HIERARCHICAL PRIORITY–BASED CONTROL OF SIGNALIZED INTERSECTIONS IN SEMI-CONNECTED CORRIDORS

FINAL PROJECT REPORT

by

Ali Hajbabaie
Mehrzad Mehrabipour
Rasool Mohebifard
North Carolina State University

Ahmed Abdel-Rahim
Sameh Sorour
Umair Yaqub

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium, the U.S. Government, and the matching sponsor assume no liability for the contents or use thereof.

Technical Report Documentation Page		
1. Report No.	**2. Government Accession No.** 01701468	**3. Recipient's Catalog No.**
4. Title and Subtitle Hierarchical Priority-based Control of Signalized Intersections in Semi-Connected Corridors	**5. Report Date**	
	6. Performing Organization Code	
7. Author(s) Ali Hajbabaie, 0000-0001-6757-1981; Mehrzad Mehrabipour, Rasool Mohebifard, Ahmed Abdel-Rahim, 0000-0001-9756-554X; Sameh Sorour, Umair Yaqub	**8. Performing Organization Report No.** 2018-M-WSU-2	
9. Performing Organization Name and Address PacTrans Pacific Northwest Transportation Consortium University Transportation Center for Region 10 University of Washington More Hall 112 Seattle, WA 98195-2700	**10. Work Unit No. (TRAIS)**	
	11. Contract or Grant No. 69A355174110	
12. Sponsoring Organization Name and Address United States of America Department of Transportation Research and Innovative Technology Administration	**13. Type of Report and Period Covered**	
	14. Sponsoring Agency Code	
15. Supplementary Notes Report uploaded at www.pacTrans.org		

16. Abstract

In this project, we developed an efficient distributed yet coordinated algorithm to control signalized intersections in connected and semi-connected corridors. The research enhances traffic signal optimization formulations to allow for the incorporation of connected vehicles and existing point detector data in the models, the distribution of decisions at both the intersection and corridor levels to reduce computational complexity, and the coordination of control decisions among various intersections by a distributed, cloud-fog-based communication network to push solutions toward system-level optimality. We utilized a hierarchical, priority-based control using fog-cloud architecture in this research. The fog component consists of micro-data centers with limited computational capabilities, collocated with the road-side units (RSUs), responsible for computing optimal timings of traffic signals in real time utilizing their limited capabilities. A cloud backbone is connected to all fog components of the city or each city zone to exchange information among neighboring fog components to enable coordinated yet distributed optimization of traffic signal timings. It also performs analytics on the collected data and decisions from all the fog components to learn spatio-temporal patterns, especially in the case of semi-connected scenarios. The proposed architecture allows intersection- and corridor-level optimization algorithms to be run and control decisions to be communicated to the traffic signal system with existing control hardware and communication technology.

Numerical results for a simulated corridor of ten intersections showed that the approach can effectively determine near-optimal signal timing parameters under different demand levels, and significant improvements in traffic operations wee observed with increased connected vehicle market penetration rates. Because of the COVID-19 pandemic, this project did not conduct any field tests, and all tests were performed with a micro-simulation testbed. Future research that includes hardware-in-the-loop and field tests is needed.

17. Keywords Signal Timing, Connected Vehicles, Distributed, Coordinated, Network-level		**18. Distribution Statement** No restrictions.	
19. Security Classification (of this report) Unclassified.	**20. Security Classification (of this page)** Unclassified.	**21. No. of Pages**	**22. Price** NA

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API	Application programming interface
BSM	Basic safety message
CCS	Centralized control system
CCU	Cloud control unit
CTM	Cell transmission model
CV	Connected vehicle
DCO	Distributed coordination and optimization
DCS	Decentralized control system
DCCS	Distributed-coordinated control system
FCU	Fog control units
GA	Genetic algorithm
HCS	Hierarchical control system
HILS	Hardware-in-the-loop simulation analysis
MILP	Mixed integer linear program
MINLP	Mixed integer non-linear program
MPC	Model predictive control
NIATT	National Institute of Advanced Transportation Technology
NTCIP	National Transportation Communications for ITS Protocol
OBU	Onboard unit
OPAC	Optimized Policies for Adaptive Control
RHODES	Real-Time Hierarchical Optimizing Distributed Effective System
RSU	Road side unit
RT-TRACS	Real-Time Traffic Adaptive Control System
SCATS	Sydney Coordinated Adaptive Traffic System
SpaT	Signal phasing and timing
TC	Traffic controller
UDP	User datagram protocol
UTOPIA	Urban Traffic Optimization by Integrated Automation

Acknowledgments

Executive Summary

Connected vehicles, the internet of things, and smart infrastructure technologies facilitate the exchange of real-time, highly granular information among individual users in transportation networks, system operators, and the supporting infrastructure through communication standards. Harnessing this emergent, ubiquitous connectivity and its resulting data stream will open unexplored possibilities to improve network mobility by optimizing the timing of signalized intersections.

This project developed a priority-based, distributed coordinated algorithm to determine signal plans in transportation networks with connected and semi-connected corridors. We proposed a fog-cloud architecture in which to implement the hierarchical, priority-based control. This algorithm decomposes the signal timing problem into intersection-level sub-problems and optimizes sub-problems simultaneously. The decomposition approach decreases the complexity of the problem because sub-problems have a smaller number of variables and can be optimized in parallel. The computation of sub-problems is assigned to fogs. The fog component finds signal plans for each sub-problem (intersection) in real time.

The sub-problems are coordinated by an exchange of information through a cloud backbone that is in connection with intersection-level fog components. The cloud backbone administers the exchange of information among the intersection-level fog components and provides requested information by fogs. The cloud also monitors changes in the transportation network and adjusts the coordination when needed.

Numerical results for a simulated corridor of ten intersections showed that the approach effectively determined near-optimal signal timing parameters under different demand levels, and

significant improvements in traffic operations were observed with increased connected vehicle

market penetration rates.

1   Introduction

1.1   <u>Background</u>

Connected vehicles, the internet of things, and smart infrastructure technologies facilitate the exchange of real-time, highly granular information among individual users in transportation networks, system operators, and the supporting infrastructure through communication standards. Harnessing this emergent ubiquitous connectivity and its resulting data stream will open unexplored possibilities to improve network mobility, specifically by optimizing the timing of signalized intersections.

Corridor-level signal timing can be formulated as an optimization model; however, this model is computationally complex (NP-complete) with nonlinear terms, mixed-integer decision variables, various connected vehicle penetration rates and data streams, and stochastic transportation demand/capacity and driver behavior. The state-of-practice signal system technology is based on point detector data. It does not utilize connected vehicle data and does not have the computational power to solve complex optimization models. This research will addressed these shortcomings by (1) formulating a mathematical optimization model for signal timing that can utilize both detector and connected vehicle data, (2) developing an efficient solution technique, and (3) executing the solution technique with a cloud-fog hierarchical architecture.

Previous research on signal timing optimization can be divided into four main groups based on the problem architecture: (a) centralized, (b) hierarchical, (c) decentralized, and (d) distributed-coordinated approaches.

The majority of signal timing optimization algorithms utilize a centralized formulation and architecture (Abu-Lebdeh and Benekohal, 1997; Beard and Ziliaskopoulos, 2006; Chang and Sun, 2004; Feng et al., 2015; Girianna and Benekohal, 2002; Hajbabaie and Benekohal, 2013; Hunt et

al., 1983; Karoonsoontawong and Waller Travis, 2010; Medina et al., 2013; Putha et al., 2012; Sun et al., 2006; Teklu et al., 2007). As a result, they optimize the various signal timing parameters (i.e., phase plan, cycle length, green times, and offsets) of all intersections at the same time in one optimization model. However, network-level signal timing optimization is an NP-complete problem (Hajbabaie, 2012; Wünsch, 2008), and a central optimization technique will not be scalable and applicable to large transportation networks.

To overcome this issue, decentralized approaches have been developed that decompose the network into regions with varying numbers of intersections. Rather than solving a single optimization problem for the entire network, several optimization problems are solved simultaneously (Gregoire et al., 2015; Lee et al., 2013; Adacher and Oliva, 2014; Goodall and Park, 2013; Porche and Lafortune, 1998; Priemer, 2009; Wongpiromsarn et al., 2014). As a result, decentralized approaches are scalable. However, fully decentralized approaches tend to lose performance because of a lack of coordination among the split optimization processes. Therefore, they mostly control signals locally and may find sub-optimal solutions rather than a global optimal one.

To overcome this issue, hierarchical approaches have been developed that decompose the network optimization problem into a multi-level control problem with distinct objectives at each level. The underlying concept of most hierarchical approaches is to make network-level decisions at the upper (or central) level, and real-time, small-area computations at the lower (or intersection) level (Gartner et al., 2001; Head et al., 1992; Mauro and DiTaronto, 1990; Gartner and Andrews, 2002; Sims and Dobinson, 1980). Designing signal control methods with a reasonable balance between these levels is challenging (Dion and Hellinga, 2001), as the upper and lower levels may

have objective functions that compete with each other. In addition, the computational complexity of the upper level may increase with the size of the network.

This project developed efficient distributed yet coordinated, priority-based algorithms to control signalized intersections in connected and semi-connected corridors. We utilized a hierarchical, priority-based control with fog-cloud architecture to achieve this objective. The fog component consists of micro-data centers with limited computational capabilities collocated with road-side units (RSUs). They are responsible for computing the optimal timings of traffic signals in real time utilizing their limited capabilities. To achieve this objective, we developed distributed coordination and optimization algorithms to effectively reduce the complexity of the problem by decomposing it into intersection-level sub-problems while still guaranteeing near-optimal solutions through effective coordination of various decisions.

On the other hand, the cloud backbone is connected to all the intersection-level fog components to provide network-level monitoring and optimal control decisions. The role of this cloud-level control is to administer the exchange of information among the intersection-level fog components and to enable distributed yet coordinated optimization. It also analyzes the data to identify any changes in the spatio-temporal patterns of the network operations that might require a change in the coordination scheme at each intersection. In the individual intersection-level fog computation, the optimization is performed with factoring data from a certain number of neighboring intersections to ensure effective coordination. The cloud-level control can adjust the number and extent of the intersections included in the fog-level optimization in cases of incidents and preemption for emergency vehicles, buses, or trucks carrying hazardous materials. This innovative, hybrid, two-level control approach ensures real-time coordinated and optimized decisions that can adapt to changes in the characteristics of network traffic operations.

The project focused on the existing traffic signal system control technology and developed formulation and solution techniques compatible with it. It also studied the impacts of various connected vehicle penetration rates.

1.2   Project Goal and Objectives

The goal of this project was to improve mobility in connected and semi-connected corridors. The main objective of this research was to develop efficient distributed yet coordinated algorithms to control signalized intersections in connected corridors and those that are semi-connected (where not all vehicles have connectivity capability or refrain from sharing intentions for privacy reasons). The research was intended to enhance traffic signal optimization formulations to allow for the incorporation of connected vehicles and existing point detector data in the models, the distribution of decisions at both the intersection and the corridor levels to reduce computational complexity, and the coordination of control decisions among various intersections by a distributed, cloud-fog-based communications network to push solutions toward global optimality.

1.3   Report Organization

This report contains eight chapters. The second chapter presents a comprehensive review of literature relevant to signal control in transportation networks. Chapter 3 introduces an optimization model to determine the timing of signalized intersections in a transportation network. Chapter 4 describes the development of a distributed optimization and coordination solution technique to find signal timing parameters in real time. Chapter 5 presents the incorporation of connected vehicle and point detector data in the formulation. Chapter 6 provides information on the cloud-fog communication architecture and how it can be implemented in a micro-simulation-based testbed. Chapter 9 includes numerical results, and Chapter 10 presents concluding remarks.

1.4   Institutional Partnerships

This project included two important parts: methodological and technological developments. The Washington State University (WSU) team led the methodological developments: problem formulation, distributed solution algorithm, and micro-simulation test-bed development.

The University of Idaho (UI) team   led the technological advancements: communication/computing infrastructure and modeling the proposed system operations in a hardware-in-the-loop simulation (HILS) environment that enabled the execution and testing of the proposed algorithms.

2 Literature Review

This chapter presents a review of literature on existing signal control approaches.

2.1 Centralized Control Systems

Centralized control systems (CCSs) include a central processor that determines system control variables. The processor sends the determined variables to individual traffic controllers. Increasing the study period, adding signalized intersections, considering turning movements, considering various phasing plans, and applying realistic traffic flow dynamics adversely affect the computational complexity of these systems. Therefore, CCSs may fail for large-scale networks, especially when CCSs use optimization-based algorithms to find optimal solutions (Chan, 1997). The algorithms employed in CCSs can be categorized into optimization-based and heuristic approaches.

2.1.1 Optimization-Based Approaches

These approaches find globally optimal solutions for an intersection, an arterial, or a network of intersections. Webster (1958) optimized the signal timing plans of an isolated intersection to minimize travel delay for the first time. Since then, many types of research in the field of signal timing optimization have been conducted.

Lo (1999) developed a cell transmission model (CTM) based, mixed-integer linear formulation to minimize total delay. In this formulation, the non-linear CTM equations were linearized to reduce the computational complexity. Green and red signal durations were optimized with a fixed cycle length. Moreover, the formulation captured traffic dynamics under different demand patterns and network conditions. A small network of 15 cells was solved centrally, and the requirement of developing efficient solution techniques for larger case studies was noted. The formulation was also simplified to cases with no turning movements. This simplification was a

7

restrictive assumption. Later, Lo (2001) proposed a CTM-based, mixed-integer linear formulation with turning movements to optimize signal timing variables. The solutions of formulation did not have the vehicle holding back problem.

Lin and Wang (2004) proposed a cell-based, mixed-integer linear formulation to optimize cycle lengths and green splits. Different objective functions, including the minimization of total delay, the number of stops, and lost time, were integrated with weight multipliers associated with each objective. The formulation was optimized by using CPLEX for a network of two intersections with 15 cells, one-way streets, and no turning movements. The researchers also modified the formulation to consider emergency vehicles in traffic streams. The main drawback was that the formulation could not be solved for larger networks unless more efficient algorithms were used.

Beard and Ziliaskopoulos (2006) proposed a mixed integer linear program (MILP) formulation to minimize total travel time that considered signal timing and dynamic traffic assignment problems simultaneously. Some important features of this formulation were modeling turning movements, protected and permissive movements, and multiple origins and destinations. The decisions about routes, cycle lengths, and splits were determined. They showed that dynamic signal timing optimization performed better than pre-timed signal timings, particularly when an incident happened in the network. However, commercial software was not able to solve the formulation with standard methods, e.g., the simplex, dual simplex, and interior point methods for large networks.

Karoonsoontawong and Waller (2010) presented a formulation that considered signal timing optimization, user equilibrium dynamic traffic assignment, and capacity expansion problems simultaneously. The formulations were cell-based, mixed-integer, linear, and non-linear. Formulations were also proposed as deterministic, stochastic, bi-level, and single-level. The

decision set included phase sequence, cycle lengths, green signal duration, and offsets. Commercial software programs such as GAMS, CPLEX, and DICOPT weer employed to solve formulations for a network of seven cells with three intersections. The complexity of proposed formulations would not allow solving them using commercial software for large-scale networks.

Guilliard et al. (2016) proposed an MILP formulation to minimize total delay and optimize signal timing parameters for a network of intersections. Non-homogeneous time steps were captured by using CTM, the link transmission model, and the queue transmission model. Networks with three, six, and nine intersections with different geometric and symmetry properties were used in numerical results. The queue transmission model decreased the computational complexity of the formulation in comparison to CTM. However, the solution technique complexity was not studied.

Wada et al. (2017) proposed an MILP formulation to coordinate signal timing plans across arterials under deterministic and stochastic demand profiles. The variational theory introduced by Daganzo (2005b, 2005a) was employed to analyze delay and create the formulation. The variational theory was based on the kinematic wave model introduced by Newell (1993). The cross-entropy method was tested on a network of three intersections. The proposed method needed substantial modifications, including the addition of turning movements, different phases, and lane properties. Moreover, the method was not generalized for different network geometries and demand profiles.

Mohebifard and Hajbabaie (2019) proposed a Benders decomposition technique to solve signal timing optimization problems for large-scale networks. The problem was formulated as a mixed-integer, non-linear program. The solution technique consisted of a master and a primal problem that were optimized iteratively until the desired gap between the upper bound and lower

bound was achieved. The primal problem was converted to a CTM simulation that led to a more efficient approach than previous techniques, and the solution optimality was guaranteed in the finite number of iterations. However, the approach was not suitable for online decision-making because of long run times for finding solutions.

2.1.2   Heuristic Approaches

Increasing the decision space by growth in network size and time study could lead to the failure of optimization-based approaches. Hence, heuristic approaches have been proposed to offer less computational complexity. The genetic algorithm (GA) has been a widely used heuristic for solving signal timing problems during the past decades (Abu-Lebdeh and Benekohal, 2000; Ceylan and Bell, 2004; Lo and Chow, 2004; Stevanovic et al., 2007). It has also been used to solve complex optimization problems in other disciplines (Hajibabai and Ouyang, 2013; Hajibabai and Saha, 2019).

Abu-Lebdeh and Benekohal (1997) developed a non-linear formulation to maximize network throughput with time-varying features to coordinate signal plans across an arterial. The signal timing problem was optimized under oversaturated conditions to determine offsets and green signal durations. A GA was developed to solve the formulation for an arterial of five intersections. They considered only one-way streets with no turning movements. The solution quality was not evaluated, and the algorithm complexity was a factor of network size.

In one of the primary works on GA to solve a signal timing optimization problem, Chang and Sun (2004) presented a maximal progression possibility operation that was a heuristics to find signal timing plans by simultaneously minimizing total delay and the number of stops. This algorithm adapted parameters for the most congested intersection of a transportation network and applied the same parameters to the rest of the network. A network of 12 oversaturated and 13

undersaturated intersections was used for testing. The results showed better performance of this algorithm in comparison with TRANSYT-7F and signal timing optimization software. The optimality gap for solutions was not examined in this study.

Zhang et al. (2010) also proposed a GA-based algorithm to solve deterministic and stochastic cell-based, mixed-integer linear programs. The problem optimized signal timing plans that included a sequence of phases, green signal duration, cycle length, and offsets. In the stochastic case, demand was uncertain. The objective was to minimize the total delay of an arterial. The algorithm was tested on an arterial consisting of five intersections. In a study by Zhang, Yin, and Chen (2013), exposure risk to traffic emissions was incorporated in the proposed model of Zhang et al. (2010) by creating a bi-level formulation. The drawbacks of these studies included unknown solution properties and substantial growth in the computational complexity of the algorithm as the network size increased.

In 2010, Lertworawanich et al. used goal programming to model and find solutions for models with three CTM-based objective functions, including minimization of spillovers, minimization of delay, and maximization of throughput. The weighted sum of the objective functions was optimized by using a GA-based algorithm. The algorithm was applied to a network of nine intersections, and vehicle holding back was not a problem in these solutions because of the use of non-linear equations for CTM. However, the computational complexity of the model was more than the conventional linearized CTM-based signal timing problems.

He, Head, and Ding (2011) developed an MILP formulation to minimize total delay by determining phase durations on an arterial. This platoon-based model considered transit and passenger cars and communication between vehicles and infrastructure. They decreased the formulation complexity by creating significant platoons, assuming first come first served, and

having limited coordination. The proposed approach to solve the formulation of prioritized platoons and determined queues and the algorithm found signal plans by using communications data. The approach performed better than SYNCHRO on an arterial of eight intersections and under different demand patterns. However, the algorithm could fail as the network size would increase because of the required computational efforts.

Hajbabaie et al. (2011) and Medina et al. (2011) proposed an approach that used a GA, evolution strategies, and approximate dynamic programming for finding signal plans in oversaturated conditions. They compared their approach performance with the approximate dynamic programming. The approximate dynamic programming performed faster than other algorithms. The GA generated better results when demand was predefined, whereas dynamic programming produced better results when demand was non-predictive. Hajbabaie et al. (2010) studied the effects of different left turn policies on network operations in the presence of optimized signal timing parameters. Hajbabaie and Benekohal (2011) also studied the effects of using a common cycle length on traffic operations in transportation networks. Hajbabaie and Benekohal (2011b) and Medina et al. (2013) studied the effects of optimal signal timing parameters on optimal metering rates in urban street networks and found significant improvements in traffic operations.

Putha et al. (2012) applied an ant colony optimization algorithm to coordinate signal plans in oversaturated conditions. They found green signal duration by maximizing a weighted network throughout and penalizing queues. The researchers compared the ant colony optimization with the traditional GA and showed that colony optimization was more efficient in terms of computation time. However, the proposed algorithm was not real time.

Hajbabaie and Benekohal (2015) proposed a formulation for simultaneous optimization of signal timing and system optimal traffic assignment problems. The formulation maximized the

weighted number of completed trips by considering uncertainty in drivers' behavior and network demand. Cycle lengths, green signal duration, and offsets were determined by optimizing this formulation. A GA was used to solve the formulation for a network of 20 intersections and showed better results than CORSIM.

2.2   Decentralized Approaches

In decentralized control systems (DCS), signal timing problems are decomposed into several independent sub-problems, and each sub-problem is assigned to a single controller. Local controllers make decisions for individual sub-problems by using a processing unit. In these systems, the determined decisions of each sub-problem are not coordinated with other sub-problems. DCSs can be applied to large-scale networks because each sub-problem has a significantly fewer number of decision variables. However, these systems may not provide good quality solutions (Chan, 1997).

Wong and Wong (2003) and Wong and Heydecker (2011) presented mixed-integer linear programs to optimize signal timing plans and movements of lanes at isolated intersections. Two objective functions used in the modeling were the maximization of capacity and the minimization of cycle lengths. Optimizing movement types for each lane improved the performance of intersections. However, this optimization required adding many decision variables, and as a result, the computational complexity of problems was increased.

In 2009, Priemer and Friedrich presented a decentralized, phase-based algorithm to find optimal phase sequences and estimate queue lengths, assuming the availability of vehicles to the infrastructure communication scheme. At each time horizon, an optimization problem was solved by using a dynamic programming algorithm and complete enumeration approaches to minimize queue lengths at intersections. In this algorithm, controllers did not communicate and collaborate.

Wongpiromsarn et al. (2012) developed a distributed algorithm to determine the signal settings of local controllers separately. The algorithm was designed on the basis of a backpressure routing concept (Tassiulas and Ephremides, 1992) used in the context of communication and power networks. Each local controller separately selected phases that satisfied several constraints for each time slot by using the information on the queue lengths of an intersection. The entering demand at each junction was obtained by using data from loop detectors, and the output of lanes was set according to the entering demand value. Even though the algorithm performed better than adaptive traffic signal controllers, the junctions were not coordinated, so the solutions were locally optimal.

Goodall et al. (2013) presented a decentralized algorithm that used locations, headings, and speed in a network that included connected vehicles. They used a rolling-horizon approach to find signal timing plans. They also predicted different objective functions for each horizon by using microscopic simulation. In low penetration rates, the algorithm did not perform well.

Adacher et al. (2014) presented a spatial decomposition algorithm to minimize the weighted sum of delays at signalized intersections that employed a platoon model. By using a distributed consensus algorithm, they segmented the transportation network into nodes. A K-means algorithm was also implemented to locate controllers in each node. Then, signal timing plans and offsets were determined with surrogate measures for intersections in a node. The information was shared among intersections of a node with neighboring intersections in a fully distributed way.

Yu et al. (2018) proposed an MILP formulation to minimize total delay by optimizing the trajectory of vehicles and signal plans for isolated intersections with CVs. They employed Newel's car-following model (Newell, 2002) to find the lane changing of vehicles, trajectories, arrivals in

14

a platoon, phase sequences, phase durations, and cycle lengths. The communication among intersections was ignored, which created problems in the generalization of this approach for the network of intersections.

2.3   Hierarchical Control Systems

In hierarchical control systems (HCS), signal timing problems are solved as a hierarchy in multiple layers. Each layer solves part of the signal timing problem.

The Sydney Coordinated Adaptive Traffic System (SCATS) was the first adaptive and real-time hierarchical approach. This approach had two levels: strategic and tactical. First, the signal timing problem for a network of at most ten intersections was optimized. Next, they determined green splits for each intersection individually by using single controllers to satisfy intersection-level requirements (Luk, 1984; Sims and Dobinson, 1980). SCATS had a time-consuming pre-processing step to specify all available split plans with many parameters to be calibrated.

The Optimized Policies for Adaptive Control (OPAC) strategy is another widely used approach proposed by Gartner (1982, 1983) to find signal timing plans in real time. This strategy had three layers: local control, coordination, and synchronization. In the first layer, they determined the best change policy for a horizon by using the arrival data of vehicles, a prediction model, and some information from the third layer. This layer could change decisions on the basis of new data feeds. In the second layer, they determined optimal offsets for each intersection. In the third layer, they found cycle lengths for all or a group of intersections. Gartner, Pooran, and Andrews (2001) used OPAC on the Reston Parkway, which had 16 signalized intersections in Reston, Virginia. OPAC was also incorporated into the Real-Time Traffic Adaptive Control System (RT-TRACS) proposed by Gartner, Stamatiadis, and Tarnoff (1995), which was a state-

of-the-art system intended to implement and select traffic control strategies according to traffic conditions.

The Urban Traffic Optimization by Integrated Automation (UTOPIA) (Mauro and DiTaronto, 1990) was another hierarchical approach with online and offline optimizations and two levels. UTOPIA considered both central and local controllers. The central controllers were responsible for the area level to optimize the signal plans in a selected area. The local controllers oversaw the local level. They created coordination among neighboring intersections and made local decisions that considered central plans and controls. The local and central controllers used a rolling horizon technique. The solutions generated by UTOPIA were sub-optimal because they combined online and offline optimizations. Note that SCATS, OPAC, and UTOPIA were not able to perform well in congested traffic networks.

Head, Mirchandani, and Sheppard (1992) proposed the Real-Time Hierarchical Optimizing Distributed Effective System (RHODES) with four levels. First, they predicted route characteristics by using a model introduced by Mirchandani and Soroush (1987). They determined a medium-term (hours, days, and weeks), stochastic traffic equilibrium state. Then, they optimized signal timing variables offline for the entire network. In the third level, they solved an intersection-level signal timing problem by signal scheduling and intersection dispatching. They then transferred the decisions to controllers. This system could respond to stochastic traffic demand, stochastic drivers' behavior, stochastic transit and pedestrian traffic patterns, speed variations, network blockages, and various network characteristics. The objective differed from user-oriented to system-oriented functions, depending on the congestion conditions of a network. Having a complicated structure that required many techniques and algorithms at each level was the main drawback of RHODES.

Later, Feng et al. (2015) developed an algorithm that optimized minimum and maximum green signal durations, phase sequences, and actual green durations in real time by using connected vehicle data. The algorithm minimized total delay and queue length by using VISSIM. The algorithm provided less delay in comparison to actuated signal control systems.

Yang and Jayakrishnan (2015) developed a network-level hierarchical signal control with two levels. At the strategy level, an integrated signal timing and traffic assignment formulation were optimized. Then, queue weight update and signal optimization problems were solved at the control level. This structure decreased the computational complexity of signal timing formulations, but the solutions might not provide a close optimality gap.

Lee, Wong, and Varaiya (2017b, 2017a) developed a hierarchical approach to find time-variant cycles and green signal durations by optimizing the group-based adaptive signal controllers of isolated intersections. They used a max-pressure control algorithm (Varaiya, 2013) that was implemented to change the signal plans by using traffic data. This approach outperformed several methods. However, information exchange among neighboring intersections was ignored, which might help significantly in improving the solution quality.

2.4   Distributed-Coordinated Control Systems

Distributed-coordinated control systems (DCCS) distribute the computations among various controllers. The controllers exchange information to determine signal plans. A similar decomposition concept is applied to traffic metering (Mohebifard and Hajbabaie, 2019b), dynamic speed harmonization problems (Tajalli and Hajbabaie, 2018a), and traffic assignment problems (Yahia et al., 2018).

Timotheou et al. (2013) proposed an algorithm with a distributed coordinated framework to solve the CTM-based signal timing problem for a network of intersections. They considered a

spatial distribution by intersection-level segmentations and temporal decomposition by discretizing the study period. The intersection-level problems exchanged information to find near-optimal solutions in possibly real time. A restrictive assumption was the consideration of only one-way streets. The sequential procedure for optimizing intersection-level problems was not also efficient.

Timotheou et al. (2015) employed the alternating direction method of multipliers (Boyd et al., 2011) to enhance the algorithm developed by Timotheou et al. (2013). This algorithm considered temporal and spatial decompositions with steps. The integer decision variables of signal timing formulation were relaxed, and the formulation was solved by using the alternating direction method of multipliers. Then, they retrieved the decision variables by rounding. The alternating direction method of multipliers was iterative with an unknown convergence rate. Hence, finding solutions in real time might fail. Also, the solutions after rounding might not satisfy all constraints.

Mehrabipour and Hajbabaie (2017) proposed a distributed and coordinated solution technique to solve a mixed-integer linear program and determine signal timing plans. The approach first distributed the formulation into intersection-level sub-problems to reduce the number of decision variables and constraints in each optimization. Then, the sub-problems shared information about their capacity and outgoing vehicles with their neighbors. The approach was real time and could generate solutions with a 2 percent gap with the optimal solution. However, the approach could not consider connected vehicles.

Islam and Hajbabaie (2017) implemented a distributed-coordinated approach for a link-level signal timing optimization formulation that assumed a 100 percent connected vehicle environment and the availability of road-side units at intersections. The objective function of the formulation was the maximization of throughput and the minimization of queue lengths. Required

information for the signal timing optimization formulation and coordination among isolated intersections were obtained by using VISSIM. However, the optimality properties of solutions were not discussed in this paper. Islam et al. (2020) extended this approach to work with various connected vehicle market penetration rates and observed significant improvement in traffic operation with an increase in the connected vehicle market penetration rate. Islam et al. (2020a) further extended this approach to minimize energy consumption while maintaining the same mobility level in a transportation network. Islam et al. (2021) studied the effects of connectivity and traffic observability on adaptive traffic control by using the approach presented in Islam et al. (2020). They also studied the effects of connectivity on travel time reliability following the approaches presented by Aghdashi et al. (2015, 2013), Hajbabaie et al. (2016), and Zegeer et al. (2014).

Mohebifard et al. (2019) presented a real-time, distributed, coordinated solution technique to find the optimal value of signal timing variables at signalized intersections and metering gates. The problem formulation was a mixed-integer, non-linear program. The approach decomposed the formulation by using intersection-level spatial decomposition. Each sub-network solved an optimization problem for an intersection and broadcasted the information to its neighboring sub-networks. This study also considered different penetration rates for connected vehicles. With a similar structure, Tajalli et al. (2020) proposed a distributed and coordinated technique to determine signal timing plans and speeds simultaneously.

Torabi et al. (2020) proposed a distributed, multi-agent approach to find signal timing plans. Each agent was in charge of one intersection, and it worked cooperatively with other agents to determine signal timing variables and consider network changes. The agent presented a new

signal plan if the intersection was congested and sent the information to other agents. Determining the severity of congestion and signal splits required determining many user input parameters.

Mirheli et al. (2018) developed a methodology to optimize the trajectory of connected automated vehicles while approaching a signal-free intersection. They found solutions by using a centralized architecture enabled with a Monte Carlo tree search algorithm following the approach presented by Mirheli and Hajibabai (2020). Mirheli et al. (2019) extended this approach by developing a distributed, consensus-based approach that utilized the existing computational power in each connected automated vehicle. As a result of using the distributed approach, trajectories were found in real time. Niroumand et al. (2020a) introduced a new phase, called the white phase, to accommodate connected automated vehicles more efficiently with fewer phase transitions. Niroumand et al. (2020b) studied the effects of the white phase on intersection operations and found significant improvements as a result of using it. Tajalli and Hajbabaie (2021) developed a lagrangian-based approach to simultaneously optimize signal timing parameters and the trajectories of connected automated vehicles and found significant improvement in traffic operations. Mohebifard and Hajbabaie (2021) developed a methodology to control the trajectory of connected automated vehicles in roundabouts in the presence of human-driven vehicles. They observed significant improvements in traffic operations when the trajectories of automated vehicles were optimized (Mohebifard and Hajbabaie, 2020).

2.5   Summary

Different approaches have been developed to solve signal timing optimization problems. CCSs are the only systems that can find globally optimal solutions. They are classified into exact and heuristic algorithms and solve signal timing problems centrally. Because network size and study periods affect the computational complexity of these systems, they are intractable even for

medium-sized networks. Then, there are HCSs and DCCSs, which are computationally more efficient, but their solutions are not optimal. Moreover, HCSs have complicated structures and need costly infrastructures, and DCCSs either have restrictive assumptions about network topology or are not real time. DCCSs can provide real-time solutions, but solutions are not near-optimal because sub-problems do not cooperate.

In this project, we proposed a distributed optimization and coordination algorithm that outperforms the existing algorithms by considering trade-offs between run-time and quality of solutions. While the solutions are generated in real-time, they are very close to optimal values. Furthermore, the algorithm can be applied to real-world and generalized networks and does not involve any restrictive assumptions about network topology and network size, assumptions about demand patterns, and requirements of costly infrastructure and complicated frameworks. The algorithm includes decomposition and coordination schemes to reduce the computational complexity of problems and avoid finding locally optimal solutions, respectively.

3   Problem Formulation

3.1   Introduction

The signal timing problem is modeled as a mixed-integer, non-linear, multi-modal formulation in this chapter. This formulation uses cell transmission model (CTM) traffic dynamics introduced by Daganzo (1994 and 1995). CTM creates a piecewise linear relationship between flow and density. CTM can simulate traffic flow across a network and capture shockwaves, queue properties, and speed. The research team had used CTM in previous work for speed harmonization (Tajalli and Hajbabaie, 2018b), traffic metering (Mohebifard and Hajbabaie, 2018a), and traffic assignment (Mehrabipour et al., 2019). In CTM, roadways or links are divided into homogeneous segments called cells, and the desired study period is discretized into several time intervals named time steps. Moreover, a cell length is equal to the distance that a vehicle can travel by free-flow speed in a single time step.

3.2   Notations

Figure 3.1shows an intersection cell representation with different types of cells, and table 3.1 shows the definitions of all sets, decision variables, and parameters used in the problem formulation.
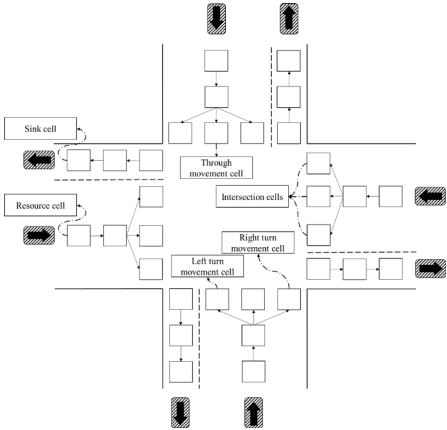

Figure 3.1. A cell-based representation of an intersection

Table 3.1. The definitions of sets, decision variables, and parameters

Sets:	
T	Set of all time steps
C	Set of all network cells
C_O	Set of all source cells
C_S	Set of all sink cells
C_I	Set of all intersection cells
I	Set of all intersections
Mo	Set of all classes of vehicle
B	Set of all buses
$P(b)$	Set of cells C in the path of bus $b \in B$
$E(k)$	Set of all intersection cells of intersection $k \in I$ with through and left-turning movements
$O(i)$	Set of all cells of intersection $k \in I$ with conflicting through and left-turning movements with movement $i \in E(k)$
R	Set of all concurrent through and right turn movement with adjacent movements
$P(i)$	Set of all predecessor cells of cell $i \in C$
$S(i)$	Set of all successor cells of cell $i \in C$
Decision Variables:	
g_i^t	A binary variable for signal status at cell $i \in C_I$ at time step $t \in T$. If green, $g_i^t = 1$. Otherwise $g_i^t = 0$.
$x_i^{t,m}$	Number of vehicles in cell $i \in C$ at time step $t \in T$ from class $m \in Mo$
$y_{ij}^{t,m}$	Number of vehicles going from cell $i \in C$ to downstream cell $j \in S(i)$ at time step $t \in T$ from class $m \in Mo$
v_i^t	Space-mean speed in cell $i \in C$ at time step $t \in T$
w_i^t	Dummy variable. It takes 1 if there is a bus in cell $i \in C$ at time step $t \in T$, and 0 otherwise
$w_{b,i}^t$	Dummy variable. It takes 1 if bus $b \in B$ is in cell $i \in C$ at time step $t \in T$, and 0 otherwise
χ_b^t	Position of a bus $b \in B$ at time step $t \in T$
$v_{b,i}^t$	Speed of a bus $b \in B$ at cell $i \in C$ at time step $t \in T$
Parameters:	
$d_i^{t,m}$	Demand on source cell $i \in C_O$ at time step $t \in T$ for class $m \in Mo$
F_i	Saturation flow rate of cell $i \in C$
f_i^t	Variable saturation flow rate of intersection cell $i \in C_I$ at time step $t \in T$
M_i	Maximum number of vehicles that cell $i \in C$ can accommodate
$r_i^{t,m}$	Turning percentage at intersection cell $i \in C_I$ at time step $t \in T$ for class $m \in Mo$
f	Reduction factor of saturation flow rate due to start-up lost time
L_i	Starting coordinate of cell $i \in C$
U_i	Ending coordinate of cell $i \in C$
χ_b^0	Initial position of bus $b \in B$ at time step $t = 0$
v_f^c	Free-flow speed of passenger cars
v_f^b	Free-flow speed of buses
ρ	Ratio of the backward shockwave speed to the free flow speed
L^m	Length of a vehicle of class $m \in Mo$
α^m	Passenger occupancy of a vehicle of class $m \in Mo$
ψ^m	Occupancy ratio of vehicle class $m \in Mo$ relative to passenger car
$\mathcal{M}$	Big number
ΔX	Length of a cell that is a space traveled by a vehicle in free-flow speed v_f^c within time horizon Δt
G_{max}	Maximum green time
G_{min}	Minimum green time
$\hat{x}_j^t$	Occupancy of cell $j \in C$ at time step $t \in T$ given from CTM simulation run
$\hat{y}_{ij}^t$	Number of vehicles flowing from cell $i \in C$ to cell $j \in S(i)$ at time step $t \in T$ obtained from CTM
δ_{ij}	Kronecker delta, $\delta_{ij} = 1$ when $i = j$. Otherwise $\delta_{ij} = 0$.
ΔT	Time horizon

3.3 Optimization Model

The objective function of the model is to minimize the total travel time when vehicle types are considered, as shown in Equation (3-1). This equation is the summation of the number of vehicles $x_i^{t,m}$ in all cells $i \in C$ except for sink cells $i \in C_S$, multiplied by the passenger occupancy α^m over all time steps $t \in T$ and vehicle types $m \in Mo$. The passenger occupancy α^m shows the possible number of people in each vehicle class.

$$\max \sum_{m \in Mo} \sum_{t \in T} \sum_{i \in C \backslash C_S} \alpha^m x_i^{t,m} \qquad (3\text{-}1)$$

Constraint (3-2) ensures the conservation of flow for ordinary cells $i \in C \setminus \{C_S, C_O\}$, source cells $i \in C_O$, and sink cells $i \in C_S$. Let δ_{ij} represents Kronecker delta, and $\delta_{ij} = 1$ when $i = j$. Otherwise, $\delta_{ij} = 0$. Entry demand to source cell $i \in C_O$ at time step $t \in T$ is denoted by d_i^t. In this conservation flow constraint, the difference in cell occupancy in consecutive time steps is equal to the difference in incoming and outgoing flows from that cell. Note that the incoming flow to source cells is demand, and sink cells do not have any outgoing flow.

$$(\delta_{il} + \delta_{is}) \sum_{k \in P(i)} y_{ki}^{t,m} - (\delta_{io} + \delta_{il}) \sum_{j \in S(i)} y_{ij}^{t,m} + \qquad \forall\, t \in T, i \in C, l \in C \setminus \{C_S, C_O\}, \quad (3\text{-}2)$$

$$d_i^t(\delta_{io}) = (\delta_{il} + \delta_{io} + \delta_{is})(x_i^{t+1} - x_i^t) \qquad\qquad o \in C_O, s \in C_S, m \in Mo$$

The outgoing flow $\sum_{j \in S(i)} y_{ij}^{t,m}$ of cell $i \in C \setminus \{C_S\}$ at time step $t \in T$ for class $m \in Mo$ should be less than or equal to the existing vehicles $x_i^{t,m}$ in that cell, as shown in Constraint (3-3). Because cell $i \in C \setminus \{C_S\}$ can have more than one successor cell $j \in S(i)$, the flow of all leaving links from the cell are accumulated to find the outgoing flow.

$$\sum_{j \in S(i)} y_{ij}^{t,m} \le x_i^{t,m} \qquad\qquad \forall\, t \in T, i \in C \setminus \{C_S\}\, m \in Mo \quad (3\text{-}3)$$

The outgoing flow $\sum_{m\in Mo}\sum_{j\in S(i)} y_{ij}^{t,m}$ of cell $i \in C$ at time step $t \in T$ is limited to the saturation flow rate F_i of the cell when no bus is available in that cell, as shown in Constraint (3-4). Note that the saturation flow rate is decreased by a factor of ρ_i when at least one bus exists in that cell. Constraint (3-5) forces a similar limitation but for the incoming flow $\sum_{m\in Mo}\sum_{i\in P(j)} y_{ij}^{t,m}$ to cell $j \in C$.

$$\sum_{m\in Mo}\sum_{j\in S(i)} y_{ij}^{t,m} \leq w_i^t \rho_i F_i + (1 - w_i^t)F_i \qquad\qquad \forall\, i \in C, t \in T \qquad (3\text{-}4)$$

$$\sum_{m\in Mo}\sum_{i\in P(j)} y_{ij}^{t,m} \leq w_i^t \rho_i F_j + (1 - w_i^t)F_j \qquad\qquad \forall\, j \in C, t \in T \qquad (3\text{-}5)$$

The incoming flow $\sum_{m\in Mo}\sum_{i\in P(j)} y_{ij}^{t,m}$ to cell $j \in C$ at time step $t \in T$ should be limited by the available capacity of that cell, $M_j - \sum_{m\in M}\psi^m \times x_j^{t,m}$, as shown in Constraint (3-6). The capacity of cell $j \in C$ is denoted by M_j. Also, the vacant capacity of cell $j \in C$ that can be occupied by vehicles is shown by $M_j - \sum_{m\in M}\psi^m \times x_j^{t,m}$. The conversion factor to determine the occupied space of the cell in terms of passenger cars is denoted by ψ^m. Note that capacity of source and sink cells is set to a big arbitrary number.

$$\sum_{m\in Mo}\sum_{i\in P(j)} y_{ij}^{t,m} \leq M_j - \sum_{m\in M}\psi^m \times x_j^{t,m} \qquad\qquad \forall\, t \in T, j \in C \qquad (3\text{-}6)$$

Equation (3-7) computes the space-mean speed v_i^{t+1} for cell $i \in C$ at time step $t + 1$. This speed is found by dividing the available space in a cell by the duration of the time step, as proposed by Aziz (2019). The available space is the difference between the cell length ΔX and the length that is occupied by vehicles. The occupied length is found by multiplying the average vehicle length L^m by the difference in the number of vehicles $x_i^{t,m}$ in cell $i \in C^n$ and the total number of vehicles $\sum_{j\in \Gamma(i)} y_{ij}^{t,m}$ that exit that cell toward cell $j \in \Gamma(i)$ at time step $t \in T$.

$$v_i^{t+1} = \frac{\Delta X - \left[\sum_{m \in M}\left(x_i^{t,m} - \sum_{j \in \Gamma(i)} y_{ij}^{t,m}\right) \times L^m\right]}{\Delta t} \qquad \forall t \in T, i \in C \qquad (3\text{-}7)$$

Previous research on multiclass CTM (Liu et al., 2015; Mesa-Arango and Ukkusuri, 2014) has assumed that buses move slower than passenger cars in the system. Equation (3-8) ensures that the space-mean speed $v_{b,i}^t$ of bus $b \in B$ in non-intersection cell $i \in C \backslash C_I$ at time step t is the minimum of the space-mean speed v_i^t in that cell and the free flow speed of buses v_f^b.

$$v_{b,i}^t = \min(v_i^t, v_f^b) \qquad \forall t \in T, i \in C \backslash C_I, b \in B \qquad (3\text{-}8)$$

Equation (3-9) finds the space-mean speed $v_{b,i}^t$ of a bus at intersection cell $i \in C_I$ based on the signal status $g_i^{t,n}$ of the corresponding cell. The speed of a bus at an intersection cell takes the minimum of space mean speed in that cell and the free-flow speed of buses if the signal is green; otherwise, it will be zero.

$$v_{b,i}^t = \min(g_i^t \times v_i^t, v_f^b) \qquad \forall t \in T, i \in C_I, b \in B \qquad (3\text{-}9)$$

Equation (3-10) updates the position χ_b^t of bus $b \in B$ at time step $t \in T$ using its position χ_b^{t-1} and speed $v_{b,i}^{t-1}$ in the previous time step. When bus $b \in B$ exists in cell $i \in P(b)$ at time step t, $w_{b,i}^t$ will be 1.0, and we ensure that the position χ_b^t is updated.

$$\chi_b^t = \chi_b^{t-1} + w_{b_i}^{t-1} \times v_{b,i}^{t-1} \times \Delta t \qquad \forall t \in T, i \in P(b), b \in B \qquad (3\text{-}10)$$

Equation (3-11) sets an initial position $\chi_b^{t=0}$ for each bus in the network.

$$\chi_b^{t=0} = \chi_b^0 \qquad \forall b \in B \qquad (3\text{-}11)$$

Constraints (3-12) and (3-13) force the value of the dummy variable w_i^t to be 1.0 when at least one bus exists in cell $i \in C$ at time step $t \in T$. Note that $w_{b,i}^t$ takes 1.0 when there is a bus $b \in B$ in cell $i \in P(b)$ at time step $t \in T$. In fact, it is 1.0 when the bus is located in interval $[L_i, U_i]$.

$$w_i^t \leq \sum_{b \in B^n} w_{b,i}^t \qquad\qquad \forall t \in T, \forall i \in C, m = \text{bus} \qquad (3\text{-}12)$$

$$w_i^t \geq w_{b,i}^t \qquad\qquad \forall t \in T, \forall i \in C, \forall b \in B \qquad (3\text{-}13)$$

Equation (3-14) finds the number of buses $x_i^{t,m=\text{bus}}$ in cell $i \in C$ at time step $t \in T$.

$$x_i^{t,m} = \sum_{b \in B} w_{b,i}^t \qquad\qquad \forall t \in T, \forall i \in C, m = \text{bus} \qquad (3\text{-}14)$$

The turning movement proportion $r_j^{t,m}$ at intersection cell $j \in C_I$ at time step $t \in T$ for vehicle type $m \in Mo$ is assumed to be given as an input. We denote C_I as the set of intersection cells. The total outflow $\sum_{k \in S(i)} y_{ik}^{t,m}$ of intersection cell $i \in C_I$ is distributed among all outgoing links $y_{ij}^{t,m}, i \in P(j)$ originating from that cell according to turning proportion $r_j^{t,m}$, as shown in Constraint (3-15). The turning percentages can be constant or variable over time steps. In the constant case, the time index should be removed.

$$y_{ij}^{t,m} \leq r_j^{t,m} \sum_{k \in S(i)} y_{ik}^{t,m} \qquad\qquad \forall\, t \in T, j \in C_I, i \in P(j) \qquad (3\text{-}15)$$

Constraint (3-16) ensures that at most two non-conflicting movements from the set of through-movements and left-turning movements $E(k)$ at intersection $k \in I$ will receive a green signal indication at time step $t \in T$. We then limit only one of the conflicting movements to have a green indication for each time step in Constraint (3-17). Equation (3-18) sets the signal indication of the concurrent through-movements and right turning movements $(i, j) \in R$ equal.

28

$$\sum_{j \in E(k)} g_j^t \leq 2 \qquad\qquad \forall k \in I, t \in T \qquad\qquad (3\text{-}16)$$

$$g_i^t + g_j^t \leq 1 \qquad\qquad \forall k \in I, i \in E(k), j \in O(i), t \in T \qquad (3\text{-}17)$$

$$g_i^t = g_j^t \qquad\qquad \forall (i,j) \in R, t \in T \qquad\qquad (3\text{-}18)$$

Equation (3-19) determines the variable saturation flow rate f_i^t of intersection cells based on signal status. The variable saturation flow rate f_i^t is set to be the constant saturation flow rate F_i when the signal of the cell is green $g_i^t = 1$.

$$q_i^t = g_i^t F_i \qquad\qquad \forall i \in C_I, t \in T \qquad\qquad (3\text{-}19)$$

Constraints (3-20) and (3-21) ensure a maximum and minimum green time for through-movements and left-turning movements in cell $i \in C_I$, respectively.

$$\sum_{\zeta=t}^{t+G_{max}+1} g_i^\zeta \leq G_{max} \qquad\qquad \forall k \in I, i \in E(k), t \in T, t \leq |T| - G_{max} \quad (3\text{-}20)$$

$$\sum_{\zeta=t+1}^{t+G_{min}} g_i^\zeta \geq (g_i^{t+1} - g_i^t) G_{min} \qquad \forall k \in I, i \in E(k), t \in T, t \leq |T| - G_{min} \quad (3\text{-}21)$$

When a signal indication changes from red to green, the saturation flow rate in intersection cells is reduced to account for the start-up lost time, as shown in Constraint (3-22). Let f denote a reduction factor in the saturation flow rate due to a start-up lost time. When $g_i^{t+1} = 1$ and $g_i^t = 0$, Constraint (3-22) is activated, and the outflow $\sum_{m \in Mo} \sum_{j \in S(i)} y_{ij}^{t+1,m}$ of cell $i \in C_I$ at time step $t \in T$ is limited to a smaller number than F_i. This means that the outflow of the corresponding cell is reduced.

$$\sum_{m \in Mo} \sum_{j \in S(i)} y_{ij}^{t+1,m} \leq F_i - F_i f(g_i^{t+1} - g_i^t) \qquad \forall t \in T, i \in C_I \qquad (3\text{-}22)$$

Constraints (3-23) and (3-24) ensure the non-negativity of the number of vehicles x_i^t in cell $i \in C$ at time step $t \in T$ and the flow of vehicles y_{ij}^t between cells for time step $t \in T$.

$$x_i^{t,m} \geq 0 \qquad\qquad \forall\, t \in T, i \in C, m \in Mo \qquad\qquad (3\text{-}23)$$

$$y_{ij}^{t,m} \geq 0 \qquad\qquad \forall\, t \in T, i \in C \backslash \{C_s\}, j \in P(i), m \in Mo \qquad (3\text{-}24)$$

3.4   Summary

This chapter presented a non-linear, mixed-integer, multi-modal optimization model for the timing of signalized intersections. The objective function and constraints of the model are described in detail in this chapter. The main decision variables are the signal timing plans for a network of intersections. The following chapters describe the development of algorithms to optimize this mixed-integer program for determining signals.

4   Distributed-Coordinated Algorithm

4.1   <u>Introduction</u>

This chapter presents a solution technique with spatial and temporal decompositions to optimize signal timing parameters in real-time and generate high-quality solutions. The proposed algorithm decomposes the mixed-integer formulation presented in Chapter 3 into several sub-problems by relaxing constraints that connect links between each pair of adjacent intersections. In other words, this decomposition transforms a network-level optimization model into several intersection-level models. Hence, this intersection-level decomposition reduces the number of decision variables in each problem and provides the possibility of parallelism.

Creating a coordination scheme among sub-problems is key to avoiding locally optimal solutions. Information for the coordination is generated by using a CTM simulation run. Hence, the entry demand for each intersection and the available capacity of receiving cells at each intersection are determined and updated on the basis of the CTM simulation outputs. Note that a lack of information on the available capacity of each intersection can create queue spillback because the objective function of the formulation is accumulated throughput maximization. The coordination among sub-problems is handled by the cloud components, while the optimization of sub-problems is taken care of by the fog components.

A rolling horizon technique is also incorporated in the algorithm to respond to online information and instantaneous changes in traffic demand and link capacity. This technique helps decrease the computational complexity of the formulation and provides real-time solutions. The decomposition of the formulation and data transition procedures are described in the following sections with a simple example, shown in figure 4-1. This figure shows a network of two intersections with cell-based discretization. The network is shown before decomposition as it is

simulated (figure 4.1a) and after decomposition as it is optimized (figure 4.1b). The following

sections are explained for sub-problem A. The same discussion is valid for sub-problem B.
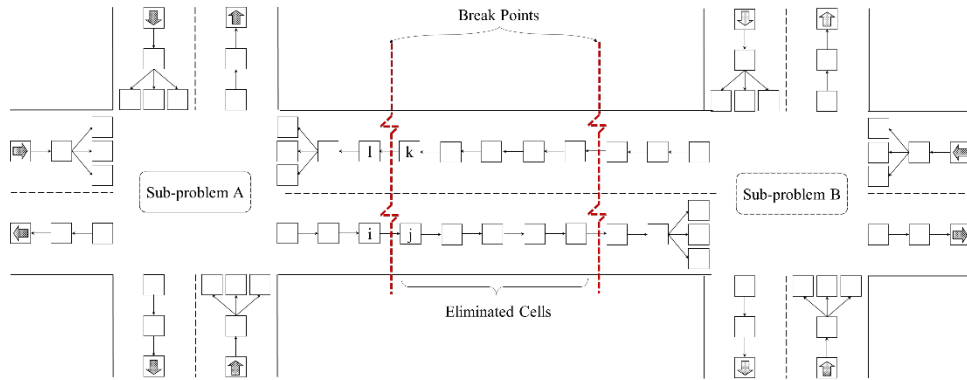
(a) Before decomposition
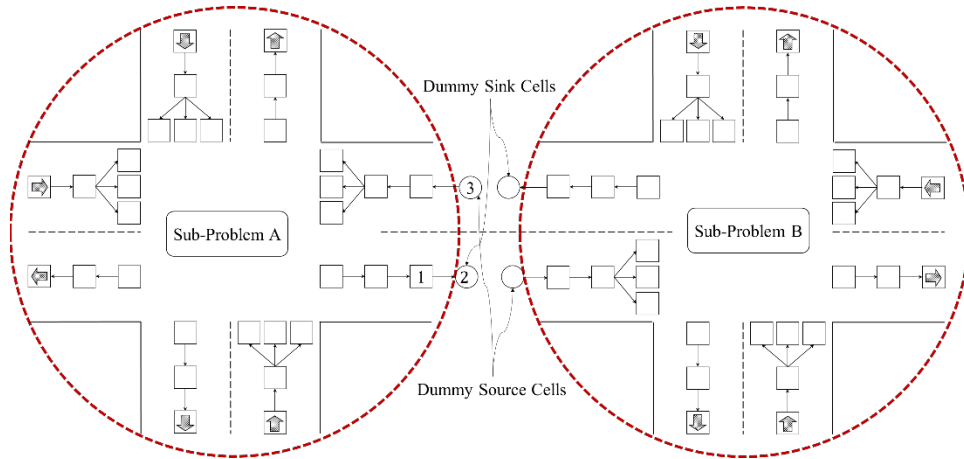
(b) After decomposition

Figure 4.1. The schematics diagram of distributed optimization and distributed coordination

4.2   Distribution of Problem Formulation

To create intersection-level sub-problems, the links between cells i and j and cells k and l

are disconnected. Then, each sub-problem represents a mixed-integer linear program for one

intersection and is optimized separately to determine the signal timing variables of the

corresponding intersection for a prediction period. Dummy source and sink cells should be added

to send in and absorb vehicles at the boundaries, respectively. To maintain the conservation of

flow, Constraint (3-2) is added to each sub-problem for new cells. New sets are defined as in table 4-1 to be able to write a generalized formulation for sub-problems.

Table 4.1. Sets for a generalized formulation for sub-problems

Sets:	
H	Set of all time steps in a time horizon
S	Set of all sub-problems
C^s	Set of all network cells in sub-problem $s \in S$
C_O^s	Set of all source cells in sub-problem $s \in S$
C_S^s	Set of all sink cells in sub-problem $s \in S$
C_1^s	Set of all cells in sub-problem $s \in S$ except for its source cells $i \in C_O^s$
C_2^s	Set of all cells in sub-problem $s \in S$ except for its sink cells $i \in C_S^s$
C_3^s	Set of all cells in sub-problem $s \in S$ except for its source and sink cells $i \in C_O^s \cup C_S^s$
$P(b)^s$	Set of cells in the path of bus $b \in B$ in sub-problem $s \in S$
$E(k)^s$	Set of all intersection cells in sub-problem $s \in S$ with through and left-turning movements
$O(i)^s$	Set of all cells in sub-problem $s \in S$ with conflicting through and left-turning movements with movement $i \in E(k)$
R^s	Set of all concurrent through and right turn movement with adjacent movements in sub-problem $s \in S$
$P(i)^s$	Set of all predecessors of cell $i \in C^s$ in sub-problem $s \in S$
$S(i)^s$	Set of all successors of cell $i \in C^s$ in sub-problem $s \in S$

The following optimization program shows the objective function and set of constraints for sub-problem $s \in S$. This program is a modified version of the problem formulation described in Chapter 3. The formulations differ in the definition of sets.

$$\max \sum_{m \in Mo} \sum_{t \in H} \sum_{i \in C_2^s} \alpha^m x_i^{t,m} \tag{4-1}$$

$$(\delta_{il} + \delta_{is}) \sum_{k \in P(i)^s} y_{ki}^{t,m} - (\delta_{io} + \tag{4-2}$$
$$\forall \, t \in H, i \in C^s, l \in C^s \setminus \{C_O^s, C_O^s\},$$
$$\delta_{il}) \sum_{j \in S(i)^s} y_{ij}^{t,m} + d_i^t(\delta_{io}) = (\delta_{il} + \delta_{io} +$$
$$o \in C_O^s, s \in C_O^s, m \in Mo$$
$$\delta_{is})(x_i^{t+1} - x_i^t)$$

$$\sum_{j \in S(i)^s} y_{ij}^{t,m} \le x_i^{t,m} \qquad \forall \, t \in H, i \in C_2^s \ m \in Mo \tag{4-3}$$

$$\sum_{m \in Mo} \sum_{j \in S(i)^s} y_{ij}^{t,m} \le w_i^t \rho_i F_i + (1 - w_i^t)F_i \qquad \forall \, i \in C_2^s, t \in H \tag{4-4}$$

$$\sum_{m\in Mo}\sum_{i\in P(j)^s} y_{ij}^{t,m} \leq w_i^t \rho_i F_j + (1-w_i^t)F_j \qquad \forall j \in C_1^s, t \in H \qquad (4\text{-}5)$$

$$\sum_{m\in Mo}\sum_{i\in P(j)^s} y_{ij}^{t,m} \leq M_j - \sum_{m\in M}\psi^m \times \qquad \qquad (4\text{-}6)$$

$$x_j^{t,m} \qquad \qquad \forall t \in H, j \in C_1^s$$

$$v_i^{t+1} \qquad \qquad (4\text{-}7)$$

$$= \frac{\Delta X - \left[\sum_{m\in M}\left(x_i^{t,m} - \sum_{j\in\Gamma(i)} y_{ij}^{t,m}\right) \times L^m\right]}{\Delta t} \qquad \forall t \in H, \forall i \in C^s$$

$$\chi_b^{t=0} = \chi_b^0 \qquad \forall b \in B \qquad (4\text{-}8)$$

$$w_i^t \leq \sum_{b\in B} w_{b,i}^t \qquad \forall t \in H, \forall i \in C^s, m = bus \qquad (4\text{-}9)$$

$$w_i^t \geq w_{b,i}^t \qquad \forall t \in H, \forall i \in C^s, \forall b \in B^n \qquad (4\text{-}10)$$

$$x_i^{t,m} = \sum_{b\in B} w_{b,i}^t \qquad \forall t \in H, \forall i \in C^s, m = bus \qquad (4\text{-}11)$$

$$\sum_{j\in E(s)} g_j^t \leq 2 \qquad \forall t \in H \qquad (4\text{-}12)$$

$$g_i^t + g_j^t \leq 1 \qquad \forall i \in E(s), j \in O(i), t \in H \qquad (4\text{-}13)$$

$$g_i^t = g_j^t \qquad \forall (i,j) \in R^s, t \in H \qquad (4\text{-}14)$$

$$q_i^t = g_i^t F_i \qquad \forall i \in C_I^s, t \in H \qquad (4\text{-}15)$$

$$\sum_{\zeta=t}^{t+G_{max}+1} g_i^\zeta \leq G_{max} \qquad \forall i \in E(s), t \in H, t \leq |T| - G_{max} \qquad (4\text{-}16)$$

$$\sum_{\zeta=t+1}^{t+G_{min}} g_i^\zeta \geq (g_i^{t+1} - g_i^t)G_{min} \qquad \forall i \in E(s), t \in H, t \leq |T| - G_{min} \qquad (4\text{-}17)$$

$$y_{ij}^{t,m} \leq r_j^{t,m} \sum_{k\in S(i)^s} y_{ik}^{t,m} \qquad \forall t \in H, j \in C_I^s, i \in P(j)^s \qquad (4\text{-}18)$$

$$\sum_{m\in Mo}\sum_{j\in S(i)^s} y_{ij}^{t+1,m} \leq F_i - F_i f(g_i^{t+1} - g_i^t) \qquad \forall t \in H, i \in C_I^s \qquad (4\text{-}19)$$

$$x_i^{t,m} \geq 0 \qquad \forall t \in H, i \in C^s, m \in Mo \qquad (4\text{-}20)$$

$$\forall t \in H, i \in C^s \backslash \{C_s\}, j \in P(i)^s, m \qquad (4\text{-}21)$$

$$y_{ij}^{t,m} \geq 0$$

$$\in Mo$$

$$v_{b,i}^t \le v_i^t \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s, b \in B \qquad (4\text{-}22)$$

$$v_{b,i}^t \le v_f^b \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s, b \in B \qquad (4\text{-}23)$$

$$v_{b,i}^t \ge v_i^t - \mathcal{M}(1 - \beta_{1i}^t) \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s, b \in B \qquad (4\text{-}24)$$

$$v_{b,i}^t \ge v_f^b - \mathcal{M}(1 - \beta_{2i}^t) \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s, b \in B \qquad (4\text{-}25)$$

$$\beta_{1i}^t + \beta_{2i}^t = 1 \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s, b \in B \qquad (4\text{-}26)$$

$$\beta_{1i}^t, \beta_{2i}^t \in \{0,1\} \qquad\qquad \forall t \in H, i \in C^s \backslash C_I^s \qquad (4\text{-}27)$$

$$\mu_i^t \le g_i^{t,n} \times v_f^b \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}28)$$

$$\mu_i^t \le v_i^t \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}29)$$

$$\mu_i^t \ge v_i^t - (1 - g_i^{t,n}) \times v_f^b \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}30)$$

$$\mu_i^t \ge 0 \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}31)$$

$$v_{b,i}^t \le \mu_i^t \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}32)$$

$$v_{b,i}^t \le v_f^b \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}33)$$

$$v_{b,i}^t \ge \mu_i^t - \mathcal{M}(1 - \gamma_{1i}^t) \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}34)$$

$$v_{b,i}^t \ge v_f^b - \mathcal{M}(1 - \gamma_{2i}^t) \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}35)$$

$$\gamma_{1i}^t + \gamma_{2i}^t = 1 \qquad\qquad \forall t \in H, i \in C_I^s, b \in B \qquad (4\text{-}36)$$

$$\gamma_{1i}^t, \gamma_{2i}^t \in \{0,1\} \qquad\qquad \forall t \in H, i \in C_I^s \qquad (4\text{-}37)$$

$$\chi_b^t = \chi_b^{t-1} + z_i^t \times \Delta t \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}38)$$

$$z_i^t \le w_{b,i}^t \times v_f^b \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}39)$$

$$z_i^t \le v_{b,i}^t \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}40)$$

$$z_i^t \ge v_{b,i}^t - (1 - w_{b,i}^t) \times v_f^b \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}41)$$

$$z_i^t \ge 0 \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}42)$$

$$L_i \le \chi_b^t + \mathcal{M}(1 - w_{b,i}^t) \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}43)$$

$$L_i \geq \chi_b^t - \mathcal{M}(w_{b,i}^t + \eta_{1i,b}^t) \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}44)$$

$$\chi_b^t < U_i + \mathcal{M}(1 - w_{b,i}^t) \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}45)$$

$$\chi_b^t > U_i - \mathcal{M}(w_{b,i}^t + \eta_{2i,b}^t) \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}46)$$

$$w_{b,i}^t + \eta_{1i,b}^t + \eta_{2i,b}^t = 1 \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}47)$$

$$\eta_{1i,b}^t + \eta_{2i,b}^t \leq 1 \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}48)$$

$$\eta_{1i,b}^t, \eta_{2i,b}^t \in \{0,1\} \qquad\qquad \forall t \in H, i \in P(b)^s, b \in B \qquad (4\text{-}49)$$

Two new sets of constraints should be added to the formulation of each sub-problem to create a coordination scheme. These constraints are defined in the next section. After the constraints have been added, they are updated before sub-problems are optimized with the newly generated information from the integrated CTM simulation.

4.3   Coordination of Sub-problems

Coordination of signal timing decisions among sub-problems is needed to push solutions toward system-level optimality. Sub-problems coordinate their decisions by exchanging information on the number of incoming vehicles, the available capacity of receiving cells, and signal timing parameters at neighboring sub-problems. The required information can be obtained by a CTM simulation run and by introducing new constraints to each sub-problem.

Three key cells in figure 4-1.b are 1, 2 and 3, and cells i, j, k, and l in figure 4-1a. Cell 2 is a dummy cell that represents a sink cell for the sub-problem A. Cell 3 is a dummy cell that represents a source cell for sub-problem A.

We define Constraints (4-50) and (4-51) to ensure that each sub-problem does not send more vehicles than the available capacity of its neighbors. Hence, the outgoing vehicles from

intersection A, y_{12}^t, are limited to the available capacity of cell $j \in C,\ M_j - \hat{x}_j^t$. We define $\hat{x}_j^t$ as the occupancy of cell $j \in C$ at time step $t \in T$ obtained from the CTM simulation run.

$$y_{12}^t \leq M_j - \hat{x}_j^t \qquad\qquad \forall\, t \in T \quad (4\text{-}50)$$

The next piece of information is about the demand at the dummy source cells of each sub-problem. Constraint (4-8) updates demand at the boundary cells of an intersection according to the number of vehicles coming from neighboring intersections. In the example of figure 4-1, the demand at cell 3 is equivalent to the flow of vehicles $\hat{y}_{kl}^t$ between cell k and l given by the CTM simulation run for time step $t \in T$. The entry demand at the dummy source cells of each sub-problem needs to be known.

$$d_3^t = \hat{y}_{kl}^t \qquad\qquad \forall\, t \in T \quad (4\text{-}51)$$

4.4   Model Predictive Control Algorithm

Solving each sub-problem for the entire study period is still computationally expensive, and the algorithm may not generate solutions in real time. Moreover, we are unable to consider dynamic changes in demand, sudden incidents, and any other network changes by considering the entire study period for optimizing sub-problems. Therefore, we use a model predictive control (MPC) algorithm to avoid these possible issues.

In this algorithm, the study period is discretized into smaller horizons, and each sub-problem is optimized only for the determined horizon. After all sub-problems have been optimized in parallel for the determined horizon, the algorithm moves the horizon forward one time-step. In the next round, the sub-problems are optimized for the new horizon. The solution of the first time-step in each horizon is stored to be used as the final solution. After putting all solutions together and covering all time-steps, we simulate the network and find the flow and occupancy of vehicles.

4.5   Overall Framework

The proposed algorithm includes three main steps:

Step 1› Initialization

Step 2› Integrated CTM simulation

Step 3› Optimization of sub-problems.

Figure 4-2 shows a schematic diagram of the proposed algorithm. The algorithm starts with $t = 1$, and initial signal timing parameters are given as user inputs. These signal timing parameters should satisfy the constraints of the signal timing optimization model. *Step 1* iterates over multiple iterations between the optimization of sub-problems and simulation steps without moving a time horizon to improve the initial signals. Next, *Step 2* receives the signal timing parameters from *Step 1* as input and simulates an entire network for horizon t to $t + \Delta T$. The outputs of this step are $\hat{x}_i^t$ for all cells $i \in C$ at time step $t \in T$ and $\hat{y}_{ij}^t$ for all links between cell $i \in C$ and cell $j \in S(i)$ at time step $t \in T$. Using these outputs, Constraints (4-50) and (4-51) are updated in the formulation of each sub-problem. Then, in *Step 3*, all sub-problems are optimized separately and simultaneously for horizon t to $t + \Delta T$. The optimized signal timing plans from the sub-problems are used for the next horizon of the simulation.

Moreover, before a time horizon is moved forward, whether the end of the study period has been reached is checked, and if it has not been reached, then the time horizon is moved a time-step forward. Otherwise, the algorithm is stopped. The signal timing parameters of the first time-step of each horizon are stored to form the final solution generated by the algorithm.
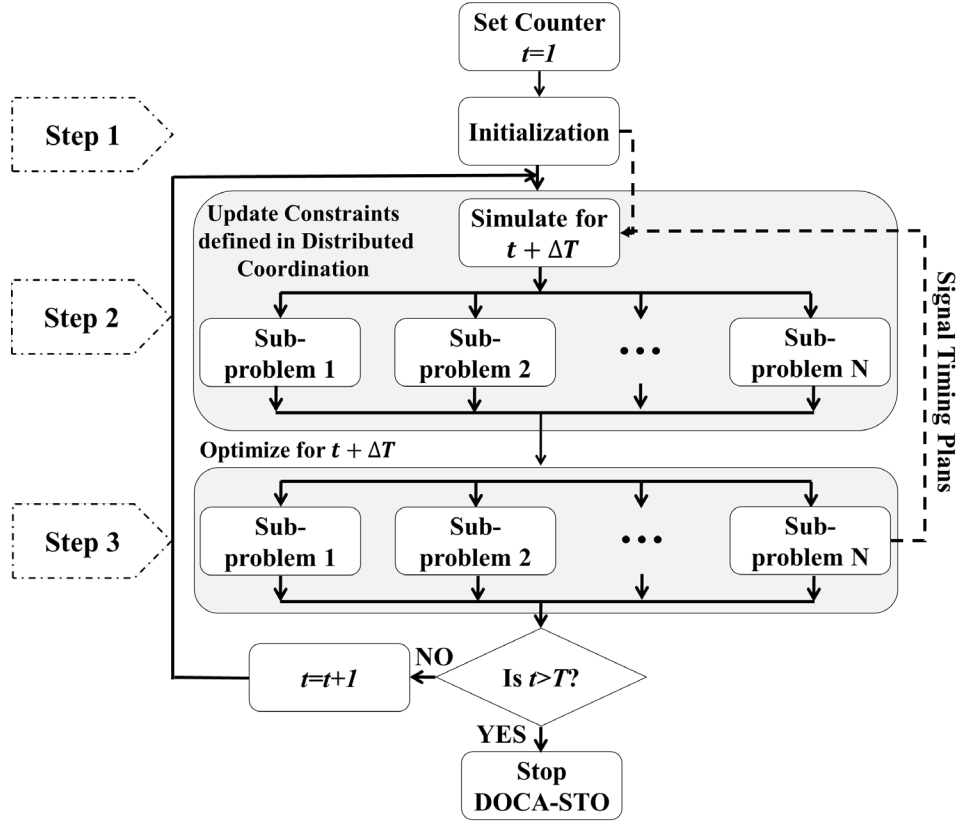
Figure 4.2. The framework of DOCA-STO

4.6   Summary

This chapter explains how the distributed-coordinated algorithm works. The two main components of the algorithm, including the distribution of the formulation and the coordination scheme, are described in detail. In the distribution section, the formulation is decomposed into several sub-problems, and each sub-problem represents one intersection. The sub-problem also has the decision variables and constants within its assigned intersection. These sub-problems are optimized in the fog components. We also proposed a temporal decomposition. In the coordination section, we exchange some information among sub-problems such as the available capacity of receiving sub-problems and incoming demand to each sub-problem. The computations are done in the cloud component. Finally, this chapter provides the details of implementing the approach in

real time by using a model predictive algorithm. The overall framework is also explained to help

understand all sections of this algorithm together.

5   Micro-Simulation Test-Bed Development

5.1   Introduction

This chapter describes the development of a microsimulation testbed to evaluate the performance of the developed methodology.

5.2   System State Estimation

The state of the system should be known at the beginning of each horizon: the occupancy of all cells at the first time step of each horizon is input and defines the initial state of the system. The locations of vehicles in the network are converted into cell occupancies for this purpose. We introduced an approach to use location information of connected vehicles (CV) and vehicle counts of loop detectors to estimate cell occupancies.

Consider the link in figure 5-1. The link includes several CVs that are equipped with onboard units and can transmit their location to road-side units in a network. If the penetration rate of connected vehicles is 100 percent, then the state estimation is straightforward. The location of vehicles will be mapped to their corresponding cells, and the cell occupancies will be the summation of vehicles in the cells. However, the locations of unequipped vehicles should be estimated when the CV penetration rate is less than 100 percent.
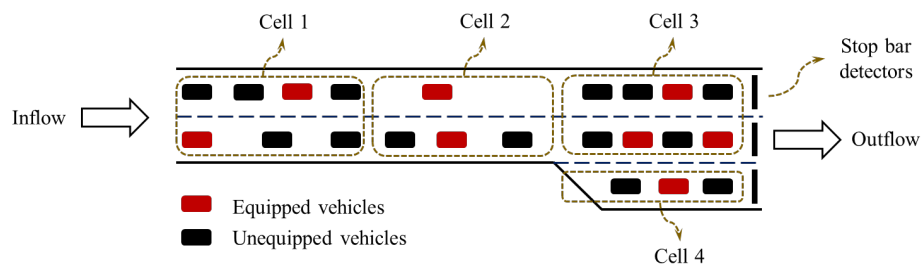

Figure 5.1. A link with the corresponding cells including both equipped (CV) and unequipped (non-CV) vehicles (Mohebifard and Hajbabaie, 2018b)

The proposed approach for system state estimation is the integration of two different estimation techniques. The first technique approximates the density distribution of vehicles in each network link with a vehicle sample that includes CVs. The second technique uses the flow feasibility and conservation constraints of CTM to estimate cell occupancies. These two estimations are averaged on the basis of the market penetration rate of CVs.

5.2.2   State Estimation Using CV Location Information

The distribution of vehicles in each network link can be estimated by using Equation (5-1). In the equation, xe_i^t is the number of CVs in cell $i \in C$ at time step $t \in T$. By dividing xe_i^t over the total number of CVs in the link $l \in L$ that contains cell i, the distribution of CVs in the link can be found. In this equation, L represents the set of all network links. By multiplying the estimated distribution of CVs by the total number of vehicles in the link V_l^t, the distribution of both CVs and non-CVs can be found. The total number of vehicles in each link can be estimated by tracking the inflow and outflow of each link such that V_l^t will be equal to the cumulative outflow minus the cumulative inflow of each link up to time step $t \in T$, assuming that all links are equipped with stop-bar detectors.

$$x_i^{t,CV} = V_l^t \frac{xe_i^t}{\sum_{i \in l} xe_i^t} \qquad\qquad \forall i \in C, l \in L, t \in T \qquad\qquad (5\text{-}1)$$

Note that as the penetration rate of CVs increases, the distribution estimation error decreases because the sample size of CVs increases and better represents the distribution of vehicles in a link. However, the estimation error increases with low CV penetration rates. To address this issue, we used the CTM flow conservation and feasibility equations to adjust the estimations given low CV penetration rates.

5.2.3 State Estimation Using CTM Flow Feasibility and Conservation Principles

In this technique, we used the flow feasibility and conservation equations to track cell occupancies in each link. For more clarification, consider figure 5-1 again. The cell occupancies will be initialized to zero at the beginning of the study period when the network is empty. As vehicles enter the network, the loop detectors can track vehicles entering and exiting each link. This information can be utilized to simulate each link with the demand rate equal to the total vehicle counts of its upstream stop-bar detectors. If we assume that the flow of vehicles follows the flow conservation and feasibility equations (3-7)-(3-13), then we can estimate the cell occupancies $x_i^{t,CTM}$ for each cell $i \in C$ and time step $t \in T$. Note that this approach does not require any information from CVs and relies on only vehicle counts of loop detectors, assuming that all links are equipped with stop-bar detectors.

5.3 Combined CV and CTM State Estimations

Once the estimations of $x_i^{t,CV}$ and $x_i^{t,CTM}$ have been found with the proposed approaches, the estimations are combined by using Equation (5-2). In this equation, the state estimations are adjusted by the penetration rate $0 \leq PR \leq 1$ of CVs. This adjustment gives higher weight to estimations based on CVs when the penetration rate is high because the estimations are more accurate, but with low penetration rates, the estimations of the CTM-based approach will be given higher weight. Moreover, the estimated cell occupancies cannot increase the capacity of the cell N_i.

$$x_i^t = \min\{PR x_i^{t,CV} + (1 - PR)x_i^{t,CTM}, N_i\} \qquad \forall i \in C, l \in L, t \in T \tag{5-2}$$

5.4 Summary

This chapter provides a detailed explanation for estimating the initial cell occupancies. Once we have the estimation, signal timings for each sub-problem can be optimized. In the

distributed optimization, a network will be divided into smaller sub-networks. The central optimization problem will be also decomposed into several sub-problems representing each sub-network. Each of the sub-problems optimizes signals for its corresponding sub-network. The distributed coordination component shares information about the predicted number of vehicles that enter the sub-networks and their available capacity for receiving vehicles between all sub-networks. This information will be incorporated as several constraints into the sub-problems so that the sub-problems can coordinate their decisions and push their local solutions toward the globally optimal solutions.

6   Implementation of a Cloud-Fog Architecture

6.1   <u>Introduction</u>

The objective of this portion of the project was to add a communication aspect to the traffic controller optimization algorithm to emulate real-life scenarios. The fog level consists of fog control units (FCUs), each connected to a traffic controller (TC) at each intersection (later on, for the CV stage, they will also be connected to an RSU). The cloud control unit (CCU) is linked to all FCUs and collects information from each periodically over the network.

We implemented three different architectures: centralized cloud-based, distributed cloud-based, and fully distributed. In the first architecture, each FCU pulls data from its associated traffic controller and communicates information such as vehicular traffic, signal phasing and timing, and intersection map to the CCU. The CCU performs the optimization and sends back the calculated results comprising the next signal phases to each FCU individually.

In the second architecture, the optimization process or computations are distributed, but the communication is centralized. In other words, at the start of each interval before the next phase calculation, the CCU pulls all the information from the FCUs. However, the FCUs also send an additional "REQUEST" packet to the CCU. This packet is one of the fields that contain the list of neighboring FCUs from which information is required to calculate the next phases. The CCU responds to each REQUEST individually by sending the information of interest to each FCU separately. Later on, multicast broadcasting may be used to make the process more efficient. The optimizer is run locally (either on the FCU or by the TC itself).

In the fully distributed architecture, each FCU sends a simple "REQUEST" packet to each FCU from which its associated traffic controller requires information. At the beginning of every cycle, the FCUs respond to the requests from neighboring FCUs. As soon as each FCU has

received the necessary information, the optimizers are run locally to calculate the next phases. Please note that both communication and computation are distributed in this scenario, and all FCUs communicate to each other separately.

Architecture-wise, the first scenario is a simple client-server architecture in which the CCU acts as the server, does the optimizations, and sends back the results to its clients, which are the FCUs (though they may be technically called fog servers). On the other hand, in the distributed architectures the FCUs are also servers that serve requests from other clients (including the CCU). Before we go to the implementation details for simulation, we would like to justify the philosophy behind testing three different architectures and the benefits and drawbacks of each.

The complete set-up is simulation-based such that the traffic controllers are virtually emulated in VISSIM. An "interface program" creates virtual clients (using multi-threading) that are linked to each FCU. The FCUs obtain the controller information via the VISSIM COM port. All FCU and CCU servers are hosted on the same machine on the local host IP address with different ports to simulate network traffic.

Wherever applicable, the NTCIP standard has been used to facilitate the incorporation of RSUs and onboard units (OBUs) in the next phase of this project. The NTCIP standard defines the user datagram protocol (UDP) for communication at the link and network levels. Therefore, the UDP is used for the communication protocol, and an object-oriented design is used to construct the packets to facilitate the smooth transition to the full NTCIP standard (currently, some methods and attributes of the NTCIP standard have been omitted for ease of implementation). The following sections describe in detail how each of the three proposed architectures was simulated on a single machine. PTV VISSIM 11 was used for traffic simulation.

6.2   Centralized Architecture

To simulate the centralized scenario, a virtual server is used to represent the CCU, which also hosts the optimization platform. Multithreaded UDP clients represent the FCUs. An additional application is deployed to interface between the clients representing the FCUs and the traffic controller in VISSIM over its available COM interface. Please refer to the flowchart in figure 6-1 for more details.
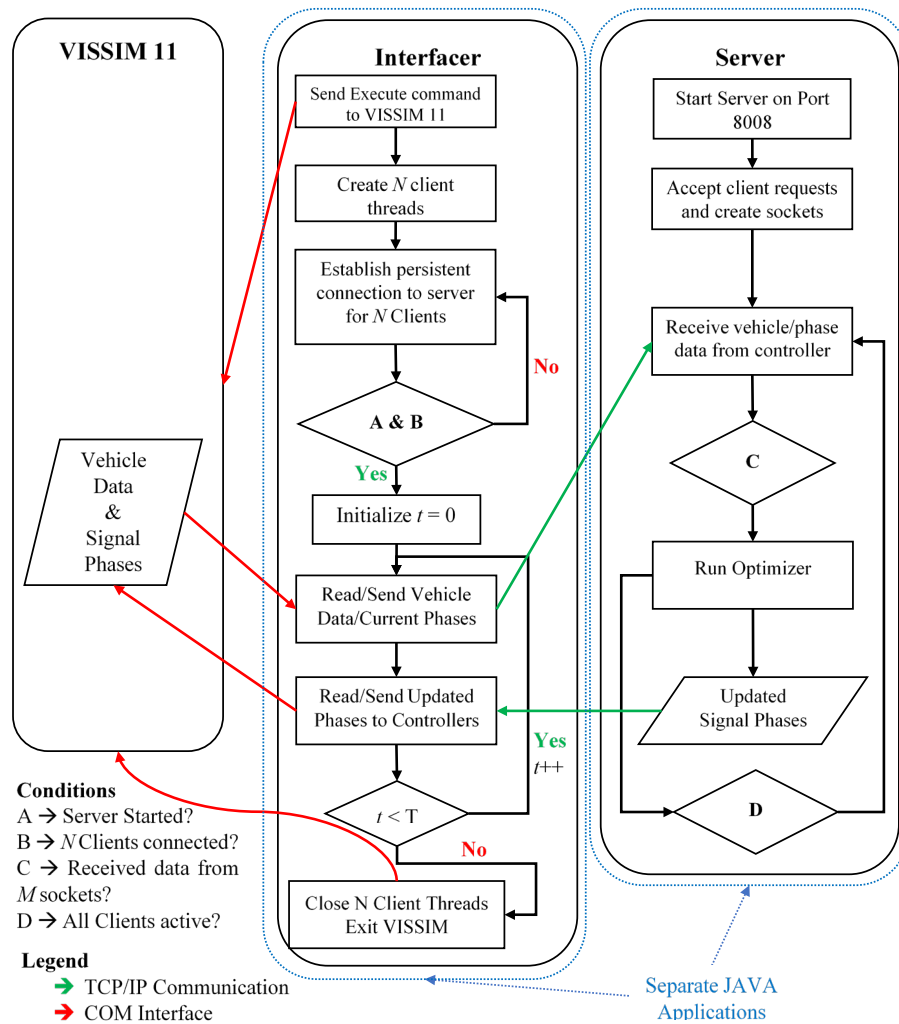

Figure 6.1. Illustration of the centralized architecture

An object for sending the vehicle and current phase information was designed such that it follows the NTCIP standard. In this architecture, the objects are converted into UDP packets and sent over the network. The server collects the information from all clients and executes the optimization algorithm to generate the new phases. The calculated phases are transmitted back in the same format.

To elaborate further, the CCU is assumed to be persistent and is running indefinitely. As soon as the VISSIM environment has been executed from the JAVA program, the initial phase information and vehicle information is converted to the object format. Following that, communication is initiated to the server. Once the server has responded, the FCUs convert the objects into serialized UDP packets and transmit them. The server handles the clients by by using separate threads until it has received information from all FCUs. It is assumed that the server has apriori knowledge of the number of clients it will be serving in any optimization iteration.

Once the information has been received at the CCU, it executes the distributed optimization algorithm and communicates the results for each FCU back by using a similar UDP packet format. The FCUs unpack that information into objects and then send each attribute back to the traffic controllers by using the COM port of VISSIM.

6.3 Distributed Architecture with Cloud Communication

To simulate a partially distributed approach, all FCUs, as well as the CCUs, are treated as servers that serve requests of either the TC connected via the COM interface or other servers, including the FCUs and other CCUs. The partially distributed simulation is different than the fully centralized case in that the FCUs are multi-threaded servers instead of clients. However, these servers are launched from an interface similar to that in the fully centralized case. This is because the FCUs only send either request packets on behalf of the TCs hosted by the interface. (Please

note that even in the fully distributed scenario, we assume all TCs are hosted by just one interface application for ease of implementation of the simulation. Multiple instances of VISSIM running the same traffic simulation are not practically relevant.)

At the beginning of each cycle, the interface starts up VISSIM, creates different objects to represent each intersection/TC, and starts up the clients (representing the FCUs). The CCU server is assumed to be always online. The FCUs wait until they sense that the CCU is ready to communicate (practically, the server is always expected to be up during normal operations). As soon as communication with the CCU has been established, the FCUs start transmitting the information to the CCU along with their requests. The CCU responds to each FCU in parallel using multiple threads by accepting the sent information and storing it on a memory stack and then forwards the relevant information according to the REQUEST packets from each FCU.

Once the FCUs have been received, all relevant information, they execute the optimizer.

In the current set-up, for both distributed cases, only in the interface, the FCU-TC communication and optimizer execution is not handled in parallel via multiple threads. Rather, it is done sequentially, and consequently, the optimizers are also run serially. For simulation purposes, it does not matter whether each FCU executes the optimizer in parallel. Any execution/communication time evaluations via Monte Carlo simulations can be averaged over the number of runs and intersections. However, in the real-life scenario, each FCU will be physically connected to the TC without any need for an interface with a traffic simulation package. Please refer to the flowchart in figure 6-2 for more details.
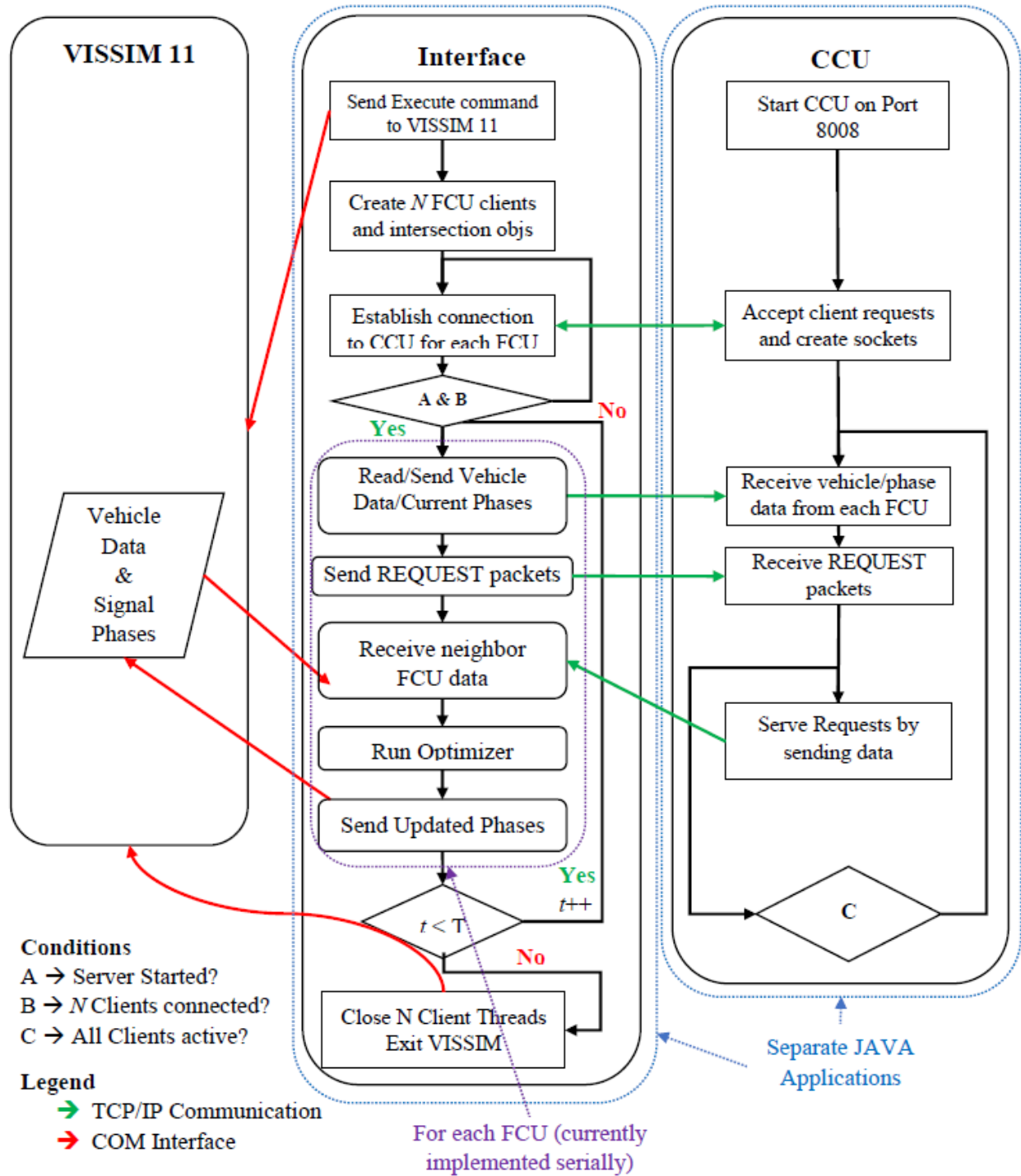
VISSIM 11

Interface

Send Execute command to VISSIM 11

Create *N* FCU clients and intersection objs

Establish connection to CCU for each FCU

A & B

No

Yes

Read/Send Vehicle Data/Current Phases

Send REQUEST packets

Receive neighbor FCU data

Run Optimizer

Send Updated Phases

t < T

Yes
t++

No

Close N Client Threads Exit VISSIM

CCU

Start CCU on Port 8008

Accept client requests and create sockets

Receive vehicle/phase data from each FCU

Receive REQUEST packets

Serve Requests by sending data

C

Vehicle Data & Signal Phases

Conditions
A → Server Started?
B → *N* Clients connected?
C → All Clients active?

Legend
→ TCP/IP Communication
→ COM Interface

Separate JAVA Applications

For each FCU (currently implemented serially)

Figure 6.2. Illustration of the partially distributed architecture

6.4   Fully Distributed Architecture

In the fully distributed scenario, the CCU is eliminated because the FCUs directly communicate with each other. The CCU may be added to represent a central supervisor that periodically gathers information or takes exceptional decisions. However, for normal operations, the FCUs and their associated controllers operate fully independently.

In this scenario, the FCUs are designed as independent server applications that field requests by either responding to REQUEST packets or they pull information from the TCs via the interface. At the start of each cycle, each TC shares the relevant information packet with the associated FCU. Following that, each TC also shares the REQUEST packets with each associated FCU. Separate SEND and REQUEST applications were designed to facilitate the communication of different types of requests. The FCUs forward these requests to the appropriate neighboring FCUs, which send back the requested information packet. This information is forwarded to the associated TC.

In this setting, it is assumed that the optimizer application is hosted at the TC and not at the FCU. For simulation purposes, this does not make a difference, but practically, it could be relevant, especially if there was a huge discrepancy between the computational capabilities of the FCU and TC processors. However, the current approach facilitates interfacing with VISSIM better. For full deployment, the optimization platform can be hosted as requested by the users.

Either way, as soon as the optimized signal phases are calculated for each intersection, they are forwarded to the appropriate TC over the COM interface. Figure 6-3 illustrates the fully distributed architecture for N intersections with a functional view of the role of the interface and the FCU server applications.
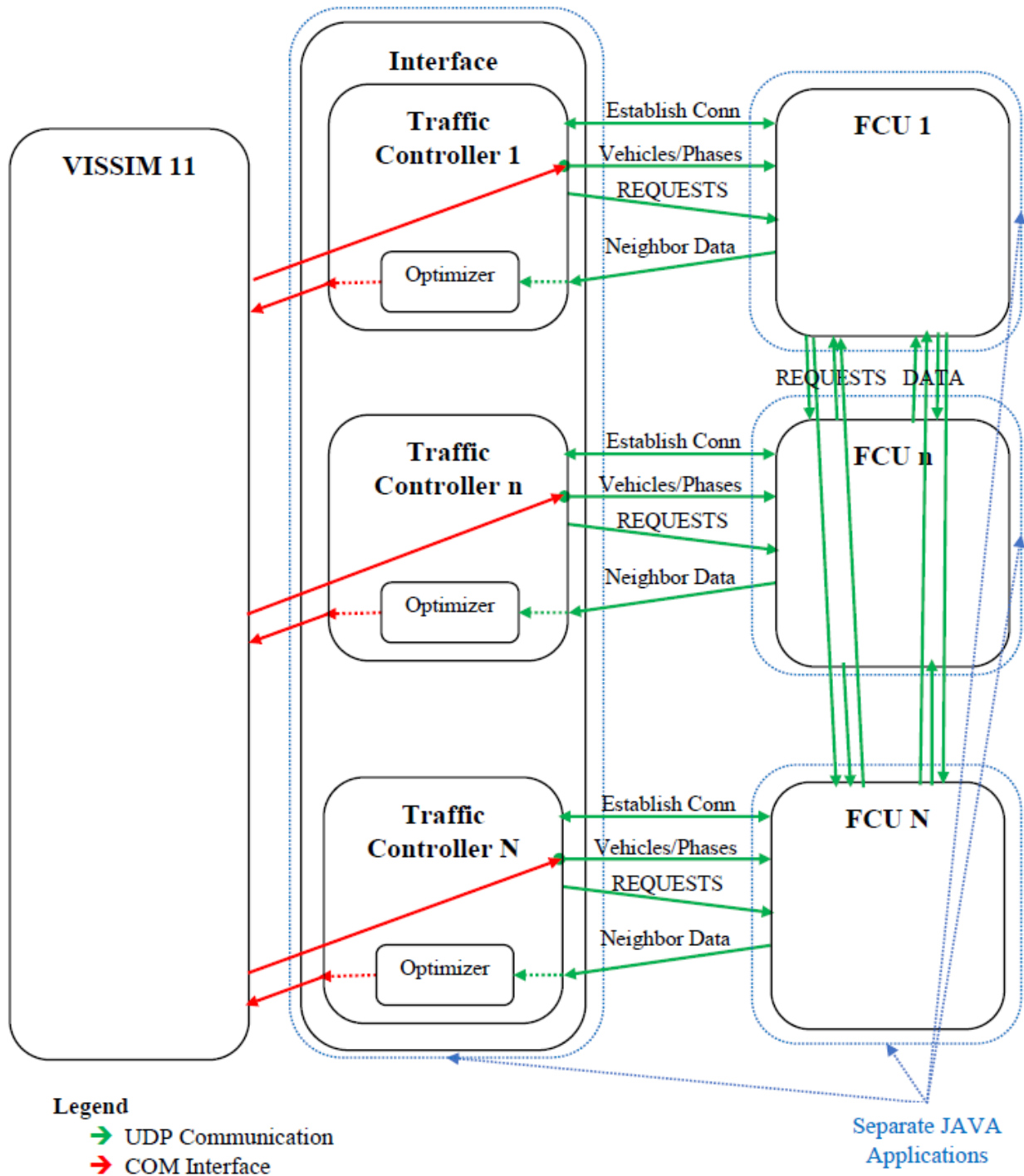
Figure 6.3. The functional view of the fully distributed architecture

7 Case Study and Numerical Results

7.1 Case Study

The case study was a section of the SR 522 corridor in Seattle, Washington, with ten intersections, 18 origins, and 18 destinations as shown in figure 7-1. This arterial street provides routes from Seattle to Kenmore and Bothell and is mainly used for daily commuting between cities. Figure 7-2 shows the intersections with more details.
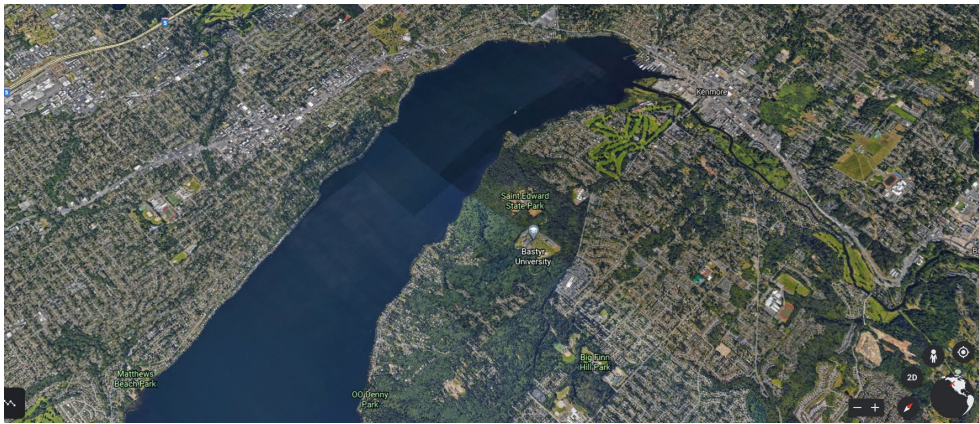

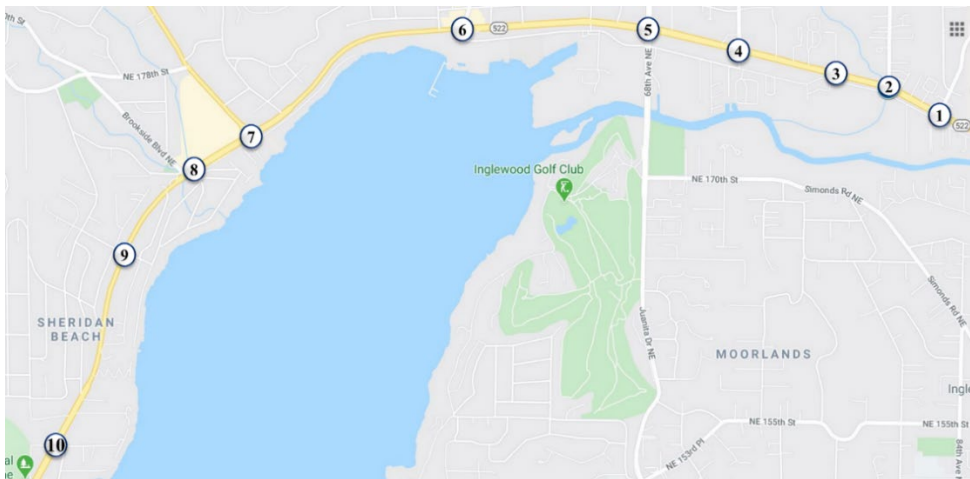**Figure 7.1.** Google Earth satellite map of the case study

Figure 7.2. Google map of a ten-intersection network on SR 522, Seattle, Washington, with intersection movements

The case study included a variety of movements, including bidirectional movements with two- to three-lane segments. Table 7-1 presents the left turning, through, and right turning volumes at all intersections. The volume data were obtained from the Washington State Department of Transportation for the PM peak. The network was loaded with 8,292 passenger cars and 59 transit buses for one hour, according to the PM peak volume data. This traffic load was considered our first demand profile. We also analyzed the corridor by increasing and decreasing the demand by 15 percent as demand profiles 2 and 3, respectively.

Table 7.1. Turning volume (veh/hour) at the intersections

Intersection	Westbound			Southbound			Eastbound			Northbound		
	Left	Through	Right	Left	Through	Right	Left	Through	Right	Left	Through	Right
1	NA	1769	20	13	NA	70	130	1366	NA	NA	NA	NA
2	NA	1638	200	75	NA	85	155	1422	NA	NA	NA	NA
3	NA	1705	18	65	NA	74	77	1513	NA	NA	NA	-NA
4	79	1490	210	153	68	127	143	1374	39	39	65	62
5	166	1399	89	107	212	75	0	1331	678	805	191	119
6	5	2160	114	82	21	335	407	1914	342	108	21	13
7	4	1576	1023	653	8	165	340	2003	5	11	11	7
8	15	1700	37	65	16	62	70	2272	6	8	20	12
9	10	1705	55	28	5	10	12	2315	28	17	12	5
10	22	1711	NA	NA	NA	NA	NA	2320	18	23	NA	35

We considered seven origin-destination pairs or seven routes for buses. Table 7-2 shows the headways used for buses for each path, and table 7-3 shows our assumptions for the characteristics of this case study.

Table 7.2. Origin-destination pairs for buses with the headways for each route

OD	1	2	3	4	5	6	7
Headways (sec)	300	300	300	300	600	600	1000

Table 7.3. Characteristics of the network on SR 522, Seattle, Washington

Parameter	Value
Time Step (s)	6
Free-flow speed of vehicles except for buses (mph)	40
Free-flow speed for buses (mph)	20
Saturation headway (s)	2
Minimum green time for through movements in major direction (s)	18
Minimum green time for left-turning movements in major direction (s)	12
Minimum green time for through movements in minor direction (s)	12
Minimum green time for left-turning movements in minor direction (s)	6
Prediction horizon (s)	9
Study period (s)	3600

7.2   Distributed-Coordinated Approach Performance

Table 7-4 shows total traveled distance, total travel time, total delay, total stop delay, and total throughput for 10 percent, 30 percent, 50 percent, 70 percent, and 90 percent CV market penetration rates. The results showed that the total travel time decreased by more than 31 percent when the CV penetration rate increased from 10 percent to 90 percent. The CV penetration rates were increased by 20 percent increments, and the biggest impact happened when the CV market share was increased to 30 percent from 10 percent. An increase in CV market share yielded improvements in all other performance measures.

Table 7.4. Performance of the distributed coordinated approach

Demand profile	Penetration Rate (%)	Total travel time (s)	Total delay	Total stops	Total delayed stops	Total throughput
1	10	1351621	784790	44670	595918	2867
	30	1007667	319810	11258	183703	3473
	50	969508	281199	11361	145884	3555
	70	956077	274739	10496	144120	3549
	90	925443	258450	10325	129066	3589
2	10	1586435	914425	70388	581170	3377
	30	1208044	423604	19119	230458	3875
	50	1141926	374643	17905	194243	3955
	70	1101272	357439	16870	181317	3967
	90	1064688	335132	16096	168198	4013
3	10	1046275	476950	9893	372039	2633
	30	859934	233667	7709	131670	3035
	50	816453	206521	6674	109357	3060
	70	811583	202169	7080	103840	3107
	90	783026	186847	6516	92024.2	3120

Figure 7.3 shows the average delays, average speeds, average number of stops, and average stopped delays for different CV penetration rates. Increasing the penetration rate of connected vehicles led to better performance of the system.
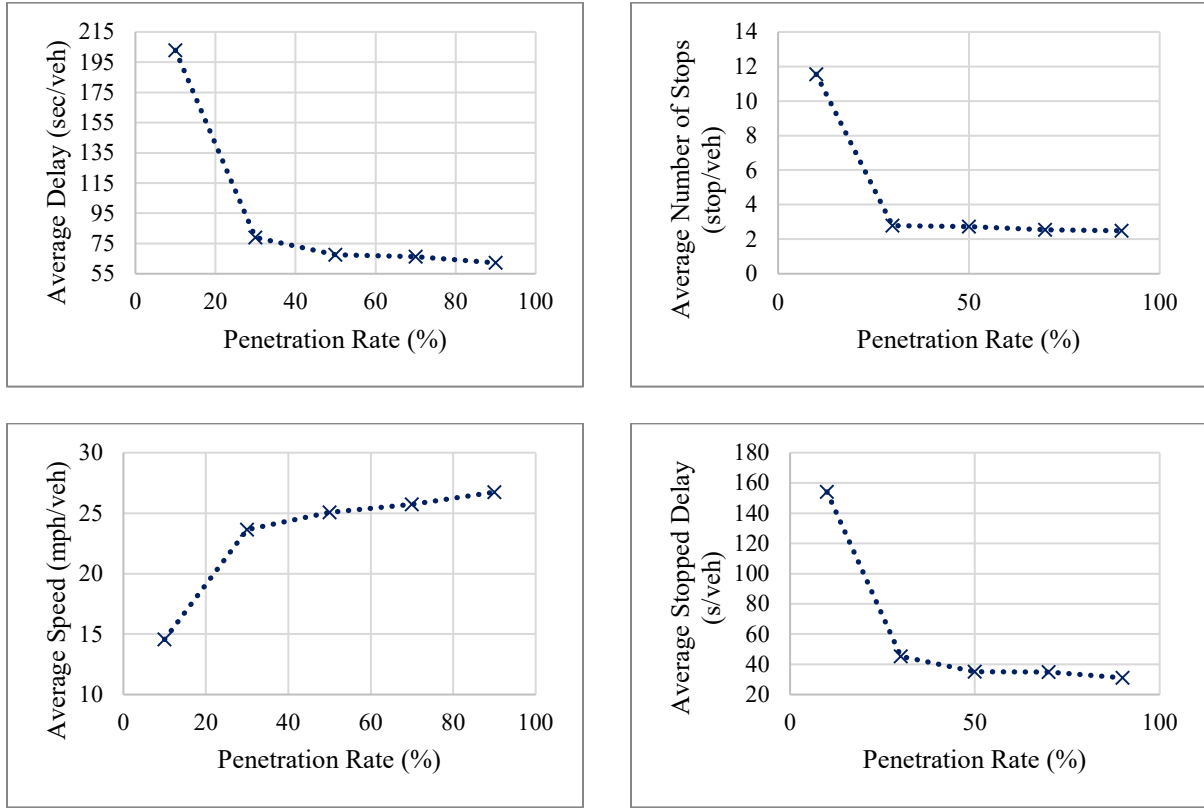
Figure 7.3. Average delays, speeds, stops, and delayed stops for different penetration rates for demand profile 1

We increased the demand level by 15 percent and measured the average delays, average speeds, average number of stops, and average stopped delays for different CV penetration rates, as shown in figure 7-4. Overall, the increase in demand level led to a decline in network performance.
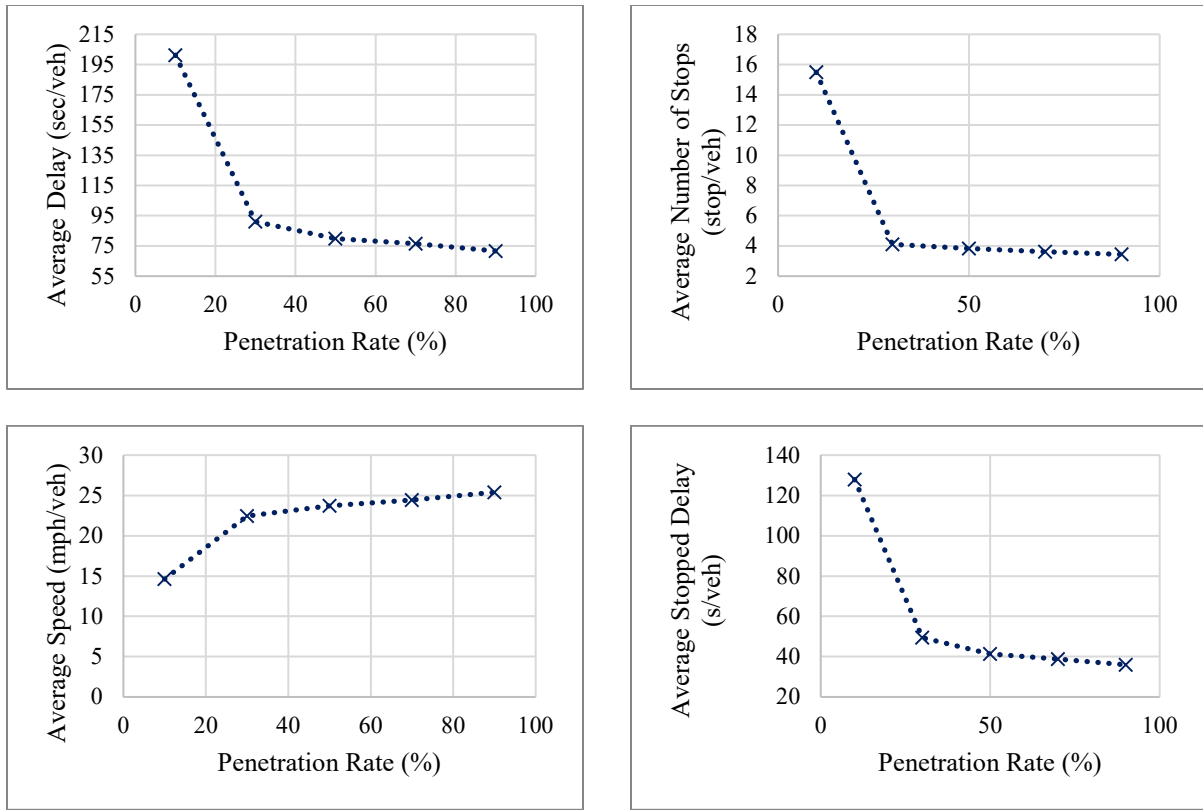
Figure 7.4. Average delays, speeds, stops, and delayed stops for different penetration rates for demand profile 2

Figure 7-5 shows the average delays, average speeds, average number of stops, and average stopped delays for different penetration rates, including 10 percent, 30 percent, 50 percent, 70 percent, and 90 percent when we decreased the demand by 15 percent. Increasing the penetration rate of connected vehicles led to better performance of the system. The average delay decreased by 52.60 percent with an increase in the CV penetration rate from 10 to 30 percent; decreased by 58.10 percent with a CV penetration rate increase from 10 to 50 percent; decreased by 59.24 percent with a CV penetration rate increase from 10 to 70 percent;, and by 62.32 percent with an increase from 10 to 90 percent. The same CV penetration rate increases led to 24.74 percent, 30.66 percent, 31.01 percent, and 36.59 percent reductions in the average number of stops, respectively.
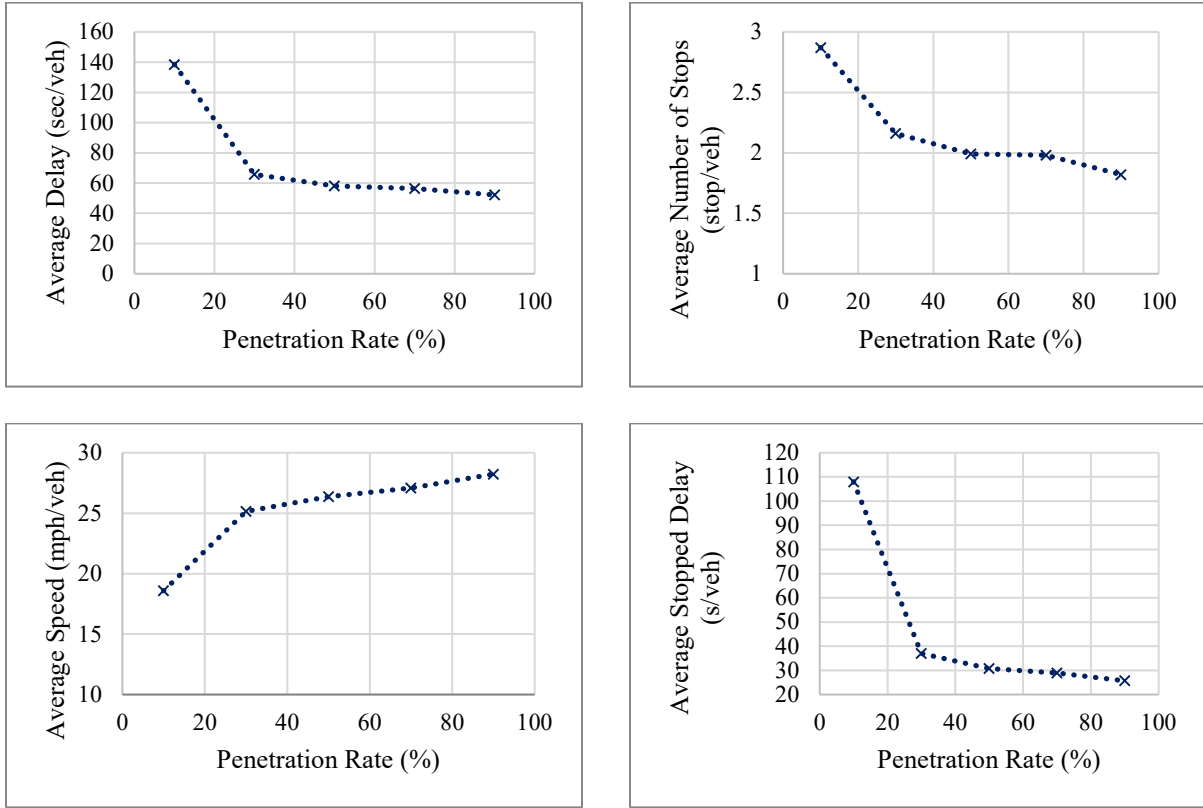
Figure 7.5. Average delays, speeds, stops, and delayed stops for different penetration rates for demand profile 3

8 Summary and Conclusions

This project developed an optimization model and a solution technique to determine signal timing parameters in real time for large transportation networks. The model can account for different market shares of connected vehicles, transit movements, and time-varying demand. The solution technique is a model predictive control and follows a distributed architecture. The distributed optimization reduces the computational complexity and allows the solution technique to be applied to transportation networks of various sizes. Effective coordination among intersection controllers is designed to create consensus among them in their decision-making process.

We developed a cloud-fog communication architecture to facilitate the execution of the solution technique. The fog level consists of fog control units (FCUs), each connected to a road-side unit (RSU) at each intersection. Each of them was designed to collect the vehicle information from its collocated RSU and perform the distributed coordination and optimization (DCO) tasks for each intersection's signal. The cloud control unit (CCU) is linked to the FCUs through wired or wireless backhaul links. This CCU is designed to both exchange information between the FCUs and enforce some parameters in their DCO processes in case of special events that require more coordination, such as traffic incidents and preemption for first responders. FCUs are implemented by using a virtual machine on a workstation.

Numerical results in a simulated corridor of ten intersections showed that the approach can effectively determine near-optimal signal timing parameters under different demand levels, and significant improvement in traffic operations was observed with increased connected vehicle market penetration rates. Because of the COVID-19 pandemic, this project did not conduct any field tests, and all tests were performed with a micro-simulation testbed. Future research that includes hardware-in-the-loop and field tests is needed.

References

Abu-Lebdeh, G., Benekohal, R., 2000. Genetic algorithms for traffic signal control and queue management of oversaturated two-way arterials. Transp. Res. Rec. J. Transp. Res. Board 61–67.

Abu-Lebdeh, G., Benekohal, R., 1997. Development of traffic control and queue management procedures for oversaturated arterials. Transp. Res. Rec. J. Transp. Res. Board 119–127.

Adacher, L., Gemma, A., Oliva, G., 2014. Decentralized spatial decomposition for traffic signal synchronization. Transp. Res. Procedia 3, 992–1001.

Adacher, Ludovica Oliva, Gabriele, A.G., 2014. Decentralization spatial decomposition for traffic signal synchronization, in: 17th Meeting of the EURO Working Group on Transportation, EWGT2014. Sevilla, Spain.

Aghdashi, S., Hajbabaie, A., Schroeder, B.J., Trask, J.L., Rouphail, N.M., 2015. Generating scenarios of freeway reliability analysis: Hybrid approach. Transp. Res. Rec. 2483, 148–159. https://doi.org/10.3141/2483-17

Aghdashi, S., Rouphail, N.M., Hajbabaie, A., 2013. Estimation of Incident Propensity for Reliability Analysis in the Highway Capacity Manual. Transp. Res. Rec. 2395, 123–131. https://doi.org/10.3141/2395-14

Aziz, H.M.A., 2019. Energy and Mobility Impacts of System Optimal Dynamic Traffic Assignment for a Mixed Traffic of Legacy and Automated Vehicles. Transp. Res. Rec. 0361198119845658. https://doi.org/https://doi.org/10.1177/0361198119845658

Bazzan, A.L.C., 2005. A distributed approach for coordination of traffic signal agents. Auton. Agent. Multi. Agent. Syst. 10, 131–164.

Beard, C., Ziliaskopoulos, A., 2006. System optimal signal optimization formulation. Transp. Res. Rec. J. Transp. Res. Board 102–112.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., others, 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends®in Mach. Learn. 3, 1–122.

Ceylan, H., Bell, M.G.H., 2004. Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transp. Res. Part B Methodol. 38, 329–342.

Chan, P., 1997. Advanced Transportation Management Technologies. Ref. Guid. Fed. Highw. 22.

Chang, T.-H., Sun, G.-Y., 2004. Modeling and optimization of an oversaturated signalized network. Transp. Res. Part B Methodol. 38, 687–707.

Daganzo, C.F., 2005a. A variational formulation of kinematic waves: Solution methods. Transp. Res. Part B Methodol. 39, 934–950.

Daganzo, C.F., 2005b. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transp. Res. Part B Methodol. 39, 187–196.

Daganzo, C.F., 1995. The cell transmission model, part II: Network traffic. Transp. Res. Part B Methodol. 29, 79–93.

Daganzo, C.F., 1994. The cell transmission model: a simple dynamic representation of highway trafc. Transp Res Part B 28 269287.

Dion, F., Hellinga, B., 2001. A Methodology for Obtaining Signal Coordination within a Distributed Real-time Signal Control System with Transit Priority, in: Transportation Research Board 80th Annual Meeting. Washington D.C.

Feng, Y., Head, K.L., Khoshmagham, S., Zamanipour, M., 2015. A real-time adaptive signal control in a connected vehicle environment. Transp. Res. Part C Emerg. Technol. 55, 460–473. https://doi.org/10.1016/j.trc.2015.01.007

Gartner, N.H., 1983. OPAC: A demand-responsive strategy for traffic signal control.

Gartner, N.H., 1982. Development and testing of a demand-responsive strategy for traffic signal control, in: American Control Conference, 1982. pp. 578–583.

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. https://doi.org/10.1109/itsc.2001.948655

Gartner, N.H., Stamatiadis, C., Tarnoff, P.J., 1995. Development of advanced traffic signal control strategies for intelligent transportation systems: Multilevel design. Transp. Res. Rec.

Girianna, M., Benekohal, R., 2002. Dynamic signal coordination for networks with oversaturated intersfile:///C:/Users/aliha/Desktop/New Text Document.txtections. Transp. Res. Rec. J. Transp. Res. Board 122–130.

Goodall, N., Smith, B., Park, B., 2013. Traffic signal control with connected vehicles. Transp. Res. Rec. J. Transp. Res. Board 65–72.

Gregoire, J., Qian, X., Frazzoli, E., Fortelle, A. de La, Wongpiromsarn, T., 2015. Capacity-Aware Backpressure Traffic Signal Control. IEEE Trans. Control Netw. Syst. 2, 164–173. https://doi.org/10.1109/tcns.2014.2378871

Guilliard, I., Sanner, S., Trevizan, F.W., Williams, B.C., 2016. Nonhomogeneous Time Mixed Integer Linear Programming Formulation for Traffic Signal Control. Transp. Res. Rec. J. Transp. Res. Board 128–138.

Hajbabaie, A., 2012. Intelligent dynamic signal timing optimization program. University of Illinois at Urbana-Champaign.

Hajbabaie, A., Aghdashi, S., Rouphail, N.M., 2016. Enhanced decision-making framework using reliability concepts for freeway facilities. J. Transp. Eng. 142. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000797

Hajbabaie, A., Benekohal, R., 2011a. Which policy works better for signal coordination? Common, or variable cycle length, in: Proceedings of the 1st ASCE T&DI Congress. pp. 13–16.

Hajbabaie, A., Benekohal, R., 2011b. Does traffic metering improve network performance efficiency?, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. pp. 1114–1119. https://doi.org/10.1109/ITSC.2011.6083011

Hajbabaie, A., Benekohal, R.F., 2015. A Program for Simultaneous Network Signal Timing Optimization and Traffic Assignment. IEEE Trans. Intell. Transp. Syst. 16, 2573–2586. https://doi.org/10.1109/TITS.2015.2413360

Hajbabaie, A., Benekohal, R.F., 2013. Traffic signal timing optimization. Transp. Res. Rec. 2355, 10–19. https://doi.org/10.3141/2355-02

Hajbabaie, A., Medina, J.C., Benekohal, R.F., 2011. Traffic signal coordination and queue management in oversaturated intersections, Purdue University Discovery Park.

Hajbabaie, A., Medina, J.C., Benekohal, R.F., 2010. Effects of ITS-Based Left Turn Policies on Network Performance, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. pp. 80–84. https://doi.org/10.1109/ITSC.2010.5625269

Hajibabai, L., Ouyang, Y., 2013. Integrated planning of supply chain networks and multimodal transportation infrastructure expansion: model development and application to the biofuel industry. Comput. Civ. Infrastruct. Eng. 28, 247–259.

Hajibabai, L., Saha, D., 2019. Patrol route planning for incident response vehicles under dispatching station scenarios. Comput. Civ. Infrastruct. Eng. 34, 58–70.

He, Q., Head, K.L., Ding, J., 2011. PAMSCOD: platoon-based arterial multi-modal signal control with online data. Procedia-Social Behav. Sci. 17, 462–489.

Head, K.L., Mirchandani, P.B., Sheppard, D., 1992. Hierarchical framework for real-time traffic control. Transp. Res. Rec. 82–88.

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Royle, M.C., 1983. The SCOOT On-line Traffic Signal Optimisation Technique. Traffic Eng. Control 23, 190–192.

Islam, S., Aziz, H., Hajbabaie, A., 2020a. Stochastic Gradient-Based Optimal Signal Control With Energy Consumption Bounds. IEEE Trans. Intell. Transp. Syst. 1–14. https://doi.org/10.1109/tits.2020.2979384

Islam, S., Hajbabaie, A., 2017. Distributed coordinated signal timing optimization in connected transportation networks. Transp. Res. Part C Emerg. Technol. 80, 272–285. https://doi.org/10.1016/j.trc.2017.04.017

Islam, S., Hajbabaie, A., Aziz, H., 2020b. A real-time network-level traffic signal control methodology with partial connected vehicle information. Transp. Res. Part C Emerg. Technol. 121. https://doi.org/10.1016/j.trc.2020.102830

Islam, S., Tajalli, M., Mohebifard, R., Hajbabaie, A., 2021. The Effects of Connectivity and Traffic Observability on an Adaptive Traffic Signal Control System. Transp. Res. Rec. Accepted.

Karoonsoontawong, A., Waller Travis, S., 2010. Integrated Network Capacity Expansion and Traffic Signal Optimization Problem: Robust Bi-level Dynamic Formulation. Networks Spat. Econ. 10, 525–550.

Lee, J., Park, B., Yun, I., 2013. Cumulative Travel-Time Responsive Real-Time Intersection Control Algorithm in the Connected Vehicle Environment. J. Transp. Eng. 139. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000587

Lee, S., Wong, S.C., Varaiya, P., 2017a. Group-based hierarchical adaptive traffic-signal control Part II: Implementation. Transp. Res. part B Methodol. 104, 376–397.

Lee, S., Wong, S.C., Varaiya, P., 2017b. Group-based hierarchical adaptive traffic-signal control part I: Formulation. Transp. Res. part B Methodol. 105, 1–18.

Lertworawanich, P., Kuwahara, M., Miska, M., 2011. A new multiobjective signal optimization for oversaturated networks. IEEE Trans. Intell. Transp. Syst. 12, 967–976.

Lin, W.-H., Wang, C., 2004. An enhanced 0-1 mixed-integer LP formulation for traffic signal control. IEEE Trans. Intell. Transp. Syst. 5, 238–245.

Liu, H., Wang, J., Wijayaratna, K., Dixit, V. V, Waller, S.T., 2015. Integrating the bus vehicle class into the cell transmission model. IEEE Trans. Intell. Transp. Syst. 16, 2620–2630.

Lo, H.K., 2001. A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transp. Sci. 35, 148–164.

Lo, H.K., 1999. A novel traffic signal control formulation. Transp. Res. Part A Policy Pract. 33, 433–448.

Lo, H.K., Chow, A.H.F., 2004. Control strategies for oversaturated traffic. J. Transp. Eng. 130, 466–478.

Luk, J.Y.K., 1984. Two traffic-responsive area traffic control methods: SCAT and SCOOT. Traffic Eng. Control 25.

Mauro, V., DiTaronto, D., 1990. UTOPIA, in: Proceedings of the 6th IFAC-IFIP-IFORS Symposium on Control Computers and Communication in Transportation. pp. 245–252.

Medina, J., Hajbabaie, A., Benekohal, R., 2013. Effects of metered entry volume on an oversaturated network with dynamic signal timing. Transp. Res. Rec. J. Transp. Res. Board 2356, 53–60. https://doi.org/10.3141/2356-0

Medina, J.C., Hajbabaie, A., Benekohal, R.F., 2011. A comparison of approximate dynamic programming and simple genetic algorithm for traffic control in oversaturated conditions - Case study of a simple symmetric network, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. pp. 1815–1820. https://doi.org/10.1109/ITSC.2011.6082999

Mehrabipour, M., Hajbabaie, A., 2017. A Cell Based Distributed-Coordinated Approach for Network Level Signal Timing Optimization. Comput. Aided Civ. Infrastruct. Eng. an Int. J. 32, 599--616.

Mehrabipour, M., Hajibabai, L., Hajbabaie, A., 2019. A decomposition scheme for parallelization of system optimal dynamic traffic assignment on urban networks with multiple origins and destinations. Comput. Civ. Infrastruct. Eng.

Mesa-Arango, R., Ukkusuri, S. V, 2014. Modeling the car-truck interaction in a system-optimal dynamic traffic assignment model. J. Intell. Transp. Syst. 18, 327–338.

Mirchandani, P., Soroush, H., 1987. Generalized traffic equilibrium with probabilistic travel times and perceptions. Transp. Sci. 21, 133–152.

Mirheli, A., Hajibabai, L., 2020. Utilization Management and Pricing of Parking Facilities Under Uncertain Demand and User Decisions. IEEE Trans. Intell. Transp. Syst. 21, 2167–2179.

Mirheli, A., Hajibabai, L., Hajbabaie, A., 2018. Development of a signal-head-free intersection control logic in a fully connected and autonomous vehicle environment. Transp. Res. Part C Emerg. Technol. 92, 412–425. https://doi.org/10.1016/j.trc.2018.04.026

Mirheli, A., Tajalli, M., Hajibabai, L., Hajbabaie, A., 2019. A consensus-based distributed trajectory control in a signal-free intersection. Transp. Res. Part C Emerg. Technol. 100, 161–176. https://doi.org/10.1016/j.trc.2019.01.004

Mohebifard, R., Hajbabaie, A., 2021. Trajectory Control in Roundabouts with a Mixed-fleet of Automated and Human-driven Vehicles. Comput. Civ. Infrastruct. Eng. Accepted.

Mohebifard, R., Hajbabaie, A., 2020. Effects of Automated Vehicles on Traffic Operations at Roundabouts., in: The IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC).

Mohebifard, R., Hajbabaie, A., 2019a. Optimal network-level traffic signal control: A benders decomposition-based solution algorithm. Transp. Res. Part B Methodol. 121, 252–274. https://doi.org/10.1016/j.trb.2019.01.012

Mohebifard, R., Hajbabaie, A., 2019b. Distributed Optimization and Coordination Algorithms for Dynamic Traffic Metering in Urban Street Networks. IEEE Trans. Intell. Transp. Syst. 20, 1930–1941. https://doi.org/10.1109/TITS.2018.2848246

Mohebifard, R., Hajbabaie, A., 2018a. Dynamic traffic metering in urban street networks: Formulation and solution algorithm. Transp. Res. Part C Emerg. Technol. 93, 161–178. https://doi.org/10.1016/j.trc.2018.04.027

Mohebifard, R., Hajbabaie, A., 2018b. Real-Time Adaptive Traffic Metering in a Connected Urban Street Network, in: Transportation Research Board 97th Annual Meeting.

Mohebifard, R., Islam, S., Hajbabaie, A., 2019. Cooperative traffic signal and perimeter control in semi-connected urban-street networks. Transp. Res. Part C Emerg. Technol. 104, 408–427. https://doi.org/10.1016/j.trc.2019.05.023

Newell, G.F., 2002. A simplified car-following theory: a lower order model. Transp. Res. Part B Methodol. 36, 195–205.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory. Transp. Res. Part B Methodol. 27, 281–287.

Niroumand, R., Tajalli, M., Hajibabai, L., Hajbabaie, A., 2020a. Joint optimization of vehicle-group trajectory and signal timing: Introducing the white phase for mixed-autonomy traffic stream. Transp. Res. Part C Emerg. Technol. 116, 102659. https://doi.org/10.1016/j.trc.2020.102659

Niroumand, R., Tajalli, M., Hajibabai, L., Hajbabaie, A., 2020b. The Effects of the "White Phase" on Intersection Performance with Mixed-Autonomy Traffic Stream, in: The 23rd IEEE International Conference on Intelligent Transportation Systems. IEEE, pp. 2795–2800.

Porche, I., Lafortune, S., 1998. Adaptive Look-ahead Optimization of Traffic Signals. Intell. Transp. Syst. J. 4, 209–254. https://doi.org/10.1080/10248079908903749

Priemer, C., 2009. A decentralized adaptive traffic signal control using V2I communication data. Intell. Transp. Syst. 1–6.

Priemer, C., Friedrich, B., 2009. A decentralized adaptive traffic signal control using V2I communication data, in: Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference On. pp. 1–6.

Putha, R., Quadrifoglio, L., Zechman, E., 2012. Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions. Comput. Civ. Infrastruct. Eng. 27, 14–28.

Sims, A.G., Dobinson, K.W., 1980. The Sydney Coordinated Adaptive Traffic (SCAT) System Philosophy and Benefits. IEEE Trans. Veh. Technol. 29, 130–137. https://doi.org/10.1109/t-vt.1980.23833

Stevanovic, A., Martin, P., Stevanovic, J., 2007. VisSim-based genetic algorithm optimization of signal timings. Transp. Res. Rec. J. Transp. Res. Board 59–68.

Sun, D., Benekohal, R.F., Waller, S.T., 2006. Bi-level Programming Formulation and Heuristic Solution Approach for Dynamic Traffic Signal Optimization. Comput. Civ. Infrastruct. Eng. 21, 321–333.

Tajalli, M., Hajbabaie, A., 2021. A Lagrangian-based Signal Timing and Trajectory Optimization in a Mix Traffic Stream of Connected Self-driving and Human-driven Vehicles. IEEE Trans. Intell. Transp. Syst. 1–14. https://doi.org/10.1109/TITS.2021.3058193

Tajalli, M., Hajbabaie, A., 2018a. Distributed optimization and coordination algorithms for dynamic speed optimization of connected and autonomous vehicles in urban street networks. Transp. Res. Part C Emerg. Technol. 95, 497–515. https://doi.org/10.1016/j.trc.2018.07.012

Tajalli, M., Hajbabaie, A., 2018b. Dynamic Speed Harmonization in Connected Urban Street Networks. Comput. Civ. Infrastruct. Eng. 33, 510–523. https://doi.org/10.1111/mice.12360

Tajalli, M., Mehrabipour, M., Hajbabaie, A., 2020. Network-Level Coordinated Speed Optimization and Traffic Light Control for Connected and Automated Vehicles. IEEE Trans. Intell. Transp. Syst. 1–12. https://doi.org/10.1109/tits.2020.2994468

Tassiulas, L., Ephremides, A., 1992. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. IEEE Trans. Automat. Contr. 37, 1936–1948.

Teklu, F., Sumalee, A., Watling, D., 2007. A Genetic Algorithm Approach for Optimizing Traffic Control Signals Considering Routing. Comput. Civ. Infrastruct. Eng. 22, 31–43.

Timotheou, S., Panayiotou, C.G., Polycarpou, M.M., 2015. Distributed Traffic Signal Control Using the Cell Transmission Model via the Alternating Direction Method of Multipliers 16, 919–933.

Timotheou, S., Panayiotou, C.G., Polycarpou, M.M., 2013. Towards distributed online cooperative traffic signal control using the cell transmission model, in: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). pp. 1737–1742.

Torabi, B., Wenkstern, R.Z., Saylor, R., 2020. A collaborative agent-based traffic signal system for highly dynamic traffic conditions. Auton. Agent. Multi. Agent. Syst. 34, 1–24.

Varaiya, P., 2013. Max pressure control of a network of signalized intersections. Transp. Res. Part C Emerg. Technol. 36, 177–195.

Wada, K., Usui, K., Takigawa, T., Kuwahara, M., 2017. An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension. Transp. Res. procedia 23, 624–644.

Webster, F. V, 1958. Traffic signal settings, road research technical paper no. 39. Road Res. Lab.

Wong, C.K., Heydecker, B.G., 2011. Optimal allocation of turns to lanes at an isolated signal-controlled junction. Transp. Res. Part B Methodol. 45, 667–681.

Wong, C.K., Wong, S.C., 2003. Lane-based optimization of signal timings for isolated junctions. Transp. Res. Part B Methodol. 37, 63–84.

Wongpiromsarn, T., Uthaicharoenpong, T., Frazzoli, E., Wang, Y., Wang, D., 2014. Throughput Optimal Distributed Traffic Signal Control. arXiv Prepr. arXiv1407.1164.

Wongpiromsarn, T., Uthaicharoenpong, T., Wang, Y., Frazzoli, E., Wang, D., 2012. Distributed traffic signal control for maximum network throughput, in: 2012 15th International IEEE Conference on Intelligent Transportation Systems. pp. 588–595.

Wünsch, G., 2008. Coordination of traffic signals in networks and related graph theoretical problems on spanning trees. Cuvillier Verlag.

Yahia, C.N., Pandey, V., Boyles, S.D., 2018. Network Partitioning Algorithms for Solving the Traffic Assignment Problem using a Decomposition Approach. Transp. Res. Rec. 0361198118799039.

Yang, I., Jayakrishnan, R., 2015. Real-time network-wide traffic signal optimization considering long-term green ratios based on expected route flows. Transp. Res. Part C Emerg. Technol. 60, 241–257.

Yu, C., Feng, Y., Liu, H.X., Ma, W., Yang, X., 2018. Integrated optimization of traffic signals and vehicle trajectories at isolated urban intersections. Transp. Res. Part B Methodol. 112, 89–112.

Zegeer, J., Bonneson, J., Dowling, R., Ryus, P., Vandehey, M., Kittelson, W., Rouphail, N., Schroeder, B., Hajbabaie, A., Aghdashi, B., others, 2014. Incorporating Travel Time Reliability into the Highway Capacity Manual. https://doi.org/10.17226/22487

Zhang, L., Yin, Y., Chen, S., 2013. Robust signal timing optimization with environmental concerns. Transp. Res. Part C Emerg. Technol. 29, 55–71.

Zhang, L., Yin, Y., Lou, Y., 2010. Robust signal timing for arterials under day-to-day demand variations. Transp. Res. Rec. J. Transp. Res. Board 156–166.