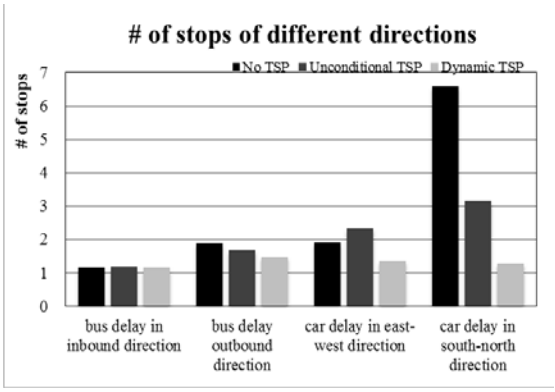
(A) Average delay of different directions



(B) Average delay of the entire network



(C) Average stopped delay



(D) Average # of stops

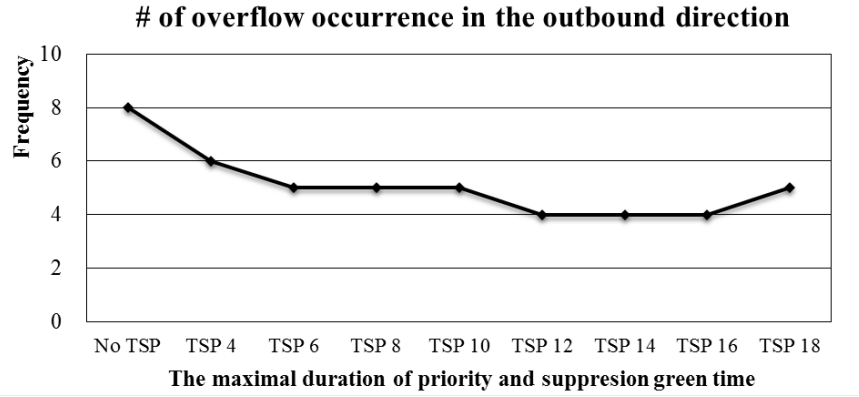
Figure 6.9: Simulation results of different directions under different control methods

Sensitivity Analysis

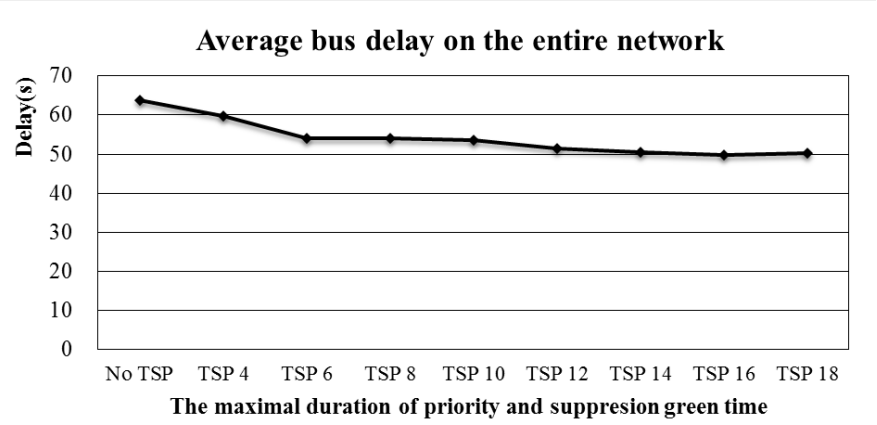
Note that the prevention of queue spillback is the most important issue in terms of traffic safety and security at the bus stations near intersections. Thus, to evaluate the reliability and effectiveness of the proposed dynamic controls, this study has tested the proposed model's sensitivity with respect to the maximal duration of priority-suppression green signals from 4s to 18s with 2s time step. As shown in Figure 6.10, regardless of the maximal duration of priority-

suppression time, the proposed real-time strategy can always reduce the occurrence of station overflow and bus average delay on the entire network, respectively.

However, the proposed controls yield a more significant reduction in overflow occurrence and bus delay if the maximal duration of priority-suppression time is set to be about 12s-14s, and such benefits will decrease with the maximal priority-suppression duration. Thus, the comparison between different priority-suppression times reveals that an overflow occurrence reduction can be achieved with the proposed strategy under the scenario of having a large priority-suppression duration. When the maximal duration is much less, this control strategy has not enough adjustment time to block much more buses into the station. Vice versa, the case of having much more adjustment time accumulates more buses into the next signal cycle and leads to station overflow.



(A) Number of events on the number of queuing buses no less than 3



(B) Average bus delay on the entire network

Figure 6.10: Performance indexes under No TSP and Dynamic TSP controls

7.0 CONCLUSIONS

7.1 KEY FINDINGS

In summary, this project developed a comprehensive CV-based traffic signal control system that can 1) reduce the CV communication delay when the system is dealing with a large number of CVs; 2) coordinate the CV flows along multiple critical paths by optimizing the intersection offsets; 3) improve both mobility and safety at intersections concurrently with optimal traffic signal timings; and 4) accommodate the operation of connected buses under a multimodal control environment.

In the Module 2 – communication network optimization, we focused on the minimization of Age of Information, which is equivalently to solve a subpacketization reduction problem for D2D coded caching while achieving the optimal rate. We first developed the packet type-based (PTB) design framework, where the subpacketization reduction problem can be formulated as an integer optimization problem with node memory constraints and proper restrictions on the set of candidates further splitting ratio vectors. Each feasible solution of the optimization problem corresponds to a valid PTB design of caching schemes. Then we focused on some special cases of node grouping (i.e., equal grouping method and certain unequal grouping methods), and proposed several classes of PTB design with constant or order reduction of subpacketization which comes from a combination of the raw packet saving gain and the further splitting ratio gain. The result showed that the previously well-known JCM scheme is not optimal in terms of subpacketization in general. We also came up with the concept of heterogeneous subpacket size to deal with node memory satisfaction under unequal node grouping. With the proposed PTB design the subpacketization complexity can be significantly reduced, which makes the employment of D2D coded caching techniques more practical.

With the development of wireless technology in Module 2, vehicles can communicate with each other and with infrastructures in the CV environment. Data collected from CVs can provide more enriched real-time information. Hence, an adaptive traffic signal control system under the CV environment can be more efficient than the traditional systems based on in-pavement loop detectors or radar sensors. This project proposed a signal progression control system to control signals dynamically, which aimed to coordinate multiple critical paths along an arterial. This control system was implemented in the CV environment. To reflect the current reality of low market penetration, the CV penetration rate was assumed as 30% in this study. Since the traffic demand may fluctuate during the coordination period, the critical paths were updated every 10 minutes based on the data collected from CVs. After determining critical paths, a model was constructed to offer a progression band for them. The objective of this model was to maximize the green bandwidth of those critical paths by adjusting offsets of intersections along the arterial. Then an algorithm solution based on dynamic programming was proposed to solve this model. To illustrate the effectiveness and potential of the proposed signal control system, an experimental simulation test was conducted in VISSIM. Results revealed that the vehicle travel times along those critical paths are reduced compared with the fixed coordination control system. And the proposed control system can reduce average delay and number of stops for both critical paths and the entire tested network.

In Module 3, this project developed an integrated system that can concurrently improve urban arterial mobility and safety performance, grounded on the same set of hardware equipment. Four control modules, dilemma zone protection, queue length estimation, signal coordination, speed harmonization, and rear-end crash prevention are integrated into the system to achieve three control objectives: proactively preventing rear-end collision, reactively protecting side-street traffic from red-light-running vehicles, and effectively facilitating speed harmonization along local arterials. Depending on the vehicle arrival time and the corresponding signal status (i.e., green, red, yellow and all-red), the system will take corresponding actions to either prevent crashes or improve signal progression. Selecting a segment of Redwood Road in Salt Lake City as our study site, our research team built a simulation platform in VISSIM. Through the VISSIM-COM interface, data were exchanged between outside computational VB.NET programs and the simulation model. Results from extensive experiments confirmed the effectiveness of the proposed system in both reducing potential intersection crash rates and improving arterial progression efficiency. The proposed control framework also proved the effectiveness of using dilemma zone protection sensors for traffic mobility improvement.

Under the multimodal environment, this project presented a real-time control system for the connected bus system of having medium-island bus stations near signalized intersections. The proposed model aims to prevent the queue spillback caused by bus overflow at the station, and reduce the bus travel time and the number of stops along the arterials as much as possible. Due to the limited storage capacity of bus stations, the conventional TSP strategy often fails to offer efficient control since high bus arrivals can form a long queue and cause a queue spillback to the neighboring signalized intersection. In response to such situations, the proposed dynamic control has integrated both priority control and suppression control strategies, where the former control is implemented to reduce the bus travel time and the latter one is adopted to prevent the bus queue spillback. Using simulation results, the proposed strategy has shown its promise in reducing total bus delay and queue spillback occurrence around the station. Our further exploration with simulation experiments for sensitivity analysis also found that the proposed method can be most effective with a large duration of priority and suppression green signal.

8.0 REFERENCES

Ahn, K., Rakha, H., & Park, S. (2013). Ecodrive application: Algorithmic development and preliminary testing. Transportation Research Record: Journal of the Transportation Research Board, (2341), 1-11.

Al-Deek, H., Sandt, A., Alomari, A., and Hussain, O. (2017). A technical note on evaluating the effectiveness of bus rapid transit with transit signal priority. Journal of Intelligent Transportation Systems, 21(3), 227-238.

Anderson P., and C. F. Daganzo. (2018). Effect of Transit Signal Priority on Bus Service Reliability. arXiv preprint arXiv:1806.09254, math.OC.

Argote-Cabañero, J., E. Christofa, and A. Skabardonis, Connected vehicle penetration rate for estimation of arterial measures of effectiveness. Transportation Research Part C: Emerging Technologies, 2015. 60: p. 298-312.

Ashraf, M. I., Bennis, M., Perfecto, C., & Saad, W. (2016, December). Dynamic Proximity-aware Resource Allocation in Vehicle-to-Vehicle (V2V) Communications. In Globecom Workshops (GC Wkshps), IEEE (pp. 1-6). IEEE.

Ashraf, M. I., Liu, C. F., Bennis, M., & Saad, W. (2017, June). Towards low-latency and ultra-reliable vehicle-to-vehicle communication. In Networks and Communications (EuCNC), 2017 European Conference on (pp. 1-5). IEEE.

Bai, F., Stancil, D. D., & Krishnan, H. (2010, September). Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers. In Proceedings of the sixteenth annual international conference on Mobile computing and networking(pp. 329-340). ACM.

Ban, X., P. Hao, and Z. Sun, Real time queue length estimation for signalized intersections using travel times from mobile sensors. Transportation Research Part C: Emerging Technologies, 2011. 19(6): p. 1133-1156.

Bang, S., & Ahn, S. (2017). Platooning Strategy for Connected and Autonomous Vehicles: Transition from Light Traffic. Transportation Research Record: Journal of the Transportation Research Board, (2623), 73-81.

Beak, B., Head, K. L., Feng, Y. Adaptive Coordination Based on Connected Vehicle Technology. Transportation Research Record: Journal of the Transportation Research Board, Vol 2619, 2017, pp. 1-12.

Boillot, F., Blosseville, J. M., Lesort, J. B., Motyka, V., Papageorgiou, M., & Sellam, S. (1992). Optimal signal control of urban traffic networks. In Road Traffic Monitoring, 1992 (IEE Conf. Pub. 355) (p. 75). IET.

Bonneson, J., Middleton, D., Zimmerman, K., Charara, H., & Abbas, M. (2002). Intelligent detection-control system for rural signalized intersections. Texas Transportation Institute, 8.

Cai, C., G. Geers, and Y. Wang, Vehicle-to-infrastructure communication-based adaptive traffic signal control. IET Intelligent Transport Systems, 2013. 7(3): p. 351-360.

Chang G-L., M. Vasudevan, and C-C. Su. (1996) Modeling and Evaluation of Adaptive Bus-preemption Control with and without Automatic Vehicle Location System. Transportation Research A, 30(4), pp. 251-268.

Chang, G. L., Franz, M., & Lu, Y. (2012). Design and Evaluation of a dynamic dilemma zone system for a high speed rural intersection (No. MD-12-SP708B4J).

Chen, Y. and L.R. Rilett, Signal Timing Optimization for Corridors with Multiple Highway-Rail Grade Crossings Using Genetic Algorithm. Journal of Advanced Transportation, 2018. 2018: p. 1-14.

Cho, W., Kim, S. I., kyun Choi, H., Oh, H. S., & Kwak, D. Y. (2009, May). Performance evaluation of V2V/V2I communications: The effect of midamble insertion. In Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on (pp. 793-797). IEEE.

Christofa E., and A. Skabardonis. (2011) Traffic Signal Optimization with Application of Transit Signal Priority to an Isolated Intersection. In Transportation Research Record: Journal of the Transportation Research Board, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., pp. 192-201.

Costa, M., Codreanu, M., & Ephremides, A. (2014, June). Age of information with packet management. In Information Theory (ISIT), 2014 IEEE International Symposium on (pp. 1583-1587). IEEE.

Costa, M., Codreanu, M., & Ephremides, A. (2016). On the age of information in status update systems with packet management. IEEE Transactions on Information Theory, 62(4), 1897-1910.

Dabiri, S. and M. Abbas. Arterial traffic signal optimization using Particle Swarm Optimization in an integrated VISSIM-MATLAB simulation environment. in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). 2016. IEEE.

Day, I., Ag, S., & Whitelock, R. (1998, July). SCOOT-split, cycle & offset optimization technique. In TRB Mid-Year Meeting and Adaptive Traffic Signal Control Workshop (Vol. 7).

Dey, Kakan Chandra, et al. (2016) "Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network–Performance evaluation." Transportation Research Part C: Emerging Technologies 68: 168-184

Diakaki, C., M. Papageorgiou, I. Papamichail, and I. Nikolos. (2015). Overview and Analysis of Vehicle Automation and Communication Systems from a Motorway Traffic Management Perspective. Transportation Research Part A: Policy and Practice, Vol. 75, 2015, pp. 147–165. https://doi.org/10.1016/j.tra.2015.03.015.

Draft, S.J.R., Jam, J2735 dedicated short range communications (dsrc) message set dictionary. 2006.

Dresner, K., & Stone, P. (2008). A multiagent approach to autonomous intersection management. Journal of artificial intelligence research, 31, 591-656.

Feng, Y., et al., Connected Vehicle–Based Adaptive Signal Control and Applications. Transportation Research Record: Journal of the Transportation Research Board, 2016. 2558(1): p. 11-19.

Feng, Y., Head, K. L., Khoshmagham, S., & Zamanipour, M. (2015). A real-time adaptive signal control in a connected vehicle environment. Transportation Research Part C: Emerging Technologies, 55, 460-473.

Flint, M., Polycarpou, M., & Fernandez-Gaucherand, E. (2002). Cooperative path-planning for autonomous vehicles using dynamic programming. IFAC Proceedings Volumes, 35(1), 481-486.

Florin, R. and S. Olariu, (2015) A survey of vehicular communications for traffic signal optimization. Vehicular Communications. **2**(2): p. 70-79.

Florin, R. and S. Olariu, A survey of vehicular communications for traffic signal optimization. Vehicular Communications, 2015. 2(2): p. 70-79.

Foley & Lardner LLP. (2017). 2017 Connected Cars & Autonomous Vehicles Survey. Retrieved from https://www.foley.com/files/uploads/2017-Connected-Cars-Survey-Report.pdf.

Gartner, N. H., Pooran, F. J., & Andrews, C. M. (2001). Implementation of the OPAC adaptive control strategy in a traffic signal network. In Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE (pp. 195-200). IEEE.

Gartner, N.H., et al., A multi-band approach to arterial traffic signal optimization. 1991. 25(1): p. 55-74.

Gates, T. J. (2007). A dynamic on-demand all-red clearance interval extension process for stochastic vehicular arrivals at signalized intersections. University of Wisconsin--Madison.

Gazis, D., Herman, R., & Maradudin, A. (1960). The problem of the amber signal light in traffic flow. Operations Research, 8(1), 112-132.

Ge, J. I., and G. Orosz. (2014) Dynamics of Connected Vehicle Systems with Delayed Acceleration Feedback. Transportation Research Part C: Emerging Technologies, Vol. 46, , pp. 46–64. https://doi.org/10.1016/j.trc.2014.04.014.

Goodall, N., Smith, B., Park, B.B. Traffic signal control with connected vehicles. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2381, 2013, pp. 65-72

Goodall, N.J., B.L. Smith, and B.B. Park, Microscopic Estimation of Freeway Vehicle Positions From the Behavior of Connected Vehicles. Journal of Intelligent Transportation Systems, 2014. 20(1): p. 45-54.

Grumert, E. F., and A. Tapani. (2017) Using Connected Vehicles in a Variable Speed Limit System. Transportation Research Procedia, Vol. 27, 2017, pp. 85–92. https://doi.org/10.1016/j.trpro.2017.12.050.

Guler, S. I., Menendez, M., & Meier, L. Using connected vehicle technology to improve the efficiency of intersections. Transportation Research Part C: Emerging Technologies, Vol. 46, 2014, pp. 121-131.

Hafeez, K. A., Zhao, L., Ma, B., & Mark, J. W. (2013). Performance analysis and enhancement of the DSRC for VANET's safety applications. IEEE Transactions on Vehicular Technology, 62(7), 3069-3083

He, Q., Head, K.L., Ding, J. PAMSCOD: Platoon-based arterial multi-modal signal control with online data. Transportation Research Part C: Emerging Technologies, Vol. 20, 2012, pp. 164-184.

Head L., D. Gettman, and Z. Wei. (2006) Decision Model for Priority Control of Traffic Signals. In Transportation Research Record: Journal of the Transportation Research Board, No. 1978, Transportation Research Board of the National Academies, Washington, D.C., pp. 169-177.

Hounsell N.B., and B.P. Shrestha. (2012) A New Approach for Co-operative Bus Priority at Traffic Signals. IEEE Transactions on Intelligent Transportation Systems, Vol. 13, No. 1, pp. 6-14.

Hounsell N.B., B.P. Shrestha, J.R. Head, S. Palmer, and T. Bowen. (2008) The Way ahead for London's Bus Priority at Traffic Signals. IET Intelligent Transport Systems, Vol. 2, No. 3, pp. 193-200.

Hu J., B. B. Park, and Y. J. Lee. (2015) Coordinated Transit Signal Priority Supporting Transit Progression under Connected Vehicle Technology. Transportation Research Part C: Emerging Technologies, 55, pp. 393-408.

Hu J., B. B. Park, and Y. J. Lee. (2016). Transit signal priority accommodating conflicting requests under Connected Vehicles technology. Transportation Research Part C: Emerging Technologies, 69, pp.173-192.

Hu, J., Park, B., & Parkany, A. (2014). Transit signal priority with connected vehicle technology. Transportation Research Record: Journal of the Transportation Research Board, (2418), 20-29.

Huang, K., Yang, X., Lu, Y., Mi, C. C., & Kondlapudi, P. (2018). Ecological Driving System for Connected/Automated Vehicles Using a Two-Stage Control Hierarchy. IEEE Transactions on Intelligent Transportation Systems.

Hurwitz, D. S., Abadi, M. G., McCrea, S., Quayle, S., & Marnell, P. (2016). Smart Red Clearance Extensions to Reduce Redlight Running Crashes (No. FHWA-OR-RD-16-10). Oregon Department of Transportation, Research Section.

Husch, D. and J. Albeck, SYNCHRO 5.0 User Guide for Windows. 2000, Traffieware Corporation, Albany, CA.

Jahangiri, A., Rakha, H., & Dingus, T. A. (2016). Red-light running violation prediction using observational and simulator data. Accident Analysis & Prevention, 96, 316-328.

Kadota, I., Uysal-Biyikoglu, E., Singh, R., & Modiano, E. (2016, September). Minimizing the age of information in broadcast wireless networks. In Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on (pp. 844-851). IEEE.

Kam, C., Kompella, S., Nguyen, G. D., & Ephremides, A. (2016). Effect of message transmission path diversity on status age. IEEE Transactions on Information Theory, 62(3), 1360-1374.

Kamalanathsharma, R. K., & Rakha, H. A. (2013, October). Multi-stage dynamic programming algorithm for eco-speed control at traffic signalized intersections. In Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on (pp. 2094-2099). IEEE

Katsaros, K., Kernchen, R., Dianati, M., & Rieck, D. (2011, August). Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform. In IWCMC 2011-7th International Wireless Communications and Mobile Computing Conference (pp. 918-923).

Kaul, S. K., Yates, R. D., & Gruteser, M. (2012, March). Status updates through queues. In Information Sciences and Systems (CISS), 2012 46th Annual Conference on (pp. 1-6). IEEE.

Kaul, S., Gruteser, M., Rai, V., & Kenney, J. (2011, June). Minimizing age of information in vehicular networks. In Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on (pp. 350-358). IEEE.

Kaul, S., Yates, R., & Gruteser, M. (2012, March). Real-time status: How often should one update?. In INFOCOM, 2012 Proceedings IEEE (pp. 2731-2735). IEEE.

Kell, J. H., & Fullerton, I. J. (1998). Manual of traffic signal design, englewood cliffs, nj: Inst. transp. eng.

Kenney, J. B. (2011). Dedicated short-range communications (DSRC) standards in the United States. Proceedings of the IEEE, 99(7), 1162-1182.

Khoudour, L., Lesort, J. B., & Farges, J. L. (1991). Traffic Management. RTS English Issue Number 6. Prodyn: three years of trials in the Zelt experimental zone. Recherche Transports Securite, (6).

Klein, A. G., Farazi, S., He, W., & Brown, D. R. (2017). Staleness bounds and efficient protocols for dissemination of global channel state information. IEEE Transactions on Wireless Communications, 16(9), 5732-5746.

Lee, J., & Park, B. (2012). Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment. IEEE Transactions on Intelligent Transportation Systems, 13(1), 81-90.

Lee, J., Park, B., Yun, I. Cumulative travel-time responsive (CTR) real-time intersection control algorithm under the connected vehicle environment. Journal of Transportation Engineering, Vol. 139, No. 10, 2013, pp. 1020-1029.

Li M., Y. Yang, W. Zhang, and K. Zhou. (2011) Modeling and Implementation of Adaptive Transit Signal Priority on Actuated Control Systems. Computer-Aided Civil and Infrastructure Engineering, 26(4), 2011, pp. 270-284.

Li, J., et al., Platoon Priority Visualization Modeling and Optimization for Signal Coordination in the Connected Vehicle Environment. Transportation Research Record: Journal of the Transportation Research Board, 2019. 2673(5): p. 36-48.

Li, M., et al., Application of vehicle infrastructure integration data on real-time arterial performance measurements. 2008.

Lin Y., X. Yang, G-L. Chang, and N. Zou. (2013) Transit Priority Strategies for Multiple Routes under Headway-based Operations. In Transportation Research Record: Journal of the Transportation Research Board, No. 2356, Transportation Research Board of the National Academies, Washington, D.C., pp. 34-43.

Ling K., and A. Shalaby. (2004) Automated Transit Headway Control via Adaptive Signal Priority. Journal of Advanced Transportation, Vol. 38, No. 1, pp. 45-67.

Little, J.D., M.D. Kelson, and N.H. Gartner, MAXBAND: A versatile program for setting signals on arteries and triangular networks. 1981.

Lowrie, P. R. (1990). Scats, sydney coordinated adaptive traffic system: A traffic responsive method of controlling urban traffic.

Lu, N., Cheng, N., Zhang, N., Shen, X., & Mark, J. W. (2014). Connected vehicles: Solutions and challenges. IEEE internet of things journal, 1(4), 289-299.

Ma W., W. Ni, K. L. Head, and J. Zhao. (2013) Effective Coordinated Optimization Model for Transit Priority Control under Arterial Progression. In Transportation Research Record: Journal of the Transportation Research Board, No. 2356, Transportation Research Board of the National Academies, Washington, D.C., pp. 71-83.

Ma W., X. Yang, and Y. Liu. (2010) Development and Evaluation of a Coordinated and Conditional Bus Priority Approach. In Transportation Research Record: Journal of the Transportation Research Board, No. 2145, Transportation Research Board of the National Academies, Washington, D.C., pp. 49-58.

Ma, J., Li, X., Zhou, F., Hu, J., & Park, B. B. (2017). Parsimonious shooting heuristic for trajectory design of connected automated traffic part II: computational issues and optimization. Transportation Research Part B: Methodological, 95, 421-441.

Ma, X., Chen, X., & Refai, H. H. (2009). Performance and reliability of DSRC vehicular safety communication: a formal analysis. EURASIP Journal on Wireless Communications and Networking, 2009(1), 969164.

Marsden, B.G., E.C.-p. Chang, and B.R.J.I.J. Derr, The Passer II-84 System: A Practical Signal Timing Tool. 1887.

Mauro, V., & Di Taranto, C. (1990). Utopia. In Control, computers, communications in transportation (pp. 245-252).

McCoy, P. T., & Pesti, G. (2003, January). Dilemma zone protection on high-speed signalized intersection approaches: Advance detection versus advance warning flashers and advance detection. In 82nd Annual Meeting of the Transportation Research Board, TRB. Washington, DC, USA.

Mirchandani P. B., A. Knyazyan, K. L. Head, and W. Wu. An Approach towards the Integration of Bus Priority, Traffic Adaptive Signal Control, and Bus Information/Scheduling Systems. Journal of Scheduling, Vol. 505, 2001, pp. 319-334.

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.

Munigety, C. R., Gupta, P. A., Gurumurthy, K. M., Peeta, S., & Mathew, T. V. (2016). Vehicle-type dependent car following model using spring-mass-damper dynamics for heterogeneous traffic (No. 16-5025)

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., & Wang, Y. (2003). Review of road traffic control strategies. Proceedings of the IEEE, 91(12), 2043-2067.

Pappas, N., Gunnarsson, J., Kratz, L., Kountouris, M., & Angelakis, V. (2015, June). Age of information of multiple sources with queue management. In Communications (ICC), 2015 IEEE International Conference on (pp. 5935-5940). IEEE.

Park, S. Y., Lan, C. L., & Chang, G. L. (2017). Integration of Dilemma Zone Protection with Speed Harmonization at High Speed Suburban Intersections (No. 17-03782), Transportation Research Board 95th Annual Meeting, Washington D.C.

Park, S., C.L. Lan, G.L. Chang, D. Tolani, and P. Huang. Design and Predeployment Assessment of an Integrated Intersection Dilemma Zone Protection System. Journal of Transportation Engineering, DOI: 10.1061/(ASCE)TE.1943-5436.0000894, 2016.

Priemer, C., and Friedrich, B. A decentralized adaptive traffic signal control using v2i communication data. In: Proceedings of the IEEE International Conference on Intelligent Transportation Systems, 2009, pp. 1-6.

Rao, B. S. Y., and P. Varaiya. (1994). Roadside Intelligence for Flow Control in an Intelligent Vehicle and Highway System. Transportation Research Part C: Emerging Technologies, Vol. 2, No. 1, 1994, pp. 49–72. https://doi.org/10.1016/0968-090X(94)90019-1.

Rim, H., et al., Estimation of Lane-Level Travel Times in Vehicle-to-Vehicle and Vehicle-to-Infrastructure–Based Traffic Information System. Transportation Research Record: Journal of the Transportation Research Board, 2011. 2243(1): p. 9-16.

Schrank, D., T. Lomax, and B.J.T.T.I. Eisele, (2012) urban mobility report. 2011: p. 1-57.

Silva, C. M., Masini, B. M., Ferrari, G., & Thibault, I. (2017). A survey on infrastructure-based vehicular networks. Mobile Information Systems, 2017.

Song X., Yuan M., Liang D., and Ma L. (2018) Optimization Method for Transit Signal Priority considering Multirequest under Connected Vehicle Environment. Journal of Advanced Transportation, 2018.

Stevanovic, A., P.T. Martin, and J.J.T.R.R. Stevanovic, VisSim-based genetic algorithm optimization of signal timings. 2007. 2035(1): p. 59-68.

Sun, Y., Polyanskiy, Y., & Uysal-Biyikoglu, E. (2017, June). Remote estimation of the Wiener process over a channel with random delay. In Information Theory (ISIT), 2017 IEEE International Symposium on (pp. 321-325). IEEE.

Sun, Y., Uysal-Biyikoglu, E., Yates, R. D., Koksal, C. E., & Shroff, N. B. (2017). Update or wait: How to keep your data fresh. IEEE Transactions on Information Theory, 63(11), 7492-7508.

Talebpour, A., & Mahmassani, H. S. (2016). Influence of connected and autonomous vehicles on traffic flow stability and throughput. Transportation Research Part C: Emerging Technologies, 71, 143-163

Tiaprasert, K., et al., Queue Length Estimation Using Connected Vehicle Technology for Adaptive Signal Control. IEEE Transactions on Intelligent Transportation Systems, 2015. 16(4): p. 2129-2140.

Transportation Research Board National Research Council. (2015). Highway capacity manual. TRB Business Office.

Truong L. T., G. Currie, and M. Sarvi. (2017). Analytical and simulation approaches to understand combined effects of transit signal priority and road-space priority measures. Transportation Research Part C: Emerging Technologies, 74, 275-294.

Wallace, C.E., et al., TRANSYT-7F user's manual. 1984.

Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2014). Rolling horizon control framework for driver assistance systems. Part I: Mathematical formulation and non-cooperative systems. Transportation research part C: emerging technologies, 40, 271-289.

Wang, M., W. Daamen, S. P. Hoogendoorn, and B. van Arem. (2016) Connected Variable Speed Limits Control and Car-Following Control with Vehicle-Infrastructure Communication to Resolve Stop-and-Go Waves. Journal of Intelligent Transportation Systems, Vol. 20, No. 6, 2016, pp. 559–572. https://doi.org/10.1080/15472450.2016.1157022.

Wang, Z., & Hassan, M. (2008, September). How much of DSRC is available for non-safety use?. In Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking (pp. 23-29). ACM.

Werbos, P.J.H.o.i.c.n., fuzzy and a. approaches, Approximate dynamic programming for realtime control and neural modelling. 1992: p. 493-525.

Wilbur S. and Associates, Westinghouse Airbrake Company, and Institute of Public Administration. (1968) Study of Evolutionary Urban Transportation, Volumes I, II and III. U.S. Department of Housing and Urban Development.

Wu G., L. Zhang, W. Zhang, and M. Tomizuka. (2012) Signal Optimization at Urban Highway Rail Grade Crossings Using an Online Adaptive Priority Strategy. Journal of Transportation Engineering, Vol. 138, No. 4, pp. 479-484.

Wu, X., Yang, J., & Wu, J. (2018). Optimal status update for age of information minimization with an energy harvesting source. IEEE Transactions on Green Communications and Networking, 2(1), 193-204.

Xia, H., Wu, G., Boriboonsomsin, K., & Barth, M. J. (2013, October). Development and evaluation of an enhanced eco-approach traffic signal application for connected vehicles. In Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on (pp. 296-301). IEEE.

Xu, B., et al., Cooperative Method of Traffic Signal Optimization and Speed Control of Connected Vehicles at Isolated Intersections. IEEE Transactions on Intelligent Transportation Systems, 2019. 20(4): p. 1390-1403.

Yagar S., and B. Han. (1994) A Procedure for Real-time Signal Control that Considers Transit Interference and Priority. Transportation Research B 28(4), pp.315-331.

Yang, K., Guler, S. I., Menendez, M. Isolated intersection control for various levels of vehicle technology: Conventional, connected, and automated vehicles. Transportation Research Part C: Emerging Technologies, Vol. 72, 2016, pp. 109-129.

Yang, X., Y. Cheng, and G.-L. Chang, A multi-path progression model for synchronization of arterial traffic signals. Transportation Research Part C: Emerging Technologies, 2015. 53: p. 93-111.

Yates, R. D. (2015, June). Lazy is timely: Status updates by an energy harvesting source. In Information Theory (ISIT), 2015 IEEE International Symposium on (pp. 3008-3012). IEEE.

Yates, R. D., & Kaul, S. (2012, July). Real-time status updating: Multiple sources. In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on (pp. 2666-2670). IEEE.

Ye Z., and M. Xu. (2017). Decision Model for Resolving Conflicting Transit Signal Priority Requests. IEEE Transactions on Intelligent Transportation Systems, 18(1), 59-68.

Zegeer, C. V. (1977). Effectiveness of green-extension systems at high-speed intersections.

Zhang, C., et al., AM-Band: An Asymmetrical Multi-Band model for arterial traffic signal coordination. Transportation Research Part C: Emerging Technologies, 2015. 58: p. 515-531.

Zhang, G., & Wang, Y. (2011). Optimizing minimum and maximum green time settings for traffic actuated control at isolated intersections. IEEE Transactions on Intelligent Transportation Systems, 12(1), 164-173.

Zhang, L., Wang, L., Zhou, K., & Zhang, W. B. (2012). Dynamic all-red extension at a signalized intersection: A framework of probabilistic modeling and performance evaluation. IEEE Transactions on Intelligent Transportation Systems, 13(1), 166-179.

Zhou, F., Li, X., & Ma, J. (2017). Parsimonious shooting heuristic for trajectory design of connected automated traffic part I: theoretical analysis with generalized time geography. Transportation Research Part B: Methodological, 95, 394-420.

Zhu, F., & Ukkusuri, S. V. (2015). A linear programming formulation for autonomous intersection control within a dynamic traffic assignment and connected vehicle environment. Transportation Research Part C: Emerging Technologies, 55, 363-378.

Zimmerman, K., Tolani, D., Xu, R., Qian, T., & Huang, P. (2012). Detection, control, and warning system for mitigating dilemma zone problem. Transportation Research Record: Journal of the Transportation Research Board, (2298), 30-37.

Maddah-Ali, M. A., & Niesen, U. (2014). Fundamental limits of caching. IEEE Transactions on Information Theory, 60(5), 2856-2867.

Yan, Q., Cheng, M., Tang, X., & Chen, Q. (2017). On the placement delivery array design for centralized coded caching scheme. IEEE Transactions on Information Theory, 63(9), 5821-5833.

Shanmugam, K., Ji, M., Tulino, A. M., Llorca, J., & Dimakis, A. G. (2016). Finite-length analysis of caching-aided coded multicasting. IEEE Transactions on Information Theory, 62(10), 5524-5537.

Shanmugam, K., Tulino, A. M., & Dimakis, A. G. (2017, June). Coded caching with linear subpacketization is possible using Ruzsa-Szeméredi graphs. In 2017 IEEE International Symposium on Information Theory (ISIT) (pp. 1237-1241). IEEE.

Shangguan, C., Zhang, Y., & Ge, G. (2018). Centralized coded caching schemes: A hypergraph theoretical approach. IEEE Transactions on Information Theory, 64(8), 5755-5766.

Tang, L., & Ramamoorthy, A. (2018). Coded caching schemes with reduced subpacketization from linear block codes. IEEE Transactions on Information Theory, 64(4), 3099-3120.

Jin, S., Cui, Y., Liu, H., & Caire, G. (2019). A New Order-Optimal Decentralized Coded Caching Scheme with Good Performance in the Finite File Size Regime. IEEE Transactions on Communications.

Ji, M., Caire, G., & Molisch, A. F. (2015). Fundamental limits of caching in wireless D2D networks. IEEE Transactions on Information Theory, 62(2), 849-869.

Woolsey, N., Chen, R. R., & Ji, M. (2017, December). Device-to-device caching networks with subquadratic subpacketizations. In GLOBECOM 2017-2017 IEEE Global Communications Conference (pp. 1-6). IEEE.

Woolsey, N., Chen, R. R., & Ji, M. (2018, May). Coded Caching in Wireless Device-to-Device Networks Using a Hypercube Approach. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1-6). IEEE.

Li, S., Maddah-Ali, M. A., Yu, Q., & Avestimehr, A. S. (2017). A fundamental tradeoff between computation and communication in distributed computing. IEEE Transactions on Information Theory, 64(1), 109-128.