DEVELOPMENT AND EVALUATION OF AN ADAPTIVE TRANSIT SIGNAL PRIORITY SYSTEM USING CONNECTED VEHICLE TECHNOLOGY

A Dissertation

by

XIAOSI ZENG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yunlong Zhang
Committee Members,	Xiubin Wang
	Luca Quadrifoglio
	Lewis Ntaimo
	Kevin Balke
Head of Department,	Robin Autenrieth

December 2014

Major Subject: Civil Engineering

ABSTRACT

Transit signal priority (TSP) can be a very effective preferential treatment for transit vehicles in congested urban networks. There are two problems with the current practice of the transit signal priority. First, random bus arrival time is not sufficiently accounted for, which've become the major hindrance in practice for implementing active or adaptive TSP strategies when a near-side bus stop is present. Secondly, most research focuses on providing bus priority at local intersection level, but bus schedule reliability should be achieved at route level and relevant studies have been lacking.

In the first part of this research, a stochastic mixed-integer nonlinear programming (SMINP) model is developed to explicitly to account for uncertain bus arrival time. A queue delay algorithm is developed as the supporting algorithm for SMINP to capture the delays caused by the interactions between vehicle queues and buses entering and exiting near-side bus stops. A concept of using signal timing deviations to approximate the impacts of TSP operations on other traffic is proposed for the first time in this research. In the second part of the research, the deterministic version of the SMINP model is extended to the arterial setting, where a route-based TSP (R-TSP) model is develop to optimize for schedule-related bus performances on the corridor level. The R-TSP model uses the real-time data available only from the connected vehicle communications technology.

Based on the connected vehicle technology, a real-time signal control system that implements the proposed TSP models is prototyped in the simulation environment. The connected vehicle technology is also used as the main detection and monitoring mechanism for the real-time control of the adaptive TSP signal system. The adaptive TSP control module is designed as a plug-in module that is envisioned to work with a modern fixed-time or adaptive signal controller with connected vehicle communications capabilities.

Using this TSP-enabled signal control system, simulation studies were carried out in both a single intersection setting and a five-intersection arterial setting. The effectiveness of the SMINP model to handle uncertain bus arrival time and the R-TSP model to achieve corridor-level bus schedule reliability were studied. Discussions, conclusions and future research on the topic of adaptive TSP models were made.

DEDICATION

I would like to dedicate this dissertation to my farther, Xianqiang Zeng, for being a good role model of discipline and dedication, and to my mother, Meijuan Cai, for her unconditional love and supporting me all the way.

ACKNOWLEDGEMENTS

I would like to thank my committee chair and adviser, Dr. Zhang, for his guidance throughout the course of my academic studies at the Texas A&M University. His invaluable advice and always considerate support help shape not only my academic career but also my personal future. I am also greatly indebted to my research supervisor and committee member, Dr. Balke, without whom this research would not be conceived at the first place. And he has provided continuous financial support for me to focus on making the best quality research and has mentored me for the past several years. Many sincere thanks also go to my committee members, Dr. Wang, Dr. Quadrifoglio and Dr. Ntaimo for their value comments in throughout the development of my dissertation. Especially thank Dr. Ntaimo to allow the use of some computer codes to facilitate the development of an important part of the simulation experiment.

I would also like to show my deepest appreciations to so many of great people with whom I have befriended during my doctoral study at the Texas A&M University and the Texas A&M Transportation Institute. Thank Dr. Songchitruksa for instilling a good work ethic in me; thank Dr. Chu for making me truly understanding critical thinking; thank Dr. Yin for discussing many technical details in this research with me; thank Mr. Sunkari for hand-on teaching me basics of signal control logics; thank Wei Lu for being a great friend who cheers me up during tough times and enriches my life experience.

Finally, thanks to my beloved parents for their unconditional love and support.

NOMENCLATURE

AASHTO	American Association of State Highway and Transportation Officials
ANN	Artificial Neural Network
AVL	Automatic Vehicle Location
BRT	Bus Rapid Transit
COM	Component Object Model
CV	Connected Vehicle
DEP	Deterministic Equivalent Program
DSRC	Dedicated Short Range Communications
GPS	Global Positioning System
ITE	Institute of Transportation Engineers
L-TSP	Localized Transit Signal Priority
MILP	Mixed Integer Linear Programming
MINP	Mixed Integer Nonlinear Programming
NEMA	National Electrical Manufacturers Association
OBU	On-Board Unit
PC	Passenger Car
PRG	Priority Request Generators
PRS	Priority Request Server
PVD	Probe Vehicle Data

RBC	Ring-Barrier Controller
RSU	Roadside Unit
R-TSP	Route-Based Transit Signal Priority
RF	Radio Frequency
SMINP	Stochastic Mixed Integer Nonlinear Programming
TSP	Transit Signal Priority
UPA	Urban Partnership Agreement
USDOT	United States Department of Transportation
VAP	Vehicle Actuation Programming

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

Urban congestion costs the United States of more than 100 billion dollars annually [1]. The cost is estimated to almost double if all public transportation services are discontinued and the public transit riders are forced to travel in private vehicles[1]. Not hard to see, public transportation services provide an efficient, economical and environmentally friendly means to move the general public around the cities without straining the transportation infrastructure. In fact, transit system is considered by the USDOT Urban Partnership Agreement (UPA) program as one of the four strategies for reducing traffic congestion [2].

Due to the importance of transit systems in reducing urban congestion, decision makers often seek to implement preferential treatments to transit vehicles to further improve their operations and thus their attractiveness to the public. There are a number of treatments available. Some treatments provide exclusive priority to buses via modified roadway segments, such as median bus-way, exclusive lanes and the like. Others furnish the priority at spot locations that yield best results, such as transit signal priority, queue jump and bypass lanes, curb extensions, and so on [3].

Among all the available preferential treatments, transit signal priority (TSP) is one of the most popular approaches in the US. A study in 2010 [3] showed close to 70% of the 64 urban areas in the survey had implemented some forms of TSP, ranking the first among all available preferential treatments. This is because providing signal priorities to the transit buses is considered as one of the easiest implementable and highly effective strategies to lower bus travel time and increase service reliability.

A very promising wireless communications technology that have been successfully applied in traffic safety, the connected vehicle (CV) technology, has great potential to improve traffic operations [4]. With much richer dataset and the capability of two-way communications, the application of the CV technology on TSP can potentially enable providing priority in a more intelligent and adaptive way.

1.1. Scope and Problem

Public transit is a broad term that generally refers a shared passenger transport service that is available for use by the general public. For example, buses, trams, ferries and light rails are modes of the public transport. In this research, the focus is placed on providing traffic signal priority to transit buses in a mixed traffic environment with buses and private vehicles. Therefore, the primary issue is to develop a TSP strategy that can improve bus's operational performance at one or multiple intersections while not inducing too much negative impacts on other private vehicles.

The problem can be further broken down into the following sub-problems:

- How to improve bus's performance at one intersection;
- How to improve bus's performance at multiple intersections;
- How to limit the negative impacts of TSP on other private vehicles; and

- How connected vehicle technologies can benefit the development of adaptive TSP models and improve real-time signal operations.

Each sub-problem is related to yet different from one another. The first problem is what methods can be deployed to improve a bus performance, such as vehicle delay or person delay, at one intersection. In practice, there are passive, active, adaptive, real-time TSP strategies that can provide signal priority conditionally or unconditionally. In the literature, some research developed using rule-based algorithm for bus priority while others applied mathematical optimization to find a better timing for both buses and private vehicles. There are also a variety of timing adjustment methods, such as phase extension, truncation, rotation and insertion. Furthermore, near-side bus stop usually presents a challenge to active TSP strategies, and how variable dwell time should be accounted for with the near-side bus stop needs to be answered. Finally, shall the operations strategy or model developed be different for multiple conflicting buses arriving the same time versus just one bus needing priority?

The second problem deals with bus priority at a signalized arterial. The first key question is whether giving priority at each intersection locally can achieve optimal performance at a corridor level. And second question to answer is if optimal schedule lateness or deviation can also be achieved by granting priority locally at each intersection?

Both the first and second sub-problems cannot avoid addressing what the impacts of the TSP strategy being developed will have on traffic in the concurrent and the conflicting phases, and how the impacts shall be quantified and measured.

Finally, the fourth sub-problem identifies the potential benefits that can be delivered by using the latest communications technology. More importantly, if the type of data available from this technology can enable new kind of TSP control system, and how the control system shall function with this technology.

1.2. Research Objectives

The overarching objective of this research is to develop a real-time control system that can easily accommodate the priority needs of multiple transit buses on a coordinated signal corridor while minimizing the impacts of the priority treatment to the general traffic. Specific objectives include the following:

- Develop a stochastic model that accounts for random bus arrival time
- Develop a route based TSP strategy that considers signal coordination
- Prototype a real-time priority-based control system and implement in a simulation environment for model evaluation

In order to successfully implement a change of timing that is not too disruptive to the general traffic, planning in advance is the key. In the most ideal situation, if all information can be precisely known ahead of time, optimal timing would be guaranteed. However, uncertainty grows as one look further into the future. Therefore, predictions are necessary. Predictions, however, are always with uncertainties. That means the quality of a planning decision is at best the quality of the prediction results, which reflect the inherent uncertainty of a random event. One either has a more polished prediction model or explicitly considers how uncertainties can be incorporated into the decision system. So the first research objective is to develop a TSP model that accounts for the stochastic bus arrival time.

Secondly, priority control for a coordinated corridor is expected to be different from that at a single intersection. This is because a well-timed signal corridor allows platoon to flow through the corridor with minimal stops. But transit vehicles do not operates under the same average corridor travel speed as the private vehicles. For a TSP model to operate at the corridor level, extra delays from disrupted signal progression due to TSP operations should be accounted for. For the control model to be used in real-time, the optimization model needs to be light-weighted and can be solved in reasonably short period of time even for multiple intersections.

Based on the models and algorithms developed for the previous two objectives, the third objective is to prototype a real-time signal control system in a simulation environment to implement the developed optimization models. The prototype system shall be light-weighted and can operate in real-time multiple bus lines simultaneously. Additionally, the simulation platform shall incorporate the connected vehicle communications protocol or at least imitate the basic functions enabled through vehicle-to-infrastructure wireless communications Evaluations of the TSP models against the state-of-the-practice TSP strategies shall be carried out to study the effectiveness and limitations of these models.

1.3. Dissertation Organization

Section 2 provides a comprehensive literature review on several major TSP strategies, their categorization and relevant studies. A review on the connected vehicle technology and its use in TSP deployment is presented. Section 3 describes a stochastic model that is used to account for random bus stop dwell times at a signal intersection running a fixed-time scheme. Section 4 expands on the discussions in the previous section and develops the local TSP model into a route-based TSP model for a fixed-time coordinated signal arterial. Section 5 documents the architecture and main components of the simulation platform built for the prototyping and evaluation of the TSP models. Section 6 and 7 provide details on various analyses and comparisons made between the models developed in this research and the TSP strategy used in practice. Section 6 focuses on the evaluation of model performance at the single intersection setting while Section 7 emphasizes on the evaluation at a coordinated signal corridor setting. Section 8 summarizes all the works done, highlights the conclusions and offers the author's perspective on future research directions.

2. BACKGROUND AND LITERATURE REVIEW

This section first provides an overview of the basics of transit signal priority, including its benefits and impacts. A comprehensive review on the literature covers a variety of TSP strategies, their categorization and relevant studies, and attempts to answer the questions, such as if TSP is effective and what are the main challenges when developing a TSP strategy. A quick review on the logical layers of a typical signal control system with TSP operations is provided. A review on the connected vehicle technology and how it is being used in current research conclude this section.

2.1. Overview of Transit Signal Priority

2.1.1. What is Transit Signal Priority?

Transit signal priority is an operational strategy that expedites the movements of in-service transit vehicles through signalized intersections. Since its earliest implementation in 1968 [5], TSP has been deployed to reduce transit delay at intersections, minimize transit travel time at arterials, improve transit service reliability, maximize intersection person throughput, and thereby increase transit quality of service. Different from signal preemption which *interrupts* normal signal processes for special events (e.g. emergency vehicles, trains passing), signal priority *modifies* the normal signal processes to better accommodate transit vehicles[6]. The difference between the two operation modes is more conspicuous in a corridor case. For signal preemption, there is no consideration for maintaining the existing signal timing for coordination; its sole purpose is to provide immediate right-of-way to the requesting entity. But for a signal priority, signal coordination may be modified only to a degree that is not significantly impeding the flow of the general traffic.

In general, a TSP control strategy is simply an additional layer in the overall traffic signal control system. And TSP controls are usually not dependent on the type of traffic control system used. But the type of TSP control can be deployed certain limited by the infrastructure availability which are typically related to the type of signal control system being used. Without loss of generality, Figure 1 illustrates an example of the main components for a typical TSP system at a local intersection level. At minimum, a fleet vehicle with some signal emitter, such as optical emitter and Radio Frequency (RF) tag, can send signal to a detector or receiver connected to the signal controller. TSP logic is processed and implemented in the controller, confirmation signals are optionally send back to the vehicle.

The signal emitted from the equipped fleet vehicle can be simply “I am a bus, and I have arrived at the detection zone”. More preferably, it could couple with other location tracking system such as AVL or GPS to allow continuous real-time tracking of the system. Note that, transit or traffic management center may also be added in the loop to form a centralized TSP architecture as described in NTCIP 1211 [7]instead of the distributed TSP architecture that Figure 1 depicts.

Besides being deployed by themselves, TSP strategies sometimes work in conjunction with other preferential treatments. The addition of a queue jump lane [8, 9] is one of the strategies used most frequently in combination with the transit signal priority. Another example is that conditional TSP can be used with bus stop holding strategy to achieve better schedule adherence [10].

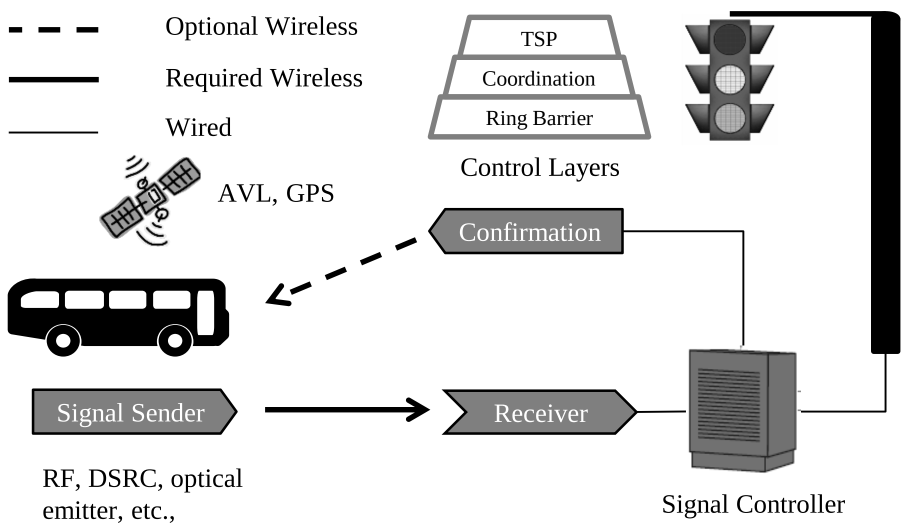


Figure 1: Example of a Transit Signal Priority System at Local Intersection Level.

2.1.2. <u>General TSP Benefits and Impacts</u>

Several benefits are generally expected when a TSP system is in place. The first most obvious benefit is reduced of transit waiting time at local intersection level, or in turn reduced travel time at network-level. Secondly, schedule reliability is also expected to benefit from a TSP system, due to the side-effect of reducing intersection wait time. Because reducing wait time actually minimizes the chance of long wait at intersections, which in turn lowers the fluctuation in actual travel time. These two benefits enabled by a TSP system are considered as two of the five categories of measures for the transit riders' perception of comfort and convenience[11].

In fact, travel time reduction and schedule reliability improvement has directly positive benefits for transit agencies as well. Faster route travel time means shorter turn-around time, which may lead to savings on transit agencies' investment by requiring lower number of buses to maintain a certain schedule. And the ability to adhering to a regular schedule will prevent buses on the same line from bunching up [11], so as to maintain the optimal level of the passenger loads in each bus for passenger comforts.

The third benefit a TSP system may provide is the reduced number of stops along an arterial corridor. Without a TSP system, transit vehicles are normally out of synchronization with the private traffic flow. Thus it is very likely that the bus needs to stop for at least a few seconds after pulling out of a bus stop to get back to the progression band. With a TSP system, more time is given to a bus to reduce the chance the bus needs to make additional stops. And this leads to a number of other benefits: increased rider comfort, lowered tail-pipe emissions, reduced wear and tear on bus mechanical components and less pavement maintenance.

The aggregated effect of these direct benefits increases the attractiveness of this mode of transportation, which indirectly draws more ridership and lowers the use of expensive and eco-destructive private vehicles. Notwithstanding its benefits, a TSP system may also induce negative impacts. And the most cited impacts are excessive delays on the side-street traffic. In some instances, extra stops could be observed in the arterial traffic flow.

As shown in Table 1, field studies have generally reported positive results for transit signal delay ranging from 6 to 46%, and impacts to other traffic varies greatly. Unfortunately, even the authors [6]admitted that these statistics are only suitable to provide a general sense of the benefits and impacts by a TSP system implemented in the field. Due to the difficulty to conduct collect before and after data in a controlled manner, it is not meaningful to use the statistics for any scientific comparisons.

Table 1:Summary of Benefits and Impacts of TSP (source [6]).

Location	Type	No. of Signals	TSP Strategy	Benefit/Impact
Portland, OR Tualatin Valley Hwy.	Bus	13	early green, green extension	• Bus travel time savings of 1.7 to 14.2% per trip • 2 to 13 seconds reduction in per intersection delay • Up to 3.4% reduction in travel time variability
Europe	Bus	Five case study sites	Various	• 10 seconds per intersection reduction in transit signal delay • 40 to 80% potential reduction in bus signal delay • 6 to 42% reduction in transit travel times in England and France • 0.3 to 2.5% increase in auto travel times
Seattle, WA Rainier Avenue	Bus	20	early green, green extension	• 24% average reduction in stops for TSP buses • 5-8% reduction in travel times • 25-34% reduction in average intersection bus delay for TSP eligible buses • 40% reduction in critically late trips (trips not completed before next trip scheduled start)
Toronto, Ontario	Street car, Bus	350	early green, green extension	• Up to 46% reduction in transit signal delay • 10 street cars removed from service • 4 buses removed from service in 2 corridors • Cross street traffic not significantly affected
Chicago, IL Cermak Rd	Bus	15	early green, green extension	• 7 to 20% reduction in transit travel time depending on time of day, travel direction • 1.5 second/vehicle average decrease in vehicular delay (range: +1.1 to –7.8) • 8.2 second/vehicle average increase in cross-street delay (range: +0.4 to +37.9)
San Francisco, CA	LRT, trolleys	16	early green green extension	• 6 to 25% reduction in transit signal delay
Los Angeles, CA Wilshire & Ventura Blvds	Bus	211	early green, green extension, actuated transit phase	• 8% reduction in average running time • 33-39% decrease in bus delay at intersections • Minimal impacts to cross street traffic: average of 1 second per vehicle per cycle increase in delay
Pierce County, WA Pacific Ave and 19th St. corridors	Bus	42	signal coordination, early green, green extension	• Signal coordination reduced total signal delay 18-70% for auto traffic, and 5-30% for transit • TSP reduced transit signal delay an additional 20-40% beyond signal coordination • TSP had little impact on traffic progression

2.2. Transit Signal Priority Strategies

There are a large number of methods to modify the signal operations to provide favorable treatments to a transit vehicle. Differing in the use of detection system, decision making mechanism, various priority strategies can be generally categorized into passive, active, and adaptive. Passive and active priorities were devised much earlier than the adaptive TSP strategies, thus they are hereby called traditional TSP strategies.

2.2.1. Conventional TSP Strategies

2.2.1.1. Passive Priority

The earliest preferential strategy ever devised for buses at an intersection is to provide signal priority regardless the presence of a transit vehicle. Transit-based signal coordination scheme is a good example where signal progressions are computed based on the speed of buses. The advantage is easy implementation, does not require a transit detection/priority request generation system [12] and once it is in place it requires no maintenance costs at all[13]. However, this treatment usually causes unnecessary or even excess delays, stops and frustrations to non-transit travelers [12].

Although easy to implement, passive priority is generally not suited for priority in mixed traffic conditions. The simulation study by [14]showed that passive priority caused long vehicle queues on the main street, which in turn delayed the priority but out of the progression that was originally optimized for unimpeded bus traveling speed.

However, passive priority may be most useful on an arterial with transit vehicles that can run on exclusive right-of-way. For example, [15] formulated an arterial progression model based on MAXBAND MILP-2 for tram vehicles. The simulation case study of 9 signalized intersections in VISSIM showed the tram vehicle progress through the corridor without any delay or a single stop. In another study involved with Bus Rapid Transit (BRT), a model was developed to provide unconditional priority to a BRT line in an arterial corridor [16]. The case study showed that unconditional TSP is feasible during off-peak hours.

2.2.1.2. Active Priority

Bus priority can also be provided only when a transit bus is detected. Comparing to passive strategies, the active priority methods require a detection mechanism to check-in and, optionally, check-out buses, and a set of decision rules to granting priorities. Depending if the decisions rules are in place, active priority can be conditional or unconditional. Unconditional priority [16, 17]normally just need to know the detection of a bus, and the system provides priority without checking if any established criteria are met. Conditional priority is more sophisticated, and decisions of granting a priority is made usually based on a bus's lateness, occupancy and other criteria [10, 12, 18, 19].

Some field studies have shown that active TSP is capable of yielding as much as 34% decrease in bus delay in large metropolitan areas such as Seattle[3]. Thus, active TSP priority strategies have gained large popularity and are generally considered as the state-of-the-practice for field TSP deployments.

There are four basic active priority treatments: phase extension, red truncation, phase insertion and phase suppression [6]. More advanced techniques, such as actuated transit phases, phase rotation, phase splitting and so on, have also been extensively documented in many other studies [11, 20]. All these priority treatments have pre-defined rules that can be activated in different combinations of traffic volume, bus arrival time, schedule lateness, and so on.

However, the primary problems with active priority strategies is its lack of ability to handle multiple priority requests simultaneously due to the First-Come-First-Serve (FCFS) policy [21]. This limitation implies that once the priority is given to the first arriving bus, no further priority can be furnished to the second arrival even if the second bus is with a higher need and the accommodation of the first priority leads to higher delay to the second one. Zlatkovic, et al. [22] have showed FCFS policy may be worse than a policy that provides no priorities at all, and they further developed an if-then algorithm to circumvent the problem for conflicting priority cases that are relatively simple.

To handle multiple priority requests at a single intersection simultaneously, AASHTO, ITE and NEMA formed a joint committee to formulate the NTCIP 1211 standard that defines the system components and prioritization process [7] for the standard signal control prioritization system. As illustrated in Figure 2, the priority request for each bus is generated by the PRG, and is processed in the priority request server using a pre-defined multi-priority strategy rule.

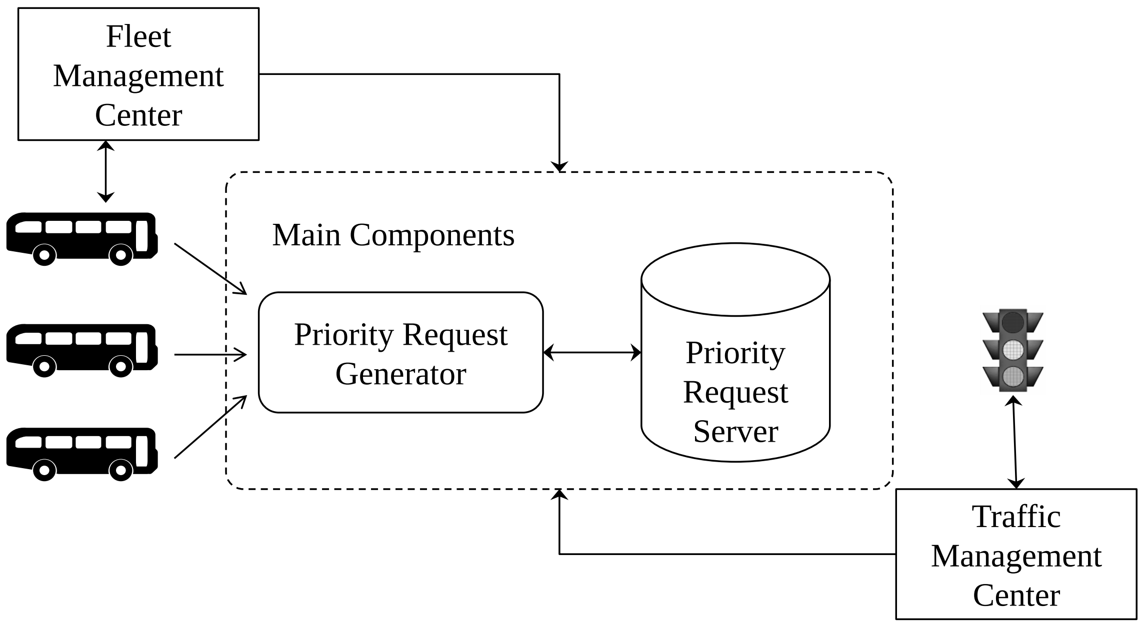


Figure 2: Signal Control Prioritization System.

Although the operations framework and system objects are well defined in the standard, the prioritization strategy is up to the users to define. Research into the specific prioritization strategies remains sparse in the literature. Handling multiple priority requests in a coordinated corridor with active TSP strategy is almost non-existent.

2.2.2. Adaptive TSP Strategies

Adaptive priority, as is sometimes called real-time priority (such as in [11]), usually refers to a TSP strategy whose control decision is derived from mathematical models that tries to optimize the signal timing settings to achieve a given criteria. These criteria may include vehicle delay, transit delay, person delay, and headway variations. These models, at minimum, use the information regarding the bus, prevailing traffic conditions and current signal timing states as inputs to either select alternative timing plans or redesign entirely the current timing parameters.

This type of priority strategies takes a systematic approach to make the best decisions that can take other traffic into simultaneous considerations. These strategies normally require the pre-existence of a functional adaptive traffic control system. Therefore, deployment costs of an adaptive TSP strategy are normally high at locations without an existing adaptive signal system. Although the numbers are increasing, the adaptive TSP deployments in the United States remain very low [6].

Despite its cost in real-world implementation, these strategies are generally considered as one of the best ways to find an optimal balance between TSP benefits and impacts. Many recent studies have shown promising results. Most of the studies focus on developing models that achieve similar effects of those treatments used in the active TSP strategy, such as green extension, red truncation and phase insertion.

According to modeling objectives, there are at least three distinct branches of the adaptive TSP strategy:

- Models that aim at minimizing vehicle delays
- Models that strive to maintaining bus headway deviations
- Models that improve bus schedule adherence

Table 2 summarizes the most recent works on the topic of transit signal priority. Detailed descriptions and the main problems with current research are provided in the following subsections.

Table 2: Summary of Recent Studies on TSP.

Method Type	Year	Authors	Model	Objective	Considerations for Transit Vehicles							Considerations for Auto Vehicles			Active TSP Capable			Controller Considerations		
					Uncertain Arrival	Interval Arrival Time	Multiple Requests	Position in Queue	Headway Deviation	Schedule Deviation	Passenger Load	Queuing	Person Delay	Stochastic Arrival	Green Extension	Red Truncation	Phase Insertion	Ring-Barrier	Coordination	Phasing Sequencing
Model-Based	2008	Q. He	MILP	Delay		x	x				x		x		x	x		x	x	
	2011	L. Meng	MINP	Delay				x			x	x			x	x		x	x	x
	2011	E. Christofa	MILP	Delay			x	x			x	x	x		x	x				
	2012	Q. He	MILP	Delay			x				x		x		x	x		x	x	
	2012	W. Ma	DP	Delay, Headway			x		x		x				x	x	x			
	2012	G. Wu	NP	Delay	x						x		x		x	x		x	x	
	2013	Y. Lin	MILP	Delay			x		x		x		x		x	x			x	
	2005	M. Vasudevan	DP	Headway			x		x		x		x	x	x	x			x	
Heuristic / AI	2008	J. Stevanovic	GA	Delay	x									x	x	x		x	x	x
	2014	M. Ghanim	GA	Delay, Schedule	x					x		x		x	x	x			x	
	2004	K. Ling	RL	Headway					x											
	2011	M. Tlig	MA	Headway		x	x	x	x						x	x	x		x	
Rule-Based	2010	W. Ma	-	Schedule						x	x				x	x	x	x	x	
	2012	N. Hounsell	-	Headway					x						x	x				
	2008	A. Skabardonis	-	Delay			x				x				x	x	x	x	x	

Note: DP (Dynamic Programming), MILP (Mixed Integer Linear Programming) , MINP (Mixed Integer Nonlinear Programming) , GA (Genetic Algorithm), RL (Reinforce Learning), MA (Multi-Agent)

2.2.2.1. Delay Minimization

Li, et al. [23] proposed an adaptive TSP model that minimizes weighted sum of transit vehicle and other traffic delays. The model optimized the green splits of three cycles for a dual-ring controlled traffic signal. By computing not only the green but also the red time for each phase, the model was able to capture the evolution of TSP-induced queues and their delays using deterministic queuing theory. Due to the nonlinear nature of phase red-time and vehicle delays, the optimization model is MINP. A field study was reported with 43% bus delay reduction and 12% of delay increase on passenger car.

Christofa and Skabardonis [24] developed a transit signal priority system that is based on the combined person delay of transit and auto vehicles. The auto vehicle delays are first estimated using deterministic queuing theory, where arrivals and departures are constant. The position of a transit vehicle in the auto vehicle queue is explicitly modeled to obtain the bus delay. In addition, the passenger load of each bus or auto vehicle is used as the weighting factor to determine the relative priority among multiple conflicting transit routes as well as between bus and auto vehicles.

Stevanovic, et al. [25] presented a genetic algorithm model that works in micro-simulation environment to optimizes four basic signal timing parameters (i.e. cycle length, offest, splits and phase sequence) and transit priority settings. The objective of the optimization is the sum of total delay and weighted number of stops for all vehicles. Two TSP strategies are made possible by optimizing the transit priority parameters: green extension and red truncation. Taking advantage of the random seeds in the micro-simulation, the stochasticity characteristics of vehicle arrivals are implicily addressed.

Ma, et al. [26] developed a TSP control framework that uses a dynamic programming approach determine a timing plan with minimal bus delays. In a multi-request scenario, each request is weighted by bus occupancy and schedule deviations. Three active priority strategies are explicitly modeled: green extension, red truncation and phase insertion. Although the delay to non-transit vehicles are not computed, the degree of saturation is set as a constraint to ensure the impact to other traffic is not too large. The framework further implements a rolling horizon approach to enhance its real-time control capability. Simulation study has shown up to 30% reductions of bus delays comparing to fixed time control with no TSP implementations.

He, et al. [27] proposed a unified platoon-based framework called PAMSCOD that considers multiple models of travel, excluding pedestrian and bicyclists. The framework includes an MILP model that searches the optimal signal plan by feeding priority requests (buses and/or vehicular platoons) and phasing data to signal controller in real-time. The objectives of the optimization model are to minimize the total of bus and platoon delays and to maximize the slack green time. The slack green is the extra green time available for a typical actuated controller to extend phases until gap-outs or max-outs. This method addresses the shortcoming that an adaptive signal controller usually operates on a fixed split basis, which cannot take advantages of industrial-standard controllers that are based on vehicle actuations.

Skabardonis and Geroliminis [28] developed an analytical model for real-time estimation of travel times for buses along a signalized arterial. At entering the network, if a bus is late, the TSP priority algorithm is carried out to first estimate the bus travel time and then find an undersaturated intersection to apply priority using green extension, red truncation or phase insertion.

Wu, et al. [29] presented a nonlinear program with linear constraints to minimize passenger delays for light rail transit at multiple grade crossings. By assuming fixed cycle length, the computation of transit delay is able to capture the random error of predicted arrival time at each intersection.

Using the concept that longer headway will attribute to higher waiting time, Lin, et al. [30] formulated the passenger waiting time at a bus stop using the forward and backward headway before TSP control. In addition bus stop wait time, a minimization model is developed to also for on-board passenger delay and person delay on the cross street. The simulation result showed some improvement. However, it is not immediately clear how the model developed based on two intersections can easily extend to a signal arterial.

2.2.2.2. Headway Adherence

In the most congested urban areas, transit lines may not necessarily have published timetable; instead, buses may be operated according to the scheduled time headway [31]. This may be efficient when time headways are very short. In high-frequency services, passengers generally do not arrive for a specific bus but expect they would wait less time than the scheduled headway. However, maintaining bus headways is notoriously difficult in mixed traffic environment. Bus bunching may result, and it decreases the bus capacity utilization and causes further delays to passengers [11]. Research has been looked into using TSP to better regulate headways.

Ling and Shalaby [32] used a reinforcement learning (RL) algorithm to adapt to an optimal set of green times to pull and push transit vehicles in order for them to gradually recover to the scheduled headway. By pairing up the current phase status and the bus schedule deviation, an RL agent can calculate the best phase duration while taking into consideration all practical phase length constraints. A simulation study was reported that the RL algorithm brings down the headway deviation by more than 20%.

Vasudevan [33] argued that signal control would actively change the headway adherence at an intersection. However, this change is only local, because the outcome from the bus priority control does not feedback to the progression control level. So how a signal control decision is changing the headway adherence on the route level is not capture. This consideration is crucial to making a bus back to schedule over several intersections.

Tlig and Bhouri [34] developed an innovative multi-agent system that simultaneously regulates general traffic and promotes bus service regularity. The system employs four agents: bus agent, bus route agent, intersection agent and stage (phase) agent. A set of protocols are established for each agent to compute their own properties and to communicate/negotiate with other agents. The priority of a bus is modeled by its schedule lateness. Four TSP strategies are possible: extension, truncation, phase insertion and rotation. A simulation study is conducted on a network with 6 intersections and 3 bus routes. The result shows the proposed method gives the lowest bus headway deviation.

Hounsell and Shrestha [31] proposed a headway-based method to determine when to provide priority and when not to. The priority granting conditions are: 1) forward headway is larger than the scheduled headway and 2) forward headway is larger than backward headway. Since bus headway regularity is directly proportional to passenger waiting time, as the study showed, minimizing headway irregularity can minimize wait time. Simulation modeling has shown about 4% improved wait time.

2.2.2.3. Schedule Adherence

As one of the main metrics for good quality of transit service, bus on-time performance is considered as the primary modeling objective in many of the literature. There may be at least two ways to define schedule adherence: schedule lateness and schedule deviation. Lateness has linearly relationship with delay while it is positive. That means, lateness minimization approach is normally very similar to delay minimization approach. However, schedule deviation is a little trickier and requires the bus to arrive within a certain window of the schedule arrival time. Thus, literature on schedule deviation is very limited.

Albright and Figliozzi [35] recognized that not all intersections along a congested arterial contribute equally to a bus schedule performance. Some intersection is either inherently or temporarily more important than the intersection for any particular bus lines. However, the stochasticity associated with bus schedules and operations make it very difficult to identify those critical intersections. The researchers applied statistical analysis on historical data on a congested corridor in Portland, Oregon to identify factors influencing TSP effectiveness on minimizing bus schedule lateness.

Furth and Muller [10] combined the bus stop holding strategy with a conditional TSP strategy to minimize schedule deviation within 10 seconds of the intended arrival time. Signal priority is provided only to late buses, while early buses will be hold at the bus stops. Although this method is not exactly adaptive TSP, their proposed strategy can shed light on how pull-and-push strategy can help improve schedule regularity. However, this strategy does need carefully tailored schedules and a cooperative process that engages bus operators and supervisors.

Ma, et al. [19] proposed a coordinated and conditional bus priority system for a group of intersections. The authors argued that granting priority at intersection level may not benefit bus arrival at downstream intersections. Thus they proposed to compare the estimated bus delay with the permitted delay specified by the system user. The bus priority is provided only when the difference is too large. And they also proposed two priority strategies to be used at each intersection: 1) normal priority for reducing bus delay for late bus, 2) reverse priority to increase buses delay for early bus. The optimal combination of priority strategies is found using a MILP model. A simulation result show significant improvement in terms of schedule deviations.

Ghanim and Abu-Lebdeh [36] developed a real-time TSP system for a coordinated arterial. The system utilizes a GA optimizer to simultaneously minimize PC delays, transit corridor travel time, and bus schedule deviation by optimizing cycle length, split and offset. The bus arrival times at the intersections along the corridor are predicted using an artificial neural network (ANN) approach. The models were implemented in VISSIM, and the simulation results showed 5-90% reduction for regular traffic delay and 15-85% reduction for transit delays. No schedule deviation results were reported.

2.2.3. Challenges with TSP Modeling

The strength of adaptive TSP strategies is to allow flexible development of a priority control model on top of a signal control system. However, this flexibility also brings about numerous challenges, which include:

- Bus arrival time at the stop bar is inherently random or uncertain
- Interactions between bus and cars confounds bus arrival time estimation
- Finding balance between TSP benefits and impacts
- Locally realized benefits may not be retained system-wide
- Arterial progression speed of buses is not compatible with that of passenger cars

2.2.3.1. Uncertain Bus Arrival Time

All TSP strategies involve making decisions based on the prediction of a future event – in this case, bus arrival at stop bar. For a complicated signal system, advance planning for timing adjustment is the key for TSP success. However, the earlier in advance the planning takes place, the more uncertainty the planning can know about its inputs. That implies ineffectiveness prevails if the planning is sufficiently early. Most of the studies in the literature are based on the assumption that bus arrival times are deterministically available.

He [37] argued that it is important to consider the fact that a bus may not arrive precisely at the time that was predicted. To consider the arrival uncertainty, an interval estimation of the arrival time was used instead of a point estimate. This is called the robust optimization, which is a good attempt to consider randomness in bus arrival time. However, the robust optimization in fact optimizes for worst case scenario. Stevanovic, et al. [25] also expressed the importance of modeling randomness in the design of bus priority schemes. But their approach requires full-scale simulations of many arrival scenarios, which is computationally cumbersome for any practical purposes. The research by Wu, et al. [29] is one of the very few studies that explicitly accounted for the random distribution of bus arrival time, in the form of prediction errors. Nonetheless, prediction errors are generally small intervals, which work very well to account for small perturbation in traffic conditions.

In the context of transit operations, one important source of uncertainty about bus arrival time, or equivalently travel time, is the bus dwell time at a bus stop. Previous simulation [38] and field [14] studies clearly showed that the dwell time variability can significantly reduce the TSP efficacy by causing buses to arrive later or earlier than predicted. Thus, a good TSP strategy shall explicitly account for such randomness.

2.2.3.2. Interactions between Bus and Cars

In mixed traffic environment, interactions between bus and passenger cars are inevitable, for example, bus movement maybe blocked by standing queue. In general, studies have to assume minimal to no interactions between bus and cars[28]. Such assumptions are reasonable if the interactions are limited to lane-changing or car-following. However, when a nearside bus stop is present, the interactions are much more than simply perturbing bus travel speeds. In such cases, interacting activities between buses and cars may have very detrimental effect on the prediction of the bus arrival time. This is because passenger cars may form queues which may extend beyond the entrance of a near-side bus stop, therein blocking the buses' entrances and exits. As observed in [14], if the vehicle queue is short when a bus approaches the intersection, the bus may exit the intersection in the same cycle; if the vehicle queue is long when a bus approaches, it may take the bus up to three cycles to exit. The determination of bus arrival time at the bus stop becomes much more difficult. It is a significant challenge that a TSP system has to address if a near-side bus stop is present at the intersection.

2.2.3.3. Modeling TSP Impacts

Favoring transit movements in a system that is designed for non-transit vehicles necessarily implies that non-transit vehicles will not be able to achieve optimal performance. The difference between optimal and sub-optimal performance of non-transit vehicles is the negative impact imposed by the TSP strategy. Normally, vehicle delay or person delay may be used to quantify this impact. But the formulations are generally second order and hard to account for platoon in a coordinated system. A good TSP strategy shall easily and truly model its impact.

2.2.3.4. Non-transferrable Local Benefits

Except for passive TSP, most modeling consideration in previous TSP studies has been placed on individual intersection level. Although it has been applied to arterials and corridors, almost all TSP strategies (active and adaptive only) are applied at local level. Meaning, the adjustment of timing made at one intersection does not need to take into account the timing states in the neighboring intersections. Looking at any single intersection may not help solve the overall optimization problem. For example, some intersection may be more saturated, so not much can be done to accommodate a bus's priority need; other intersection may be exactly the opposite. In fact, more and more research starts to acknowledge this problem [19, 27, 29, 30, 36].

This problem may not be too obvious when the TSP objective is to minimize delay. After all, any delay savings to the transit vehicle will certainly amount to the total travel time reduction. The real problem is with schedule adherence metrics. Making at one intersection on time may not mean any benefits. But making it on time to the critical intersection, such as the last intersection in the corridor, is the only one that counts. So if the performance measure, such as schedule deviation, is calculated at the corridor level, a good TSP strategy should be applied at the corridor level too.

2.2.3.5. Incompatible Progression Speed

A well-coordinated arterial benefits passenger cars by creating a time-space channel for non-stop traversal of multiple intersections. This time-space channel is called signal progression and is designed for passenger vehicles only. Transit vehicles are inevitably pulled out of signal progression for loading and unloading passengers. That is to say, it is almost impossible to keep the bus within the green band created by the signal progression. Hence, providing priority to buses along the corridor means breaking the optimal signal progression for passenger vehicles; or conversely, keeping optimal progression means not providing priority to buses. This is an important and difficult challenge that a good TSP strategy, especially the one applied on a corridor level, has to address.

2.3. Basics of Signal Control Logics

Since one of the research objectives is to design a real-time signal control system for transit signal priority, it is important to review some basic principles of signal control logics. In the context of transit signal priority, there are generally three layers of signal control logics in the signal control hierarchy. As shown in Figure 3, the three layers from bottom to top are ring-barrier control layer, coordination control layer and transit signal priority control layer.

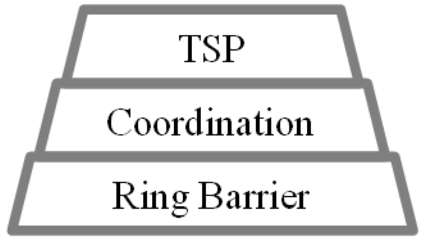


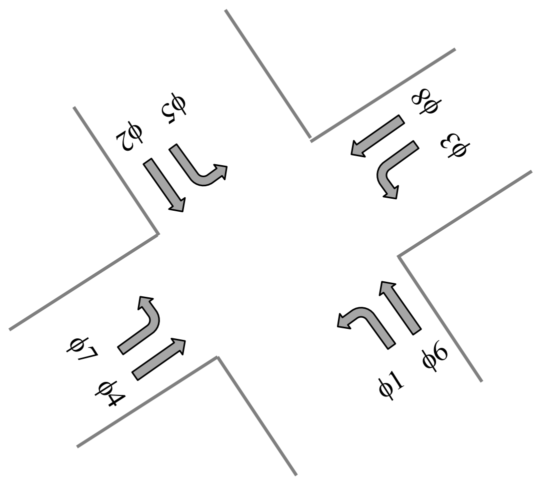
Figure 3: Layers of Signal Control Logics for a Real-Time Transit Priority Signal Control System

2.3.1. Ring Barrier Control Layer

Signal phasing at most intersections in the United States makes use of a standard National Electrical Manufacturers Association (NEMA) ring-and-barrier structure[39]. This structure is embedded in the bottom layer of the control hierarchy so that it defines the basic signal control behaviors that the signal system has to conform with at all time without exceptions. Figure 4-(a) shows a typical four-leg intersection, and each movement is labeled with a number that is called a signal phase. The arrangement of the phases is according to the NEMA standard.

Figure 4-(b) illustrates the ring-and-barrier structure in the ring-barrier diagram. As shown, there are two rings, and each ring consists of a sequence of phases, greens of which display sequentially in time. Phases in the same rings are conflicting phases, greens of which cannot be displayed in the intersection at the same time. For example, phase 2 and 1 conflict each other. Phases in different rings are concurrent phases. For example, phase 2 and 6 can be green at the same time. The barrier is the time when

right-of-way switches between the main and the minor streets; and the barrier ensures that no phases at the main street can be displayed at the same time as the phase at the cross street.



(a) Typical Four-Legged Intersection with NEMA Phasing

Barrier

Ring 1	**φ2**	**φ1**	**φ4**	**φ3**
Ring 2	**φ6**	**φ5**	**φ7**	**φ8**

(b) Ring-Barrier Control Diagram

Figure 4: Illustration of Ring-Barrier Control Logic

2.3.2. Coordination Control Layer

The second control layer is the coordination layer, which is responsible to define coordination behaviors for the intersections in a coordinated arterial system. Figure 5 shows an illustration of an example of the signal coordination logic in a time-space diagram. Note that the discussion for the coordination layer is meant to be brief here, only the main point about signal coordination is to be made. More details, design philosophies, caveats and so on can be found in [40].

The main purpose of the coordination layer is to align the start and/or end times of the coordinated phases in such a way that these coordinated phases in different intersections form a time-space channel that allows main-street vehicles to progress downstream non-stop. For instance in Figure 5, vehicles that travel on phase 2 will be able to travel in the phase 2 progression without being stopped by any intersections. Such a progression is created by properly setting up the offset, Δ_i for the *i*-th intersection. The offset is the time difference between the yield point at the *i*-th intersection and that at the master intersection. The yield point, YR_i, is a fixed point in the cycle of intersection *i*. In this research, Type 170 is used for the definition of yield point [40]. By changing the offset for an intersection, especially the critical intersection, will likely result in shrinking or expanding of the progression band.

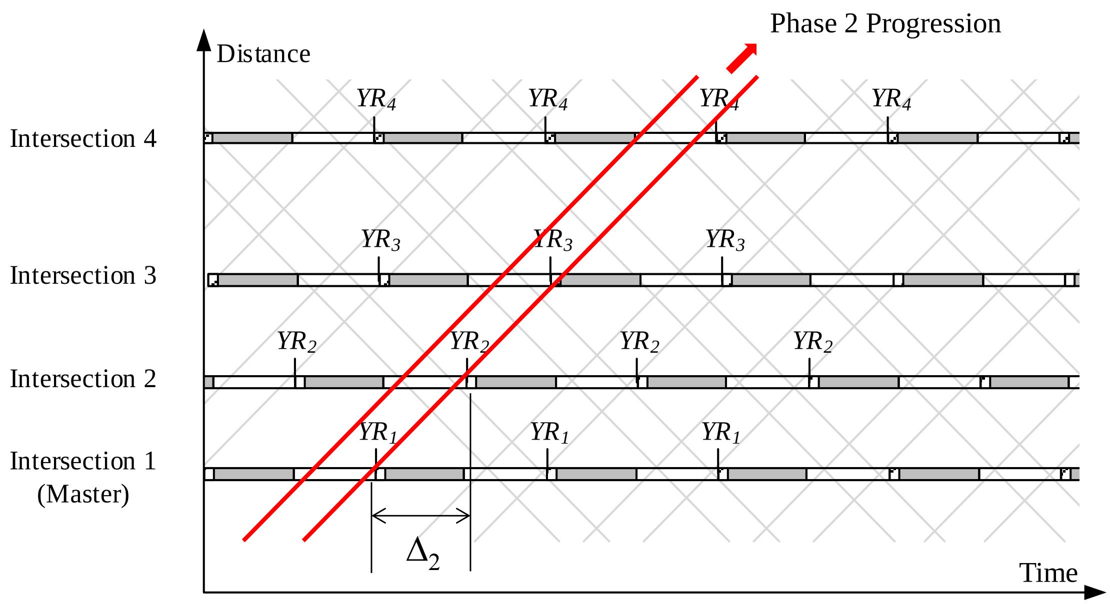


Figure 5: Illustration of Simplified Signal Coordination Control Logic

2.3.3. TSP Control Layer

The transit signal control layer defines special control logics that works to benefit the movement of transit vehicles. For example, in a typical active TSP strategy, the algorithm for granting priority based on bus detection is implemented in this layer. In this research, the adaptive TSP models are implemented in this layer.

2.4. TSP Control using Connected Vehicle

One of the most critical components to a TSP implementation is communications. Sound selection of a communication means could mean the success and failure of a TSP project[6]. Different types of communications exist and are usually coupled with different kinds of detection systems. Recent years, due to the advancement in wireless communications and development of mobile technologies, many more ways for detecting and communicating with a transit vehicle become available. As a result, wireless technologies become cheaper, and communication messages become richer. For example, 4G cellular network enables long distance, high-bandwidth, and relatively low-latency point-to-point communications. This technology is very suitable for the centralized TSP architecture described in NTCIP 1211. For another example, local area network (LAN) using local wireless sensors, such as DSRC and Bluetooth, can enable the development of the distributed TSP system also described in NTCIP 1211.

2.4.1. Connected Vehicle Technologies

The Connected Vehicle (CV) technology refers to the suite of technologies that employ the Dedicated Short-Range Communications (DSRC) protocols for relaying information between vehicles and any DSRC-enabled devices. Due to its capability of two-way transmissions of digital contents over the air with low-latency, high reliability and industry-grade security, the CV technology has become the increasingly popular technology that powers a new wave of traffic safety and mobility applications[4].

Vehicle speed, position and other vehicle parameters that are otherwise difficult to collect can be continuously obtained using the CV technology on a real-time basis. The communications range of a typical DSRC unit is about 3,000 feet (1,000 meters), but the actual range may be less than 1,000 feet (300 meters) due to line-of-sight obstructions and other environmental varieties [41]. This necessarily implies a CV-enabled traffic controllers can easily perceive the ever-changing traffic conditions in its vicinity and make intelligent decisions accordingly.

In order to promote the standardization of using the DSRC protocol, the Society of Automobile Engineers [42] complied a message set dictionary. The probe vehicle data (PVD) message is a standard message in the dictionary, and is used for a vehicle to send vehicle attributes and a snapshot of the recent vehicle's running status to a roadside DSRC unit. Each snapshot can support up to 42 vehicle data elements, including basic and customized vehicle operational statistics.

2.4.2. Application of Connected Vehicle on Adaptive TSP

One way is to use the enriched dataset to provide more accurate arrival predictions for both vehicles [27, 43] and buses. Another direction is to propose a completely different signal control paradigm along with the TSP algorithms that fully utilizes the tracking ability of the connected vehicle on individual vehicles [43]. Adopting a new signal control paradigm that fully makes use the CV technology may still be unrealistic for at least a while. However, it is conceivable that a TSP model on top of the existing signal paradigm could take advantage of the DSRC technology.

In this research, we assume 100% buses are equipped with CV technology and all traffic signals can communicate with these buses and process the data transmitted via the CV technology. There are two critical usages of the CV technology in this research:

- **Continuous Monitoring:** The two-way communications capability of the CV technology allows a traffic signal controller to track all the buses within its range in real-time. To the authors' best knowledge, not many other technologies currently existing can provide such capability. Nonetheless, this is a critical capability for the adaptive TSP models developed in this research to work in a real-time signal control system. This is because a real-time adaptive system requires constant updates of the signal timing plan to ensure the priorities provided to buses meet the optimum policy. The capability of monitoring bus running status at all time is particularly useful for the rolling-horizon optimization scheme described in section 5.5.1.
- **Enriched Dataset:** The fast and high-bandwidth connections among CV-enabled devices lead to more frequent transmissions of larger chucks of high-resolution vehicle data. Data include bus speed (both current and past), location, route, complete schedule information, number of passengers on-board, whether the bus will skip bus stop, and so much more. The formulations of the TSP models and the queue delay computation algorithm are all based on the assumption that such data are available.

However, although the real-time adaptive TSP system so developed in this research makes use of the advanced features of the CV technology, we don't foresee this system be limited to using only the CV technology. In fact, any other wireless communications technologies which have high-bandwidth and two-way communications capabilities could be employed in this TSP system. For example, 4G LTE cellular communications technology is an alternative candidate.

2.5. Summary

This section presented some basic background on topic of transit signal priority. Literature review provided insights on a plethora of TSP related research, which was followed by a summary on the main challenges seen throughout these studies. A quick overview was provided for the three logical layers of a general TSP signal control system. The section finally ended with a brief overview on the connected vehicle technology, and how it is used to benefit the development of the real-time TSP system developed later.

3. TSP MODEL FOR ISOLATED INTERSECTION

This section* discusses the development of a stochastic mixed-integer nonlinear programming (SMINP) model for real-time TSP control at a single intersection. A basic SMINP formulation based on the two stage stochastic programming theory is presented. The model minimizes the deviation of green time from optimal background green time, as a proxy to the impacts of the priority timing on non-transit vehicles. Second, the basic model is improved by explicitly accounting for the interactions between a bus arriving at a near-side bus stop and the standing queues waiting at the signal. A queue delay computation algorithm is also developed to calculate the critical parameters used for the enhanced SMINP model.

3.1. Introduction

Different from the signal preemption operations, a priority-capable signal control strategy does not interrupt but only modify normal signal operations in favor of the priority vehicles[6]. An effective TSP strategy minimizes the interruptions to other traffic while attempting to provide priority to transit vehicles. In the literature, mathematical models have been proposed to find optimality for various objective functions. Ma, et al. [26] formulated the objective using bus delays, while Li, et al. [23] added the auto delays in the formulation. Christofa, et al. [44] multiplied the estimated

---

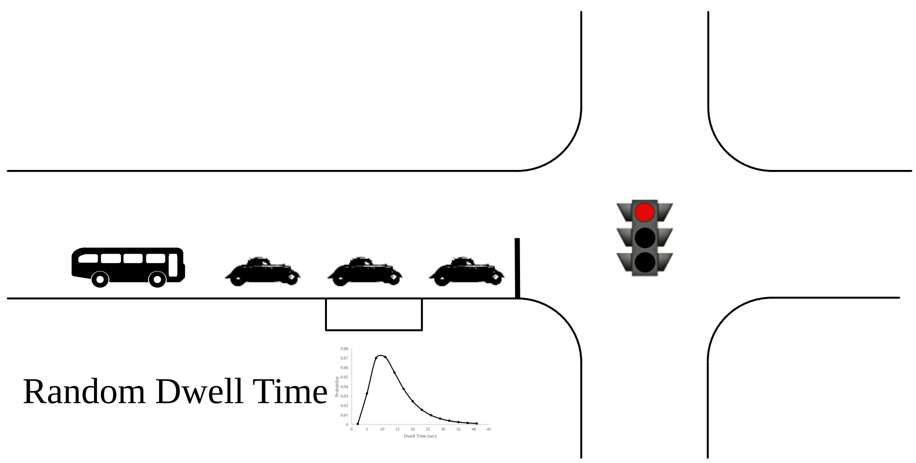
* Part of this section is reprinted with permission from "A Real-Time Transit Signal Priority Control Model Considering Stochastic Bus Arrival Time" by X. Zeng, Y. Zhang, K. Balke and K. Yin, 2014. IEEE Transactions on Intelligent Transportation Systems, volume 15(4), p1657-1666. Copyright 2014 IEEE.

number of occupants in each type of vehicles in order to minimize the person delay. He, et al. [27] took a different approach that minimizes bus delays while maximizing the green times on the phases with higher traffic demands. These studies all confirmed the fact that the attempt to reduce bus delay will necessarily increase the delay to other vehicles, especially those on the conflicting phases. Many of these models give the users the ability to assign weights to the respective traffic flows.

Another key design factor for successful transit priority implementation is the ability to accurately predict the arrival time of the bus at the stop bar [45]. Models have been developed to estimate vehicle arrival times at bus stops or along a corridor [46-48]. Inaccurate predictions would generally result in failed treatments for transit vehicles. Some failed implementations may not be obvious especially in low volume conditions, whereas others may cause bus delays to rise. Stevanovic, et al. [25] also expressed the importance of modeling randomness in the design of bus priority schemes.

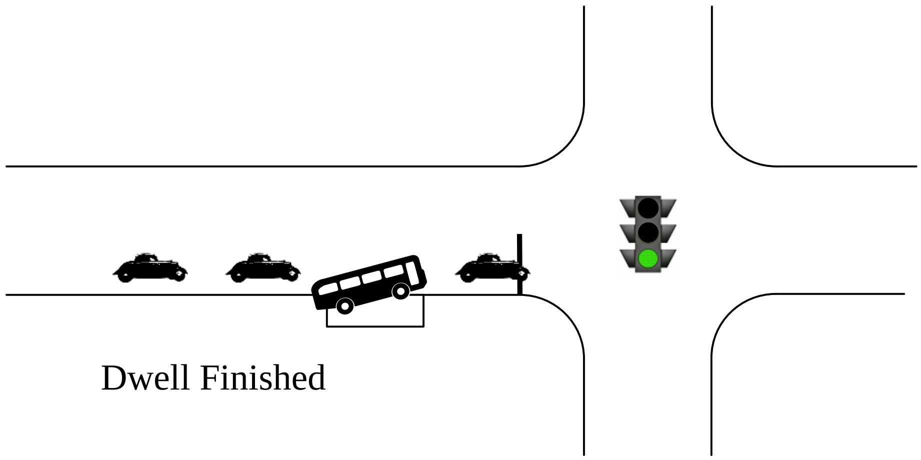
Nevertheless, due to its complexity, there is only few models robustly account for the uncertainty of bus arrival time at the stop bar. A mixed-integer linear programming (MILP) formulation developed by He [37] explicitly allows an input of interval arrival time instead of a point arrival time. Wu, et al. [29] formulated a nonlinear program to minimize the expected delays for light rail transit vehicles at multiple grade crossings. The delay is integrated over a normally distributed function of predicted arrival time, which closed-form is obtained.

The approaches used in previous studies only allowed the consideration of the uncertainty of bus arrival time within a small interval, but large variations of arrival times render great difficulties in seeking for an optimal solution. Furthermore, the prediction of bus arrival time at the stop bar is confounded even more by the presence of a near-side bus stop. Figure 6 illustrates the likely interactions between a transit bus and a vehicle queue: a) long standing queue blocking bus from entering the bus stop; b) after the door closed, if queue exist, this queue will further delay bus from pulling out of the bus stop. Both cases represent extra delay that need to be accounted for when providing signal preferential treatment.



(a) Queue Blocking Bus Entry to Bus Stop

Figure 6: Signal Control Prioritization System.



(b) Queue Delaying Exit of Bus from Bus Stop

Figure 6 Continued.

Due to its complexity, current practice typically ignores or circumvents the problem. For example, the user-manual of the Ring-Barrier Controller (RBC) in VISSIM [49] recommends that a common practice is to place a detector at the exit of the bus stop to detect the departure of a bus. Such an approach eliminates the need to consider bus dwell time. However, this strategy leaves very little time for any control strategies to implement a good timing plan. A good TSP strategy is the one that can provide sufficient priority with low impacts to other traffic; and early detection of a transit vehicle is the key to achieve that [12]. So another solution is to place the bus stop at the far side of the intersection. A need exists to not only be able to capture this impact of this randomness to the prediction accuracy be also to use this uncertainty to our advantage to devise an expectedly optimal timing plan.

3.2. Two Stage Stochastic Programming

A stochastic mathematical program [50, 51] finds an optimal solution to a problem by explicit modeling of the uncertainties of input parameters. This technique has been applied in many areas, including vehicle routing[52], fleet assignment [53], and production planning[54].

In its simplest forms, a stochastic program typically consist of two stages, each of which can be thought of a particular timeline in a decision making process. Stage one is the "now" stage that corresponds to the time that one has to make a decision on a set of decision variables. Let x denote an n_1-element vector of first stage decision variables. All parameters associated with x are collected prior to decision making, and can be deterministically formulated in the "now" stage. Stage two is the "future" stage that represents processes that would occur after the decision making process. Because these "future" processes have not been observed yet, these parameters are inherently random and may take a variety of values when the future unfolds.

Every "now" decision, x, has consequences on the future processes. For every "now" decision that is incompatible with the "future" process, one pays a "cost". This cost is generally termed as the recourse cost, quantified by the second stage decision variables. Let z denotes an n_2-element vector of second stage decision variables. If we can summarize the recourse costs as a function of the "now" decision and the "future" processes (which is called a recourse function, denoted as $f(x,\tilde{\omega})$), then we can find the best “now” decision that minimizes the recourse costs under all “future” scenarios. Here, we present a generic mathematical description for two-stage stochastic program model as in the following:

Stage 1:
$$\begin{aligned} \text{Min} \quad & c^T x + \mathrm{E}\,[\,f\,(x,\tilde{\omega})] \\ \text{s.t.} \quad & Ax \geq b \\ & x \geq 0 \end{aligned} \tag{3-1}$$

where *c*, *A* and *b* are parameters with known values at the timing of decision making, while $\tilde{\omega}$ is a random parameter defined on a probability space (Ω, F, P) . $f(\cdot)$ is the recourse function that gives the penalty of a selection of second stage decision variable on the first stage objective function. $\mathrm{E}\,[\cdot]$ denotes the expectation function. For a given x and an outcome $\omega \in \Omega$, the recourse function can be written as:

Stage 2:
$$\begin{aligned} f\,(x,\tilde{\omega}) = \text{Min} \quad & q^T z \\ \text{s.t.} \quad & Wz \geq r(\omega) - Tx \\ & z \geq 0 \end{aligned} \tag{3-2}$$

where *q*, *W*, and *T* are parameter matrices that do not vary according to the realization of scenario ω, while r is the parameter matrix that do vary for scenario ω. For interested readers, Birge and Louveaux [55] provided an excellent introduction to the fundamentals of stochastic programming.

3.3. Concepts of Timing Deviations

In the theory of traffic signal control, optimal signal timing is normally defined as the combination of signal phase sequence and duration that yields the lowest average delay for passenger car equivalents under a generally stable traffic condition. That is, no alternative timing can produce lower vehicle delay on average under current traffic conditions. In practice, traffic engineers routinely find good timings to lower vehicle delays for different traffic patterns, or implement adaptive traffic signal system to automate the process. Regardless, the goal of a signal retiming or the use of an adaptive signal system is to adjust the timing in order to maintain its optimality for a traffic flow pattern. It is practical to assume that optimal or near-optimal timing is generally available for a well-attended traffic signal infrastructure.

These timings are generally optimal for only passenger vehicles and do not consider real-time arrivals of transit vehicles. So when a TSP operation modifies the signal timing to improve the optimality of the transit operations, it inevitably forfeits the optimality of the signal timings for non-transit vehicles. And the loss of optimality (or increase of vehicle delay) can be loosely estimated using the positive deviation (i.e. compression) of the green time for a phase because of the monotonic relationship between the delay and the green duration of a phase. And this monotonic relationship holds given a fixed cycle length for both under-saturated and oversaturated conditions.

In under-saturated conditions, the uniform delay proposed in HCM [56] is often used to determine the delay to a particular signal phase. The delay can be written in the following:

$$d = \frac{0.5C(1-g/C)^2}{1-v/S} \tag{3-3}$$

where d is the uniform delay, C denotes the cycle length, g and v are the green duration and vehicle volume of the phase respectively, and S is the saturation flow rate. It is obvious, the decrease of green time, g, from C to 0 leads to the increase of d. To account for oversaturated traffic conditions, average overflow delay can be calculated using different models. Figure 7 shows a hybrid model for overflow delay, which increases as the v/c ratio (i.e. degree of saturation) increases.

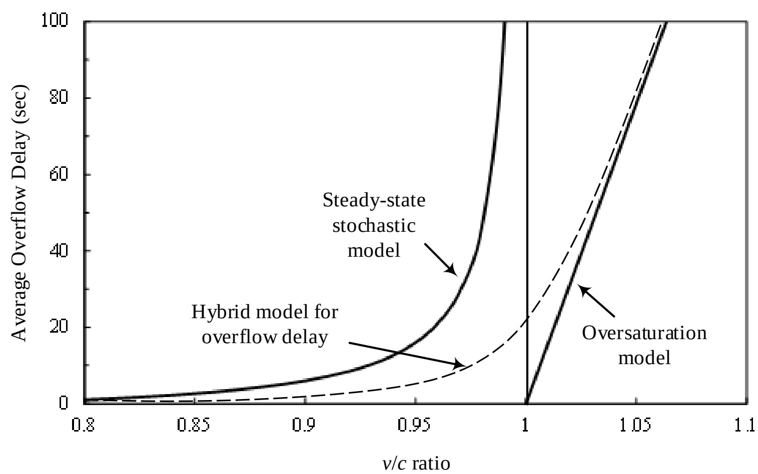


Figure 7: Monotonic Relationship between Overflow Delay and *v/c* Ratio.

Therefore, when the green time of a phase is shortened from the optimal green duration due to a TSP operation, the vehicle delay for that phase is bound to increase. That is to say, it is possible to use the green time deviation (compression in particular) to loosely approximate the increase of PC delay by a TSP operation.

One immediate benefit of using this heuristic approach to approximate TSP impact is to eliminate the need of explicitly writing out a second order delay formulation. Secondly, even if a second-order function is imposed on the deviation term to penalize higher deviation values, the objective function will remain positive semi-definite as long as the coefficients are non-negative. This results in a convex program that can be easily solved by any standard commercialized solvers.

3.4. Formulations for Transit Signal Priority

In a TSP control system, advance planning is the key to any successful strategies. Once detected upstream, the signal control system may need to decide if timing adjustments will be needed so as to prepare for the arrival of a priority vehicle. However, the arrival time of the bus is not certain, and the decision for a certain timing to be implemented "now" may or may not be consistent with the actual bus arrival time in the "future". It is easy to compute the extra bus delay that would occur if we choose a timing that is inconsistent with the actual bus arrival time. Therefore, the bus delay can be thought of the recourse cost, which is a function of the "now" decisions of signal timing and the "future" bus arrival time. Following this logic, we can build a stochastic two-stage mixed integer nonlinear program (SMINP) for a typical TSP problem.

3.4.1. Notations

Sets

J the set of all phases

K the set of cycles within the planning horizon

Decision variables

t_{jk} the start time for phase *j* of cycle *k*

g_{jk} the green time for phase *j* of cycle *k*

v_{jk} the split for phase *j* of cycle *k*

y_{jk} the deviation of green time on phase *j* of cycle *k* from optimal green time

d_j the priority delay of a bus requesting for phase *j*

d_{jk} queue delay for the bus requesting phase *j* of cycle *k*

θ_{jk} priority service decision for a bus at phase *j* of cycle *k*

Parameters

C cycle length

c_{jk} weight for green deviation of phase *j* of cycle *k*

Y,R yellow time and red clearance time

V_{jk} the average flow rate for phase *j* in cycle *k*

S_j the saturation flow rate on phase *j*

X_{jk} the degree of saturation for phase *j*

G_{jk}^{opt} the background green time for phase *j* of cycle *k*

$g_{jk,min}$ the minimum green time for phase *j* of cycle *k*

D_{dwell} dwell time at the bus stop

BR_j the time within a cycle that a bus arrives on phase *j*

$\overline{BR}_j$ Projected bus arrival time on phase *j* excluding possible delays

$\underline{BR}_{jk}$ Latest time to start green on phase *j* of cycle *k* without causing queue delay to the bus on cycle *k*

3.4.2. First-Stage Objective Function

The first stage objective function is also the overall objective function that considers the expected recourse cost computed from the second-stage objective function. Let $\boldsymbol{t}$, $\boldsymbol{v}$ be the vectors of start times and splits of all phases respectively, and BR be unknown bus arrival time. The overall objective function can be formulated as follows:

$$\text{Minimize:} \qquad \sum_{k \in K} \sum_{j \in J} c_{jk} y_{jk}^2 + \mathrm{E}\,[Q(\mathbf{t}, \mathbf{v}, BR)] \tag{3-4}$$

The first term is the sum of the non-expanding changes in green times, as defined in section 3.4.3.3. This timing deviation is used to approximate the impacts of the TSP operations on other traffic. The second term is the expected delay of the priority request, which can be easily evaluated given the distribution of the bus arrival time is known, as in section 3.4.4. Generally speaking, the optimality of the overall objective is found at *the signal timing that cuts down the most bus delays* while *deviates the least from the timing that is optimal for the general traffic*.

Each weight on the first term, c_{jk}, determines how much one phase should be penalized when compared with another phase. In effect, the weight parameter controls the distributions of priority needs in terms of seconds among all the conflicting phases. The weight parameter can be formulated as a function of the congestion level on each phase. The idea is that phases that are more congested shall deviate less from the optimal green time comparing to those less congested phases.

3.4.3. First-Stage Constraints

The formulation in this stage shall realistically model the behaviors and the characters of the signal controller in question. Since this is a single intersection setting, there is no coordination layer. Hence the constraints in this stage defines the logics that reside in the ring-barrier layer, which include precedence relationship, critical degree of saturation requirement, and timing deviation.

3.4.3.1. Precedence Relationship

Head, et al. [57] proposed a precedence relationship to model the standard ring-barrier signal timing structure, illustrated by the precedence diagram shown in Figure 8. Later, He, et al. [27] applied the framework to develop a deterministic priority model which minimizes the delay of priority requests. We applied this precedence framework in the formulations of the first stage constraints.

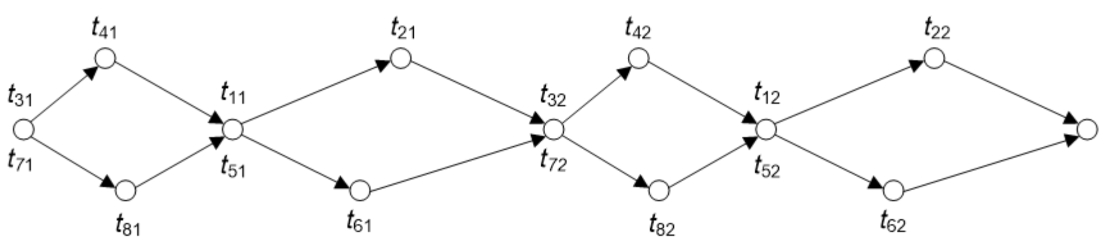


Figure 8: Precedence Diagram (source: Head, et al. [57]).

The precedence diagram is a schematic representation of the mathematical model of the ring-barrier control logic. In Figure 8, the *t* variable at a node represents the start time of a phase; the time difference between two *t* variables is the split of the enclosing phase, which is represented by a *v* variable. To completely define the ring-barrier control logic using this precedence model for the next two cycles in the planning horizon, the following constraints are required:

$$t_{1,k} = 0; \qquad \forall k \tag{3-5}$$

$$\begin{array}{l} t_{2,k} = t_{1,k} + v_{1,k}; \quad t_{3,k} = t_{2,k} + v_{2,k}; \quad t_{4,k} = t_{3,k} + v_{3,k} \\ t_{6,k} = t_{5,k} + v_{5,k}; \quad t_{7,k} = t_{6,k} + v_{6,k}; \quad t_{8,k} = t_{7,k} + v_{7,k} \end{array} \qquad \forall k \tag{3-6}$$

$$t_{1,k} = t_{5,k}; \; t_{7,k} = t_{3,k}; \; t_{6,k} = t_{2,k} \qquad \forall k \tag{3-7}$$

$$t_{4,k} + v_{4,k} = kC \qquad \forall k \tag{3-8}$$

$$v_{jk} = g_{jk} + Y + R \qquad \forall j, \forall k \tag{3-9}$$

$$g_{jk} \geq g_{jk,\min} \qquad \forall j, \forall k \tag{3-10}$$

$$t_{jk}, g_{jk}, v_{jk} \geq 0 \qquad \forall j, \forall k \tag{3-11}$$

This formulation explicitly models the ring-barrier control structure [58] that is widely used in North America. Constraint (3-6) defines the timelines and sequences of all the phases in both rings. Constraint (3-7) indicates which phases are serving as barriers.

Constraint (3-8) defines the end time of the planning horizon as a multiple of cycle length. This would allow the optimization get back to the normal cycle start time if the intersection is a part of a coordinated corridor. The minimum green requirement is defined in Constraint(3-10).

3.4.3.2. Critical Degree of Saturation

Given the average saturation flow rate for a phase S_j, the average flow rate V_{jk}, for the phase in cycle k, and the effective green time g_{jk} and cycle length C_k for each cycle, one can ensure the degree of saturation for the phase (X_j) over the planning period to be less than the maximum allowable value, X_c:

$$X_j = \frac{\sum_{k \in K} V_{jk} C_k}{\sum_{k \in K} S_j g_{jk}} \le X_C \qquad \forall j \tag{3-12}$$

Note that, inequality (3-12) only restricts the overall degree of saturation of a phase X_j, but not X_{jk}. This implies, if the green a phase j in one cycle k is too short (e.g. rendering oversaturation), then the green for the same phase in the other cycles within the planning cycles have to be long enough to clear the excessive queue from cycle k. This formulation renders additional flexibility in adjusting the timing in favor of the transit bus, but the resulting temporary oversaturation may have undefined behavior. One of the most infamous consequences is left-spillback or blockage [59]. To avoid undefined behaviors, additional constraints can be imposed to ensure that the minimum green time for a phase has to meet the maximum degree of saturation at every cycle. All the above inequalities jointly define the limits for timing adjustment.

3.4.3.3. Timing Deviations

With basic signal control logics defined as in above, it is important to quantify the impacts of a TSP operation on non-transit vehicles. As argued in section 3.3, the signal timing deviation can be used to loosely approximate the vehicle delay increase due to signal timing changes. In particular, not just any signal timing changes will increase the average vehicle delay on a phase, but only the changes that causes a phase duration to be shortened. Constraint (3-13) defines the deviations of new green times g_{jk} from optimal background green times G_{jk}^{opt}.

$$\begin{aligned} & y_{jk} \geq G_{jk}^{opt} - g_{jk} \\ & y_{jk} \geq 0 \end{aligned} \qquad \forall \boldsymbol{j}, \forall \boldsymbol{k} \qquad (3\text{-}13)$$

The two inequalities effectively dictate that only the positive deviations are penalized, and any increase of g_{jk} from G_{jk}^{opt} has no direct costs to the objective function. However, it should be noted, given a fixed planning horizon and the precedence relationship, the expansion of a phase necessarily leads to the compression of the conflicting phases. So the penalty for the expansion of a phase is simply the sum of penalty of the compression of all its conflict phases. Penalizing an expanded phase is considered a double count.

3.4.4. Second-Stage Objective Function

For given ***t***, ***v*** and a number of random events $\omega \in \Omega$, the recourse function, $Q(\bullet)$, is deterministically computable. With a well-defined discrete probability space $(\Omega, \mathcal{F}, P)$, the expectation can be evaluated by $E(Q) = \sum_{\omega \in \Omega} p(\omega) Q(\omega)$. For a given discrete random event (ω), the second stage recourse function of a classical two-stage stochastic program can be formulated as the following:

$$Q\left(t, v, BR_j^s\right) := \min \sum_{j \in J} o_j d_j^s \tag{3-14}$$

BR_j^s denotes a realized scenario, *s*, of bus arrival time on phase *j* out of all the possible arrival scenarios in Ω . d_j^s denotes the delay to the priority request placed on phase *j*, which is a function of the bus arrival time and current signal timings. The weight, o_j, of the priority delay determines the level of priority for a bus. This priority can be formulated based on need, for example, as a function of the bus passenger loads or bus schedule lateness.

3.4.5. Second-Stage Constraints

The constraints in the second-stage mostly concerns with the computations of bus priority delay using the timing variables from the first stage and the random arrival times as the input parameters. This stage enumerates all the scenarios in Ω . For each scenario *s*, a set of constraints (from (3-15) to (3-20)) are defined to compute the bus delay. For notation convenience, superscript *s* is dropped for all these constraints:

$$BR_j \geq t_{j,k-1} + g_{j,k-1} - (1-\theta_{jk})M \qquad \forall k \in K \setminus \{1\}, \forall j \qquad (3\text{-}15)$$

$$BR_j \leq t_{jk} + g_{jk} + (1-\theta_{jk})M \qquad \forall k, \forall j \qquad (3\text{-}16)$$

$$\sum_{k \in K} \theta_{jk} = 1 \qquad \forall j \qquad (3\text{-}17)$$

$$\theta_{jk} = \{0,1\} \qquad \forall j, \forall k \qquad (3\text{-}18)$$

Where θ_{jk} is a binary variable identifies which phase and cycle the bus will be served. For example, if bus arriving after end of phase *j* of cycle *k*-1 (i.e. inequality (3-15)) and before the end of phase *j* of cycle *k* (i.e. inequality (3-16)), then θ_{jk} is 1. For all other cycles, θ_{jk} are 0's. *M* is a large constant that can be set as the end time of the planning horizon (i.e. $|K|C$).

Assuming no delays caused by vehicle queues dissipating before the bus, the delay to the bus is simply $d_j = \max\{t_{jk} - BR_j, 0\}$ if the bus is to be served at phase *j* of cycle *k* (i.e. $\theta_{jk} = 1$), which can be equivalently expressed as:

$$d_j \geq t_{jk} - BR_j - (1-\theta_{jk})M \qquad \forall j, \forall k \qquad (3\text{-}19)$$

$$d_j \geq 0 \qquad \forall j \qquad (3\text{-}20)$$

The formulation to compute priority delay via constraints (3-15) to (3-20) is for a single bus. For multiple buses, each bus requires a separate set of these constraints. So if there are N buses, each bus have S number of arrival scenarios. There are a total of $N \cdot S$ sets of these constraints required.

3.5. Bus Delay from Interacting with Vehicle Queues

3.5.1. Formulation Change to Consider Queue Delay

A critical issue arises when it comes to determine the arrival time of the bus, BR_j, at the stop bar. Current practice generally assumes a constant travel time from the detection time of the bus. However, if a near-side bus stop is present, buses may interact with standing queues when it enters or exits the bus stop. These interactions complicate the estimates bus arrival times at the stop bar.

Figure 9 illustrates a possible bus trajectory when the bus is approaching the intersection with a near-side bus stop. It is not unlikely that a bus needs to stop as many as three times at an approach with a near-side bus stop, even under unsaturated traffic conditions. To develop a robust optimization scheme, the computation of priority delay needs to consider the bus interactions with vehicle queues and the bus stop. For all practical purposes, it is assumed that vehicle arrival rates are constant, acceleration and deceleration for bus are negligible, and the bus dwell time is known.

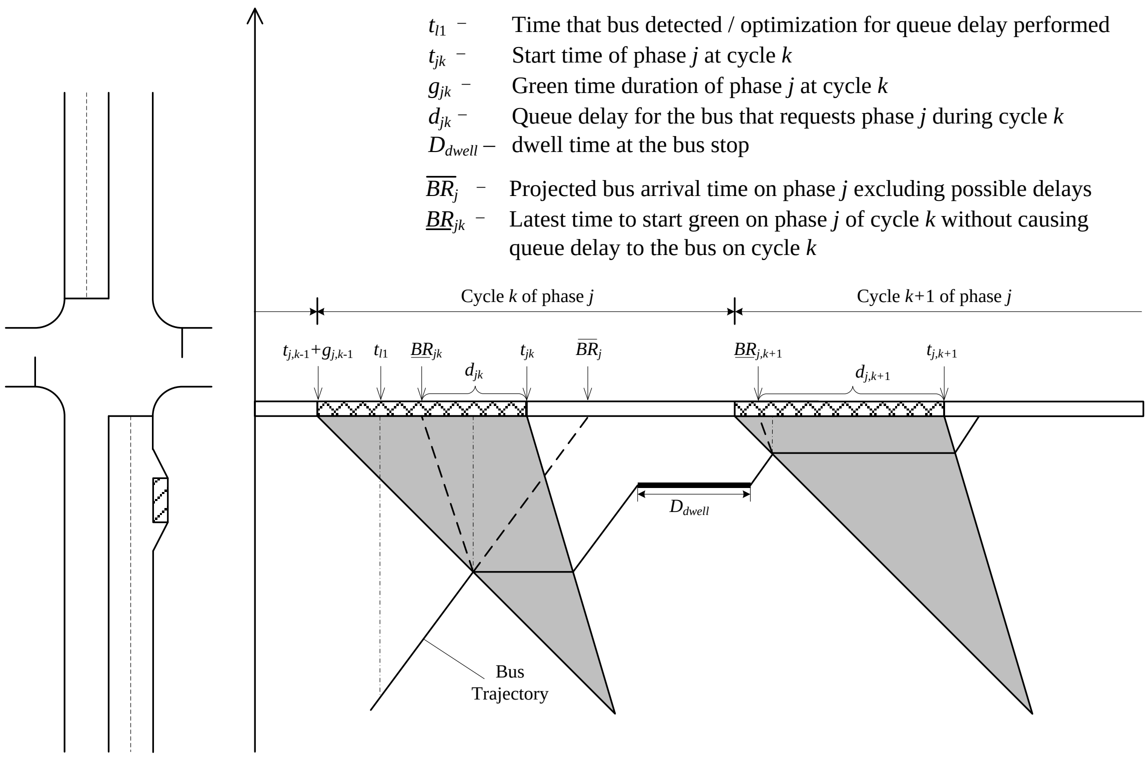


Figure 9: General Architecture of the Simulation Evaluation Platform

First of all, recognize that the summation of all stopping time minus the dwell time is the queue delay time, which is controllable through the start and end time of phase *j*. Let the projected arrival time of the bus under free flow conditions be $\overline{BR}_j$, which can be calculated with the location of the bus and its running speed. Denote d_{jk} as the queue delay on cycle *k* for the bus requesting phase *j*. And let D_{dwell} be the dwell at the bus stop. The actual bus arrival time $BR_j = \overline{BR}_j + D_{dwell} + \sum_{i=1}^{|K|} d_{ji}$. Replace the arrival time of the original formulation (inequality (4-16) and (4-17)), we get:

$$\overline{BR}_j + D_{dwell} + \sum_{i=1}^{|K|} d_{ji} \le t_{jk} + g_{jk} + (1-\theta_{jk})M \qquad \forall k, \forall j \tag{3-21}$$

$$\overline{BR}_j + D_{dwell} + \sum_{i=1}^{|K|} d_{ji} > t_{j,k-1} + g_{j,k-1} - (1-\theta_{jk})M \qquad \forall k \setminus \{1\}, \forall j \tag{3-22}$$

Further let $\underline{BR}_{jk}$ be the latest time to start phase *j* green of cycle *k* so that there would be no queue delay on cycle *k* for the bus. And this means except *k* = 1, all $\underline{BR}_{jk}$ will be dependent on all d_{jk} from previous cycles. This implies if |*K*| is large, the number of constraints will increase exponentially. Fortunately, |*K*| is generally small. Therefore, the priority delay with consideration of queue delay is computed as:

$$d_{j,k-r} \ge t_{j,k-r} - \underline{BR}_{j,k-r} - (1-\theta_{jk})M \qquad \begin{array}{l}\forall j, k \in K \setminus \{1,...,r\},\\ r \in \{0,..,K-1\}\end{array} \tag{3-23}$$

$$d_{j,k-r} \ge 0 \qquad \begin{array}{l}\forall j, k \in K \setminus \{1,...,r\},\\ r \in \{0,..,K-1\}\end{array} \tag{3-24}$$

$$d_j = \sum_{k=1}^{K} d_{jk} \qquad \forall j \tag{3-25}$$

Minimizing the overall bus delay due to queue, d_j, will result in minimal queue delays in all cycles. And this means except *k* = 1, all $\underline{BR}_{jk}$ will be dependent on all d_{jk} from previous cycles. This implies if |*K*| is large, the number of constraints will increase exponentially. Fortunately, |*K*| is generally small.

3.5.2. <u>Computing Critical Parameters</u>

To enable the estimation of queue delays at current and future cycles from the queuing diagram, the most critical time points to be computed are $\underline{BR}_{jk}$. To do this, it is important to first examine the scenarios of a bus trajectory when approaching an intersection stop bar. Figure 10 simplifies all possible cases of the interactions between a bus and queues over several cycles in a time-space queuing diagram. These cases are summarized as following:

- Case 1: the bus will meet the end of the queue before arriving at the bus stop or the intersection stop bar (e.g. bus No.1 trajectory in cycle *k* of phase *j*).
- Case 2: the bus arrives at a bus stop and dwells for a short duration that it leaves the bus stop and joins the queue downstream (e.g. bus 1 trajectory after leaving the first queue it met in cycle *k* of phase *j*).
- Case 3: the bus arrives at a bus stop and dwells for a long duration that the queue backs up to the bus stop and the bus closes its door before the queue dissipates (e.g. bus No.2 trajectory after leaving the first queue it met in cycle *k+1* of phase *j*).
- Case 4: the bus arrives at a bus stop and dwells for a long duration that the queue backs up to the bus stop and the bus closes its door after the queue dissipates (e.g. bus No.3 trajectory).

Case 2 and 4 are actually variations of case 1 and 3, respectively. Therefore, studying case 1 and 3 are sufficient for capturing all possible scenarios. Figure 10 demonstrated the principle and the critical time points to give a reasonable estimate to the bus delay caused by queue. However, for computing these time points, it is necessary to make some simplifying assumptions as follows:

- The timing about the start (t_{jk}) and end ($t_{jk} + g_{jk}$) times of phase *j* in the immediately past, the current and the next few cycles are known.
- The bus travels at the desired speed (v_1) as soon as it is not dwelling at a bus stop or within a blocking queue. Note, use $-v_1$ for all computations using bus speed.
- The speeds of queue forming (v_2) and dissipating (v_3) shockwaves are known and are relatively stable.

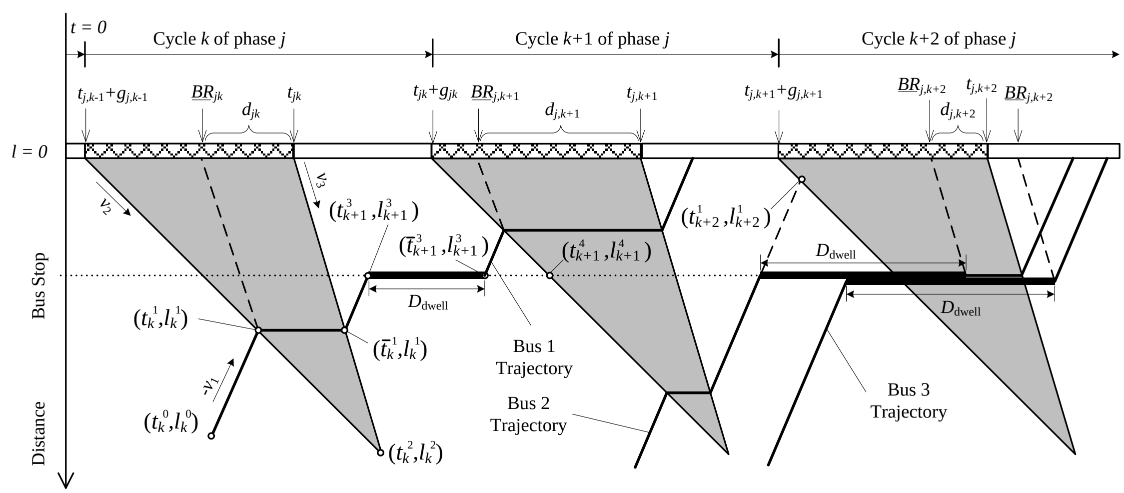


Figure 10: Critical Points for Uniquely Define A Bus Trajectory in Four Cases.

To be more specific, there are a total of five critical time-space pairs that needs be computed for every cycle of phase j in order to determine the case with which the bus is projected to encounter. These pairs are dubbed as follows: (t_k^0, l_k^0) denotes the initial state of bus at cycle k; (t_k^1, l_k^1) denotes the intersection of bus and queue trajectories; (t_k^2, l_k^2) denotes when and where the signal queue is expected to dissipate completely; (t_k^3, l_k^3) denotes when bus will arrive at the bus stop, or stop bar if no bus stop downstream; (t_k^4, l_k^4) denotes when the queue will back up to where the bus stop is, with l_k^4 always equals to the distance of the bus stop from stop bar l^{bus} .

Let k denote the cycle of phase j when the computation of bus trajectory is to be performed. And it is convenient to set the most recent end time ($t_{j,k-1}+g_{j,k-1}$) of phase j in the past as time zero. Let us consider the trajectory of bus No. 1. Given the bus No.1 is detected at (t_k^0, l_k^0) , it is obvious:

$$t_k^1 = \frac{l_k^0 + v_1 t_k^0 + v_2(t_{j,k-1} + g_{j,k-1})}{v_1 + v_2} \quad \text{and} \quad l_k^1 = -v_1(t_k^1 - t_k^0) + l_k^0 \tag{3-26}$$

$$t_k^2 = \frac{v_3 t_{jk} - v_2(t_{j,k-1} + g_{j,k-1})}{v_3 - v_2} \quad \text{and} \quad l_k^2 = v_3(t_k^2 - t_{jk}) \tag{3-27}$$

By comparing t_k^1 and t_k^2, the bus is projected to be blocked by the queue for d_{jk} (case 1). It is then very easy to compute $\underline{BR}_{jk}$ and $\overline{t}_k^1$. Case 2 starts immediately following the bus being released from the queue at $(\bar{t}_k^1, l_k^1)$. Given there is a bus stop (l^{bus}) downstream of l_k^1 and the bus is not skipping this stop, it immediately follows for the cycle k+1 of phase j:

$$l_{k+1}^3 = l^{\text{bus}} \quad \text{and} \quad t_{k+1}^3 = \frac{l_{k+1}^3 - l_k^1}{-v_1} + \bar{t}_k^1 \quad \text{and} \quad \bar{t}_{k+1}^3 = t_{k+1}^3 + D_{\text{dwell}} \tag{3-28}$$

$$t_{k+1}^4 = \frac{l_{k+1}^3}{v_2} + (t_{jk} + g_{jk}) \tag{3-29}$$

By comparing $\bar{t}_{k+1}^3$ and t_{k+1}^4, the bus is projected to be able to leave the bus stop before the queue starts to back up to the bus stop again. Then, using $(\bar{t}_{k+1}^3, l_{k+1}^3)$ as starting point, the computation for the next segment of the bus trajectory is the same as that of the beginning segment.

On the other scenario when $\bar{t}_k^3 > t_k^4$, it results in case 3. Two points are to make for this case: (a) it is not certain at the time of computation that whether the bus will meet the queue first or the bus stop first; (b) the part of dwell time that extends into the duration of queue blockage shall not be counted as the delay to be minimized. The former point implies that it is necessary to compute the projected point (t_{k+2}^1, l_{k+2}^1) intersecting by the free flow trajectory as if no bus stop and the backward forming queue starting from $t_{j,k+1} + g_{j,k+1}$. Point (b) suggests $d_{j,k+2}$ be the duration between when the bus is ready to exit the bus stop to when the queue dissipates to the bus stop. Additionally, (b) further implies $d_{j,k+2}$ be negative if the bus is ready to exit after the queue has dissipated, as in case 4.

To summarize, a recursive procedure is developed to compute $\underline{BR}_{jk+1}$ for all four cases over several consecutive cycles starting from the time the bus is first detected at (t_k^0, l_k^0):

Step 1: Find the immediate past end time of phase *j*, ($t_{j,k-1} + g_{j,k-1}$), set it as zero and compute all future timings about phase j in reference to $t_{j,k-1} + g_{j,k-1}$. $\underline{BR}_{jk} = +\infty$.

Step 2: Compute critical time points for different cases

- If bus stop downstream of l_k^0 and no skipping set $l_k^3 = l^{\text{bus}}$ otherwise $l_k^3 = 0$
- Compute t_k^3 as if free flow for bus to bus stop, and (t_k^1, l_k^1), (t_k^2, l_k^2) as well
- If $t_k^1 \le t_k^2$, $t_k^1 \le t_k^3$, $l_k^1 > 0$ and $t_k^1 > t_k^0$ Then [// equivalently case 1]
    - $l_k^{\text{ready}} = l_k^1$ and $t_k^{\text{ready}} = t_k^1$,
    - Compute $\overline{t}_k^1$, and set $t_{k+1}^0 = \overline{t}_k^1$, and $l_{k+1}^0 = l_k^1$
- Else
    - If $l_k^3 = 0$
        - If $t_k^3 \le t_{jk} + g_{jk}$
            - Go to step 5 [//There is no queue delay from cycle *k* onward]
        - If $t_k^3 > t_{jk} + g_{jk}$

- Set $\underline{BR}_{jk} = +\infty$ [// no queue delay for current cycle]
- Go to step 4 [//but there will be delay for next cycle]

○ If $l_k^3 > 0$,

▪ compute $\bar{t}_k^3$ and t_k^4

▪ If $\bar{t}_k^3 < t_k^4$ Then [// equivalently case 2]

- Update $t_k^0 = \bar{t}_k^3$ and $l_k^0 = l_k^3$
- Go back to step 2 for current cycle

▪ If $\bar{t}_k^3 \geq t_k^4$ Then [// equivalently case 3 or 4]

- Set $l_k^{\text{ready}} = l_k^3$ and $t_k^{\text{ready}} = \bar{t}_k^3$
- Compute $\bar{t}_k^1 = l_k^{\text{ready}} / v_3 + t_{jk}$
- Set $t_{k+1}^0 = \max\{\bar{t}_k^1, \bar{t}_k^3\}$, and $l_{k+1}^0 = l_k^{\text{ready}}$

<u>Step 3</u>: Compute $\underline{BR}_{jk} = t_k^{\text{ready}} - l_k^{\text{ready}} / v_3$

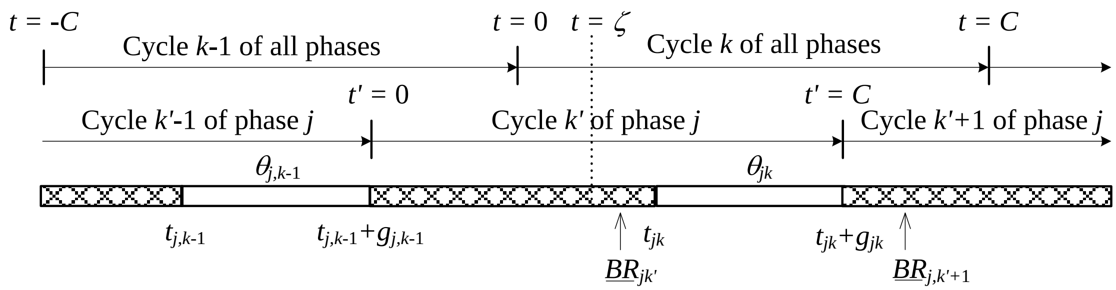
<u>Step 4</u>: Set $k = k+1$, if $k \leq K$, continue from step 1

<u>Step 5</u>: Terminate.

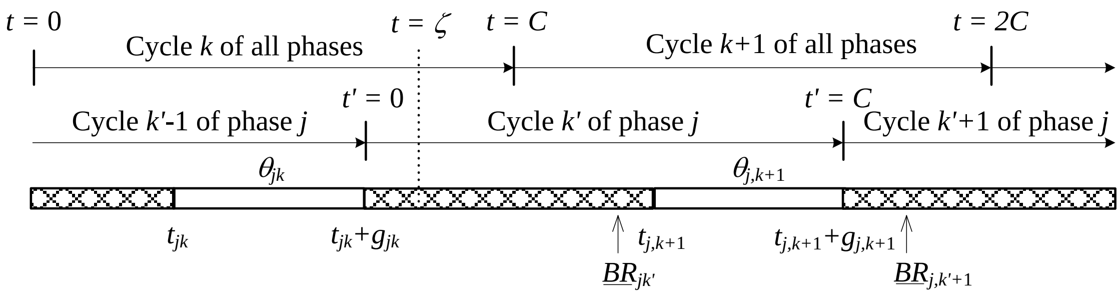
3.5.3. <u>Cycle Definitions</u>

In real-time implementation, an optimization session maybe triggered at any moment of a cycle. For some phases, queues are starting to build up and any incoming buses can only pass the intersection in the next cycle; however, for other phases, queues are dissipating, incoming buses may or may not pass the intersection in the current cycle. When calculating the bus delays, the algorithm needs to find the last end of green for the phase it is requesting (i.e. $t_{jk-1} + g_{jk-1}$). That time point may reside in this or the last cycle of all phases. Therefore, the indexing of cycle for priority delay (d_{jk}) may be inconsistent for different buses at different phases. Hence, it is essential to clearly define the cycle of phase *j* and how it is related to the cycle of all phases.

Figure 11 explains the how are the definitions of the cycles related to the optimization time in a cycle. Let ζ denote the cycle time of the current cycle *k*. The time zero for all phases (i.e. $t = 0$) shall refer to the beginning of the first phase in the cycle. If the optimization time occurs before the end of the green time of phase *j* at cycle *k*, the index of cycle for phase *j* (i.e. k') is the same as that for all phases (i.e., $k = k'$). This case is demonstrated in Figure 11-(a). However, if in the other case that the optimization time occurs after the end of the green time of phase *j* at cycle *k*, the indices of cycles are different. That is $k = k' - 1$, as shown in Figure 11-(b).



(a) Bus Detection before Phase j Green Ends in Cycle k of all phases



(b) Bus Detection after Phase j Green Ends in Cycle k of all phases

Figure 11: Cycles of All Phases versus Cycles of Phase *j*.

3.6. Accounting for Nonlinear Bus Trajectory

The objective for computing the five critical time-space pairs is to provide estimates of a set parameters $\underline{BR}_{jk}$ and one parameter $\overline{BR}_j$, from which the queue delay d_j of a bus can be computed. It is important to recognize that the definition of these points remain valid even if the linear assumption about the bus trajectory is relaxed, but the computation procedures for these points may not. Therefore, adjustments on the computation procedures help improve the estimation of the critical parameters to better represent realistic bus trajectories.

3.6.1. Computation of Nonlinear Bus Trajectory without Queue Delay

The parameter $\overline{BR}_j$ is most critical in the computation of queue delay, because it is the reference time when the bus actually needs the green time. Estimation of this parameter needs to be as accurate as possible. Fortunately, this parameter is defined by assuming no interactions of the bus with the queue of other vehicles. Hence, it can be easily computed using either free flow travel time if no bus stop is present or parabolic vehicle trajectory in and out of the bus stop. For the latter case, the exact locations before and after the bus stop where the bus starts to decelerate and accelerate at a constant rate can be easily determined. The bus trajectory is nonlinear in the area enclosed by these two locations and is linear outside of it. The standard rates 1.2 and 1.3 m/s are used for acceleration and deceleration respectively.

3.6.2. Adjustments for Bus Stop Entry Speed

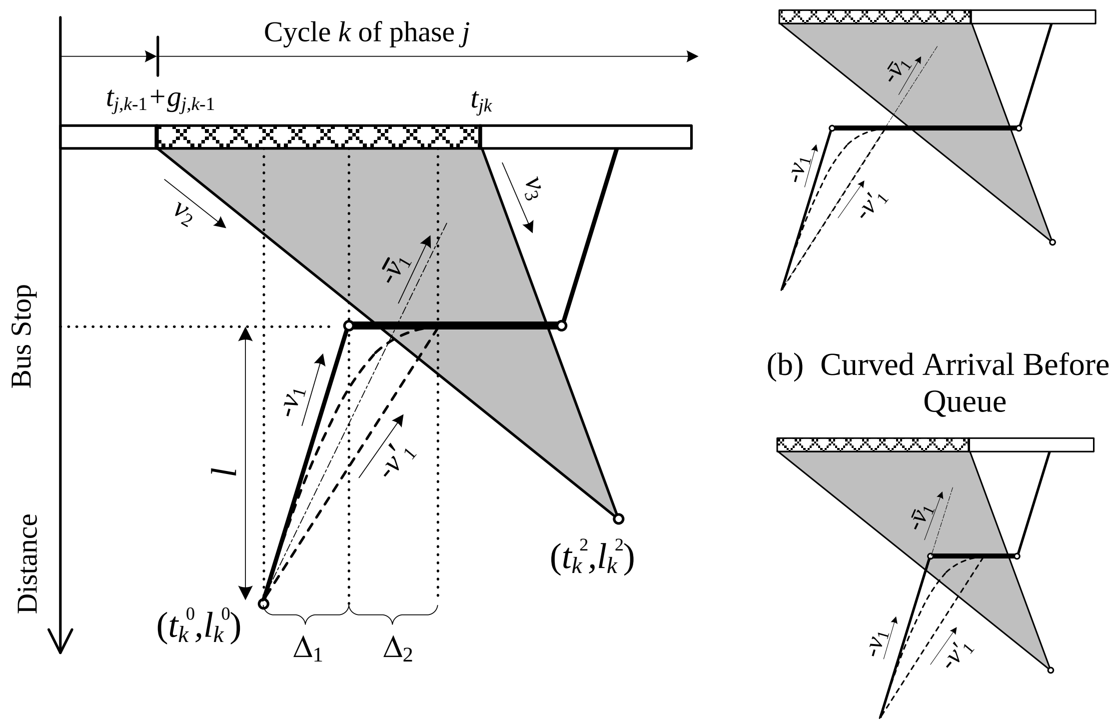
Figure 12-(a) illustrates the difference between linear and nonlinear trajectories. Let Δ_1 denote the time from the detection of the bus to when the bus arrives at the bus stop; let $\Delta_1 + \Delta_2$ denote the time the bus would actually need to apply constant deceleration rate to stop at the bus stop, assuming no queue blockage. It is possible that the bus is projected to meet bus stop first, which does not incur any queue delay according to the discussions before. On a real situation, the curved trajectory implies that the bus may actually meet the queue first, which would be delayed until the queue dissipates. Therefore, a fine-

tuning of bus arrival time is needed. To equate the two trajectories with the same distance *l*, and assumes constant *a*, we have:

$$v_1\Delta_1 = l = \frac{{v_1}^2}{2a} \quad => \quad \Delta_1 = v_1 / 2a$$
$$v_1 = a(\Delta_1 + \Delta_2) \quad => \quad \Delta_2 = v_1 / 2a \tag{3-30}$$

Therefore, v_1' is exactly half of v_1 (i.e. $v_1' = v_1 / 2$) . That means using a constant entry speed between 50-100% of the detected speed can give a good approximation to the nonlinear bus trajectory.

However, 50% range is still wide. To select a better percentage, we break it down into three cases based on whether the backward forming queue shockwave is projected to arrive at the bus stop before or after the linear and nonlinear bus trajectory. Figure 12-(a), (b) and (c) clearly illustrate the three cases: (a) queue shockwave arrives at bus stop between the arrival times projected by both trajectories; (b) queue shockwave arrives after the nonlinear trajectory; (c) queue shockwave arrives before the linear trajectory. 75% of the detected speed should be used as the entry speed for the bus for case (a), 50% should be used for case (b) (i.e. v_1'), and 100% be used for case (c) (i.e. v_1).



(a) Queue between Linear And Curved Arrivals

(b) Curved Arrival Before Queue

(c) Linear Arrival After Queue

Figure 12: Adjustment for Nonlinear Bus Trajectory

3.7. Summary

This Section documented a two-stage stochastic optimization model developed to minimize bus priority delay and signal timing deviations simultaneously. Brief discussion was given to justify the use of timing deviations to approximate TSP impacts. A queue delay algorithm to compute critical time points as inputs for the optimization to compute bus priority delay in according to signal timing changes was presented. Finally, a discussion was made on how to account for nonlinear bus trajectory.

4. TSP ON A COORDINATED SIGNAL CORRIDOR

Based on the model developed in the single intersection case previously, this section establishes a deterministic MILP model called route-based TSP (RTSP) for granting bus signal priority on a coordinated signalized arterial. A brief literature review on TSP studies on a signalized arterial is first presented. The RTSP model is then formulated with well-defined assumptions and notations. Components of the optimization model are explained in details. Then a heuristic path-projection algorithm is presented to reduce the number of decision variables in the RTSP model in run time.

4.1. Introduction

For single intersections, a variety of priority models have been proposed to minimize different forms of delays, such as the total delay of all detected buses [26, 57], the total vehicle delay [23, 60], and the total person delay of buses and passenger car [24, 44, 61]. For TSP on a signalized corridor, however, much fewer studies can be identified in the literature. There are some studies looked into the developments and evaluations of active TSP priority on a signalized corridor [10, 19, 35]. Some other research attempted integrating active TSP strategies into some of the popular network-based adaptive signal control systems, such as SPPORT [62, 63], RHODES [60], UTOPIA[64], SCOOT [14] , and SCATS [65, 66] and so forth. But only a few studies have developed their own adaptive TSP models. He, et al. [27] developed a platoon-based arterial signal control model under mixed-integer linear programing framework for multi-modal traffic, including transit buses.

Despite all these studies, the priority granting process is still mostly done at individual intersection level in an isolated fashion. However, it is argued that local prioritization may not positively contribute to system-wide performance, such as schedule and travel time [19]. This is due to the fact bus stops unavoidably pull buses out of signal progressions, and the priority received from an upstream intersection may be negated by the priority denied in a downstream intersection. Therefore, relying on providing TSP at individual intersections to facilitate quick passage of a transit bus through a corridor may very likely to result in one of the three outcomes: 1) not achieving targeted performance, 2) causing too much disruption on side-street traffic, 3) both.

It should be clearly pointed out that, transit vehicles are often expected differently from private vehicles. Bus delay is not always the subject of interest. Instead, bus service reliability is a critical metric for quality of the service from the customers' point of view [11]. For service reliability, there are generally two schools of approaches: 1) headway regularity, and 2) schedule regularity.

Headway regularity is important in high-frequency bus routes where on-time schedule arrival is not important due to very short headway [31]. Research [31, 61, 67, 68] has attempted to use TSP as a pull-and-push tool to regulate headways.

On the other hand, studies on schedule-based TSP models and algorithms remain very sparse. Ma, et al. [19] is one of the few studies attempted to find the best timing for minimizing schedule deviations over several intersections. The priority strategies devised in the study allow not only decreasing but also increase bus delay at a specific intersection to achieve final on time arrival at the last intersection. Ghanim and Abu-Lebdeh [36] included the schedule deviation formulation in a second-order objective function and applied GA algorithms to find optimal split, cycle length and offset for a signal arterial. Wadjas and Furth [69] developed a signal control algorithm that controls several intersections simultaneously to give signal priority for a light rail line, and the objective was to reduce crowding of on-board passenger and to improve its schedule regularity.

4.2. Route-Based TSP Formulation

For a route-based model, decisions on the timings at all intersections that a route would traverse need to be made at the same time so that a clear trajectory for the vehicle in question can be planned out. By doing so, increases of delay to other traffic are inevitable if the normal signal operation without priority is optimal for the prevailing traffic conditions. Delay formulations may be used to quantify those increases in order to find a balance between priorities and disruptions. However, a delay formulation is usually second order with cross-product terms. Such formulation may not be positive semi-definite quadratic functions that can be solved by any standard MINP solver, and optimal solution is either hard to obtain or takes a lot of time, which may not be suitable for online implementation.

As suggested in [20], timing deviations can be used to approximate the impacts of signal timing changes to non-transit vehicles. Extending this idea, the RTSP can be formulated in relatively straightforward fashion. Several assumptions are required:

- Current corridor timing is optimal for the prevailing traffic condition and is available as an input to the RTSP model;
- The scheduled arrival time at the next main station is available from communicating with a bus OBU;
- Bus dwell time along the corridor is not random;
- Bus interactions with standing queues are minimal before entering the bus stop.

For a modern arterial signal system, there are three fundamental parameters: cycle length, offset and split [40]. Regardless the coordinated signal system is either fixed-time or actuated, traffic engineers have to provide these three parameters in order for them to operate. More often than not, these parameters are optimized offline using a number of optimization software, such as SYNCHRO, PASSER series, TRANSYT series and many more [70]. Different optimization models may employ different objectives for their offline arterial performance optimization, such as delay-based minimization [71, 72] and bandwidth-based maximization[73]. The end result is usually a set of optimized timing plans for different times of the day (i.e., TOD plans). Hence, it is normally not difficult to obtain optimal corridor timing for use as input to the RTSP model.

As mentioned before, bus routes have main and minor stations. It is more important from both a customer and a transit agency standpoint that on-time arrival is observed at the main stations. In light of this principle, we introduce the concept of intersection group, as illustrated in Figure 13. One intersection group for a bus is enclosed by the two main stations along the route of this bus. Different buses may or may not have the same intersection group. An intersection group enclosed all the intersections whose signal timings are used in formulating a RTSP model to carry out a priority control. In addition, the schedule arrival time at the exiting station needs to be known. Since each bus has its own route information, the control system should be able to obtain the schedule time for the next main station from the bus requesting the priority via V2I communications. V2I messaging will be discussed more in detail in section 5.2.1.

The third assumption states that random variability is not allowed. This assumption in particular is needed to ensure the solvability of the MILP program. Although the MILP program is capable to allow random variability in theory, it is not expected to be real-time practical. As mentioned in the previous section, as the number of possible outcomes increases for dwell duration, the number of decision variables for the SMINP program will increase dramatically. With multiple dwell durations along the bus route, the combinations across dwell durations will quickly make the optimization model become too large to solve in real-time.

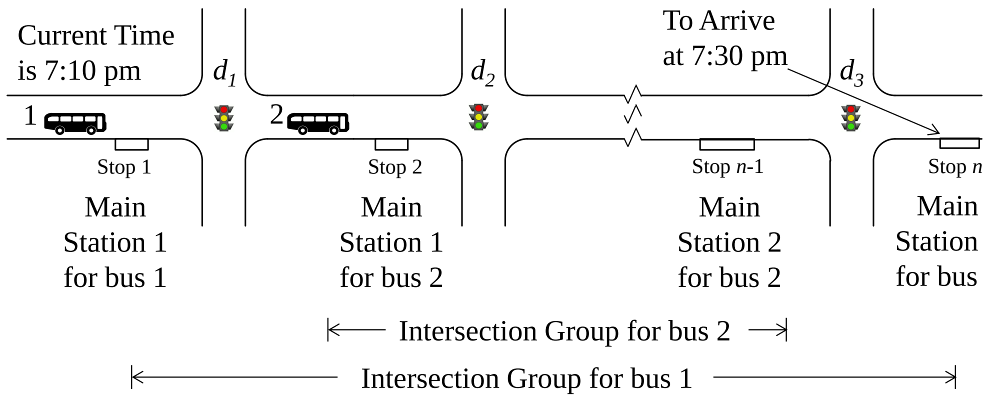


Figure 13: Illustration of The Concept of Intersection Group.

The last assumption is also a simplifying assumption. SMINP utilizes the queue delay calculation algorithm to determine the critical time points to compute the potential delays incurred by standing queues when the bus is blocked from entering the bus stop. Therefore, the potential delay is minimized. However, the queue delay computation cannot be carried out for the following intersection within the intersection group before the bus's delay at the first intersection is precisely known. Note that RTSP only applies the queue delay procedure to minimizing the bus delay at the first intersection, and not any of the other intersections due to queue delays are assumed to be minimal by this assumption.

4.2.1. <u>Notations</u>

Sets

I all intersections within one intersection group; if different buses have different intersection groups, it is then the union of all these intersection groups.

K_i cycles within the analysis horizon for intersection i

L all rings

N all buses under consideration

I_n the remaining intersections that bus n will pass through, $I_n \subseteq I$

I_n^c critical intersections to which bus n is assigned with an exit schedule, $I_n^c \subseteq I_n$

J_i all phases for intersection $i \in I$

J_n all phases that bus n will request, usually just one phase a bus will request

J_i^{yield} the coordinated phase to which the yield point is reference

J_i^{coord} the coordinated phases, $J_i^{yield} \subset J_i^{coord} \subset J_i$

$J_{il}^{barrier}$ the first phase after the signal barriers in ring l at intersection i

J_{il}^{first} the first phase in the phasing sequence in ring l at intersection i

J_{il}^{last} the last phase in the phasing sequence in ring l at intersection i

J_i^{cur} the phases that are currently in green signals at intersection i

J_i^{past} all past phases in current cycle at intersection i

Decision variables

x_{ijk} the amount of changes in the start time of coordinated phases

y_{ijk} the amount of changes in the green duration for all phases

z_{ijk} the amount of changes in the end time of the coordinated phases

τ_n the performance index of bus n

t_{ijk} the start time for phase index j of cycle k at intersection i, (local time)

g_{ijk} the green time for phase index j

v_{ijk} the split for phase index j

γ_{ik} the amount of time that exceeds the yield point of cycle k

r_{in} exit time of bus n requesting phase index j at intersection i, (local time)

$d_{in}^{\pm}$ the difference between bus arrival time and green start time at the cycle when the bus n is expected to pass through, a free variable,

d_{in} the priority delay of bus n requesting for phase index j at intersection i

β_{in} binary variable indicating if $d_{in}^{\pm}$ is negative, or equivalently if bus is not delayed

θ_{ijkn} the priority service decision for bus n at phase index j of cycle k at intersection i

Parameters

C common cycle length shared by all intersections

Y,R yellow time and red clearance time

YP_i yield point of a cycle at intersection i, (local time)

Δ_i offset at intersection i, reference style type 170 (beginning of yellow), for simplicity, the end of all red is actually used

Δ_i^o time difference between the start time of the current cycle and the selected system time zero.

Δ_i^I time difference between the start time of the first implementation cycle and the selected system time zero.

X_c user-defined critical degree of saturation

G_{ijk}^{opt} optimal green duration without priority for phase index j

G_{ijk}^{min} minimum green duration for phase index j

G_{ijk}^{curr} elapsed green times for phase index $j \in J_i^{curr}$ at current cycle

G_{ijk}^{past} elapsed green times for all phase index $j \in J_i^{past}$ at current cycle

N_{ij} number of lanes for phase index j at intersection i

V_{ij} traffic volume for phase index j at intersection i

S_{ij} the saturation flow rates for phase j at intersection i

P_{ijk} calculated platoon arrival time at coordinated phase index j at cycle k, (local time)

B_{in} detection time of bus n at intersection i, (local time)

R_{in} calculated bus arrival time to stop bar at the intersection i, $i = I_n\{1\}$, (local time)

R_{in}^{plan} planned exit time of bus n at the intersection i, (system time)

$D_{\{i-1,i\},n}$ dwell time for bus n at bus stop on link $\{i\text{-}1, i\}$

$T_{\{i-1,i\},n}^{Bus}$ travel time for bus n on link $\{i\text{-}1, i\}$

$T_{\{i-1,i\}}^{Veh}$ private vehicle travel time on link $\{i\text{-}1, i\}$, assuming all speeds are uniform

T_{ijk}^{opt} optimal start time without priority for phase index j

Q_i the estimated initial queue delay for buses at intersection i

4.2.2. Model Formulation

4.2.2.1. Basic Control Logic at Each Intersection

The first set of constraints is concerned about the behavior of the signal timing after optimization. Using the precedence relationship proposed by Head, et al. [57], the standard ring-barrier signal timing structure can be easily modeled in Equations (4-1) through (4-5). The limitation of the precedence structure is that phasing sequence cannot be easily changed without adding a large number of binary variables.

$$t_{ijk} = 0 \qquad \forall i, k = 1, \forall l, j = J_{il}^{first} \tag{4-1}$$

$$t_{ij'k} = t_{ijk} + v_{ijk} \qquad \begin{array}{l} \forall i, \forall k, \forall l, \; j \in J \setminus J_{il}^{last} \\ j' = J\{\text{orderof}(j) + 1\} \end{array} \tag{4-2}$$

$$t_{ij,k+1} = t_{ij'k} + v_{ij'k} \qquad \begin{array}{l} \forall i, k \in K_i \setminus |K_i|, \forall l, \\ j = J_{il}^{first}, \; j' = J_{il}^{last} \end{array} \tag{4-3}$$

$$t_{ijk} = t_{ij'k} \qquad \begin{array}{l} \forall i, \forall k, \forall l, \\ j \neq j' \in J_{il}^{barrier} \end{array} \tag{4-4}$$

$$t_{ijk} + v_{ijk} = kC \qquad \forall i, k = |K_i|, \forall l, j \in J_{il}^{last} \tag{4-5}$$

Constraint (4-1) and (4-5) define the start and end times of the planning horizon. The length of the planning horizon is fixed and equals to $|K_i|$ cycles. Constraint (4-2) and (4-3) define the precedence relationship among phases within the planning horizon. Equality (4-4) defines the barriers. The i^{th} set of these constraints defines the sequence and timing of the signal at the i^{th} intersection. For a coordinated corridor, a fixed reference point in a cycle needs to be enforced to guarantee synchronization among all intersections. However, a fixed point on every cycle will greatly limit the flexibility of changing timing in response to a priority requests. Inspired by the actuated-coordinated control scheme, a yield point is employed to allow the coordinated phase to have more flexible termination time in some cycles, as described in Equation 7 and 8. This is useful to enable green extension operations, typically needed for as a TSP strategy.

$$t_{ijk} + v_{ijk} \geq (k-1)C + YP_i \qquad \forall i, k \in K_i \setminus |K_i|, \; j \in J_i^{yield} \qquad (4\text{-}6)$$

$$t_{ijk} + v_{ijk} = (k-1)C + YP_i \qquad \forall i, k = |K_i|, j \in J_i^{yield} \qquad (4\text{-}7)$$

Equality (4-7) is applied to only the last cycle in the planning horizon. It ensures the signal system to recover back to synchronization after the cycle where priority is granted.

In addition to defining phase start times, green time duration for a phase is equally critical to guarantee proper signal control behavior.

$$v_{ijk} = g_{ijk} + Y + R \qquad \forall i, \forall j, \forall k \qquad (4\text{-}8)$$

$$g_{ijk} \geq G_{ijk}^{min} \qquad \forall i, \forall j, \forall k \qquad (4\text{-}9)$$

$$g_{ijk} \geq \frac{V_{ij}C}{N_{ij}S_{ij}X_C} \qquad \forall i, \forall j, \forall k \qquad (4\text{-}10)$$

$$t_{ijk}, g_{ijk}, v_{ijk} \geq 0 \qquad \forall i, \forall j, \forall k \qquad (4\text{-}11)$$

Constraint (4-8) is the relationship between green time and split of a phase. Inequality (4-10) computes a min green time that ensures a phase stays under-saturated after optimization. Temporary over-saturation may be allowed by modifying this constraint. Zeng, et al. [74] provided details on how this is achievable.

4.2.2.2. Definition of Deviations

As proven in [74], the deviation of the new timing from the original timing is a good measure of TSP impacts on the general traffic. We extend this concept onto this corridor case. In a synchronized arterial signal system, at least three types of deviations are formulated as follows:

$$y_{ijk} \geq G_{ijk}^{opt} - g_{ijk} \qquad \forall i, \forall k, j \in J_i \setminus J_i^{coord} \qquad (4\text{-}12)$$

$$x_{ijk} \geq t_{ijk} - T_{ijk}^{opt} \qquad \forall i, \forall k, j \in J_i^{coord} \qquad (4\text{-}13)$$

$$z_{ijk} \geq T_{ijk}^{opt} + G_{ijk}^{opt} - t_{ijk} - g_{ijk} \qquad \forall i, \forall k, j \in J_i^{coord} \qquad (4\text{-}14)$$

$$y_{ijk}, x_{ijk}, z_{ijk} \geq 0 \qquad \forall i, \forall j, \forall k \qquad (4\text{-}15)$$

The inequalities definitions allow the objective function to penalize only one sided signal timing changes. According to (4-12) - (4-15), y_{ijk} represents the compression of green durations, x_{ijk} is the delay for starting the coordinated phases, and z_{ijk} denotes the earliness for terminating the coordinated phases. They can also be easily formulated as deviations to the respective time points or lengths. Also, not all phases or intersections need to have the same definitions. Depending on how the progression bands binding to the coordinated phases at each intersection, different definitions of deviations may be applied. This provides a simple framework for a variety of model variations.

4.2.2.3. Bus Delay at Each Intersection

A bus could arrival at an intersection virtually in any cycle. The bus priority delay is the time it is stopped by the red signal. To compute its stop time, we need to first determine which cycle the bus is to be served based on the start and end time of the phase the bus is requesting. Given the bus arrival time, R_{1n}, at the first intersection, the selections of service cycle for all intersections are as follow:

$$r_{in} > t_{ij,k-1} + g_{ij,k-1} - (1-\theta_{ijkn})M \qquad i \in I_n, j \in J_n, k \in K_i \setminus \{1\}, \forall n \tag{4-16}$$

$$r_{in} \le t_{ijk} + g_{ijk} + (1-\theta_{ijkn})M \qquad i \in I_n, j \in J_n, \forall k, \forall n \tag{4-17}$$

$$r_{in} = R_{in} \qquad i = I_n\{1\}, \forall n \tag{4-18}$$

$$\sum_{k\in K} \theta_{ijkn} = 1 \qquad i \in I_n, j \in J_n, \forall n \tag{4-19}$$

$$\theta_{ijkn} = \{0,1\} \qquad i \in I_n,\ j \in J_n, \forall k, \forall n \qquad (4\text{-}20)$$

$$r_{in} \geq 0 \qquad i \in I_n, \forall n \qquad (4\text{-}21)$$

For a given cycle that is selected for serving the bus, the priority delay is a function of arrival time and service time, or simply $d_{in} = \max\{t_{ijk} - r_{in}, 0\}$ where $\theta_{ijk} = 1$. The following two inequalities effectively linearize the computation of d_{in}.

$$d_{in} \geq t_{ijk} - r_{in} - (1 - \theta_{ijkn})M \qquad i \in I_n,\ j \in J_n, \forall k, \forall n \qquad (4\text{-}22)$$

$$d_{in} \geq 0 \qquad i \in I_n, \forall n \qquad (4\text{-}23)$$

However, it should be noted that this constraint is effective and binding only when the d_{in} is present in the objective function. In other words, only when the objective is to minimize bus delays, the inequalities are representative of the true delay a bus will experience under the optimized timing. However, when the objective is to minimize other performance measures such as schedule deviations, then the delay inequality may allow alternate timing solutions that may not represent the true delay. An alternate formulation will be discussed later to address this issue.

4.2.2.4. Projecting Bus Trajectory

The critical part of the formulation for this route-base TSP strategy is the projected bus trajectory within the signalized corridor. The arrival time at any intersection is a direct result of the arrival times and signal delays at upstream intersections, travel times at upstream links, and dwell times at upstream bus stops. This relationship can be easily written out using one equality constraint:

$$(r_{in} + \Delta_i^o) = (r_{i-1,n} + \Delta_{i-1}^o) + d_{i-1,n} + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} \qquad i \in I_n \setminus I_n\{1\}, \forall n \tag{4-24}$$

Where Δ_i^o is the offset of the current optimization cycle start time. It is important to point out that the offset used to convert from local time to system time is not the same as the signal offset Δ_i used in arterial signal optimization. In this paper, we call this offset, Δ_i^o, as the *optimization offset*.

Figure 14 shows the difference between the signal offset and the optimization offset and how the local times are converted to system times. Due to signal offsets Δ_i, each intersection starts its cycle at different time instant. Some intersections have their current cycles started after the system time zero, resulting positive Δ_i^o, as in intersection 1 and 3; other intersections have their cycles started before system time zero, resulting negative Δ_i^o. Since t_{ijk} defined in this research always reference to the beginning of a cycle, we use Δ_i^o instead of Δ_i to convert local time to system time, as $T_{ijk} = t_{ijk} + \Delta_i^o$.

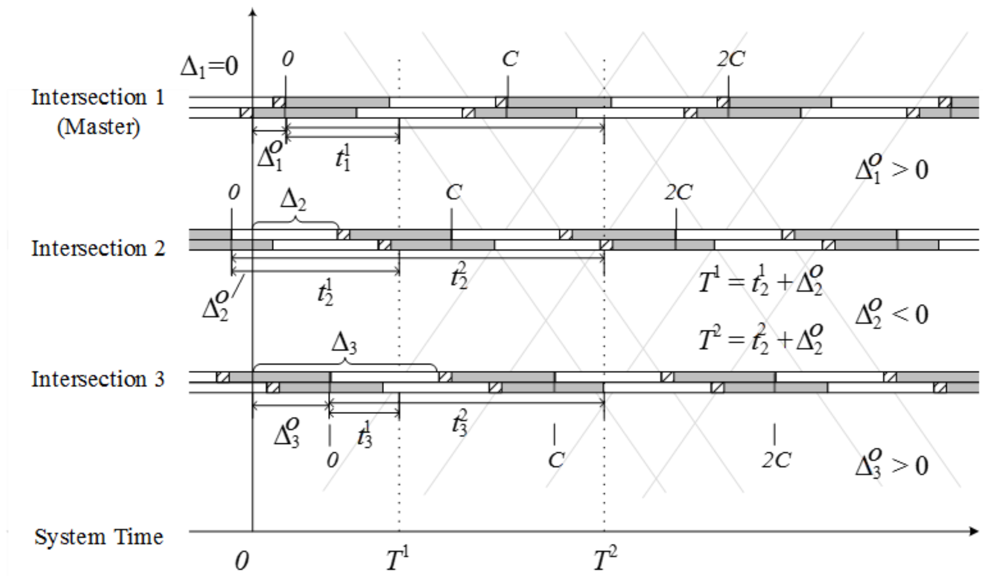


Figure 14: Conversion between Local and System Times with Optimization Offset.

4.2.2.5. Defining Corridor Performance Measure for Buses

The definition of corridor performance measure (i.e. the third term τ_n in the objective function) is an important factor that directly influence how the signal timings should respond to varying situations of bus arrivals. At least two performance measures can be considered: (a) corridor bus delay and (b) schedule lateness at reaching certain bus stations along the corridor. The corridor delay for all buses is the most straight forward definition. It is simply the summation of delay at individual intersections:

$$d_n \geq \sum_{i \in I} d_{in} \qquad \forall n \tag{4-25}$$

$$\tau_n = d_n \qquad \forall n \tag{4-26}$$

Minimizing corridor delay is an important objective to achieve for travelers already on the bus, since it minimizes their travel times. However, another equally, if not more, important objective is to adhere to a predetermined bus schedule at specified stations - schedule lateness. One advantage a route-base TSP model has is the ability to truly optimize a bus vehicle's schedule adherence. The majorities of TSP models are intended for single intersections. Many models [60, 75] incorporate the schedule adherence metrics, such as lateness or deviation, as weight coefficients of the bus priority delays. Buses that fall behind schedule get higher priorities when optimizing for their delays. Optimization models recognize them as static parameters but the objectives are typically not to minimize them.

By directly subtracting the scheduled arrival times from the actual arrival time, we can write the bus schedule lateness as the following:

$$\delta_{in} \geq r_{in} + d_{in} + \Delta_i^o - R_{in}^{plan} \qquad i = I_n^c, \forall n \tag{4-27}$$

$$\delta_{in} \geq 0 \qquad i = I_n^c, \forall n \tag{4-28}$$

$$\tau_n = \sum_{i \in I_n^c} \delta_{in} \qquad \forall n \tag{4-29}$$

where δ_{in} is the schedule lateness. I_n^c is usually a small subset of all intersections along the corridor because transit agencies generally do not need the buses to be exactly on time at every bus station, but only at some of the stations. In this research, we only assign one scheduled time per bus on its exiting time from the system.

4.2.2.6. Objective function

To be consistent with the overall goal of any TSP strategies, the objective function aims to find a balance between priority buses and regular traffic. Similar to the isolated intersection case, too much compression on any given phase causes excessive delay and should therefore be penalized. However, for a well-coordinated arterial system, the quality of signal progression is directly tied to the level of service for the corridor through traffic. Strategies that lead to late return to greens, for example, may be detrimental to system performance. Therefore, the proposed model incorporates a deviation term that approximates the degradation of signal coordination. The proposed objective function is formulated as Equation 1.

$$\text{Minimize:} \quad \sum_{i\in I}\sum_{l\in L}\sum_{j\in J}\sum_{k\in K} c_{ijk}\, y_{ijk} + \sum_{i\in I}\sum_{l\in L}\sum_{j\in Z_{il}^{2}}\sum_{k\in K} c'_{ijk}(x_{ijk} + z_{ijk}) + \sum_{n\in N} o_n \tau_n \tag{4-30}$$

Definitions of each of the three terms will be discussed in later section. In general, the first term, y_{ijk}, is used to limit the signal changes to the cross street phases. The second term, $x_{ijk}+z_{ijk}$, is intended to keep the coordinated phases in their original time windows. The last term, τ_n, is a performance index (e.g., corridor delay or schedule lateness) for bus *n* on the corridor. Detailed definitions for these terms are discussed later.

4.2.3. Exact Formulation for Schedule Lateness

In the formulation for schedule lateness, bus priority delays at individual intersections are not penalized in the objective function, only the schedule lateness is. The schedule lateness is computed as the sum of corridor travel time and corridor total signal delays. Therefore, delays at individual intersections can be of any combinations without affecting the value of lateness, δ_n, as long as the total signal delay stays the same. That means the program may results in an optimized timing based on an alternate phantom path that a bus will never travel. Figure 15 demonstrates the problem using only three possible delay solutions in each intersection that give rise to the same arrival time at the end of the trip. The real problem is that the arrival time of a bus at a particular intersection may influence its delay at that intersection. For example, path I arrives the earliest, so it is expected to be at the front of the queue; while path III arrives the latest, so it may be delayed further by the queue. So path I and III are the same as path II from the optimization standpoint, but they are not equal in real-world operations.

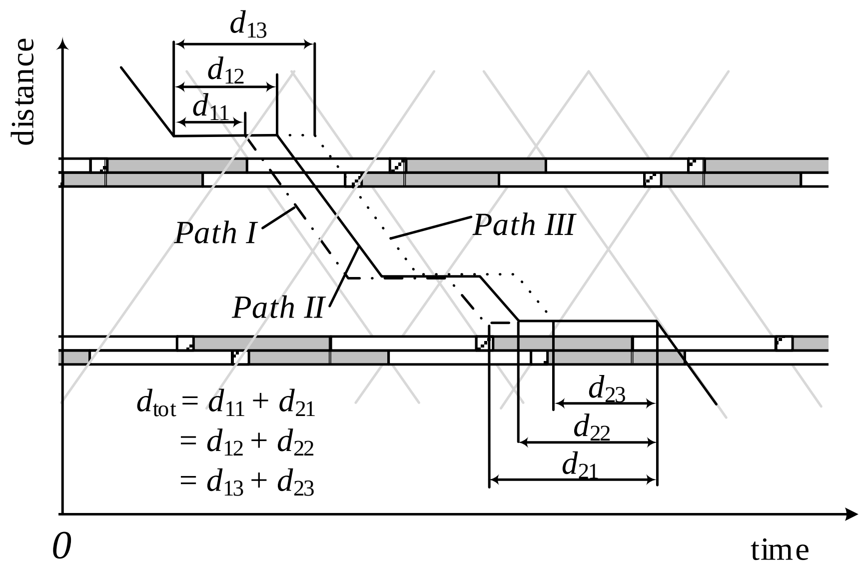


Figure 15: Alternate Solutions with Basic Formulation for Schedule Lateness.

This discrepancy between the projected path by the optimization and the actual path that will be travelled by the bus is an unnecessary error. This error will accumulate as the number of intersections grows also as the number of buses to be considered increases. As a result, a large number of alternate solutions may exist, yielding phantom bus trajectories erroneously projected by the optimization program. To eliminate potential projection errors, the calculation of bus delay at each intersection needs to be exact. That is, inequality (4-22) needs to be replaced by:

$$d_{in} = t_{ijk} - r_{in} - (1 - \theta_{ijkn})M \qquad i \in I_n,\ j \in J_n, \forall k, \forall n \qquad \text{(4-31)}$$

From this equality combined with inequality (4-16) and (4-17), one can derive that the value of d_{in} could take either positive or negative values. However, by definition, delays are non-negative, therefore a free variable must be introduced. Let $d_{in}^{\pm}$ be the deviation from the start of the green time on which cycle the bus will pass through, also we expand equality (4-31) into two inequality constraints with opposite senses:

$$d_{in}^{\pm} \geq t_{ijk} - r_{in} - (1-\theta_{ijkn})M \qquad i \in I_n, j \in J_n, \forall k, \forall n \quad (4\text{-}32)$$

$$d_{in}^{\pm} \leq t_{ijk} - r_{in} + (1-\theta_{ijkn})M \qquad i \in I_n, j \in J_n, \forall k, \forall n \quad (4\text{-}33)$$

A positive $d_{in}^{\pm}$ means that the bus arrives before the start of the green time, while a negative $d_{in}^{\pm}$ indicates the arrival is well into the green time. An indicator variable, β_{in}, can capture this difference and be used by other constraints to access the state of the free variable and perform appropriate calculations accordingly.

$$d_{in}^{\pm} \leq (1-\beta_{in})M \qquad i \in I_n, j \in J_n, \forall n \quad (4\text{-}34)$$

$$d_{in}^{\pm} \geq -\beta_{in}M \qquad i \in I_n, j \in J_n, \forall n \quad (4\text{-}35)$$

$$d_{in}^{\pm} \quad free \qquad i \in I_n, j \in J_n, \forall k, \forall n \quad (4\text{-}36)$$

With the indicator variable β_{in} , the arrival time relationship constraints need to be rewritten to either include the deviation (i.e. $d_{in}^{\pm}$) when it is positive or exclude it when it is negative. Similarly, expand equality (4-48) into the following four inequalities:

$$(r_{in} + \Delta_i^o) \geq (r_{i-1,n} + \Delta_{i-1}^o) + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} - (1 - \beta_{i-1,n})M \quad i \in I_n \setminus I_n\{1\}, \forall n \qquad (4\text{-}37)$$

$$(r_{in} + \Delta_i^o) \leq (r_{i-1,n} + \Delta_{i-1}^o) + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} + (1 - \beta_{i-1,n})M \quad i \in I_n \setminus I_n\{1\}, \forall n \qquad (4\text{-}38)$$

$$(r_{in} + \Delta_i^o) \geq (r_{i-1,n} + \Delta_{i-1}^o) + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} + d_{i-1,n}^{\pm} - \beta_{i-1,n}M \quad i \in I_n \setminus I_n\{1\}, \forall n \qquad (4\text{-}39)$$

$$(r_{in} + \Delta_i^o) \leq (r_{i-1,n} + \Delta_{i-1}^o) + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} + d_{i-1,n}^{\pm} + \beta_{i-1,n}M \quad i \in I_n \setminus I_n\{1\}, \forall n \qquad (4\text{-}40)$$

One can see that when $\beta_{in} = 0$, indicating bus arriving at red time (i.e. $d_{in}^{\pm} \geq 0$), the arrival time at the next intersection will be larger than the free-flow scenario and constraint (4-39) and (4-40) are in effect. When $\beta_{in} = 1$, bus arriving in green time (i.e. $d_{in}^{\pm} \leq 0$), so the bus will not experience any delay, constraint (4-37) and (4-38) will be in effect.

Finally, the schedule lateness should be written as follows:

$$\delta_n \geq r_{in} + d_{in}^{\pm} + \Delta_i^o - R_n^{plan} - \beta_{in}M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}41)$$

$$\delta_n \geq r_{in} + \Delta_i^o - R_n^{plan} - (1 - \beta_{in})M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}42)$$

In this way, the schedule lateness formulation guarantees a one-to-one correspondence between arterial signal timing and the bus trajectory expectedly projected by this timing. Nevertheless, there is a downside to this formulation. The dimension of the feasible region has been increased, making it slightly more computationally expensive to find the optimization solution.

4.2.4. Formulation for Schedule Deviations

The formulation for schedule deviation is similar to that for schedule lateness except that the lateness variable, δ_n, needs to be replaced by $\delta_n^+ + \delta_n^-$ in the objective function, and $\delta_n^+ - \delta_n^-$ in the constraints as the following:

$$\delta_n^+ - \delta_n^- \geq r_{in} + d_{in}^{\pm} + \Delta_i^o - R_n^{plan} - \beta_{in} M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}43)$$

$$\delta_n^+ - \delta_n^- \leq r_{in} + d_{in}^{\pm} + \Delta_i^o - R_n^{plan} + \beta_{in} M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}44)$$

$$\delta_n^+ - \delta_n^- \geq r_{in} + \Delta_i^o - R_n^{plan} - (1 - \beta_{in}) M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}45)$$

$$\delta_n^+ - \delta_n^- \leq r_{in} + \Delta_i^o - R_n^{plan} + (1 - \beta_{in}) M \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}46)$$

$$\delta_n^+, \delta_n^- \geq 0 \qquad i = I_n\{|I_n|\}, \forall n \qquad (4\text{-}47)$$

4.3. Path Projection Heuristic Algorithm

Different from the isolated intersection case, the corridor TSP formulation needs to optimize for multiple intersections at the same time. For a bus that may take *m* cycles to pass through a corridor of *n* intersections, the optimization program need to formulate a total of $m \times n$ cycles. If the number of phases at each cycle is *c*, then the total number of timing variables are $c \times m \times n$. Since approximately $m \approx n$, the total number of timing variables to be optimized is approximately cn^2. Therefore, the complexity of the MILP problem is $O(n^2)$, where *n* is the number of intersections to be included in the formulation.

However, a bus normally spends only 1 or 2 cycles per intersection. So many other cycles are mostly unaffected by the TSP priority request that takes place within the 1-2 cycle period. That means, if those irrelevant cycles can be eliminated from the optimization problem, the complexity of the MILP problem can be reduced. Therefore, a path projection heuristic algorithm is developed.

The objective of the algorithm is to find the relevant cycles that are likely to be affected by the TSP priority request along the corridor. In essence, the algorithm projects the bus path all the way downstream to the last intersection, and finds the lower-bound and the upper-bound for the path of the bus. Figure 16 illustrates the two paths representing the boundary conditions on a time-space diagram. The lower bound is used to determine when the start time of the cycle to be included is, while the upper-bound is to determine the length of the planning horizon. Therefore, instead of optimizing every single cycle within the time-space diagram, only the cycles affected needs to be considered. This reduces the complexity of the problem from $O(n^2)$ to $O(n)$.

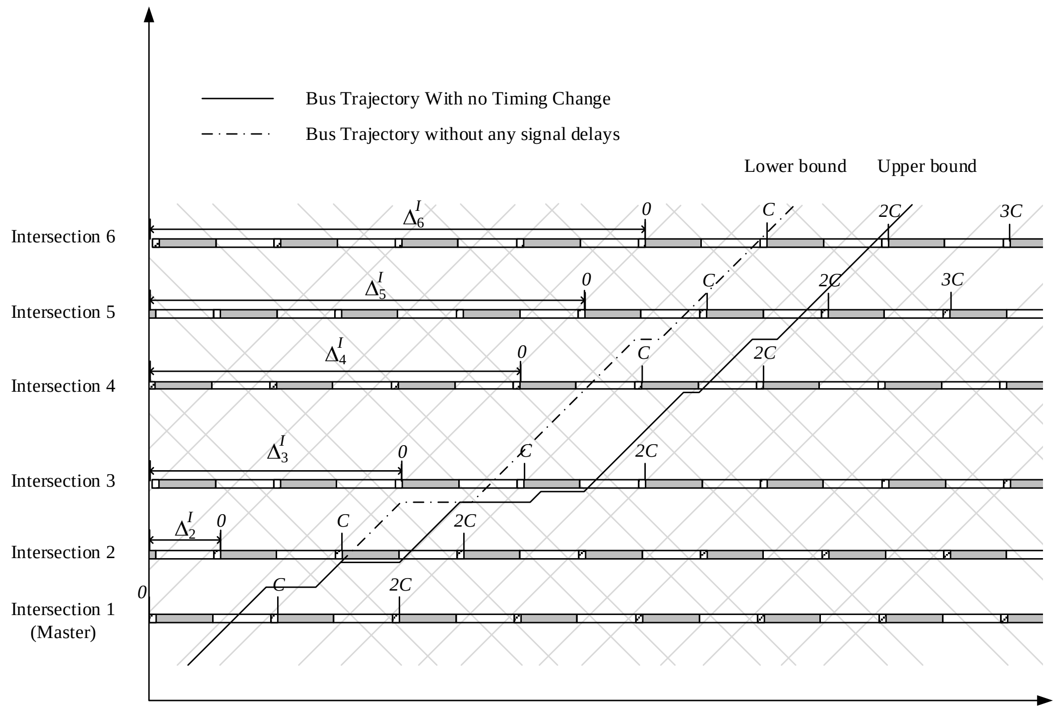


Figure 16: Example of Bounds Used to Reduce Decision Variables.

The lower bound is achieved when there are no delays at any intersection and the upper bound is simply the current path without any timing changes. The algorithm goes as follows:

Step 1: initialize the start system time and location (T_0, L_0) of the bus in question

Step 2: for intersection i, compute the arrival times, T_i^L and T_i^U, at intersection L_i

Step 2a: $T_i^L = T_{i-1} + T_{\{i-1,i\}} + D_{\{i-1,i\}}$, where $T_{\{i-1,i\}} = L_i / v$

Step 2b: $T_i^U = T_i^L + d_{i-1}$

Step 3: loop integer k_1 from system time 0, until $\Delta_i^o + k_1 C \geq T_i^L$ then $\Delta_i^I = \Delta_i^o + (k_1 - 1)C$

Step 4: loop integer k_2 from Δ_i^I until $\Delta_i^I + k_2 C \geq T_i^L$, then planning horizon is k_2

Step 5: repeat step 2 to 4 until all intersections are visited

It should be noted that an *implementation offset* Δ_i^I is introduced. The offset defines the start time of the cycles when the new timing for the i^{th} intersection to be implemented. The implementation offset replaces the optimization offset in constraint (4-24):

$$(r_{in} + \Delta_i^I) = (r_{i-1,n} + \Delta_{i-1}^I) + d_{i-1,n} + D_{\{i-1,i\},n} + T_{\{i-1,i\},n}^{Bus} \qquad i \in I_n \setminus I_n\{1\}, \forall n \qquad (4\text{-}48)$$

4.4. Summary

This section first presented a review on TSP literature with emphasis on TSP algorithms and models developed for signalized arterials. The RTSP model was then formulated with well-defined assumptions and notations. Components of the optimization model were explained in details. Finally, a heuristic path-projection algorithm was presented to reduce the number of decision variables in the RTSP model in order to integrate with any real-time control system.

5. SIMULATION EVALUATION PLATFORM

This section provides details on the development of the simulation platform which is used to evaluate the TSP models in an isolated intersection and a signalized corridor. The general architecture of the simulation platform is first outlined. Several main modules of the platform are described in more detailed. Some important features for real-time control capability of the TSP system are selectively presented.

5.1. Architecture Overview

A simulation platform is developed to implement the proposed models and to evaluate its performance against current state-of-the-practice signal system with active TSP strategies. The platform is coded and complied in the Microsoft Visual Studio C++. Figure 17 illustrates the general architecture of the simulation platform, which consists of the following main modules:

- Simulation: VISSIM traffic simulator for generating traffic and bus flows and a VAP signal controller using fix-time control logics for running arterial signal operations when no TSP implementations are in session.
- Optimization: the IBM CPLEX solver is used to execute a well-formed optimization model for TSP implementation.
- Signal Control: the main module developed in this research, which hosts a corridor controller/manager and a set of intersection controllers. This is also the interface between the simulation and the optimization modules.

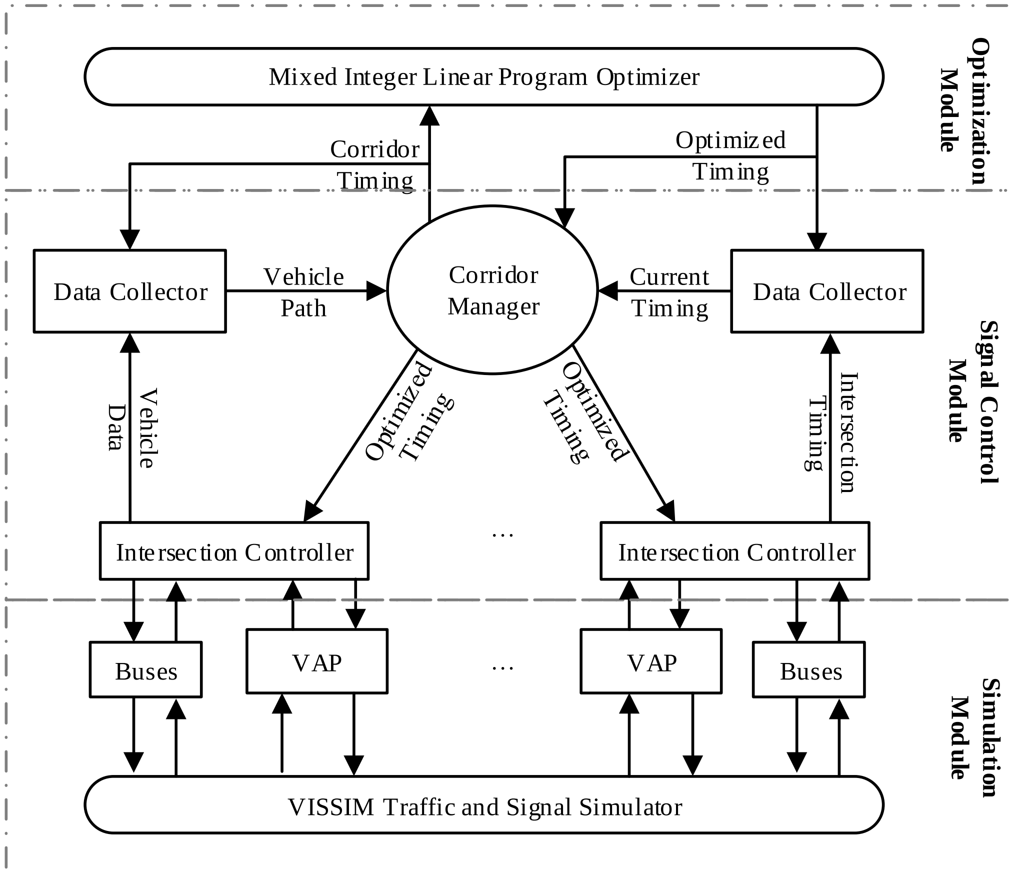


Figure 17: General Architecture of the Simulation Evaluation Testbed.

5.2. Simulation Module

The simulation module mainly functions as a vehicle generator and performance monitor, both of which can be easily setup in the standard VISSIM 5.x package. In addition, to simulate the connected vehicle environment, the component object model (COM) is setup to implement simulated wireless communications.

5.2.1. Wireless Communications between OBU and RSU

Via the COM programming language, the wireless communications is modeled via two types of vehicles: (1) Buses with connected vehicle onboard unit (OBU), (2) a parked vehicle on the roadside simulating a connected vehicle-enabled roadside unit (RSU), and each intersection has one and only one RSU.

At a 2-second interval, OBUs on buses in the network collect data related to its current state and stores it as a snapshot data. Five snapshots are recorded at any given time, and the oldest snapshot will be deleted automatically when new snapshot is available. In this research, only the latest snapshot is used. Every time the OBU makes a snapshot, it also listens to the DSRC open communications channel, to check if a RSU is nearby. If there is one, the OBU initiates a two-way communications with the RSU.

Table 3: Requirement for Static Data Stored in Each Origin.

Origin	Name	Type	Description
OBU	VehID	String	Unique identifier for the OBU
	VehClass	Integer	Vehicle class, e.g. bus, tram, light rail
	Route	Map <String, String>	A mapping between the intersection ID and turn specification
	Dwell	Map <String, Array>	A mapping between the bus stop ID and dwell distribution data
RSU	RSUID	String	Unique identifier for the RSU
	IntersectionID	String	Unique identifier for the intersection that RSU belongs to
	Layout	MAP	Map data specifying the intersection layout, including location of near and far side bus stops

Table 3 contains the minimal data requirement for both the OBU and the RSU units. Figure 18 illustrates the data flow between an OBU and an RSU. An RSU broadcasts the geometry of the intersection in terms of way points in the standard connected vehicle MAP message [76]. Any OBU come into range of the broadcast will be able to pick up the data, and determine its relative location in the intersection. The MAP message also contain phase mapping to each lane or approach. The OBU uses the MAP data from the DSRC broadcast and the route data stored in its internal database to determine the phase number it is going to request. Afterwards, the OBU takes the latest snapshot data, formats it into a PVD message and sends it to the RSU. The RSU may have its own decision mechanism or it may relay the PVD message to the corridor manager for deciding a system-wide signal optimization. Finally, either an optimized timing or a confirmation for the receipt of the PVD message is sent back to the OBU. In the simulation testbed, all communications between an OBU and the RSU is assumed to occur instantly without any delays.

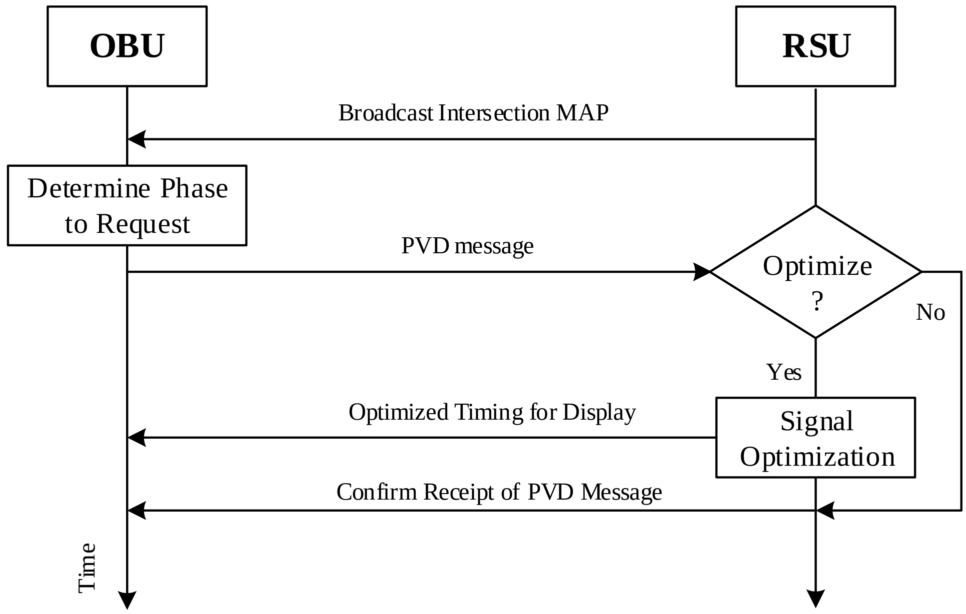


Figure 18: Data Flow Between an OBU-Equipped Bus and an RSU.

With an On-Board Unit (OBU), the transit bus theoretically can collect up to 42 short messages to form a PVD data that complied with the SAE-2735 standard [42]. Table 4 listed the main elements and their short descriptions in the PVD message transmitted from the OBU to the RSU. These data are essential for the real-time TSP system implementing the proposed models to function. In particular, the latitude and longitude of the vehicle current position can be replaced by the way point ids specified in the MAP data the OBU received earlier. In any cases, the OBU should collect its location data from its GPS or AVL system at least every 2 seconds. And the data along with other updated data should be sent to the RSU every 2 seconds for monitoring purposes after the initial communications. The GPS accuracy requirement is estimated to be 10 meters.

Table 4: List of Basic Elements in the PVD Message.

Name	Type	Description
VehID	String	Vehicle unique identifier
VehClass	Integer	Vehicle class, e.g. bus, tram, light rail
Passenger	Integer	Number of passengers currently onboard
CurSpd	Float	Current speed of the vehicle
Lat	Float	Latitude of the vehicle current position
Long	Float	Longitude of the vehicle current position
PhaseID	Integer	Phase number the vehicle is requesting
Skip	Boolean	Whether skipping the next bus stop
DwellDist	Array	Estimated dwell time distribution (outcomes)
DwellProb	Array	Estimated dwell time distribution (probabilities)
Schedule	Integer	Target time to reach a target station (will be relayed to corridor manager)

5.2.2. Normal Signal Operations in Simulation

A fixed-time signal control is implemented in VISSIM using the built-in vehicle actuation programming (VAP) language. The control runs on a standard eight phase two ring timing structure. When no optimization routine is performed, the VAP control runs as fixed-time control. The design of the TSP control system is independent to the design and logic used in the VAP control. In another words, VAP control can be replaced by another fixed-time signal controller. In fact, other types of signal control systems, such as actuated or adaptive, may be used as the main controller without TSP requests. This is because the TSP control system developed in this research implements two universal signal control command – force-off and hold, see section 5.3.2.

5.2.3. Calibration for Saturation Flow Rate

Saturation flow rate is one of the most important parameters in the simulation that affects the computation of queue delay, degree of saturation, objective function weighting factor and so on. This parameter is not always in agreement over various traffic simulation and/or optimization packages. In SYNCHRO, the saturation flow rate was determined as 1624 vehicle per hour per lane (vphpl). The default acceleration rate in VISSIM renders a higher saturation flow rate at about 1800 vphpl. Calibrating the vehicle acceleration rate alone is sufficient to adjust the saturation flow rate to a desired value (e.g. 1624 vphpl in this case).

5.3. Signal Control Module

The signal control module can be viewed as the additional logic on top of the normal (i.e., no TSP) signal control logics at each intersection. And this module is where the proposed TSP models and algorithms are implemented. This standalone design allows the module to work as a plug-in element to any standard controllers. The signal control module consists of two main controllers: 1) corridor controller, and 2) intersection controller.

5.3.1. Corridor Controller

For a set of coordinated intersections, there is only one corridor controller instance. The corridor controller is responsible of making system-wide optimizations, implementation, event logging and data archiving. Figure 19 lays out the architecture of a corridor controller.

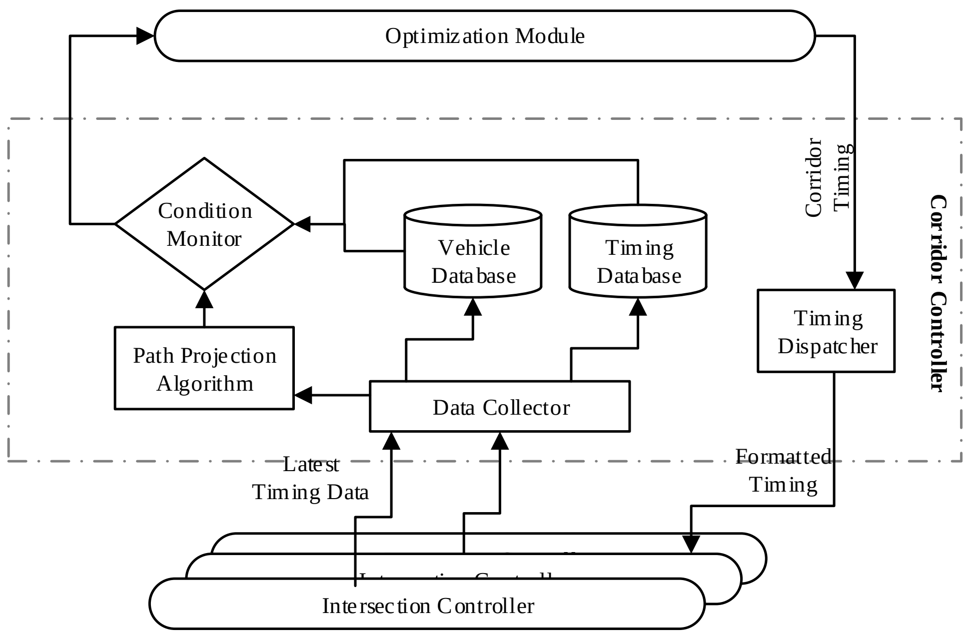


Figure 19: Architecture of an Instance of the Corridor Controller.

The controller first has a data collector, it collects vehicle data and new timing parameters from each of the intersection controllers as soon as they become available. The data are stored in its databases. Meanwhile, both the vehicle and the timing data are fed to a condition monitoring mechanism. Combined with the output from the path projection algorithm, the corridor controller determines the need for optimization. See section 4.3 for the path projection algorithm and section 5.5.1 for details on event-based optimization. Once a TSP optimization is finished, the optimized timings are returned from the optimization module to the timing dispatcher. The dispatcher will determine if implementation of the new timing should take place. For example, if it has taken too long for the optimization to complete and the timings are not long valid. Otherwise, the dispatcher sends the formatted timing to each intersection controller for implementation.

5.3.2. Intersection Controller

Each intersection has one intersection controller. Figure 20 shows the main components residing within an intersection controller. The controller is a functional module residing within the RSU of the intersection. The controller receives vehicle data from the RSU; it continuously records the actual timings from the signal controller, and store the old timings for up to one complete cycle in the past, which is a critical piece of input for the queue delay algorithm. The updates of both the vehicle and timing data are reported to the corridor controller periodically. In this research, one second update frequency was used. The signal command generator converts the new timing data into force off/hold commands.

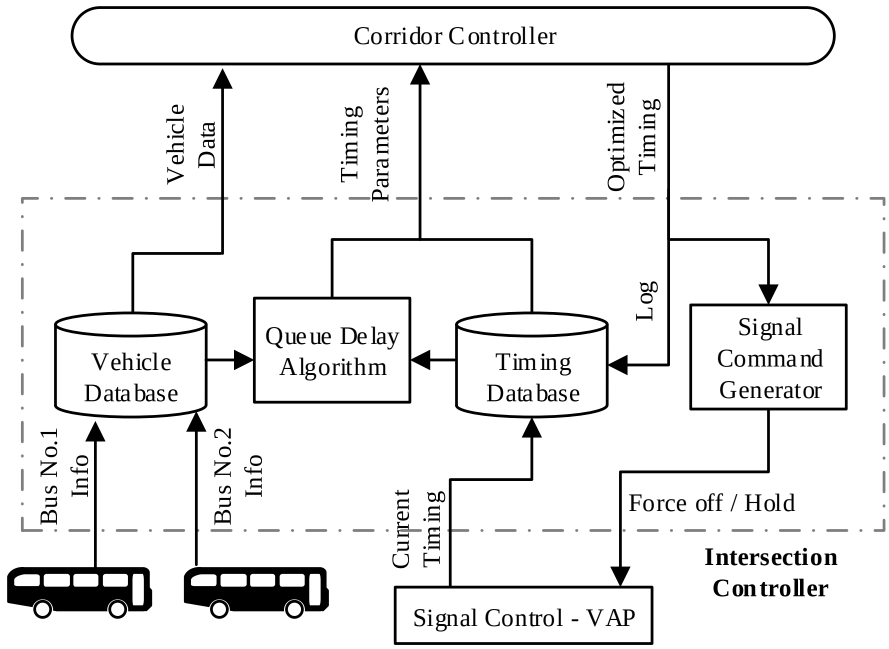


Figure 20: Architecture of An Instance of the Intersection Controller.

The intersection controller is in one of the two states at any given moment: 1) TSP control active, and 2) TSP control inactive. Figure 21 describes the control flows in these two states. The main control processes include:

- TSP Control checker – at every pass of the signal control process, a process checks if there exists an active command sent from the intersection controller. If there is, the TSP control is currently active, countdown timer will be continued. Otherwise, the TSP control is in active, proceed with normal signal operations.
- Countdown timer – Each phase in the planning horizon has one timer. For example, if there are 8 phases per cycle, and the planning period has two cycles,

there are 16 timers. This process is called every second to countdown the active timers.

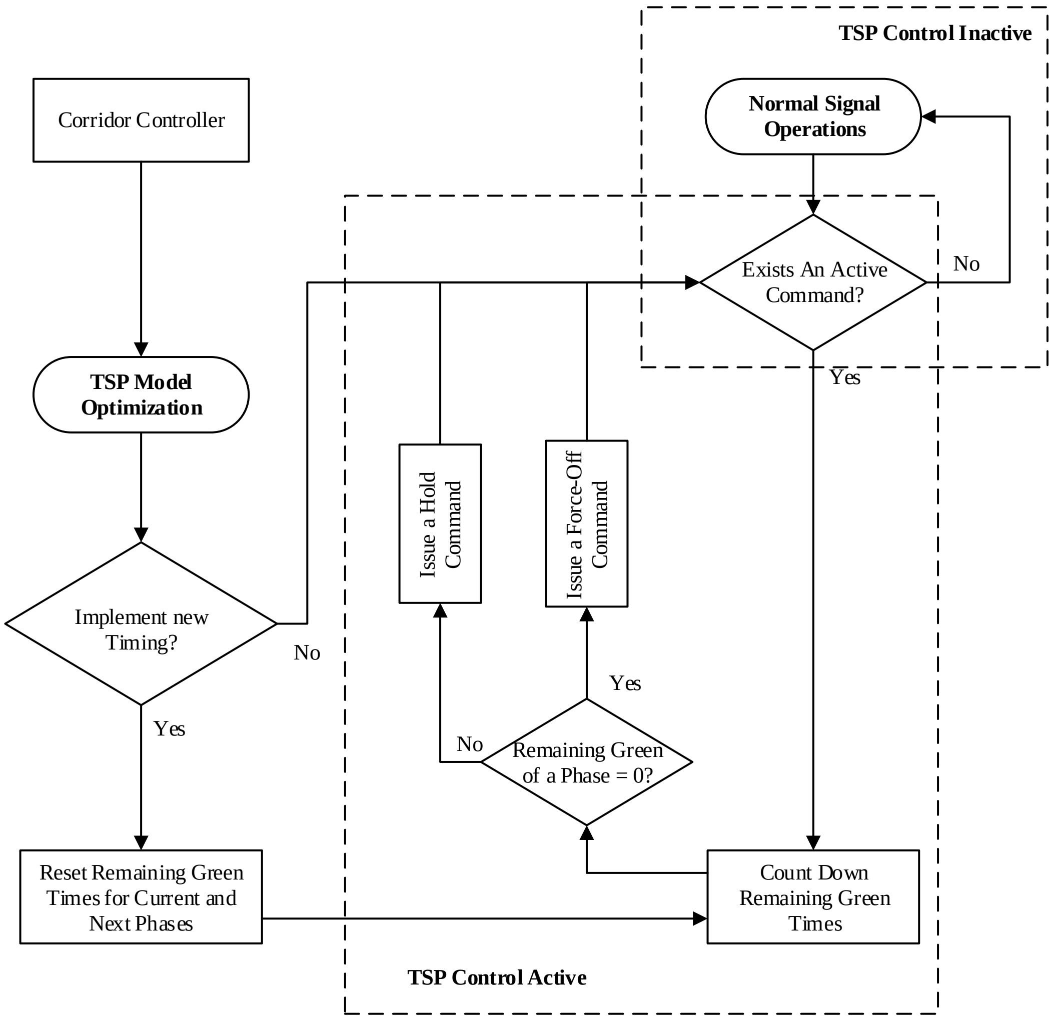


Figure 21: TSP Control Logic in the Intersection Controller.

- Set and reset timer – After an optimization routine, the new signal timings are implemented in the timers. So this process set or reset the timers accordingly.

- Issue control command – During the implementation of optimized timing, the signal control module take full control of the signal system. When the countdown timers of the active phases reach zeros, the force-off commands are issued. Otherwise, hold commands are placed every second until the force-off commands. The use of only two commands (i.e. force-off and hold) allows the system to be easily extended to any other types of signal controllers.

5.4. Optimization Module

The optimization is the core module where the TSP models are implemented. Upon receiving the bus data and the signal timing data from the corridor controller, this module decides what optimization model to use. Basically two main components in this module as described in Figure 22. The first component is a model formatter. This formatter takes all traffic, timing, and bus data from the corridor controller and formats them into CPLEX readable file. Then the second component, CPLEX solver, is directly called to find an optimal solution for the model. The CPLEX Callable Library was used to develop the optimization routines used in the second component.

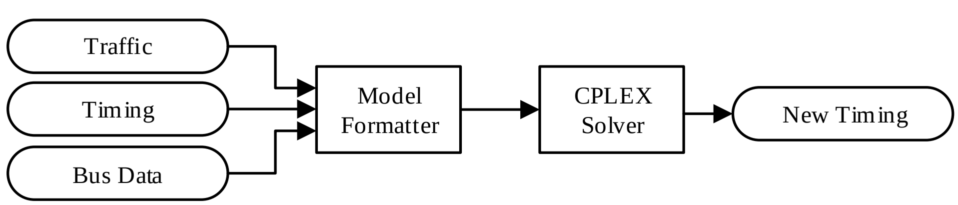


Figure 22: Flow Chart for the Optimization Module.

Since the RTSP model (see section 4) is deterministic, the formatting for this model is straight forward. But the SMINP model (see section 3) is a stochastic model, and CPLEX solver does not directly solve for a stochastic model. So the following routine is exercised:

Step 1: an initial SMINP model with only one stochastic scenario is formulated as a deterministic model;

Step 2: reformulates the SMINP model into its deterministic equivalent program (DEP), (see [55] for basic theory and examples);

> Step 2-a: add second stage constraints to the initial formulation in step 1 by enumerating all possible combinations of stochastic scenarios, one set of constraints for each combination;
>
> Step 2-b: modify the objective function by adding the second stage objective function with probabilities as weights;

Step 3: solve the DEP program as a deterministic MILP program

It is well recognized that any stochastic program is of large-scale in nature. This is because the number of combinations described in step 2 could grow exponentially. Using the bus dwell time as an example, if we discretize the dwell time of one bus into S distinct outcomes (assigning each a probability) and we have N such buses arriving at the same time, then the total number of scenarios is S^N. Each scenario corresponds to a set of second-stage constraints, m_2. That means the DEP will grow into a large program with $(m_1 + m_2 S^N)$ number of constraints, where constant m_1 is the number of first-stage constraints.

In this research, we attempt to limit the number of scenarios anywhere we can. And all the analyses conducted in this research had not generated a DEP program too large to be solved in reasonable time.

5.5. Real-Time Control Capability

Real-time capability is an important design factor for any adaptive TSP control system. A real-time control system shall be able to continue to operate regardless when and how many buses arrive at one or multiple intersections. Specifically, the system shall be able to conduct any number of optimization sessions, and to implement timing results at any point on a time horizon. The rolling optimization scheme has been employed in many adaptive traffic signal system, such as SPPORT [63] and OPAC [77]. Those real-time optimization schemes were fixed-interval, as shown in Figure 23-(a). Instead of fixed interval, another way to advance the planning horizon is when some events occur, as shown in Figure 23-(b).

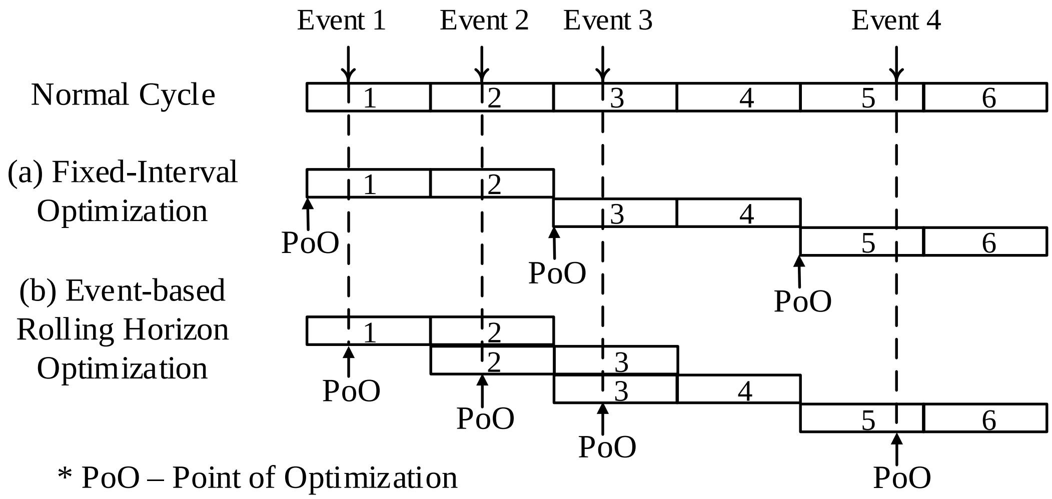


Figure 23: Types of Real-Time Optimization Schemes.

5.5.1. Event-Based Rolling Optimization Scheme

In this research, the event-based rolling-horizon optimization scheme is employed, and two events will trigger a new optimization session: 1) detection of a new bus and 2) an existing bus passes through an intersection.

The event-based scheme is more flexible than the fixed-interval rolling scheme. And it can limit the negative effect of dwell time variability by constantly monitoring the conditions and set off an event should any unexpected conditions take place. However, there are two main issues that have to be addressed in implementation:

- What formulation change needs to be made to allow optimizations to be conducted at any point over the time horizon instead of just at the beginning of a cycle?
- How to handle multiple bus arrivals either simultaneously or separately before the end of a planning horizon?

5.5.1.1. Formulation Addition

It turns out that it is quite simple to allow optimization to be conducted at any moment. The trick is by adding the following constraints to define the feasible region of the green times of each phase within the planning horizon:

$$g_{ijk} \geq G_{ijk}^{curr} \qquad \forall i, \forall k, j \in J_i^{cur} \tag{5-1}$$

$$g_{ijk} = G_{ijk}^{past} \qquad \forall i, \forall k, j \in J_i^{past} \tag{5-2}$$

G_{ijk}^{curr} and G_{ijk}^{past} denote the elapsed green time of current and past phases respectively within the current cycle. These two input parameters can be obtained by directly recording the actual timings from the beginning of the cycle to the point when the optimization occurs. These additional constraints are applicable to the RTSP formulation. But the same constraints can be written for the SMINP model by simply dropping the intersection index *i*.

5.5.1.2. Handling Multiple Buses

Any good real-time TSP control systems need to handle the arrival of multiple buses. But this is usually difficult when buses do not arrive at the same moments. This is because it is possible that the optimal timing is being implemented for the first bus may not be optimal at all for the second bus. To avoid the optimal timing for the first bus being overwritten by the arrival of the second bus, an active TSP strategy normally enforces a recovery period, during which no new TSP requests will be processed. This First-Come-First-Serve (FCFS) strategy renders very inefficient use the signal timing adjustment, which is the biggest disadvantage of the active TSP strategies.

In this research, since the mathematical formulation developed above allows the arrival inputs from multiple buses, there are two situations in handling multiple buses:

- Situation 1: When buses are detected simultaneously, one optimization session that uses all bus arrival times is needed.
- Situation 2: When buses are detected separately during the planning horizon, the optimizations are done multiple times. Each time, the optimization will include the new buses and update the trajectories of the existing buses.

One question is, however, how to define the background optimal timing if the second situation, the more likely situation, is encountered. The concept of deviation as we introduced earlier uses the background optimal timing as the reference point, and any deviation from that point is considered as impacts to other traffic. In situation 2, the second bus arrives within the period when the optimal timing for the first bus is being implemented. There are two ways to define the background optimal timing:

- Definition 1: the timing that yields minimal delay to the prevailing passenger traffic conditions.
- Definition 2: the timing that yields minimal delay to buses arrived earlier but sub-minimal delay to the passenger vehicles.

Using definition 1, the background timing remains strictly tied to the passenger vehicles only and it is the same regardless how many TSP optimization sessions have been conducted. Using definition 2, the background timing adapts to the fact the currently running timing considers both the prevailing traffic conditions and the priority needs of the existing buses.

From a glance, definition 2 seems to fit in better with the adaptive theme of the overall control philosophy. And for single intersection case, there is not much of difference between definition 1 and 2, since all the changes are made locally and do not have a wide-ranging effect. However, a revisit on the concept of deviation suggests otherwise. The concept of deviation uses the optimal background timing to approximate the optimal performance that can be achieved under the prevailing traffic conditions without bus. That implies, if the prevailing traffic condition is not changed, the timing to achieve the optimal performance for all the traffic is not changed. So, regardless how many times the timing is adjusted to fit priority needs of different buses, the optimal background timing remains unchanged as long as the prevailing traffic condition is not changed. In fact, the evaluation studies in section 7.2.5 will reveal that definition 1 is a better choice.

But the problem for using definition 1 as the background optimal timing is that it excludes the considerations of the buses which arrived previously but have not left yet. The connected vehicle technology counters this problem. Each bus OBU is constantly communicating with the RSU. So when the second bus is detected, the arrival information the first bus can be recaptured. The TSP models use the arrival information of both buses as if they are detected at the same time.

5.5.2. Variable Cycle Length in Rolling Optimization Scheme

When rolling optimization scheme is implemented for multiple buses, a practical issue emerges for allowing the cycle lengths to vary. When the start of a cycle is not fixed, it is possible that after a few optimizations, the optimized timing will completely fall out of sync with the background cycle timing. To ensure the synchronization of the ends of the optimized and the background cycles, it is important to make a book-keeping about the amount of offset between the expected and the actual start times of the optimization cycle. Figure 24 illustrates how the variable cycle length procedure can be implemented.

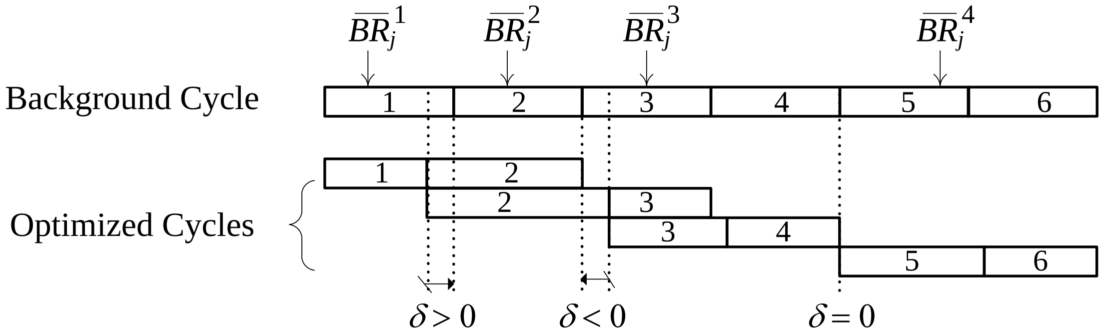


Figure 24: Variable Cycle Length Implementation in a Rolling Horizon Optimization Scheme.

Assume the optimization takes the timings of the next two cycles into consideration. When bus 1 arrives after phase j in cycle 1, the optimization program cuts the entire cycle 1 short to bring up phase *j* in cycle 2 earlier. Normally, the optimized timing will bring the timing back to normal in the third cycle. However, if a bus 2 arrives in the second cycle when the optimized timing is being implemented, the difference between the background start time of cycle 2 and the actual start time of cycle 2 shall be added to the cycle length constraint:

$$t_{jk} + v_{jk} = kC + \delta; \qquad j = J\{last\}, k = |K| \qquad (5\text{-}3)$$

δ is positive if actual start time of the cycle is earlier than the background start time of the cycle, and it is negative otherwise. This procedure can be applied to any number of look-ahead cycles. If multiple buses arrive in consecutive cycles, this procedure ensures the intersection timing returns back to normal synchronization after all look-ahead cycles are implemented.

5.5.3. Mitigating the Effect of Dwell Time Variability on a Corridor

One of the assumptions for developing the RTSP model is constant bus dwell time. Passive TSP systems generally work under this assumption as well [14]. But in any real-world situations, this assumption can hardly hold up. As also shown in preliminary studies, the variability of dwell time may sometimes have very detrimental effect on the bus as well as other traffic. The event-based rolling horizon optimization scheme in conjunction with the V2I communications works as an important updating mechanism to adjust the timing based on the realized dwell time.

The following lays out the steps to be taken in the updating system:

Step 1: The corridor manager uses the V2I technology to establish communications with the buses of interest and to continuously monitor their locations and speeds.

Step 2: Comparisons are made at a fixed interval to check if the current location of a bus falls on the projected path estimate that was used in optimizing for the current signal timings.

Step 3 – a: If yes, go back to step 2 and wait for the next check.

Step 3 – b: If no, a request for RTSP optimization is generated along with the necessary data, and a new RTSP model based on the latest bus information is solved to find a better timing that accommodate the bus path changes.

In this research, however, we did not use path deviation as the triggering event. Instead, the triggering event was every time a bus passes a signalized intersection. This is because we want to avoid subjectively defining how much deviation is considered to be significant enough to warrant a new optimization. Instead, this triggering event is inspired by the traditional TSP system with a check-out detector. Borrowing the terminology, the triggering event used in this research is that a bus checking-out of an intersection.

5.6. Summary

This section provided details on the development of the simulation testbed which is used to evaluate the TSP models in an isolated intersection and a signalized corridor. The simulation testbed consists of a simulation module, signal control module, and optimization module. The latter two modules form the basis for a signal control system with adaptive TSP models. A rolling optimization scheme was also developed to enhance the real-time control capability of the system. How to handle multiple buses were also explored

6. SIMULATION STUDIES FOR SINGLE INTERSECTION

This section* presents a series of simulation studies to test the SMINP model's performance in a single intersection setting. The test intersection and traffic conditions are first described. Analyses of model parameters are conducted; comparisons among three control models are then made. The final subsection investigates in more details what impacts stochastic bus dwell times have on the performance of the proposed model.

6.1. Simulation Test Setups

6.1.1. Test Intersection

A hypothetical four-leg intersection, as shown in Figure 25, was setup to evaluate the performance of the SMINP model. Three routes were setup for different testing scenarios. Route No. 1 enters from the eastbound approach, encountering a nearside bus stop at about 60 meters (196 feet) from the stop bar. Route No. 2 enters from the northbound approach and a bus stop locates roughly 80 meters (261 feet) from the stop bar. Route No.3 travels northbound and exits westwards without any bus stops. It is assumed that the intersection is equipped with wireless communication equipment that can detect the presence of the approaching bus and obtain relevant bus data, see Table 4 for details.

---

* Part of this section is reprinted with permission from "A Real-Time Transit Signal Priority Control Model Considering Stochastic Bus Arrival Time" by X. Zeng, Y. Zhang, K. Balke and K. Yin, 2014. IEEE Transactions on Intelligent Transportation Systems, volume 15(4), p1657-1666. Copyright 2014 IEEE.

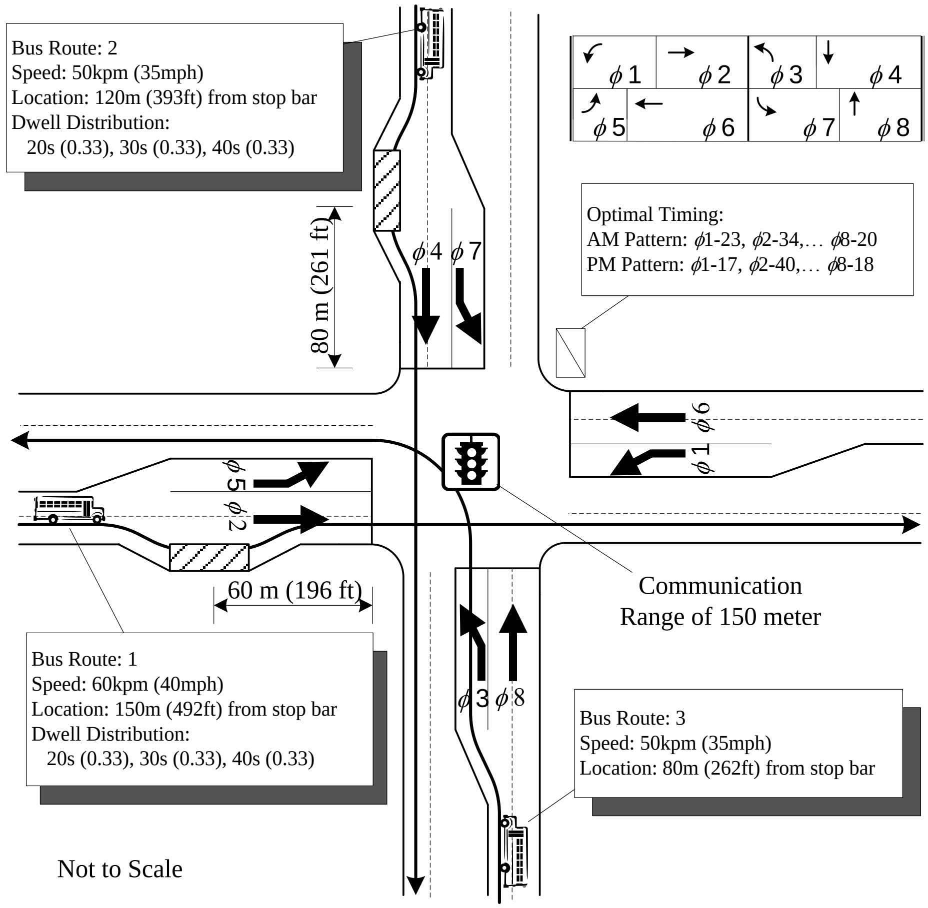


Figure 25: Hypothetical Intersection.

6.1.2. <u>Traffic Conditions</u>

Table 5 shows the setup of three congestion levels represented by the volume-to-capacity (V/C) ratios. The cycle length, 110 seconds, is optimal for the v/c = 0.9 level, and it is not optimal for other two levels. Using one common cycle length rather than one in each volume level is to test the ability of the control models to give priorities under various v/c levels. Without fixing the cycle length, any reasonable offline optimization models will produce a cycle length by attempting to achieve a v/c level at around 0.9. By fixing this cycle length, all splits are optimized in SYNCHRO [71] with the respective volume levels. The dwell time is assumed to be discretely uniformly distributed with possible outcomes of 20, 30 and 40 seconds for each bus line with a near-side bus stop.

Table 5: Parameter Setup for Simulation Evaluations.

Background Timing: Cycle Length = 110 sec Dwell Time Distribution: 20 sec (0.333), 30 sec (0.333), 40 sec (0.333)								
Phase	$\phi 1$	$\phi 2$	$\phi 3$	$\phi 4$	$\phi 5$	$\phi 6$	$\phi 7$	$\phi 8$
# of lanes	1	2	1	2	1	2	1	2
V/C = 0.5								
Volume (vph)	112	616	90	381	78	784	101	280
Opt. splits (s)	23	40	20	27	19	44	21	26
V/C = 0.7								
Volume (vph)	156	858	125	530	109	1092	140	390
Opt. splits (s)	22	44	17	27	16	50	19	25
V/C = 0.9								
Volume (vph)	200	1100	160	680	140	1400	180	500
Opt. splits (s)	21	46	15	28	14	53	17	26

6.1.3. Signal Control Models for TSP

Signal control models dictate how a transit priority is granted. Many state-of-the-art models are too complicated to be solved in real-time, have stringent requirements, such as high market penetration rates of specialized vehicles, or cannot handle stochastic arrivals. For all practical purposes, the SMINP model is designed to be a fast solvable and easily implementable model. Therefore, it is more suited to compare the SMINP against some state-of-the-practice methods of providing transit signal priority.

6.1.3.1. Fixed Time Control

The fixed time control model simply implements the offline optimized green splits for each of the v/c levels. Both passenger car and bus delays generated from this control model are considered as the benchmarks for all other models to be compared with. The fixed time control do not consider bus arrival rates or actual bus arrival time in any way. So it can be think of an optimized signal control without TSP. So for any other models with a TSP component, it is expected their passenger car delays will be higher than that observed in this control model, especially when the level of congestion is closer to saturation (e.g. v/c = 0.9).

6.1.3.2. RBC-TSP Control

The Ring Barrier Controller (RBC) in VISSIM [49] is a unified signal control emulator, which has implemented many of the most significant features of a real-world signal controller. Therefore, although it is not developed to exactly replicate the interface of a certain signal control model, its features are realistic enough to represent the existing functionalities of a typical modern signal controller.

The RBC uses a pair of check-in and check-out detectors to enable its TSP feature, namely RBC-TSP control model. Upon the detection of a bus at the check-in detector, a constant travel time with a constant slack time is applied to estimate its arrival time interval at the stop bar and performs either green extension or red truncation. With a nearside bus stop, the check-in detector is recommended to be placed at the exit of the bus stop [49] or to be coupled with a detector for door closing. This way, the need to account for the random dwell time is eliminated. In fact, there is no control parameter to allow a wider range of arrival time than a few seconds. So putting the check-in detector before the near-side bus stop will almost guarantee a failed TSP implementation. Therefore, in the comparison study, there is no choice but to place the check-in detector at the exit of the bus stop.

It is possible to enable the fully actuated control module within the RBC controller. But we are interested in the ability and limitation of the check-in-and-check-out TSP strategy furnished by the RBC controller, not its performance in actuated control. So as not to confusticate the comparison, we used the RBC controller in a fixed-time mode by setting max-recalls on all phases.

6.1.3.3. SMINP Control

The SMINP model is implemented in the simulation testbed as described in section 5 and the flow of signal control is completely illustrated in Figure 21. In the preliminary study, as in section 6.2, four variations of the SMINP model were developed but one is selected to compare with the fixed-time and RBC-TSP control models in section 6.3. In both cases, a uniform distribution was observed for bus dwell time. The optimization performed once when a bus come into the range of the intersection. But if there is more than one bus within the communications range, the most recently entered bus will be the one which initiate an optimization session that includes all buses. The new timing will be carried out until the end of the planning horizon or when another bus comes into range. The SMINP control in section 6.4 is exactly the same except the dwell time distribution is different.

6.2. Preliminary Analyses

6.2.1. Determining Weights for Deviations

The first-stage objective function controls the balance between the phase green time deviations and the bus delay. For each phase, the weight *c* determines the distribution of the deviations among all the phases. It is reasonable that a phase shall be panelized higher if it has already suffered from congestion than the one that is relatively less saturated. In this research, we compared four different ways to compute the weights.

Option 1: $c_{jk} = \dfrac{X_{jk}}{\sum_{j \in J} X_{jk}}$ (Weight -1)

Option 2: $c_{jk} = \dfrac{X_{jk}{}^{p}}{\sum_{j \in J} X_{jk}{}^{p}}$, set $p = |J|$ (Weight -2)

Option 3: $c_{jk} = \dfrac{1/(1 - X_{jk})}{\sum_{j \in J} \sum_{k \in K} 1/(1 - X_{jk})}$ (Weight-3)

Option 4: $c_j = \dfrac{1/(1 - X_j)}{\sum_{j \in J} 1/(1 - X_j)}$ where $X_j = \dfrac{V_j \sum_{k \in K} C_k}{S_j \sum_{k \in K} g_{jk}}$ (Weight -4)

First, one shall notice that each weight is normalized by the sum of all weights. The normalization ensures the weights only dictate the relative importance among different phases, not the relative importance between the total phase deviations and the bus delay. Therefore, the changes made here only affects the ways the program distribute the total deviations that are needed to reduce certain amount of the bus delay.

Option 1 and option 2 base the importance of each phase directly on the values of the degree of saturations. Specifically, option 2 makes the linear proportionality nonlinear in order to magnify the significance of higher X_{jk} values. And the polynomial order used in option 2 is set as equal to the number of phases in consideration.

Option 3 and option 4 base the importance of each phase on the reciprocal of the remaining under-saturation, which is defined as 1-X. One can easily see that option 3 is problematic if the degree of saturation is equal to or larger than 1. Option 4 rectifies the problem by computing the degree of saturation over the entire planning horizon. That means, if the underlying prevailing traffic condition is under-saturated, the optimization program will ensure under-saturation after the end of the planning horizon, and it does allow temporary oversaturation.

Option 4 requires some modifications of the original objective function,

$$\text{Minimize:} \qquad \sum_{j \in J} c_j y_j^2 + E[Q(\mathbf{t}, \mathbf{v}, BR)] \tag{5-4}$$

and it subjects to one additional constraint for each phase in a cycle as follow:

$$y_j = \sum_{k \in K} y_{jk} \tag{5-5}$$

Figure 26 shows a comparison of the performance of all four weight formulation options under the same network and traffic condition setups. In general, their ability to give priority to buses under various traffic conditions are very comparable. However, Weight-3 and Weight-4 seems to give the lowest impact on general traffic under low to medium degree of saturation levels, while Weight-1 and Weight-4 are less disruptive to general traffic on the high degree of saturation level. Weight-4 appears to be the most robust because it consistently performs above average to the best over all traffic conditions.

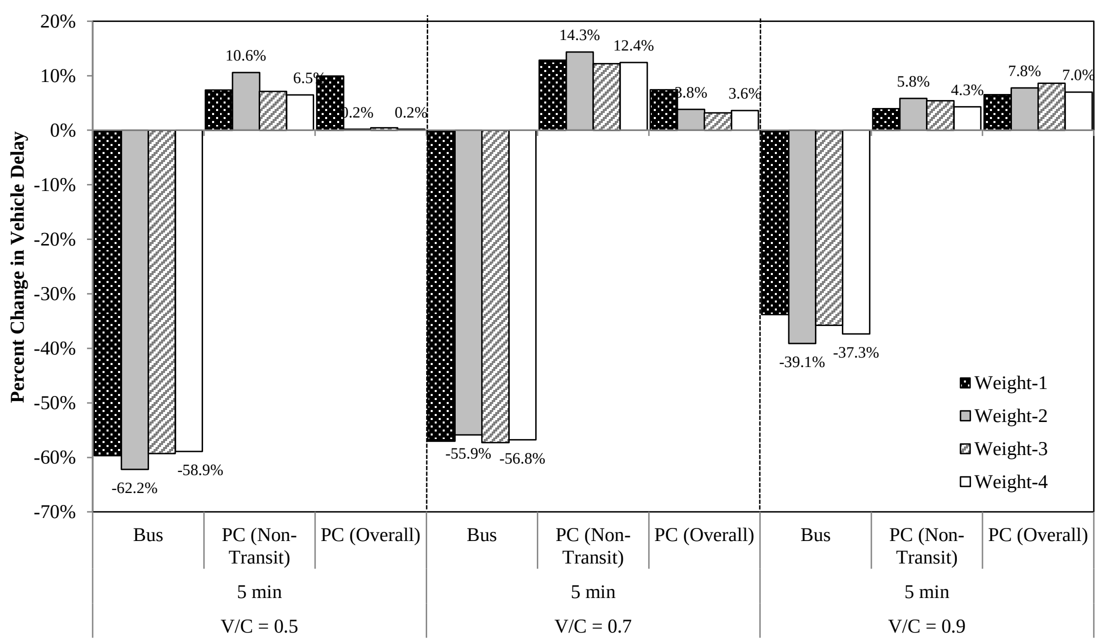


Figure 26: Comparisons of Weight Formulations.

It is also worth mentioning that a constant value can be applied to the weights of each phase. This constant value (e.g. the number of lanes) is tied to the significance of the phase. In this test run, we did not apply any constant value. Therefore, the importance of the phase is completely determined by the degree of saturation. That is, if two seconds of extensions are needed by the bus phase, the conflicting phases with the same degree of saturation on the same ring will shorten their green time by equal amount.

6.2.2. Level of Bus Priority

The weight of the priority delay, o_{jn}, is a crucial factor that allows the user to define the level of importance for a priority bus request. Different values of this weight (i.e. *priority coefficient*) may change the outcome of signal timing. For all three congestion levels, we varied the ratio of the priority coefficients $\sum_{j,k} o_{jk} / \sum_{j,k} c_{jk}$ from 0.1 to 10 at an increment of 1. Only bus route 1 is active, and the bus arrival headway is set to 5 minutes. Five random simulation seeds are used across all priority scenarios.

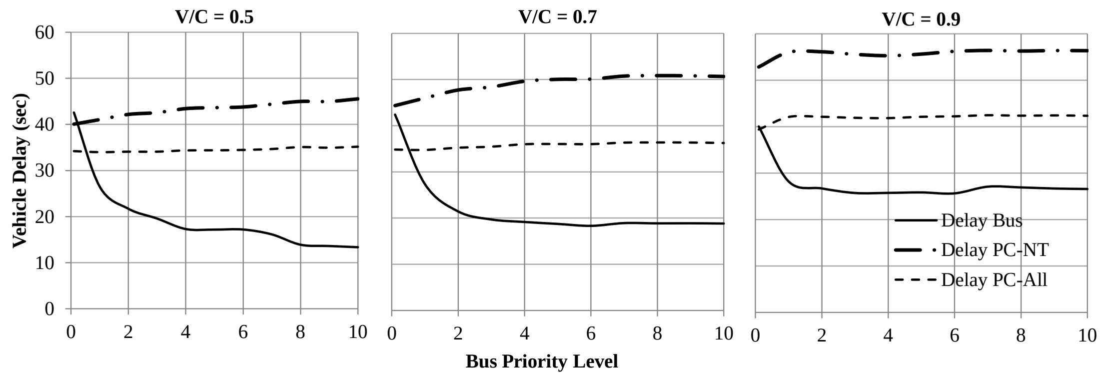


Figure 27: Comparisons of Weight Formulations.

Figure 27 illustrates the general trend of the bus delays and the passenger car delays with respect to different levels of priority. As expected, the increase of the priority weights for the bus decreases its delay and increases the delay for traffic on conflicting phases. The decrease of bus delay is particularly notable when V/C is at 0.5; the decrease is most significant from 0.1 – 1, then delay continues to decrease slowly as the priority coefficient increases. At this congestion level, the signal timing can be adjusted significantly to accommodate the bus priority without considerably impacting other traffic. However, as the intersection gets more congested, the impact of adjusting signal timing on the general traffic becomes more significant. This is manifested by the jump of non-transit PC delay from 52 – 56 when the level of priority increases from 0.1 ~ 1 at the highest V/C level. Another remarkable feature of the program is the ability to recognize the level of congestion automatically by restricting the amount of changes of the signal timing. For example, when V/C = 0.7, the decrease of bus delay levels out at about 4 or 5; when V/C = 0.9, the valley of bus delay comes much earlier at around 1. These indicate that the program will automatically cap out the maximum priority allowed to a bus priority request depending dynamically on the prevailing traffic conditions.

6.3. Comparison of Control Models

Using the Ring Barrier Controller feature in VISSIM [49], we compared the proposed model with traditional TSP operations in a standard traffic signal controller. The RBC uses a pair of check-in and check-out detectors to enable its TSP feature. Upon the detection of a bus at the check-in detector, a constant travel time with a constant slack time is applied to estimate its arrival time interval at the stop bar and performs either green extension or red truncation. With a nearside bus stop, the check-in detector is placed at the bus stop [49], to avoid accounting for the randomness in dwell time.

The RBC-TSP and the SMINP are compared with the baseline fix-time do-nothing control strategy. To compare these three control types on fair ground, fixed cycle splits are implemented in the RBC controller as well. Five random seeds are simulated for each of the volume and arrival frequency combination. A fixed priority coefficient (i.e. 5) for the SMINP was used for all cases.

6.3.1. <u>Simulation Evaluation for Single Bus Line</u>

Assuming only bus route No. 1 in Figure 25 has regular arrival at the intersection, we tested two arrival frequencies under all three congestion levels shown in Table 5. The bus headways for both frequency scenarios are larger than the planning horizon (i.e. two cycles of 110 seconds). That implies there will be no overlapping period between two consecutive optimization sessions. Therefore, the impacts of priority services are independent from one another.

Figure 28 illustrates the changes of vehicle delays comparing to the baseline fix-time control for each combination of volumes and arrival frequencies. It can be seen that both RBC-TSP and SMINP gives signal priority to the bus, resulting in much lower bus delay across all scenarios. The SMINP generally renders lower bus delay comparing to the RBC-TSP at all scenarios. In some scenario, the difference is as large as 30% improvement from the RBC-TSP, and 60 % improvement from the baseline do-nothing scenario. This means that the proposed model was able to better capture the bus arrival time and adjust the timing to favor the bus more. Another reason for the significant improvement is due to the ability of the proposed model to plan ahead. The optimization was done at the time of the bus was detected before the bus stop, while the RBC-TSP only performs adjustment of signal timings for the bus at the time it is leaving the bus stop. There are about 30-50 seconds more time for SMINP to adjust the timing. The benefits of this are that not only the bus delay has reduced significantly, the disturbance to other traffic are comparable or smaller.

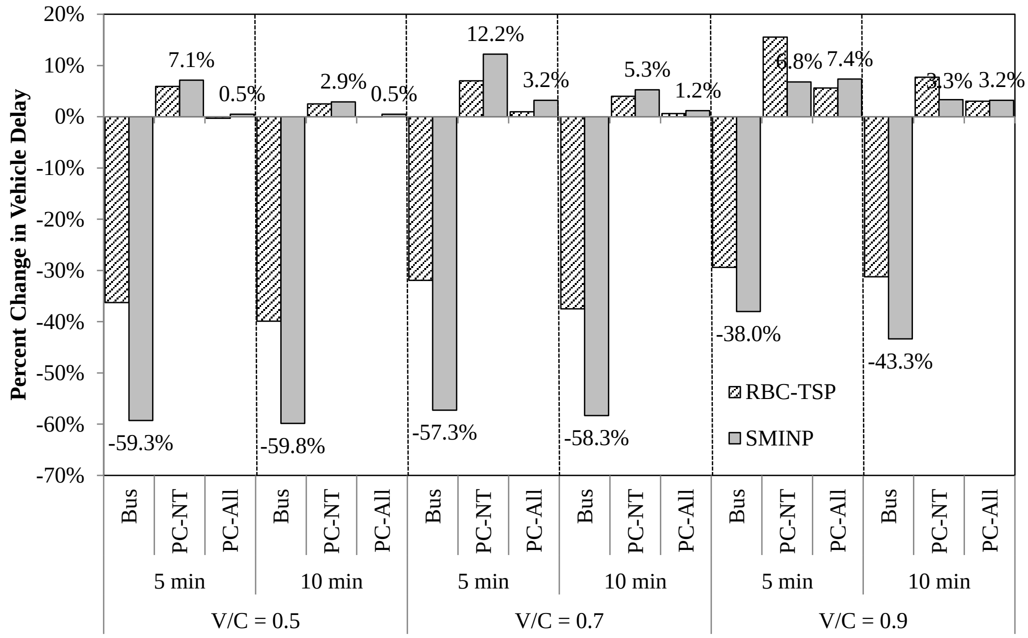


Figure 28: Percent Change in Vehicle Delays for RBC and SMINP under Single Bus Arrival Scenario.

On the other hand, the SMINP is much more responsive to the expected traffic conditions than the RBC-TSP. This is especially evident at high volume conditions (i.e. V/C = 0.9). At this volume level, when the bus arriving less frequently, the delay of traffic on non-transit phases are about 8% better than the RBC-TSP. When a bus arrives at about 5 minute interval, the delay to non-transit vehicles have increased to about 18% more than the baseline fix time control, while the SMINP maintains only about 7% increase from baseline. The ability to be responsive to the traffic condition is due to the mathematical model uses the normalized degree of saturation for each phase to spread-out the total number of seconds across all phases in the planning horizon to satisfy the bus priority needs. In this way, the start time of the phases may change significantly but the duration of the phase tends to be kept at their optimal values. The result is a much improved bus delay with a much less cost to the traffic on its conflicting phases. The delay values of the all compared scenarios are shown in Table 6.

6.3.2. Simulation Evaluation for Multiple Bus Line

Assuming more than one bus route runs through the intersection regularly, we varied the number of conflicting bus routes (i.e. two and three) under all three degree of saturation levels as in Table 5. The headways for bus route No. 1, 2 and 3 as in Figure 25 are set to 5, 6 and 8 minutes respectively. Consequently, in any one scenario, the timing optimization process for one priority service is inevitably affected by the timing changes for another priority service request. Therefore, the impacts of priority services are dependent from one another. In these complicated cases, the rolling optimization scheme has to be deployed to ensure the priority signal control can be performed continuously.

Table 6: Vehicle Delays by Control Types For Single Bus Scenario.

Intersection Degree of Saturation	Arrival Frequency	Vehicle Delay Type	Control Model		
			Fixed	RBC - TSP	SMINP
V/C = 0.5	5 min	Bus	40.3	25.7	16.4
		PC (Overall)	34.2	34.1	34.3
		PC (Non-Transit)	40.1	42.4	42.9
		PC (Transit)	30.3	28.6	28.7
	10 min	Bus	42.9	25.8	17.2
		PC (Overall)	34.1	34.1	34.3
		PC (Non-Transit)	40.1	41.1	41.2
		PC (Transit)	30.2	29.5	29.7
V/C = 0.7	5 min	Bus	39.6	27.0	16.9
		PC (Overall)	34.9	35.3	36.1
		PC (Non-Transit)	43.6	46.6	48.9
		PC (Transit)	29.2	27.8	27.6
	10 min	Bus	42.6	26.6	17.8
		PC (Overall)	34.9	35.1	35.3
		PC (Non-Transit)	43.5	45.2	45.8
		PC (Transit)	29.2	28.5	28.4
V/C = 0.9	5 min	Bus	42.3	29.8	26.2
		PC (Overall)	39.2	41.4	42.0
		PC (Non-Transit)	51.3	59.3	54.8
		PC (Transit)	31.2	29.7	33.7
	10 min	Bus	44.3	30.5	25.1
		PC (Overall)	39.0	40.2	40.2
		PC (Non-Transit)	51.3	55.3	53.0
		PC (Transit)	30.9	30.3	31.9

Note: PC (Overall) – All passenger cars on all approaches
PC (Non-Transit) – Passenger cars on phases conflicting with the bus requested phase
PC (Transit) – Passenger cars on phases concurrent with the bus requested phase

In particular, the SMINP model in this experiment used the rolling optimization method. The priority level for each route is set to 5, 3 and 2 respectively such that Route 1 has the highest priority and Route 3 has the lowest because it is a cross-street left-turn phase.

Route 1 and 2 have to come to a stop at their respective bus stops before arriving at the stop bar while route 3 do not need to stop at any bus stops. The dwell time for both route 1 and 2 follow the same discrete uniform distribution with equiprobable outcomes of 20, 30 and 40 seconds. A rule was applied in the system to prevent the rolling optimization from continuing indefinitely. The rule ignores the all the priority requests after the dynamic planning horizon being extended to 10 cycles or more. After the timing recover back to the background optimal timing at the end of the 11 cycle, new priority requests will be considered.

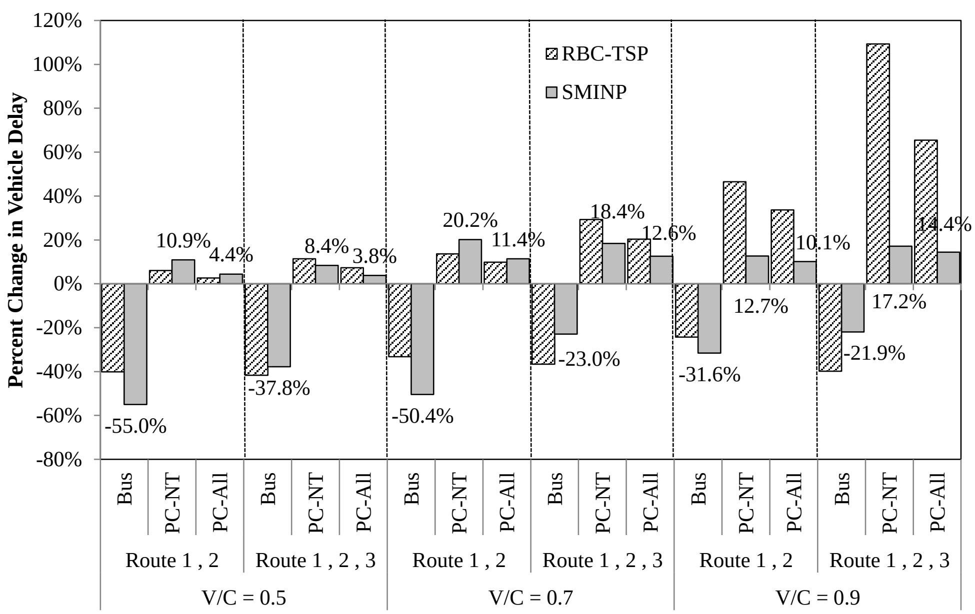


Figure 29: Percent Change in Vehicle Delays for RBC and SMINP under Multiple Bus Arrival Scenarios.

Figure 29 illustrates the changes in vehicle delays in terms of percentage when comparing the RBC-TSP and SMINP controls with the fixed-time control, and Table 7shows the absolute delay values. From the figure, several interesting observations can be drawn immediately. First of all, the RBC-TSP is slight better than SMINP in terms of non-transit phase delay and overall PC delay in low to medium degree of saturation levels when only route 1 and 2 are running. In all the other cases, the RBC-TSP under-perform the SMINP. Especially when V/C = 0.9, the RBC-TSP has failed to maintain the impacts of the priority service to an acceptable level, yielding 50 ~ 110% increase in terms of overall PC delay and 40 ~ 70 % increase in terms of non-transit phase delay. This is because, in high V/C cases, the RBC-TSP has no mechanism to capture the intensity of traffic therefore to dynamically underplay the importance bus priority requests in real-time. It is possible, in an offline setting, to fine-tune some of the RBC-TSP settings [49] such as the priority min green, recovery min green and so on. But even by doing this, numerous settings need to be refined in order to adjust the RBC-TSP setting in response to the changing traffic conditions. On the contrary, the SMINP can intelligently recognize the degree of saturation for each phase, and automatically finds the balance between the general traffic and the buses in real-time for multiple bus routes.

Table 7: Vehicle Delays by Control Types For Multiple Bus Scenario.

Intersection Degree of Saturation	Running Bus Routes	Vehicle Delay Type	Control Model		
			Fixed	RBC - TSP	SMINP
V/C = 0.5	Route 1, 2	Bus	47.3	28.3	21.3
		PC (Overall)	34.5	36.6	38.3
		PC (Non-Transit)	34.3	35.2	35.8
		PC (Transit)	33.9	33.2	32.3
	Route 1, 2, 3	Bus	46.8	27.3	29.1
		PC (Overall)	33.8	37.7	36.6
		PC (Non-Transit)	34.3	36.8	35.6
		PC (Transit)	34.8	35.8	34.3
V/C = 0.7	Route 1, 2	Bus	46.4	31.0	23.0
		PC (Overall)	35.4	40.3	42.6
		PC (Non-Transit)	35.0	38.4	39.0
		PC (Transit)	34.3	35.8	33.9
	Route 1, 2, 3	Bus	48.0	30.4	37.0
		PC (Overall)	34.3	44.4	40.7
		PC (Non-Transit)	35.0	42.2	39.4
		PC (Transit)	35.8	39.4	37.9
V/C = 0.9	Route 1, 2	Bus	48.9	37.0	33.4
		PC (Overall)	40.8	59.8	46.0
		PC (Non-Transit)	39.2	52.4	43.2
		PC (Transit)	37.0	41.9	39.3
	Route 1, 2, 3	Bus	64.3	38.7	50.2
		PC (Overall)	38.1	79.7	44.6
		PC (Non-Transit)	40.1	66.3	45.8
		PC (Transit)	42.6	49.8	47.4

Note: PC (Overall) – All passenger cars on all approaches

PC (Non-Transit) – Passenger cars on phases conflicting with the bus requested phase

PC (Transit) – Passenger cars on phases concurrent with the bus requested phase

6.4. Impacts of Stochastic Bus Dwell Time on Control Models

In any TSP signal control strategies, one key issue is to predict bus arrival time at the stop bar accurately [78, 79]. That is the time that a bus is ready to move on. If the green indication for the phase to which the bus is requesting is not turned on, the bus is delayed. Unlike passenger vehicles who are always ready to move on, transit buses are ready only after finishing loading and unloading passengers. This dwell time at a bus station is always a random factor, which makes it difficult to predict a bus arrival time at the stop bar precisely[80]. Without knowing the precise time a bus needs a priority, it is difficult for any signal control system which is deterministic in nature to provide effective signal priority. Previous simulation [38] and field [14] studies clearly showed that the dwell time variability can significantly reduce the TSP efficacy by causing buses to arrive later or earlier than predicted. Current practice often ignores the duration when bus pauses at the bus stop. For example, bus detectors are recommended to be placed at the exit of a bus stop in VISSIM user manual [81]. When the bus exits, however, there may be very little time for the signal system to react to the bus approaching. This problem is worsen if the bus stop is near to the intersection, because there will not be sufficient time for timing adjustment [82].

Since bus dwell time is a key in determining the bus travel time, many studies have taken a closer look into the bus dwell time distribution. Bertini and El-Geneidy [83] used TriMet data to study the bus travel time and found dwell time distribution is lognormally distributed statistically according to the recorded data in 459 stops. Li, et al.

[84] gathered dwell time information in rapid bus stations and found that lognormal function fits the collected data reasonably. As concluded by the study [85], Wakeby is the best distribution in terms of fitting dwell time data, and lognormal distribution still fits the data reasonably well.

6.4.1. Comparison Setup

In previous analyses in section 6.3, it was shown that the SMINP model is better than the check-in-and-check-out approach for giving priority even if the stochastic distribution was simply a uniform distribution with only 3 possible outcomes. Therefore, the comparisons to be performed here focus on a vary degree of knowledge about the underlying dwell time distribution.

First, the lognormal distribution is selected as the underlying dwell time distribution. As shown in Figure 30, the location (μ) and scale (σ) parameters for the distribution are 2.5 and 0.5 as suggested in [84]. Using this distribution, the resulting dwell time ranges from 2 seconds to 43 seconds, with a mean at 14 seconds. Note that, the SMINP model is not limited to any particular distributions. The log-normal distribution is chosen merely due to the consensus found in the literature. The complete distribution is setup on the bus stop on the eastbound approach for bus route No.1 (see Figure 25).

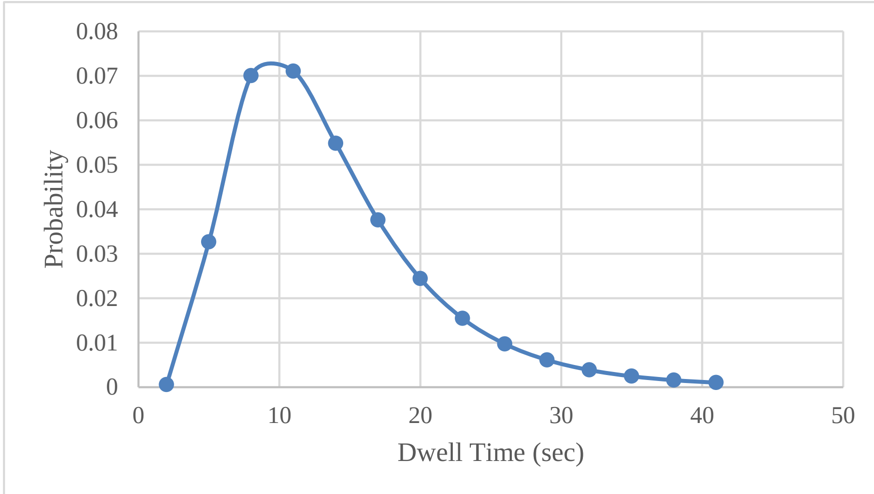


Figure 30: Log-Normal Distribution Used in Stochastic Evaluation.

Second, we vary the amount of distributional information provided the SMINP model to simulate varying degree of knowledge about the underlying distribution. Accordingly, three levels of availability for bus dwell time distribution are evaluated: A) complete knowledge, B) partial knowledge, and C) single point estimate. Case A represents a case where a lot of historical data were collected to form a complete picture of the underlying distribution. This is the best case scenario. Case C is the case where the engineers do not have enough data to build a complete distribution but only some data to give a best guess on the average dwell time. This is the worst case scenario that can still make use the SMINP model. Note that in this case the SMINP model is effectively reduced to a deterministic MINP model. This case actually represents most other model-based TSP strategies, which do not utilize any stochastic dwell time inputs. Case B is an in-between case where not only an average dwell time but also a reasonable range can be established. The distributional inputs are discretized as shown in Table 8 in order be used in the SMINP model.

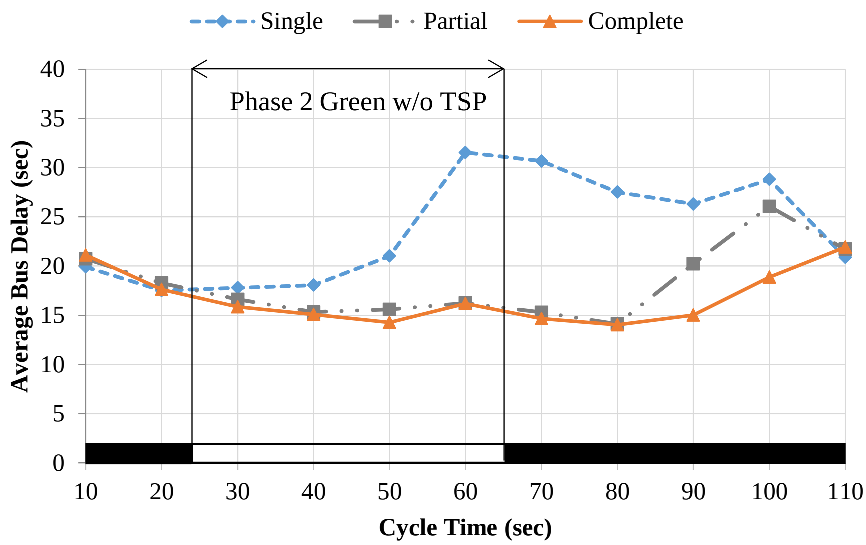
Table 8: Dwell Time Probability Inputs for Each Case.

<table>
<tr><td rowspan="2">A</td><td>Dwell (s)</td><td>2</td><td>5</td><td>8</td><td>11</td><td>14</td><td>17</td><td>20</td><td>23</td><td>26</td><td>29</td><td>32</td><td>35</td><td>38</td><td>41</td></tr>
<tr><td>Probability ($*10^{-3}$)</td><td>3</td><td>76</td><td>194</td><td>216</td><td>173</td><td>121</td><td>79</td><td>51</td><td>32</td><td>20</td><td>13</td><td>8</td><td>5</td><td>3</td></tr>
<tr><td rowspan="2">B</td><td>Dwell (s)</td><td colspan="5">5</td><td colspan="5">15</td><td colspan="4">30</td></tr>
<tr><td>Probability</td><td colspan="5">0.3</td><td colspan="5">0.6</td><td colspan="4">0.1</td></tr>
<tr><td rowspan="2">C</td><td>Dwell (s)</td><td colspan="14">14</td></tr>
<tr><td>Probability</td><td colspan="14">1</td></tr>
</table>

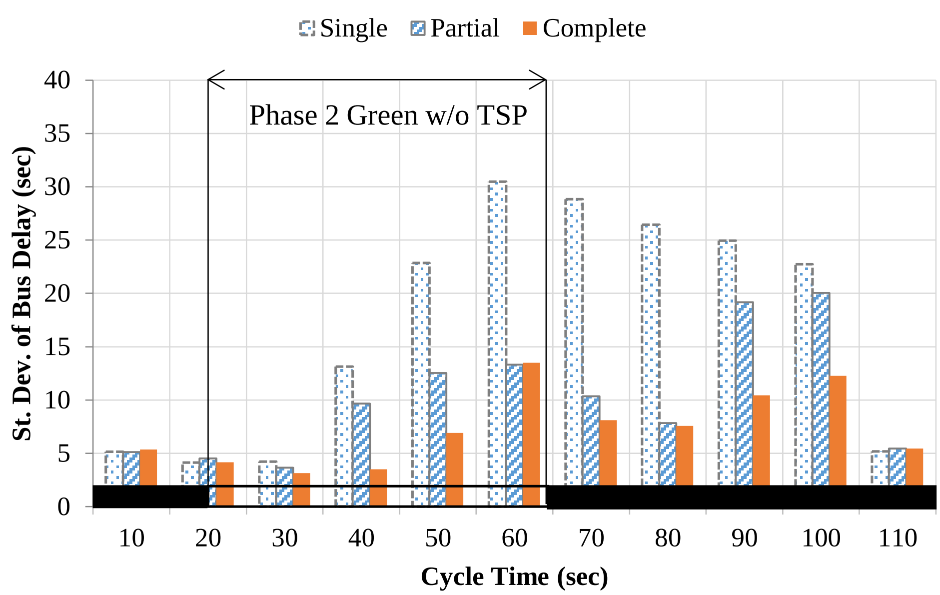
It is conceivable that bus arriving at some particular time point in a cycle may benefit more or less from a TSP operations. To identify when in a cycle that a bus arrives can benefit the most in each case, we varied the expected arrival time of the bus at the stop bar throughout the entire cycle in a 10 second increment. In each arrival time case, two types of variability are encoded in the simulation. First, the simulation is run for a period of 7800 seconds, and every 3 cycles (i.e. 330 second) one new bus will enter to network at a time instant that allows it arrives time at the stop bar at the target cycle time if it would dwell precisely the average duration. However, the bus will dwell at the bus stop for a log-normally distributed duration. Second, the simulation is repeated five times with five random seed numbers. In this way, a total of about 120 data points are collected for each targeted cycle time for expected bus arrival. And there are a total of 11 targeted cycle times for bus arrival for each of the three cases.

6.4.2. Performance Evaluation

Figure 31 and Figure 32 show the performance of the SMINP model when given varying degrees of stochastic inputs under moderate (v/c=0.7) and moderate (v/c=0.9) to heavy traffic conditions respectively. Generally speaking, the SMINP provided much better bus delay reduction almost at all targeted cycle times for bus arrival. In addition, the standard deviations of bus delays are almost uniformly better when complete distribution information is used in deciding the signal timing for bus priority. A lower standard deviation for bus delay indicates a more reliable service of the TSP strategy provided to the bus. It should be mindful that the standard deviation shall not be associated with any probability measures such as the one used to derive the confidence interval for the mean. This is because the bus delay at this intersection is a function of link travel delay, bus stop dwell time and signal delay. Although the dwell time is a log-normal distribution, the resulting bus delay is not.

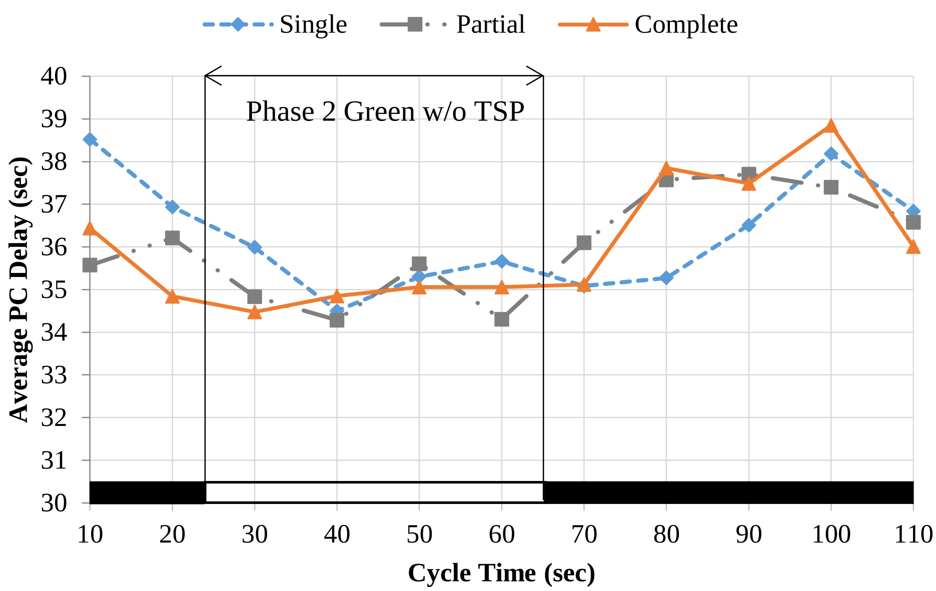


(a) Average Bus Delay



(b) Standard Deviation of Bus Delay

Figure 31: Impact of Distribution Inputs on SMINP Performance at V/C = 0.7.



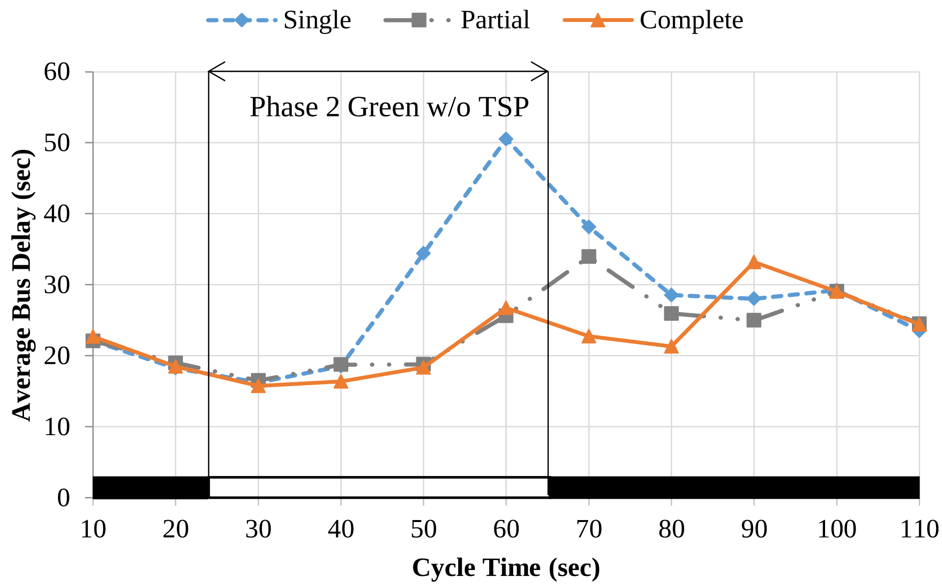
(c) Average Passenger Car Delay

Figure 31 Continued.

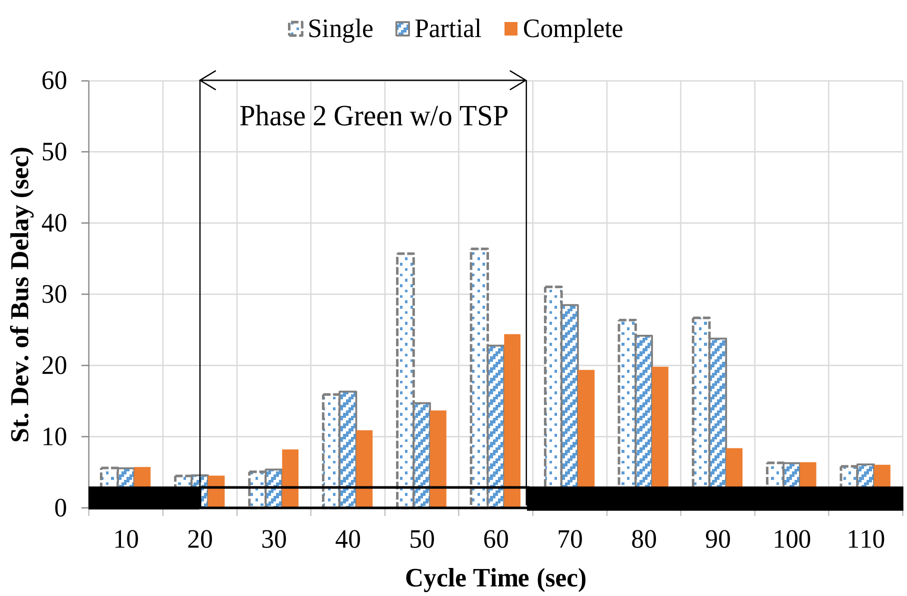
A closer examination on Figure 31 and Figure 32 reveals some interesting observations. First, Figure 31-(a) and Figure 32-(a) both show that the bus delay peaks in case A (single). Because based on a single estimate that the bus will arrive before the end of the green signal, there is no reason for any deterministic TSP strategy to extend the green signal for a priority. However, due to random chance, the bus may stay longer than expected. If the bus stay for too long, the consequence is that it may just miss the end of green time and is forced to wait until the next cycle. The chance of this happening increases as the expected bus arrival time goes nearer to the end of the green time. By taking into account the distributional information as in case C (complete), the SMINP model will extend the signal timing to cover as many dwell time cases as it can while managing the impacts to other traffic. And from the trend, one can tell that SMINP model is not simply adding a few seconds to the expected arrival time to cover the a few extra cases of buses dwelling longer than expected. Because if that is the case, the delay profile will be very similar to that of case A but only slightly shifted to the right.

Second, the random chance of bus arrival time doesn't seem to affect model performance too much in case A if the bus is expected to arrive at the beginning of or seconds before the start of the green time. This observation implies that deterministic TSP strategies are still effective in real-world implementation given its nature of randomness, if and only if the expected arrival time is close to the beginning of the green time of the priority phase. Nonetheless, Figure 31-(c) and Figure 32-(c) show that, during these same expected arrival periods, using the SMINP model with complete distribution information allows the bus priority to cause less impact on passenger cars. And the saving in PC delays is much higher when the congestion level is lower.

Lastly, supplying partial distribution information to SMINP seems to work just as well as complete information in terms of average bus delay and its deviations. This demonstrates the usefulness of the SMINP model since it is much more likely to have partial distribution information than complete information in any practical situations.

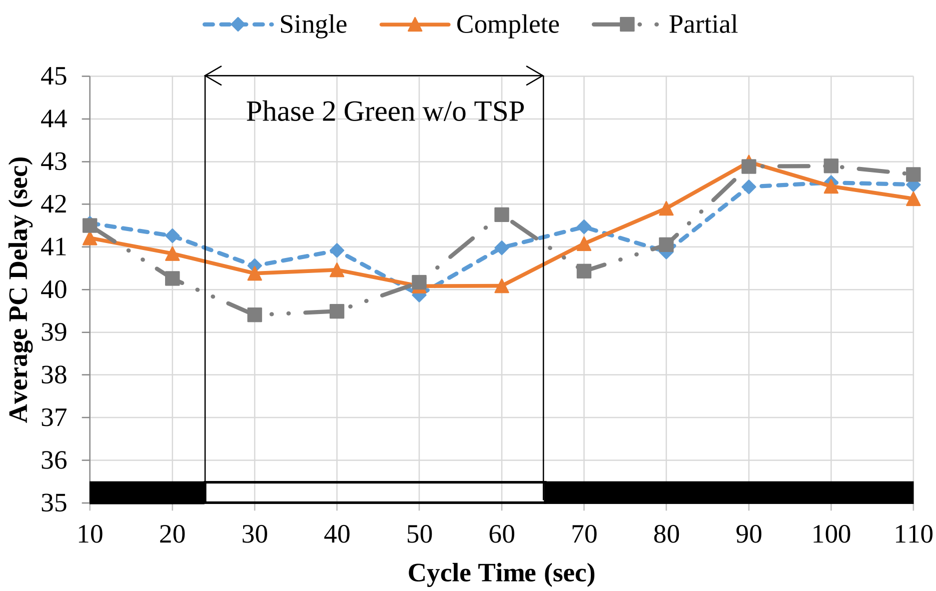


(a) Average Bus Delay



(b) Standard Deviation of Bus Delay

Figure 32: Impact of Distribution Inputs on SMINP Performance at V/C = 0.9.



(c) Average Passenger Car Delay

Figure 32 Continued.

6.5. Summary

This section evaluated in detail the proposed SMINP model in a single intersection setting. The model was implemented in a real-time transit signal priority control system. Analyses were performed to compare the proposed control model SMINP with the state-of-the-practice active TSP strategy. Both control models are compared with the fixed-time-do-nothing approach using the same hypothetical intersection in a simulation environment. In the case of no competing bus routes, the SMINP resulted in as much as 30% improvement of bus delay in low to medium congestion conditions. The comparison also indicated that the SMINP model can recognize the level of congestion of the intersection and automatically give less priority to the bus so as to minimize impact to the traffic on conflicting phases. In the case when there are three competing bus routes, SMINP handles multiple bus priority much better. The SMINP automatically adjust the relative importance of bus priority without the need to manually change the priority weighting factor, and it provides more balanced timings for both bus and the general traffic.

An analysis on the impact of stochastic bus dwell time was made. The results showed that the SMINP model performs the best when complete distribution was used as input. When only the average arrival time is available, the SMINP model worked well only if the bus arrived around the beginning of the green time. When partial information about the distribution was used as input, the SMINP's performance approximates the performance when complete distribution was used.

7. SIMULATION STUDIES FOR COORDINATED SIGNAL ARTERIAL

This section presents a series of simulation studies to test the performance of the RTSP and a number of its variation in a signalized corridor setting. First, the setup of the arterial with five signals is described. Several model variations and a localized version of the RTSP model, called Localized-TSP model, are compared. A sensitivity analysis is conducted for the weight coefficients of the objective functions. Finally, the R-TSP and the L-TSP models are evaluated for superiority on schedule-based performances.

7.1. Test Arterial Setup

Figure 33 shows the hypothetical corridor with five intersections that is used in testing the RTSP model. The intersection spacing is randomly selected, with two shorter ones on west half of the corridor and two longer ones on the east half of the corridor. Most approaches at the intersections are with two through lanes and one exclusive left-turn lane. General traffic volume is randomly selected at each intersection. But aggregately, the westbound is the peak direction of travel. Similarly, bus traffic is heavier on westbound than on eastbound, manifested by smaller headway and longer dwell duration per stop on the westbound bus route. Near-side bus stops are on the first intersections in both directions. Other bus stops are located on the far-side of the intersections or mid-block of a link to avoid interactions with standing queues.

Two bus routes are setup on the corridor. Buses on the eastbound route enter the network every 6 minutes, and they traverse from intersection 1 to 5 and stop at each bus stop for

30 seconds. These buses are scheduled to exit this group of 5 intersections 420 seconds after their entry. The minimum possible travel time, accounting for only the travel time and the dwell time, is 380 seconds. There is about a total of 40 seconds delay or 8 seconds per intersection allowed in order to be on time or early. The westbound bus route is busier with 4-minute headway. Buses on this route dwell 40 seconds per stop, and have a scheduled exit time of 500 seconds after entry, with a minimum possible travel time of 450 seconds. All buses have a desired running speed of 50 kph (31 mph), 10 kph (6.2 mph) slower than passenger cars.

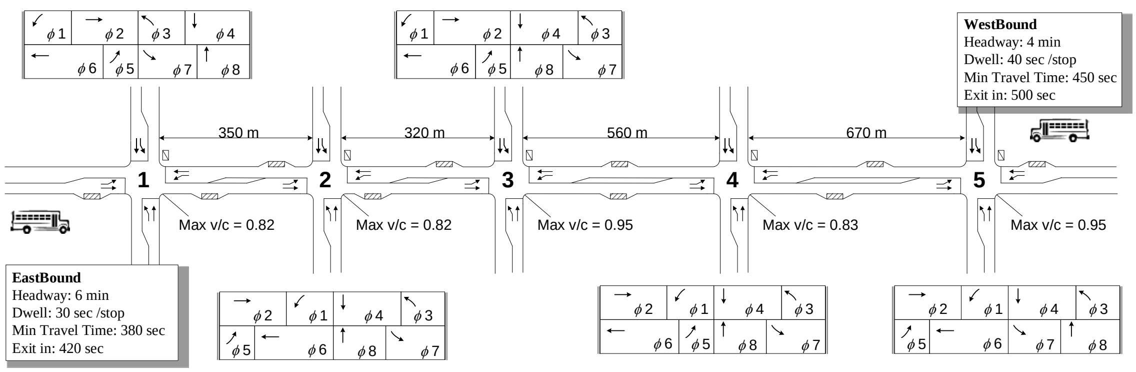


Figure 33: Hypothetical Test Corridor with Five Coordinated Intersections.

Popular signal optimization software, SYNCHRO, is used to optimize the corridor signal timing offline based on the selected traffic volume, excluding bus traffic. Both leading and lagging phasing are allowed for left-turns on all intersections. The optimized phasing sequence is shown in Figure 33. And Table 9 summarizes the traffic volumes, the optimized splits and the optimized offsets for all five intersections.

Table 9: Traffic Volume and Signal Timing Setups for Simulation Evaluation.

Intersection 1 (offset = 0)								
Phase	ϕ1	ϕ2	ϕ3	ϕ4	ϕ6	ϕ5	ϕ7	ϕ8
# of Lanes	1	2	1	2	2	1	1	2
Volume (vph)	156	858	125	530	1092	109	140	390
Optimal Splits (s)	19	40	16	25	44	15	19	22
Intersection 2 (offset = 59)								
Phase	ϕ2	ϕ1	ϕ4	ϕ3	ϕ5	ϕ6	ϕ8	ϕ7
# of Lanes	2	1	2	1	1	2	2	1
Volume (vph)	942	100	550	70	50	1220	450	150
Optimal Splits (s)	44	16	29	11	11	49	22	18
Intersection 3 (offset = 56)								
Phase	ϕ1	ϕ2	ϕ4	ϕ3	ϕ6	ϕ5	ϕ8	ϕ7
# of Lanes	1	2	2	1	2	1	2	1
Volume (vph)	130	982	400	140	1470	180	250	100
Optimal Splits (s)	19	48	20	13	51	16	20	13
Intersection 4 (offset = 16)								
Phase	ϕ2	ϕ1	ϕ4	ϕ3	ϕ6	ϕ5	ϕ8	ϕ7
# of Lanes	2	1	2	2	2	1	2	1
Volume (vph)	882	180	340	300	1210	130	500	20
Optimal Splits (s)	42	22	20	16	48	16	26	10
Intersection 5 (offset = 61)								
Phase	ϕ2	ϕ1	ϕ4	ϕ3	ϕ5	ϕ6	ϕ7	ϕ8
# of Lanes	2	1	2	1	1	2	2	2
Volume (vph)	592	150	720	80	160	1270	240	250
Optimal Splits (s)	42	21	27	10	15	48	16	21
Common Cycle Length = 100 sec								
Veh Clearance = 4 sec								

7.2. Formulation Comparisons

The model formulated earlier provides a framework for a large variation of models. For example, how to define timing deviations, whether allow variable cycle length and so on. Variations of the R-TSP model are developed and compared. In addition, for each variation of the R-TSP model developed, a corresponding model is developed that adopts a localized TSP approach. A localized TSP (L-TSP) model is essentially a route-based TSP (R-TSP) model applied to a single intersection by removing the bus route information. An L-TSP model is formulated only for one intersection as soon as a bus is detected locally. When a bus pass through multiple intersections, then multiple L-TSP models are to be created one at a time when the bus approach each intersection. The L-TSP models represent the state-of-the-practice approach of providing transit signal priorities. Comparing variations of R-TSP models to variations of L-TSP models can give us insights on the limitation of current state-of-the-practice TSP approaches.

Five variations were developed for both L-TSP and R-TSP models, i.e. TSP_L1 through TSP_L5 and TSP_S1 through TSP_S5. A summary of the setup parameters along with their performances evaluated in the test corridor are tabulated in Table 10. All TSP variations were compared against the baseline TSP_0, which was running the fixed-time timing optimized by SYNCHRO. Five identical simulation seeds were used for each model variation for a simulation period of 7800 seconds. Number of stops and delays per vehicle by vehicle classes are recorded for each intersection, but only the corridor metrics are reported. The percent changes in delays for different types of traffic are reported in Figure 34. Detailed comparisons are presented in the following subsections.

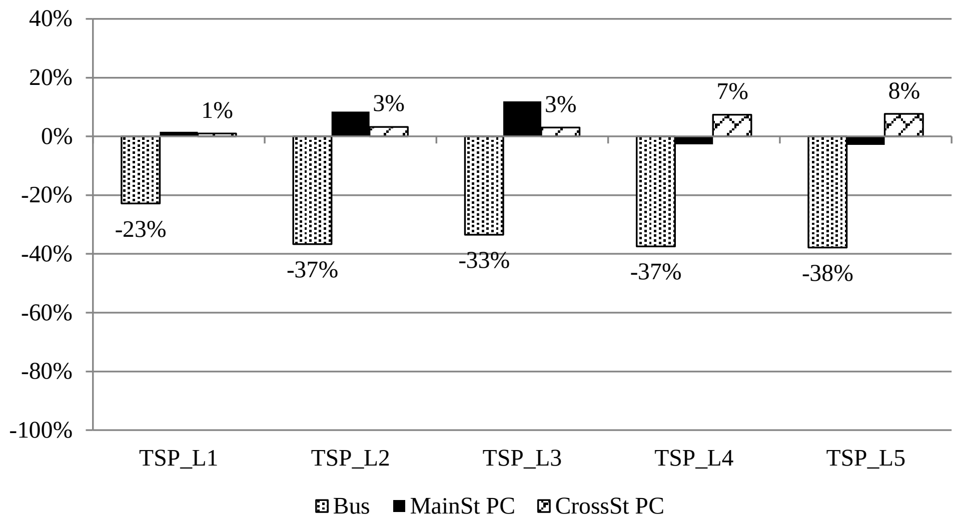
Table 10: Model Setup Parameters and Evaluation Test Results.

Model Variations	Optimization Strategies	Objective Bus	Objective Timing	Max Impl. Cycle	Allow Variable Cycle Len	Bus		All PC		Bus	Main St. PC	Cross St. PC	All PC
						Stops (/veh)	Delay (s/veh)	Stops (/veh)	Delay (s/veh)	Delay Change (%)			
TSP_0	No TSP	-	-	-	-	0.86	34.4	0.51	24.4	-	-	-	-
TSP_L1	L-TSP	Delay	SG	10	0	0.73	26.6	0.52	24.7	-23	1.5	1.0	1.2
TSP_S1	R-TSP	Delay	SG	10	0	0.67	25.1	0.53	25.7	-27	3.0	6.6	4.9
TSP_L2	L-TSP	Delay	SG	10	1	0.67	21.8	0.53	25.8	-37	8.4	3.3	5.6
TSP_S2	R-TSP	Delay	SG	10	1	0.49	17.6	0.54	26.5	-49	7.0	9.3	8.2
TSP_L3	L-TSP	Delay	LG	-	1	0.69	22.9	0.54	26.2	-33	12.0	3.1	7.1
TSP_S3	R-TSP	Delay	LG	-	1	0.27	8.0	0.59	28.6	-77	23.1	11.9	16.9
TSP_L4	L-TSP	Delay	LG+ER	-	1	0.66	21.5	0.52	25.1	-37	-2.6	7.4	2.8
TSP_S4	R-TSP	Delay	LG+ER	-	1	0.26	8.0	0.56	27.4	-77	12.2	12.3	12.2
TSP_L5	L-TSP	Delay	GD+ER	-	1	0.65	21.4	0.52	25.2	-38	-2.9	7.7	2.9
TSP_S5	R-TSP	Delay	GD+ER	-	1	0.40	12.0	0.53	25.9	-65	3.8	8.0	6.1

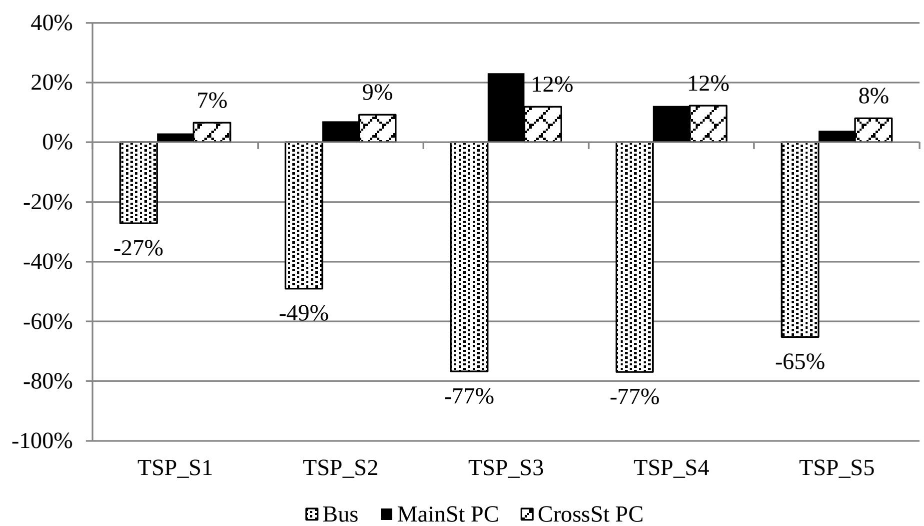
* SG – short green; LG – late green; ER – Early Red; GD – Green Start Deviation.

7.2.1. Localized TSP versus Route-Based TSP

A localized TSP (L-TSP) model is analogous to the greedy algorithm, which searches for local best at every step of execution. On the contrary, an R-TSP model takes a holistic approach to plan out the system optimum as soon as it knows the route for which the bus is taking. This difference is evident when comparing the reductions in bus delays between Figure 34-(a) and Figure 34-(b). Due to neglecting the timings and their constraints in other intersections, local optimization was not able to produce more than 37% reduction in bus delay, while the system optimization approach could yield more than 70% reduction in bus delay. To achieve similar level of bus delay reduction using local optimization routine, each intersection will need to allow temporary phase oversaturation strategy as in TSP_L6, which produces more than 60% reduction in bus delay. But the side-effect is serious disruptions to the side-street traffic (23% delay increase).



(a) Using Localized TSP Strategy



(b) Using Route-Based TSP Strategy

Figure 34: Percentage of Delay Changes Comparing against No TSP Control.

7.2.2. Fixed Cycle Length

The first variation of an RTSP model is to not allow cycle length to change. In the spirit of signal coordination, fixed cycle length guarantees that timings between intersections be synchronized over time. Hence, the only changeable timing parameters are the green splits. Both TSP_L1 and TSP_S1 penalize the compression of green phases. So when a bus requires a priority, the most either program can do is to push the green times of the conflicting phases to the minimum green, which is constrained by the max degree of saturation allowed by user inputs. Both strategies provided some priority to the buses, inducing about 15 – 22 % bus delay reduction with only 1-5% delay increase to other traffic.

7.2.3. Variable Cycle Length

Fixed cycle length is too restrictive for any decently timed signalized corridor. To allow more flexibility, TSP_L2 and TSP_S2 allow the coordinated phase to time past the yield points, effectively allowing the cycle length to change. This is an operation that resembles the actuated-coordinated signal control. However, to prevent the corridor from losing synchronization over time, it is also implemented with a max number of continuously affected cycles. When the max number is reached, any further priority requests are ignored until signal timing recovers to normal. The max number of affected cycles is set to 10 for both TSP_L2 and TSP_S2. As a result, more reduction of bus corridor delay is observed for TSP_L2 (i.e. 37%). However, variable cycle length induced 8% and 3 % more delays onto vehicles on both the main and the cross streets, and 6% increase overall. On the other hand, TSP_S2 was able to yield even more bus delay reduction (i.e. 49%) by recognizing current timings and their constraints from other coordinated signals. However, the further reduction of bus delay is not without a price. 8% of delay increase was resulted for other traffic.

7.2.4. Penalizing Timing Asynchronization

The restriction applied to the number of affected cycles is somewhat arbitrary. It is not easy to find the best threshold for this input parameter. Case in point, huge decrease in bus delay (i.e. 77%) is observed in TSP_S3 by removing the max number of affected cycle restrictions. This indicates the restriction severely limits the potential of the system optimization model in furnishing priorities to buses. To avoid using this constraint for keeping the corridor timings in sync, several model variations are tested using both optimization routines.

First variation is to use the start time of coordinated phases as a reference point. Any new timing that results in late start of the phases is discouraged. TSP_L3 and TSP_S3 both penalize the late start of coordinated greens. However, this penalization option alone does not work too well and it has led to increases of the overall vehicle delays, especially for TSP_S3. Surges in main street traffic delays of 12% and 23% were also observed for both routines respectively.

To further discourage the moving of the coordinated phases, variable z is introduced to penalize the early terminations of coordinated phases. This addition was implemented in

TSP_L4 and TSP_S4. As a result, the bus delay decrease for TSP_L4 is comparable to TSP_L2, while the increase in overall traffic delay has dropped to 3% from 6%. For TSP_S4, bus delay reduction is maintained at 77% as in TSP_S3, while the increase of main-street traffic delay is dropped from 23% to 12 %. Both TSP_L4 and TSP_S4 show the importance of penalizing late green and early red together to keep the corridor signals in sync.

Other studies have also suggested early green sometimes is detrimental to corridor signal progression. Therefore, a third option is to penalize the deviation from the scheduled start time of coordinated phases. x^+ and x^- are introduced to linearize the nonlinear deviation function (i.e. |*x*|). In the objective function, x_{ijk} is replaced with $x_{ijk}^+ + x_{ijk}^-$, and the following equality is added to the constraint set:

$$x_{ijk}^+ - x_{ijk}^- = t_{ijk} - T_{ijk}^{opt} \qquad \forall i, \forall k, j \in J_i^{coord} \qquad \text{(5-6)}$$

This formulation clearly minimizes the TSP impacts on other traffic, incurring 0% and 6% increase on delay to overall traffic for TSP_L5 and TSP_S5 respectively. But the tradeoff is the bus delay reduction. In both cases, bus delay reduction is less comparing to TSP_L4 and TSP_S4. The worse of the two, TSP_L5 yields only 17% of bus delay reduction comparing to TSP_0.

7.2.5. Allowing Temporary Over-Saturation (TOS)

All the model variations discussed so far are bounded by the minimum green times that would not result in any oversaturated phases. Although pedestrians are not explicitly modeled, these minimum green times are usually long enough to allow full walk phases for pedestrians. Therefore, the performance of those model variations could well approximate true performances in realistic traffic conditions with pedestrian traffic in the mix.

However, in some other intersections with limited pedestrian traffic, it may not be very critical to maintain a certain level of green durations at all time. Also, as Zeng, et al. [74] demonstrated in a single intersection case that allowing phases to go temporarily oversaturated in one or two cycles may not cause too much delay increase on other traffic. Similar formulation adaptations can be applied to both the L-TSP and the R-TSP models. Details of such formulation alteration for single intersection are documented in Zeng, et al. [74]. In essence, the constraint is dropped for the minimum green times that prevent the degrees of saturation from going over 1 in one cycle; instead, an alternate constraint is added to prevent the overall degree of saturation over the planning horizon for a particular phase from going above 1.

Table 11 listed the performances of the model variations with and without the TOS strategy. Generally speaking, allowing TOS gives more priorities to buses but causing more delays to other PC at the same time. This effect is most apparent for the model variations using the L-TSP optimization strategy. For example, model variation TSP_L5 can produce only about 38% bus delay reduction with merely 3% increase in PC delay, whereas TSP_L6 can use 12% more PC delay to exchange for about 70% bus delay reduction. Similar trend with lesser extents can be observed for the model variations using R-TSP strategy.

Table 11: Model Performances with and without TOS.

Model Variations	Allow Temp oversat.	Deviation of Green Duration*	Bus Delay Penalty Weight	Bus		All PC		Bus	All PC
				Stops (/veh)	Delay (s/veh)	Stops (/veh)	Delay (s/veh)	Delay Change (%)	
TSP_L5	0	COG	10	0.65	21.38	0.52	25.16	-37.90	2.92
TSP_L7	0	BOG	10	0.66	21.60	0.52	25.26	-37.27	3.33
TSP_L6	1	COG	10	0.36	10.36	0.56	28.12	-69.91	15.03
TSP_L8	1	BOG	10	0.36	10.50	0.56	28.14	-69.50	15.11
TSP_S5	0	COG	1	0.29	8.62	0.55	26.96	-74.96	10.28
TSP_S7	0	BOG	1	0.28	8.60	0.54	26.66	-75.02	9.05
TSP_S6	1	COG	1	0.19	6.30	0.59	30.42	-81.70	24.44
TSP_S8	1	BOG	1	0.20	6.14	0.55	27.64	-82.17	13.06

* BOG – Background Optimal Green; CG – Current Optimal Green;

7.2.6. Definition of Green Durations

We also explored different definitions of timing deviations. In addition to the timing deviation of signal coordination (i.e. definition of variables, x_{ijk} and z_{ijk} in the formulation), a fundamental timing deviation is the deviation from optimal green duration (i.e., definition of variable y_{ijk}). There are at least two types of optimal green durations: background and current. The background optimal green (BOG) duration refers to the green duration optimized based on the prevailing traffic conditions. In this study, it is the green duration that is obtained from SYNCHRO offline optimization, and it does not change as long as the prevailing traffic condition has not changed drastically. On the other hand, the current optimal green (COG) duration refers to the green duration that is currently being implemented. When two TSP optimization sessions have overlapping planning horizons (which is possible under the rolling horizon optimization scheme), the timing input for the second session is the new timing outputted from the first session. Therefore, the COG is different from BOG for the second session, while they are the same for the first session.

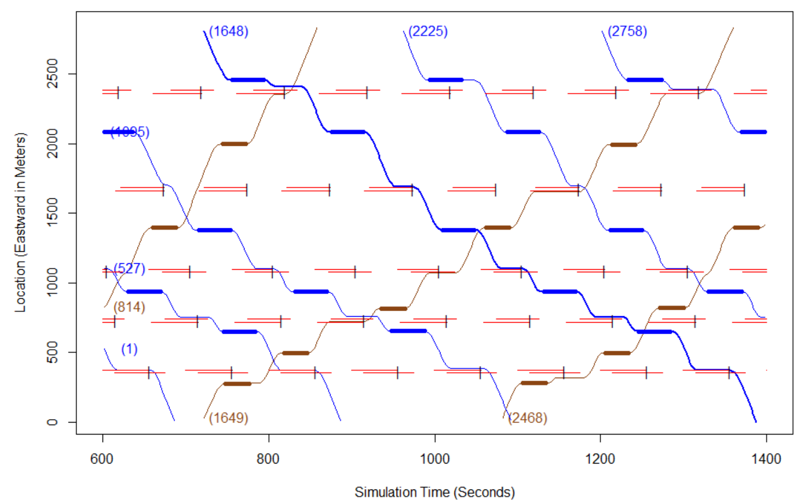
Table 11 listed the model performances of those model variations using either COG or BOG to define timing deviations. Interestingly, The definition of timing deviation almost has no effect for L-TSP models. TSP_L5 and TSP_L6 are mostly identical to TSP_L7 and TSP_L8 respectively in terms of both bus delay reduction and PC delay increment. But for R-TSP models, this definition is very critical. This is especially true when the TOS strategy is in place. By comparing TSP_S6 with TSP_S8, the increase of PC delays has drop from 24% to about half at 13% with even 1% more reduction in bus delays. This improvement is due to the tendency of long planning horizon in the R-TSP models. Because the R-TSP models need to plan for not only one but multiple intersections, it is likely they have longer planning horizons that span multiple cycles. In these cases, the chances of overlapping planning horizons from two or more optimization sessions have increased. As the overall planning horizon keep rolling longer and longer due to more and more buses appear in any parts of the corridor, the corridor timing could drift away from BOG further and further if COG is used in defining timing deviations. To prevent this, BOG is used as the timing deviations for all later model variations for both L-TSP and R-TSP.

7.2.7. Superiority of R-TSP models over L-TSP models

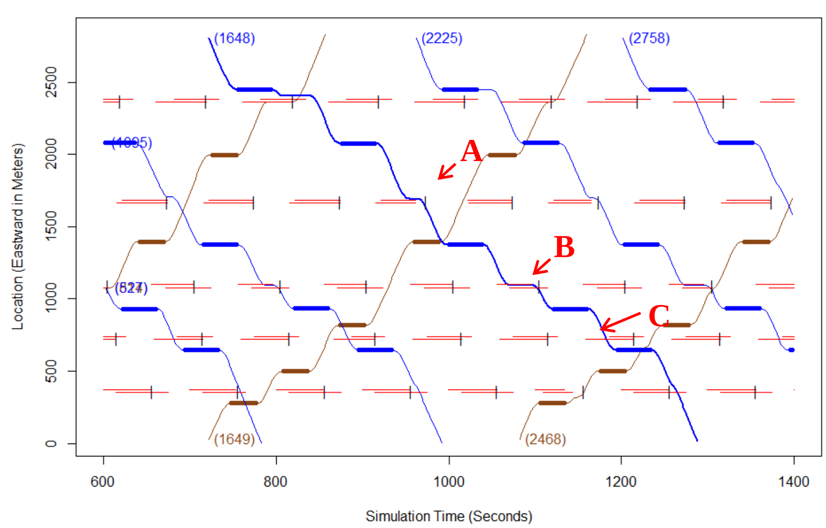
One final clue that is revealed by Table 11 is the superiority of the R-TSP models over the L-TSP models. Even with relatively high weights on bus delay penalization and more flexibility to adjust timing with the TOS strategy in place, TSP_L6 can produce at most 69.9% of bus delay reduction with more than 15% delay increase on PC traffic. As contrast, TSP_S7 could produce more than 75% bus delay reduction with only about 9% delay increase on PC traffic. This means, even in corridors with limited flexibilities in adjust timings, the R-TSP model could coordinate timing adjustments at all intersections to better serve the bus priority needs with lower impacts on other traffic. Being able to perform well when the range of adjustable green times are limited is a crucial feature for the model to be practically useful.

7.2.8. Trajectory Analysis

In addition to the overall delay analyses, we selected one particular bus line for a trajectory analysis, as shown in Figure 35. Several Points are to be made. Without any TSP strategies, the bus (ID: 1648) is stopped by the signal at every intersection. As the average stops per intersection being close to 0.9 for all other buses in TSP_0 case, this bus trajectory is representative of other bus trajectories. When TSP strategy is enabled, both L-TSP and R-TSP models are able to cut down its corridor time from close to 660 seconds to about 525 seconds on average.

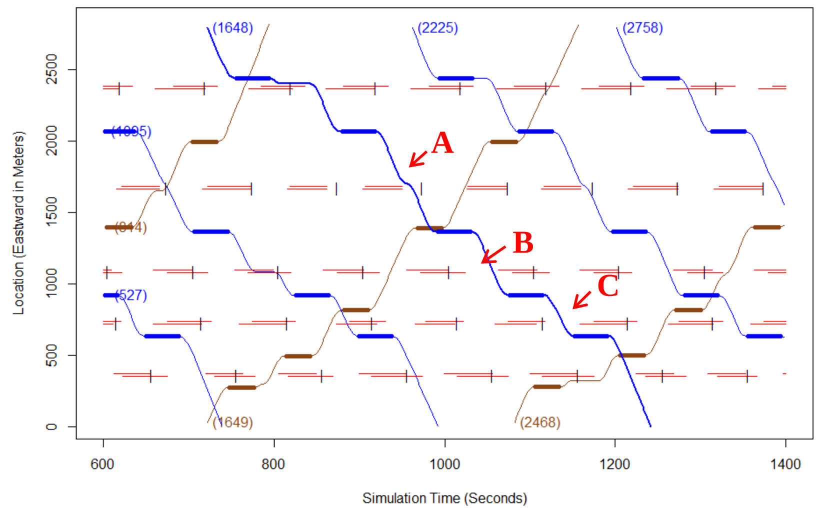


(a) Fixed Time No TSP



(b) Localized TSP (TSP_L4)

Figure 35: Trajectory Analyses for Bus Line 1648.



(c) Route-based TSP (TSP_S4)

Figure 35 Continued.

Comparing between the TSP_L4 and TSP_S4 at point A, it is interesting to see that TSP_S4 allows the bus to experience less delay by pushing ahead not only the beginning of the current cycle but also that of the previous cycle. This helps minimize the impacts to cross street traffic by distributing this priority need across two cycles instead of just one. In contrast, TSP_L4 has only managed to push back the beginning of the current cycle because the bus at the second intersection is not detected only after the beginning of the current cycle.

From a systematic perspective, allowing the bus to pass through the second intersection at A just a few seconds earlier helps the bus pass through the next two intersections much easier. TSP_L4 does not take into account the timing at all downstream intersection, it minimizes the priority delay only from a local perspective. Therefore, TSP_L4 has just missed the opportunity to extend the green time at B, while it has to extend the green at C by a large amount. In contrast, TSP_S4 plans ahead to squeeze more green time at the upstream intersection at A so as to make it through the next two intersections with no green extension needed at the third intersection.

7.3. Sensitivity Analysis

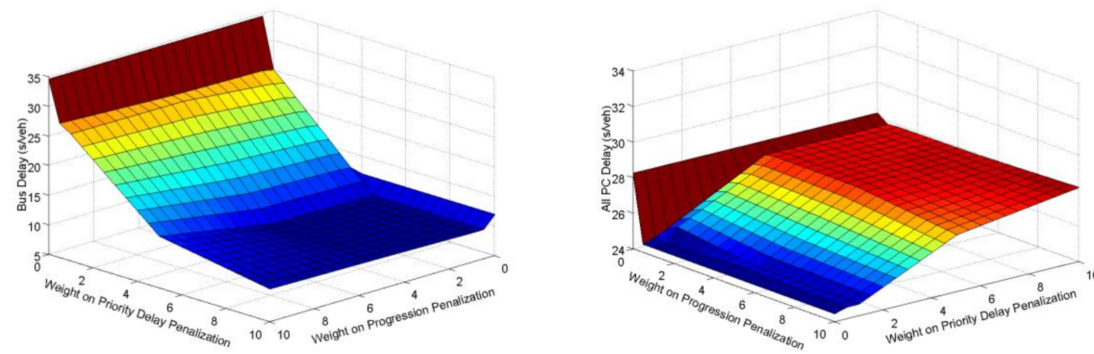
In the objective function, there are three competing terms: (1) timing duration deviations (e.g. y_{ijk}), (2) signal progression deviations (e.g. x_{ijk}) and (3) bus priority delays (e.g. d_{ijk}). Although all three types of terms are in the same unit (i.e. second), one second in signal deviation does not equal to one second to bus priority delay, due to the difference in their physical meanings. So it is important to study how the weight coefficients will change the performance of these models.

Firstly, due to different number of entries for each of the three terms, the summation of one term may dominate another one. To prevent this, the coefficients of all entries of a particular term are normalized according to the following:

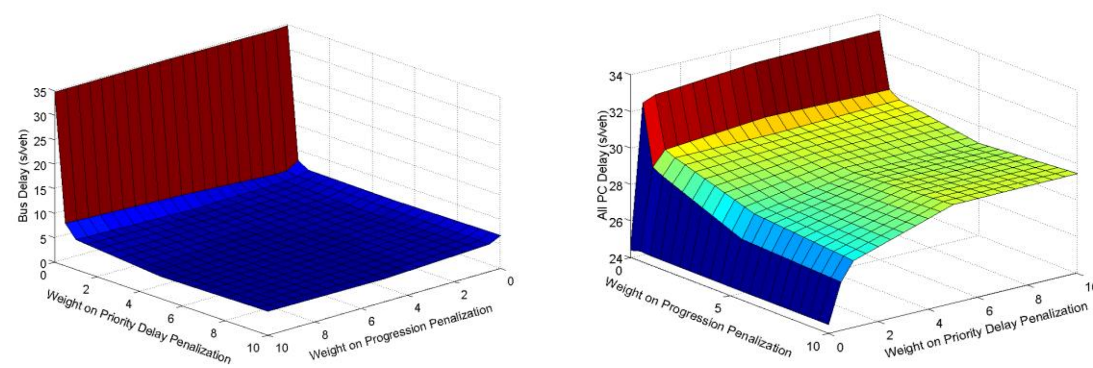
$$\begin{aligned}
&\text{Term } y: \quad c_{ijk} = X_{ijk} \Big/ \sum_{i\in I}\sum_{l\in L}\sum_{j\in J}\sum_{k\in K} X_{ijk} \\
&\text{Term } x: \quad c'_{ijk} = X_{ijk} \Big/ \sum_{i\in I}\sum_{l\in L}\sum_{j\in Z_{il}^2}\sum_{k\in K} X_{ijk} \qquad \forall i, \forall j, \forall k, \forall n \qquad (5\text{-}7) \\
&\text{Term } d: \quad o_n = AP_n \Big/ \sum_{n\in N} AP_n
\end{aligned}$$

Where X_{ijk} is the degree of saturation of a phase before optimization, and AP_n is the assigned priority level for bus *n*. The assigned priority is a static input parameter provided by the operator of the system. This is needed because some buses may be more important than other buses. For example, all buses on a busy bus line can be assigned higher static priority; or a light rail vehicle can receive higher static priority than a regular bus. Of course, these static assignments based on operators' known only serve as *a priori* to tell the model the relative importance of all the competing priority requests. But a bus with higher static priority does not mean it will get faster service time. The actual priority of all competing vehicles are determined by the outcome of the optimization model.

After normalization, all three terms have the same importance on the objective function. But there may be compelling reason that one term is more important than the other term, any changes to the values of that term should be penalized more in the objective function. To customize each term's relative importance, the normalized terms can each be multiplied by a multiplier. Let α_y, α_x and α_d denote the multipliers for term *y*, *x* and *d* respectively. In this sensitivity analysis, we keep $\alpha_y = 10$ as a constant and vary both α_x and α_d from 0 to 50 at [0, 0.1, 0.5, 1, 5, 10, 20, 50]. Therefore, 64 combinations for both R-TSP and L-TSP models are tested. A total of 128 models with different multipliers were applied to the test corridor of five intersections, as in Figure 33, for a period of 2 simulation hours for each of the five random seeds. From these 64 data points, a surface plot of bus delay and another plot of all PC delay are made for each model in Figure 36.



(a) Localized TSP (TSP_L8)



(b) Route-based TSP (TSP_S8)

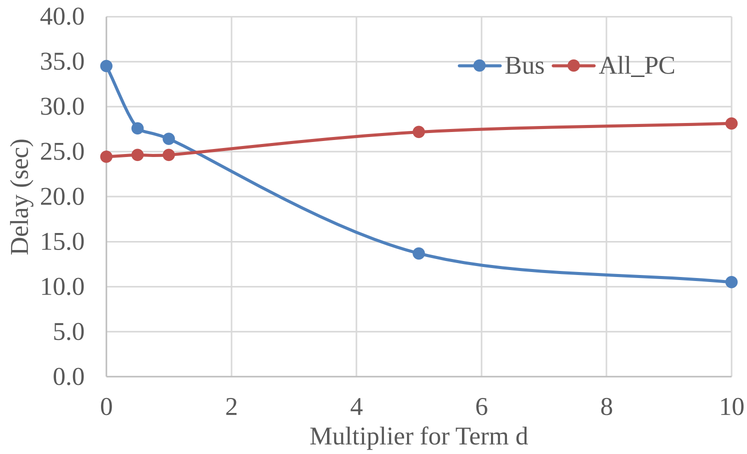
Figure 36: Delay Surfaces on Different Progression and Bus Delay Multiplier Values.

The first thing to be noted is the importance of having term *x* to limit the impacts of any TSP models to other traffic. If term *x* is absent, such as setting the multiplier for *x* (i.e. α_x) to zero, then the delay of all PC for TSP_L8 is dramatically increased to 28 sec per vehicle, and that for TSP_S8 is increased to 32 sec per vehicle as long as the multiplier for term *d* is not zero.

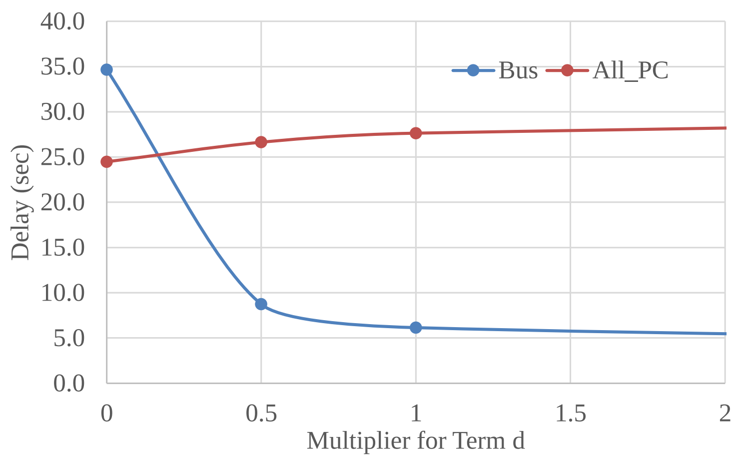
Another insight is that both models are not sensitivity to the change of α_x as long as it is not zero. At any α_d level, both the PC and bus delays are very similar when α_x varies at the positive value range. That means we can reduce the 3d delay surface to a 2d delay curve by picking any positive α_x value. Figure 37 shows such delay curves for both models. We define that a range effective if the changes of multiplier within this range also causes the change in bus or PC delays. Therefore, the effective range of α_d for TSP_L8 is (0,10], while that for TSP_S8 is (0,1] when both α_y and α_x are set to 10. This empirical range can be used as a reference for future analyses.

It is hypothesized that the effective range is related to the number of entries in each term. For TSP_L8, the numbers of entries for term *y*, term *x* and term d are on average 40, 10 and 1 respectively. For TSP_S8, the numbers are on average 320, 80 and 4 respectively.

More needs to be done to test this hypothesis.



(a) Localized TSP (TSP_L8)



(b) Route-based TSP (TSP_S8)

Figure 37: Delay Curves on Different Bus Delay Multiplier Values.

7.4. Optimize for Schedule Related Metrics

So far, analyses were done on the basis of bus delay. As mentioned before, schedule adherence is also a very critical performance measure for transit vehicles. In this study, we investigated two schedule adherence metrics: schedule lateness and deviations.

Table 12 shows some basic statistics for the bus routes on both EB and WB directions. EB buses can travel through all 5 intersections at a minimum of about 380 seconds given no signal delays at all. On the other extreme, they can be delayed at every signal if no priority strategies are in place, resulting in as much as 640 seconds in corridor travel time. For WB buses, the free flow corridor travel time is 450 seconds. But their actual travel time ranges from 578 to 678 seconds. Overall, 95.6% of the buses are late if no TSP strategies are in place. For WB buses, all of they are late by at least 30 seconds.

Table 12: Basic Statistics for Bus Routes.

Direction	Free Flow Travel Time (sec)	Headway (min)	Min Travel Time w/o TSP (sec)	Max Travel Time w/o TSP (sec)	On-Time Travel Time (Sec)	% Late Bus w/o TSP by x Seconds				
						0	10	20	30	40
EB	380	6	420	640	450	89	84	75	67	63
WB	450	4	578	678	540	100	100	100	100	87.3
Overall	-	-	-	-	-	95.6	93.6	90	86.8	77.58

7.4.1. Formulation Change for L-TSP

When optimize for bus priority intersection by intersection (as in the L-TSP models), the only possible way to optimize for schedule-adherence is to plan out a schedule for every single intersection. With a given schedule of when a bus should exit the last intersection, we establish a schedule for each intersection based on the ratio of the free-flow link travel time to the free-flow corridor travel time.

To minimize for the schedule related measure, the calculation of delay needs to be precise for the same reason in the R-TSP formulation. The delay formulations using one inequality as in [74] need to be changed into the following:

$$d_{j,k-r,n} \geq t_{j,k-r} - \underline{R}_{j,k-r,n} - (1-\theta_{jkn})M - (1-\phi_{j,k-r,n})M \qquad \begin{aligned}\forall j, k \in K \setminus \{1,\ldots,r\},\\ r \in \{0,..,|K|-1\}\end{aligned} \qquad (5\text{-}8)$$

$$d_{j,k-r,n} \leq t_{j,k-r} - \underline{R}_{j,k-r,n} + (1-\theta_{jkn})M + (1-\phi_{j,k-r,n})M \qquad \begin{aligned}\forall j, k \in K \setminus \{1,\ldots,r\},\\ r \in \{0,..,|K|-1\}\end{aligned} \qquad (5\text{-}9)$$

$$d_{jkn} \geq 0 \qquad \forall j, \forall k, \forall n \qquad (5\text{-}10)$$

Where θ_{jkn} indicates which cycle *k* within the planning horizon that bus *n* arriving at phase *j* will pass through. With only θ_{jkn}, the bus delay may result in a negative value. Therefore, ϕ_{jk} is used to determine the sign of $t_{j,k-r} - \underline{BR}_{j,k-r,n}$, and force $d_{j,k-r,n}$ to be zero when the difference is negative. To do this, two more sets of constraints are needed for this disjunctive case:

$$\underline{R}_{jkn} > t_{jk} - \phi_{jkn}M \qquad \forall j, \forall k, \forall n \qquad (5\text{-}11)$$

$$\underline{R}_{jkn} \leq t_{jk} + (1-\phi_{jkn})M \qquad \forall j, \forall k, \forall n \qquad (5\text{-}12)$$

$$d_{jkn} \geq 0 - \phi_{jkn}M \qquad \forall j, \forall k, \forall n \qquad (5\text{-}13)$$

$$d_{jkn} \leq 0 + \phi_{jkn} M \qquad \forall j, \forall k, \forall n \qquad (5\text{-}14)$$

This set of inequalities precisely formulated the relationships of signal timings, bus arrival times and delays regardless whether the delay term is present in the objective function. Therefore, schedule-related metrics can be easily derived from the delay computed above:

$$\delta_n \geq \sum_{k \in K} d_{jkn} + \overline{R}_{jn} - R_{jn}^{\text{planned}} \qquad \forall j, \forall k, \forall n \qquad (5\text{-}15)$$

$$\delta_n \geq 0 \qquad \forall j, \forall k, \forall n \qquad (5\text{-}16)$$

Where $\overline{R}_{jn}$ is the free-flow arrival time of bus n at phase *j*, and R_{jn}^{planned} is the scheduled exit time of bus *n* at phase *j* at the current intersection; similar to the R-TSP formulation, δ_n is the schedule lateness for bus n. The following δ_n^+ and δ_n^- denote the positive and negative deviations from the established exit time schedule:

$$\delta_n^+ - \delta_n^- = \sum_{k \in K} d_{jkn} + \overline{R}_{jn} - R_{jn}^{\text{planned}} \qquad \forall j, \forall k, \forall n \qquad (5\text{-}17)$$

$$\delta_n^+, \delta_n^- \geq 0 \qquad \forall j, \forall k, \forall n \qquad (5\text{-}18)$$

7.4.2. Comparing L-TSP and R-TSP Family Models

A total of 12 model variations are formulated based on their objective for bus performance and the timing flexibilities. Table 13 compiled the results of simulation runs of each model variation. Table 14 summarizes the model variations that are developed specifically for schedule adherence of buses in both eastbound (EB) and westbound (WB) directions. Figure 38 showed the schedule performances of individual buses when different variations of both the L-TSP and R-TSP models are used as the signal control model in the same simulation setting. Table 16 to Table 19 in the appendix documents the schedules performance in more detail. The following subsections discusses the main insights gained from these simulation studies.

Table 13: Performance of Each of the R-TSP and L-TSP Family Models.

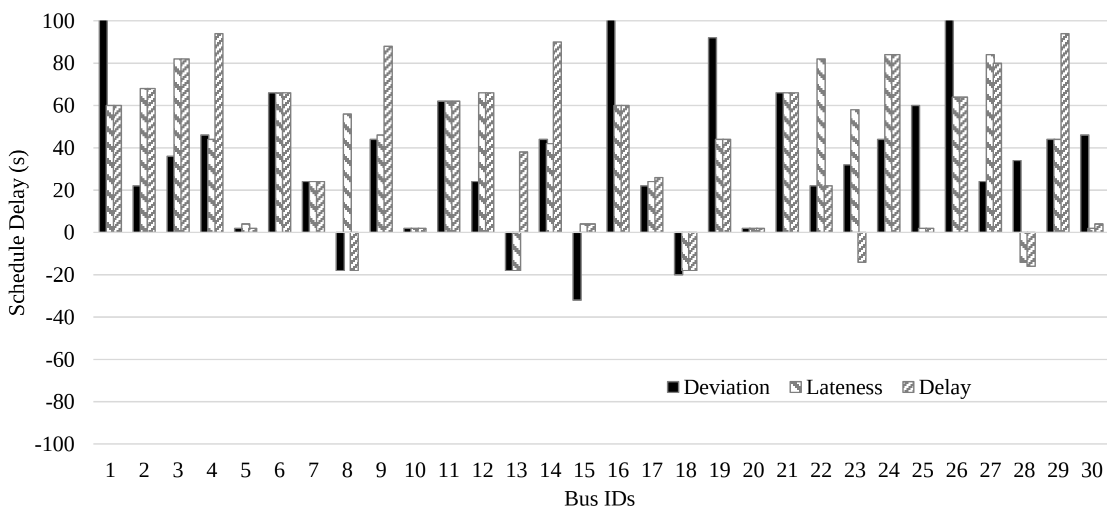
Model Variations	Objective Bus	Allow Temp oversat.	Bus		All PC		% Bus Delay Reduction	% Late Bus	Schedule Deviation (Sec)		% All PC Delay Increase
			Stops (/veh)	Delay (s/veh)	Stops (/veh)	Delay (s/veh)			Average	Std Dev.	
TSP_0	-	-	0.86	34.4	0.51	24.4	-	95.6	79.5	43.5	-
TSP_L7	Delay	0	0.66	21.6	0.52	25.3	**-37.3**	62.4	35.6	26.6	3.3
TSP_L7_L	Lateness	0	0.67	21.4	0.52	25.3	-37.9	**63.2**	33.9	25.7	3.3
TSP_L7_V	Deviation	0	0.72	23.2	0.54	26.0	-32.6	72.8	**33.7**	**28.6**	6.4
TSP_S7	Delay	0	0.25	7.5	0.58	28.3	**-78.3**	2.8	55.2	22.3	15.6
TSP_S7_L	Lateness	0	0.43	13.2	0.53	25.8	-61.7	**1.6**	27.6	17.6	5.5
TSP_S7_V	Deviation	0	0.59	19.7	0.54	26.0	-42.7	46	**5.6**	**9.6**	6.4
TSP_L8	Delay	1	0.37	10.4	0.56	28.3	**-69.9**	7.2	44.4	19.0	15.6
TSP_L8_L	Lateness	1	0.53	14.9	0.55	27.4	-56.8	**10**	23.4	14.2	12.3
TSP_L8_V	Deviation	1	0.61	18.1	0.60	30.2	-47.5	24	**10.8**	**14.1**	23.7
TSP_S8	Delay	1	0.13	4.4	0.58	29.8	**-87.2**	0.4	70.1	18.4	21.9
TSP_S8_L	Lateness	1	0.45	13.3	0.53	26.3	-61.4	**2.4**	28.4	18.1	7.7
TSP_S8_V	Deviation	1	0.57	19.3	0.55	27.3	-43.9	32.8	**3.7**	**7.3**	11.7

Note: Exit time at the last intersection is scheduled to be 450 (EB) and 540 (WB) seconds after a bus entering the network.

Table 14: Performance of R-TSP and L-TSP Models for Schedule Adherence.

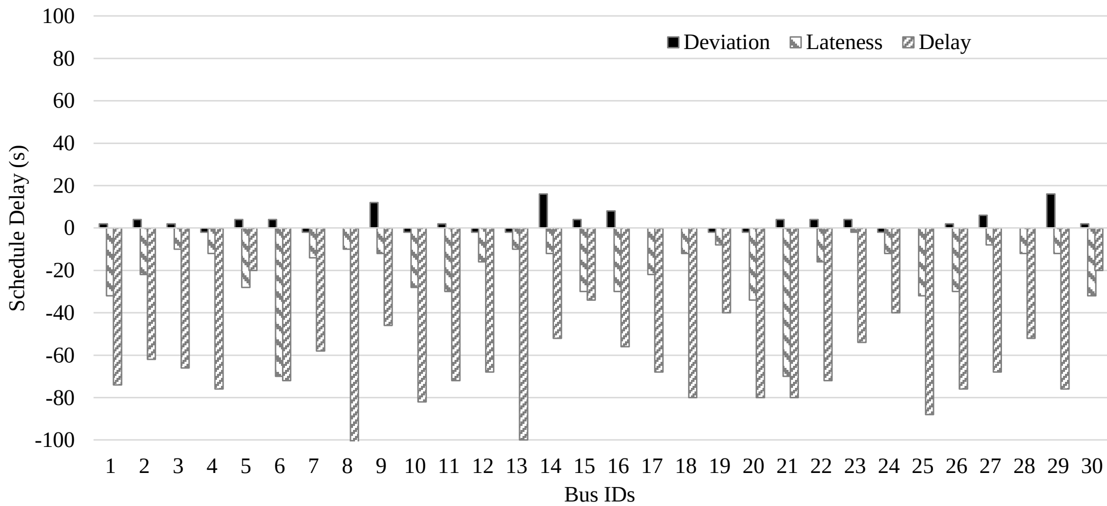
Model Variations	EB			WB			ALL			
	% Late Bus	Schedule Deviation (Sec)		% Late Bus	Schedule Deviation (Sec)		% Late Bus	Schedule Deviation (Sec)		% All PC Delay Increase
		Average	Std Dev.		Average	Std Dev.		Average	Std Dev.	
TSP_0	89	76.7	59.4	100	81.3	28.6	95.6	79.5	43.5	-
TSP_L7_L	**35**	25.7	17.0	**83**	38.5	29.6	**63.2**	33.9	25.7	3.3
TSP_L7_V	59	**22.0**	**19.6**	82	**42.4**	**30.5**	72.8	**33.7**	**28.6**	6.4
TSP_S7_L	**1**	35.2	18.2	**2**	22.5	15.2	**1.6**	27.6	17.6	5.5
TSP_S7_V	36	**7.3**	**12.0**	54	**4.9**	**8.6**	46	**5.6**	**9.6**	6.4
TSP_L8_L	**5**	25.2	16.5	**13**	22.1	12.6	**10**	23.4	14.2	12.3
TSP_L8_V	8	**6.6**	**8.0**	35	**13.7**	**16.5**	24	**10.8**	**14.1**	23.7
TSP_S8_L	**0**	34.8	19.3	**4**	24.0	16.0	**2.4**	28.4	18.1	7.7
TSP_S8_V	28	**1.7**	**1.4**	35	**5.1**	**9.1**	32.8	**3.7**	**7.3**	11.7

Note: Exit time at the last intersection is scheduled to be 450 (EB) and 540 (WB) seconds after a bus entering the network.



(a) TSP_L7 Family Models

Figure 38: Bar Plots of Bus Schedule Performances.



(b) TSP_S7 Family Models

Figure 38 Continued.

7.4.2.1. Schedule Lateness

With limited timing adjustability, TSP_L7_L cannot do much to reduce the buses' tardiness. More than 60% of the buses on average (35% EB and 83% WB) were late than the schedule exit time at the last intersection. Figure 38-(a) visualizes the lateness of individual buses in one of the simulation seed settings. Using the same L-TSP model but allowing more timing adjustability, TSP_L8_L could reduce the percentage of late buses to only 10% (5% EB, 13% WB). But by doing so, however, the impacts to non-transit vehicles worsened to 12.3 % from 3.3 % when the TSP_L7_L was implemented.

In comparison, TSP_S7_L and TSP_S8_L helped respectively 98.4% and 97.6% buses arrive either on-time or early. And their impacts to other traffic were only about 5.5%

and 7.7% of delay increase. Although these impacts were not better than those caused by the TSP_L7_L model variation, the improvement of schedule adherence is definitely much better than any variations of the L-TSP models can offer.

7.4.2.2. Schedule Deviation

Conclusions can be similarly drawn in terms of schedule deviations when comparing between R-TSP and L-TSP models. Limited timing adjustability virtually rendered TSP_L7_V impossible to achieve any schedule adherence – the average and standard deviation of schedule deviations are almost the same as those of the TSP_L7_L and TSP_L7 variations. In order to achieve better schedule adherence, some timing constraints have to be relaxed such as allowing temporary oversaturation. By so doing, TSP_L8_V was able to narrow the deviation within 11 seconds of the target schedule. However, the price was high; serious disruptions to other traffic were resulted, i.e., 24% of increase of all PC delays. That means the cost to benefit ratio of TSP_L8_V is much higher than of the TSP_S8_V and even worse than that of TSP_S7_V.

This is not surprising, because TSP_L7_V is to minimize its schedule deviation to current intersection, and the model considers early or late release from the current intersection as exactly the same. In many scenarios, early and late releases, however, have very different impacts for the downstream intersections, especially when timing adjustability is limited. But there is no way to know if early or late release will allow more or less delays at downstream intersections unless the timings from all intersections work together cooperatively. This is exactly what R-TSP is doing. Because of this ability to orchestrate all timings in a single optimization session, TSP_S7_V and TSP_S8_V can drop the average schedule deviation to within only 6 and 4 seconds of the target schedule respectively. And their impacts to other traffic are only about 6.4 and 11 % of delay increases to all PC.

Also interesting to point out, it seems that the improvements on bus performances gained from allowing timings to go oversaturated temporarily are not really worthwhile. The improvements are relatively marginal comparing to their impacts to other traffic. Furthermore, allowing green durations to always be able to run less than pedestrian walk times (as is the case when TOS is allowed) is not very realistic. Therefore, the TSP_S7 family of model variations represents the best overall performance of an R-TSP model can produce under the most realistic conditions.

7.4.3. Solution Time

All computer simulations were conducted on a desktop computer with Intel® Core ™ 2 Quad CPU of 2.4GHz and 8 GB RAM. The operating system was Windows 7 64-bit. Table 15 shows a summary of solution times for each variation of the TSP models. In general, it took less than 1 second to complete an optimization session. Although not many, there were optimization sessions which took up to 3.5 seconds to finish in the corridor optimization cases with R-TSP family models. Box plots in Figure 39 in the appendix reveal more information about the outliers.

Table 15: Summary of Optimization Solution Times.

Objective	Allow Temp Oversat.	LTSP			RTSP		
		Average (sec)	Standard Deviation (sec)	Count	Average (sec)	Standard Deviation (sec)	Count
Min Delay	0	0.140	0.037	1315	0.379	0.141	1322
Schedule Deviation	0	0.120	0.043	1309	0.516	0.208	1283
Schedule Lateness	0	0.271	0.086	1311	1.088	0.510	1305
Average		**0.177**	**0.089**	**3935**	**0.661**	**0.450**	**3910**
Min Delay	1	0.070	0.043	1311	0.414	0.149	1307
Schedule Deviation	1	0.127	0.046	1312	0.574	0.235	1291
Schedule Lateness	1	0.230	0.058	1315	0.517	0.243	1314
Average		**0.142**	**0.082**	**3938**	**0.502**	**0.223**	**3912**

The time needed for solving an L-TSP model was in general shorter than that was needed for solving an R-TSP model. That was because the L-TSP models were for single intersections only and they had much fewer variables than those in the R-TSP model formulated for the multiple intersections. It is hypothesized that the solution time may also be affected by the number of buses included in the optimizations.

7.4.4. A Remark on Schedule-Related TSP Operations

One implementation issue emerges in these experiments is labor intensity of making good schedules for every intersection for the L-TSP models and by extension for all locally based TSP strategies. When a TSP strategy to be conducted locally, it provides relatively higher or lower priority by conditioning on the difference between actual schedule lateness and the target schedule lateness. Consider an example, a bus's published schedule is to arrive at the next bus stop in 5 minutes which is 3 intersections away from current bus stop. To use any locally based TSP strategy, the schedule to arrive at intersection 1, 2 and 3 needs to be established. However, such a schedule is not favorable for two reasons.

- Setting a schedule for every intersection even if there is no near or far side bus stop is additional work that is usually not immediately available from the transit agency. And establishing such schedules is redundant because the true objective of any transit operations is to meet the schedule not at individual intersections but at bus stops that may span several intersections.
- Even with enough resources, the design of such schedules is non-trivial, because there is so many combinations of designs to choose from. Use the example above, possible schedule combinations for the three intersections include 2/5-2/5-2/5 minutes, 2-1-2 minutes, 1-1-3 minutes and so on. Some intersections are inherently busier than other intersection, so their timing adjustability is lower than others. That means in addition to travel time, traffic conditions and timing adjustability need to be accounted for when designing optimal schedule at individual intersections. And this is not a trivial task.

It requires a large amount of efforts to design optimal schedules for every single bus line at every single intersection before any locally based conditional TSP strategies can be confidently implemented. Even with the best designed schedules, it still does not change the fact that locally based TSP strategies/models can merely allow timings to work to their best in isolation. The comparison results showed that the L-TSP models can hardly achieve any corridor-level schedule deviations when the timing adjustability is restrictive. In contrast, the R-TSP considers the route nature of transit schedules, and incorporates that with the inner workings of signal coordination, which produced superior results.

7.5. Summary

This section detailed the simulation evaluation studies of the R-TSP model conducted on a coordinated signal arterial with five intersections. Comparisons over different model variations were conducted so as to refine the formulations of the R-TSP model. A Localized-TSP (L-TSP) model was adapted from a typical R-TSP model formulation based on the conventional concept of providing TSP at each intersection locally. Comparisons in terms of bus delay, schedule lateness, deviations, and delay to passenger cars were made between the R-TSP and L-TSP. The results showed that R-TSP outperformed L-TSP in that it produced greater benefits to buses while causing lower delays to other passenger cars. The solution time study showed the L-TSP models can be solved faster than the R-TSP models. A sensitivity analysis revealed that both models are somewhat sensitive to the weight parameters setup.

8. SUMMARY AND CONCLUSIONS

8.1. Contributions

This research investigates the applications of mathematical programming to adaptive transit signal priority. A real-time signal control system with adaptive transit signal priority model is developed and evaluated in both isolated intersection and arterial. Comparing to previous research and the state-of-the-practice, this research has the following contributions:

- Quantifying the impacts of a priority strategy on non-priority traffic is the most important aspect in the development of any transit signal priority strategies. This research proposed a novel and simple way of approximating such impacts using the concept of signal deviation.
- The accuracy of bus arrival time prediction and advance planning for signal adjustments are the keys for efficient and effective priority strategies. Conventional TSP strategies and models normally can only make very late planning to ensure low uncertainty about bus arrival time. A stochastic model (SMINP) was developed in this research to explicitly account for uncertain arrival time so that planning for timing adjustment can be advanced much earlier.
- One main issue revolves TSP is the handling of near-side bus stop where queue and bus have heavy interactions before entry and after exit of the bus stop. This research developed a queue delay algorithm, which quantifies the delay brought

about by the interactions. And the algorithm, in turn, enables the adaptive TSP model to explicitly minimize this interaction whenever necessary.

- The connected vehicle technology which can provide a much richer data set that enhances the model capabilities is explored. First, it enables the proper handling of multiple buses when the detections of their arrivals are not at the same time. Different from the traditional point detection system, bus arrival time can be predicted and re-predicted at any time within the communications range. So the control system can make sure all buses which haven't left the intersection or the corridor are not excluded in making a priority decision. Second, the adaptive TSP model is developed based on a much richer dataset, which include occupancy, route and schedule info and so on. This research showed these additional dataset helped improved the performance of the adaptive TSP model against the traditional active TSP strategy.
- Most of existing TSP research, active or adaptive, focuses on improving TSP performance locally. This research empirically demonstrated the significant divergence of two principles of granting priority on a signal corridor: a) granting TSP locally, and b) granting TSP on a corridor level. The simulation studies on the arterial showed granting TSP locally is difficult to achieve schedule adherence across multiple intersections. This result points to the need to that corridor-level active or adaptive TSP strategies should be implemented if transit reliability performance is to be achieved. The RTSP model developed in this research is one of such models.

- For any adaptive TSP model to be used in a real-world environment, real-time capability is crucial. The key is re-optimization. This research developed an event-based rolling horizon scheme, which takes advantage of the monitoring capability of the connected vehicle technology. And this scheme, which works in conjunction with the proposed models, allows continuous re-optimization. The real-time signal control system with adaptive TSP strategy described in the simulation platform section can be used as a working prototype for real-world implementation.

8.2. Key Findings

There are two major findings in this research. First, the comparison study of the SMINP model and the RBC-TSP strategy confirmed that advance planning for signal adjustment can yield much higher benefits to bus while limiting the impacts to other traffic. Second, the comparison between RTSP and LTSP models showed that granting TSP locally is very difficult to achieve schedule adherence on the arterial level. Other findings are listed in the following:

- A numerical experiment to evaluate the effects of priority weighting factor on the performance of the SMINP model was conducted on a hypothetical intersection. The results indicated that the user could adjust the priority to the bus by solely changing the priority weighting factor from 0.1 to 10. It also showed the model has the ability to prevent the user from using a priority level that is too high to cause oversaturation to the intersection.

- Comparison analyses were performed to compare the proposed control model SMINP with the state-of-the-practice active TSP strategy. In the case of no competing bus routes, the SMINP resulted in as much as 30% improvement of bus delay in low to medium congestion conditions. The comparison also indicated that the SMINP model can recognize the level of congestion of the intersection and automatically give less priority to the bus so as to minimize impact to the traffic on conflicting phases. In the case when there are three competing bus routes, SMINP handles multiple bus priority much better. The SMINP automatically adjust the relative importance of bus priority without the need to manually change the priority weighting factor, and it provides more balanced timings for both bus and the general traffic.
- In the arterial simulation study, it is found that penalizing timing deviation from the start and end points of the coordinated phases can help greatly limit the delays of the main-street traffic incurred by providing priorities to buses. Secondly, the R-TSP models has the ability to yield much higher bus delay reduction than that of the L-TSP models even when timing adjustability is very limited. In some instance, 65% of reduction in bus delay was produced by an R-TSP model compared to 37% reduction by an L-TSP model.
- Another more remarkable ability of the R-TSP models is bus schedule regulation. When the objective is to reduce schedule lateness, an R-TSP model can easily

produce as much as 98.6% on-time or early arrival with as little as 5.5% delay increase on other traffic. As a contrast, an L-TSP model has to use more than 12% delay increase on other traffic in exchange of no more than 90% bus schedule adherence. The comparison outcome is similar when the objective is to minimize schedule deviation. It can be easily concluded that the route-based TSP models are superior to the localized TSP models.

8.3. Future Research

Although a real-time control system with adaptive TSP models was prototyped and encouraging results were obtained in this research, there are still many interesting aspects can be further explored:

- Another important future research is the robustness of the R-TSP models or, in the other words, the susceptibility of the R-TSP models to the stochastic nature of the traffic system. In this study, the stochasticity of traffic conditions had been tested in the form of random seed numbers and long simulation periods. However, buses need to stop at bus stops for passenger loading and unloading, which time is subject to larger randomness. It is hypothesized that performance of the R-TSP models will degrade if more randomness is introduced. The rate of degradation in model performance can serve as a good indicator for the model robustness. If the robustness of the R-TSP model is low, a stochastic R-TSP model may need to be developed to explicitly account for the dwell time randomness.

- Alternatively, another way of handling stochastic traffic condition is to extend the stochastic formulation to multiple intersections. This may be the nature next step because 1) that optimizing corridor-level schedule-related performance requires corridor-level priority models, and 2) that it is feasible and necessary to explicitly account for variable bus dwell time using stochastic formulation. However, the large-scale nature of any stochastic programming approach may cause difficulty in finding optimal solutions. A branch-and-cut algorithm based on disjunctive decomposition technique [86] may be needed to provide optimal solutions.
- One important aspect of the R-TSP model requires further investigation is relationship of the effective range of the multipliers for the terms in the objective function and the number of intersections. In this research, the sensitivity analysis established the range of the multipliers for term *x* and *d* in which the model operator may make changes to influence the model outcome. This is good to allow some forms of inputs by the system manager to give more or less *a priori* priorities to certain bus lines. But the authors noticed that the effective range seems to tie to the number of intersections and the number of conflicting phases. Analytical models may be developed to give better guidance in determining the effective range.

- Another possible extension is to integrate the model with an adaptive signal control system where additional information about the development of vehicle queues at an approach can be estimated in real-time. The additional information relaxes the assumption about constant vehicle arrival and further improves the ability of the SMINP to predict the arrival time distribution of the bus to the stop bar.
- A better queue prediction model could help give a better starting point for drawing the queue diagram in the queue delay algorithm. This change may be incremental, but it may result in more consistent performance by the R-TSP model. In some cases, the author observed that the some buses couldn't pass through the intersection as planned in the model because the initial queue was much longer than expected. Connected vehicle technology can serve as a better detection mechanism to estimate the initial queue [87].
- A more sophisticated prediction models for predicting bus arrival time at each intersection can be employed to replace the naïve path project approach developed in this research. The multi-class cell-transmission model (M-CTM) developed by [88] appears to be a good candidate due to its efficiency of making predictions for traffic with different speeds, such as bus and passenger car. Also the M-CTM could account for platoon dispersion, which is typical on a long stretch of signal arterial.
- The RTSP model is designed to optimize any bus routes. However, this research only tested routes that are on a linear coordinated signal arterial. In a real-life

setting, different bus routes may get on only different portion of the arterial. In theory, there is no obvious reason why RTSP model cannot work under those circumstances. But a rigorous simulation study can provide better understanding if any adjustments to the RTSP model can better work in those cases.

REFERENCES

[1] D. Schrank, B. Eisele, and T. Lomax. *2012 Urban Mobility Report.* Texas A&M Transportation Institute, College Station, Texas, 2012.

[2] D. Jackson, Z. W., M., S. Peirce *et al.*, "Urban partnership proposals: Review of domestic and international deployments and transit impacts from congestion pricing," in the 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.

[3] A. R. Danaher. *TCRP Synthesis 83: Bus and Rail Transit Preferential Treatments in Mixed Traffic.* Transportation Research Board, Washington, D.C., 2010.

[4] X. Zeng, K. N. Balke, and P. Songchitruksa. *Potential Connected Vehicle Applications to Enhance Mobility, Safety, and Environmental Security.* Southwest Region University Transportation Center, Texas Transportation Institute, Texas A&M University System, College Station, Texas, 2012.

[5] H. Evans, and G. Skiles. (1970). Improving Public Transit through Bus Preemption of Traffic Signals. *Traffic Quarterly*. *vol.* 24 *(4)*, pp. 531-543.

[6] R. J. Baker, J. Collura, J. J. Dale *et al. An Overview of Transit Signal Priority - Revised and Updated.* ITS America, Washington, D.C., 2004.

[7] *National Transportation Communications for ITS Protocol: Object Definitions for Signal Control and Prioritization.* AASHTO, ITE, NEMA, Washington, D.C., 2008.

[8] P. V. Allen, "Jumping The Queue - Implementing Transit Signal Priority and Queue Jumpers for Bus Rapid Transit," in ITE 2012 Annual Meeting & Exhibit, Atlanta, Georgia, 2012, pp. 1-21.

[9] G. Zhou, and A. Gan. (2009). Design of Transit Signal Priority at Signalized Intersections with Queue Jumper Lanes. *Journal of Public Transportation*. *vol.* 12 *(4)*, pp. 117-132.

[10] P. G. Furth, and T. H. J. Muller. (2000). Conditional Bus Priority at Signalized Intersections: Better Service with Less Traffic Disruption. *Transportation Research Record: Journal of the Transportation Research Board (1731),* pp. 23-30.

[11] Kittleson & Associate, KFH Group Inc, Parsons Brinckhoff Quade & Douglass *et al. Transit Capacity and Quality of Service Manual, 2nd Edition.* Transportation Research Board, Washington, D.C., 2003.

[12] H. R. Smith, B. Hemily, and M. Ivanovic. *Transit Signal Priority (TSP): A Planning and Implementation Handbook.* United States Department of Transportation, Washington, D. C., 2005.

[13] T. Urbanik. (1977). Priority Treatment of Buses at Traffic Signals. *Transportation Engineering*, pp. 31-33.

[14] F. M. Oliveira-Neto, C. F. G. Loureiro, and L. D. Han. (2009). Active and Passive Bus Priority Strategies in Mixed Traffic Arterials Controlled by SCOOT Adaptive Signal System: Assessment of Performance in Fortaleza, Brazil. *Transportation Research Record: Journal of the Transportation Research Board (2128),* pp. 58-65.

[15] Y. Jeong, and Y. Kim, "Tram Signal Priority Strategy Using Bandwidth Model," in 19th ITS World Congress, Vienna , Austria, 2012, pp. 5.

[16] Z. Zuo, and G. Yang. (2013). Feasibility of Unconditional Transit Signal Priority Considering Delay Savings at Signalized Intersections: A Case Study of Dalian BRT Line No.1. *Procedia - Social and Behavioral Sciences*. *vol.* 96, pp. 828-837.

[17] H. Xu, J. Sun, and M. Zheng. (2010). Comparative Analysis of Unconditional and Conditional Priority for Use at Isolated Signalized Intersections. *ASCE Journal of Transportation Engineering*. *vol.* 136 *(12),* pp. 1092-1103.

[18] A. Skabardonis. (1998). Control Strategies for Transit Priority. *Intellimotion*. *vol.* 7 *(2),* pp. 6-7, 10-11.

[19] W. Ma, X. Yang, and Y. Liu. (2010). Development and Evaluation of a Coordinated and Conditional Bus Priority Approach. *Transportation Research Record: Journal of the Transportation Research Board (2145)*, pp. 49-58.

[20] K. Balke, "Development and Laboratory Testing of An Intelligent Approach for Providing Priority to Buses at Traffic Signalized Intersections ", Zachry Department of Civil Engineering, The Texas A&M University, College Station, TX, 1998.

[21] Econolite, "Transit Signal Priority (TSP) User Guide for Advanced System Controller," Econolite Control Products, Inc., June, 2009.

[22] M. Zlatkovic, A. Stevanovic, and P. T. Martin. (2012). Development and Evaluation of Algorithm for Resolution of Conflicting Transit Signal Priority Requests. *Transportation Research Record: Journal of the Transportation Research Board (2311)*, pp. 167–175.

[23] M. Li, Y. Yin, W.-B. Zhang *et al.* (2011). Modeling and Implementation of Adaptive Transit Signal Priority on Actuated Control Systems. *Computer-Aided Civil and Infrastructure Engineering*. *vol.* 26 *(4)*, pp. 270-284.

[24] E. Christofa, and A. Skabardonis. (2011). Traffic Signal Optimization with Application of Transit Signal Priority to an Isolated Intersection. *Transportation Research Record: Journal of the Transportation Research Board (2259)*, pp. 192-201.

[25] J. Stevanovic, A. Stevanovic, P. T. Martin *et al.* (2008). Stochastic optimization of traffic control and transit priority settings in VISSIM. *Transportation Research Part C: Emerging Technologies*. *vol.* 16 *(3)*, pp. 332-349.

[26] W. Ma, Y. Liu, and X. Yang. (2013). A Dynamic Programming Approach for Optimal Signal Priority Control Upon Multiple High-Frequency Bus Requests. *Journal of Intelligent Transportation Systems*. *vol.* 17 *(4)*, pp. 282-293.

[27] Q. He, K. L. Head, and J. Ding. (2012). PAMSCOD: Platoon-based arterial multi-modal signal control with online data. *Transportation Research Part C: Emerging Technologies*. *vol.* 20 *(1)*, pp. 164-184.

[28] A. Skabardonis, and N. Geroliminis. (2008). Real-Time Monitoring and Control on Signalized Arterials. *Journal of Intelligent Transportation Systems*. *vol.* 12 *(2)*, pp. 64-74.

[29] G. Wu, L. Zhang, W.-b. Zhang *et al.* (2012). Signal Optimization at Urban Highway Rail Grade Crossings Using an Online Adaptive Priority Strategy. *ASCE Journal of Transportation Engineering*. *vol.* 138 *(4)*, pp. pp 479-484.

[30] Y. Lin, X. Yang, L. Jia *et al.* (2013). Development of Model-based Transit Signal Priority Control for Local Arterials. *Procedia - Social and Behavioral Sciences*. *vol.* 96, pp. 2344-2353.

[31] N. Hounsell, and B. Shrestha. (2012). A New Approach for Co-Operative Bus Priority at Traffic Signals. *IEEE Transactions on Intelligent Transportation Systems*. *vol.* 13 *(1)*, pp. 6-14.

[32] K. Ling, and A. Shalaby. (2004). Automated Transit Headway Control via Adaptive Signal Priority. *Journal of Advanced Transportation*. *vol.* 38 *(1)*, pp. 45-67.

[33] M. Vasudevan, "Robust Optimization Model for Bus Priority under Arterial Progression," University of Maryland, College Park, Maryland, 2005.

[34] M. Tlig, and N. Bhouri. (2011). A Multi-Agent System for Urban Traffic and Buses Regularity Control. *Procedia - Social and Behavioral Sciences*. *vol.* 20, pp. 896-905.

[35] E. Albright, and M. Figliozzi. (2012). Factors Influencing Effectiveness of Transit Signal Priority and Late-Bus Recovery at Signalized-Intersection Level. *Transportation Research Record: Journal of the Transportation Research Board (2311)*, pp. 186–194.

[36] M. S. Ghanim, and G. Abu-Lebdeh. (2014). Real-Time Dynamic Transit Signal Priority Optimization for Coordinated Traffic Networks Using Genetic Algorithms and Artificial Neural Networks. *Journal of Intelligent Transportation Systems (Available Online Only)*.

[37] Q. He, “Robust-Intelligent Traffic Signal Control within a Vehicle-To-Infrastructure And Vehicle-To-Vehicle Communication Environment,” Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ, 2010.

[38] F. Dion, and M. Ghanim, “Impact of Dwell Time Variability on Transit Signal Priority Performance at Intersections with Nearside Bus Stop,” in Transportation Research Board 86th Annual Meeting, Washington, D.C., 2007.

[39] B. E. Chandler, M. C. Myers, J. E. Atkinson *et al. Signalized Intersections Informational Guide.* Report No. FHWA-SA-13-027, Federal Highway Administration, Washington, D.C., 2013.

[40] P. Koonce, L. Rodegerdts, K. Lee *et al. Traffic Signal Timing Manual.* Report No. FHWA-HOP-08-024, Federal Highway Administration, Washington, D.C., 2008.

[41] S. Andrews, and M. Cops. *Vehicle Infrastructure Integration Proof of Concept - Vehicle.* Report No. FHWA-JPO-09-003, FHWA-JPO-09-017, FHWA-JPO-09-043, US DOT Research and Innovative Technology Administration, Washington, D.C., 2009.

[42] Society of Automobile Engineers. "Dedicated Short Range Communications (DSRC) Message Set Dictionary," http://standards.sae.org/j2735_200911.

[43] B. L. Smith, R. Venkatanarayana, H. Park *et al. IntelliDrive*^{SM} *Traffic Signal Control Algorithms: Task 2: Development of New Traffic Control Signal Algorithms under IntelliDrive.* University of Virginia, Charlottesville, Virginia, 2010.

[44] E. Christofa, I. Papamichail, and A. Skabardonis. (2013). Person-Based Traffic Responsive Signal Control Optimization. *Intelligent Transportation Systems, IEEE Transactions on*. *vol.* 14 *(3)*, pp. 1278-1289.

[45] T. Chin-Woo, P. Sungsu, L. Hongchao *et al.* (2008). Prediction of Transit Vehicle Arrival Time for Signal Priority Control: Algorithm and Performance.

Intelligent Transportation Systems, IEEE Transactions on. *vol.* 9 *(4)*, pp. 688-696.

[46] F. W. Cathey, and D. J. Dailey. (2003). A Prescription for Transit Arrival/Departure Prediction Using Automatic Vehicle Location Data. *Transportation Research Part C: Emerging Technologies*. *vol.* 11 *(3–4)*, pp. 241-264.

[47] D. Dailey, S. Maclean, F. Cathey *et al.* (2001). Transit Vehicle Arrival Prediction: Algorithm and Large-Scale Implementation. *Transportation Research Record: Journal of the Transportation Research Board*. *vol.* 1771, pp. 46-51.

[48] S. I.-J. Chien, Y. Ding, and C. Wei. (2002). Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*. *vol.* 128 *(5)*, pp. 429-438.

[49] PTV America, "Ring Barrier Controller User Manual," PTV America, August 2010, 77 p.

[50] G. B. Dantzig. (1955). Linear Programming under Uncertainty. *Management Science*. *vol.* 1 *(3-4)*, pp. 197-206.

[51] E. M. L. Beale. (1955). On Minimizing a Convex Function Subject to Linear Inequalities. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 173-184.

[52] A. S. Kenyon, and D. P. Morton. (2003). Stochastic Vehicle Routing with Random Travel Times. *Transportation Science*. *vol.* 37 *(1)*, pp. 69-82.

[53] A. R. Ferguson, and G. B. Dantzig. (1956). The Allocation of Aircraft to Routes—An Example of Linear Programming under Uncertain Demand. *Management science*. *vol.* 3 *(1)*, pp. 45-73.

[54] A. Charnes, W. W. Cooper, and G. H. Symonds. (1958). Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management science*. *vol.* 4 *(3)*, pp. 235-263.

[55] J. R. Birge, and F. V. Louveaux. *Introduction to Stochastic Programming*, p. 421, Springer, New York, New York, 1997.

[56] "Highway Capacity Manual 2010," Transportation Research Board, National Research Council, Washington, D.C., 2010.

[57] L. Head, D. Gettman, and Z. Wei. (2006). Decision Model for Priority Control of Traffic Signals. *Transportation Research Record: Journal of the Transportation Research Board (1978)*, pp. 169-177.

[58] "NEMA Standards Publication TS 2-2003: Traffic Controller Assemblies with NTCIP Requirements ", National Electrical Manufacturers Association, 2003.

[59] K. Yin, Y. Zhang, and B. X. Wang. (2010). Analytical Models for Protected plus Permitted Left-Turn Capacity at Signalized Intersection with Heavy Traffic. *Transportation Research Record: Journal of the Transportation Research Board (2192)*, pp. 177-184.

[60] P. Mirchandani, A. Knyazyan, L. Head *et al.* (2001). An Approach Towards the Integration of Bus Priority, Traffic Adaptive Signal Control, and Bus Information/Scheduling Systems. *Computer-Aided Scheduling of Public Transport*, pp. 319-334.

[61] Y. Lin, X. Yang, G.-L. Chang *et al.* (2013). Transit Priority Strategies for Multiple Routes Under Headway-Based Operations. *Transportation Research Record: Journal of the Transportation Research Board (2356)*, pp. 34–43.

[62] M. Conrad, F. Dion, and S. Yagar. (1998). Real-Time Traffic Signal Optimization with Transit Priority: Recent Advances in the Signal Priority Procedure for Optimization in Real-Time Model. *Transportation Research Record: Journal of the Transportation Research Board (1634)*, pp. 100-109.

[63] S. Yagar, and B. Han. (1994). A Procedure for Real-Time Signal Control That Considers Transit Interference and Priority. *Transportation Research Part B: Methodological*. *vol.* 28 *(4)*, pp. 315-331.

[64] C. d. Taranto. (2007). Absolute Priority: Providing the Ability to Monitor Real-Time Traffic Flow Across the Whole Network. *Traffic Technology International*, pp. 100-101.

[65] M. Fellendorf, "Public Transport Priority Within SCATS - A Simulation Case Study in Dublin," in Institute of Transportation Engineers 67th annual Meeting, Boston, Massachusetts, 1997.

[66] M. Figliozzi, C. Monsere, C. Slavin *et al. Evaluation of the Performance of the Sydney Coordinated Adaptive Traffic System (SCATS) on Powell Boulevard in Portland, OR.* Report No. OTREC-RR-13-07, Portland State University, Portland, Oregon, 2013.

[67] N. B. Hounsell, B. P. Shrestha, J. R. Head *et al.* (2008). The Way Ahead for London's Bus Priority at Traffic Signals. *IET Intelligent Transport Systems*. *vol.* 2 *(3)*, pp. 193-200.

[68] K. Ling, and A. Shalaby. (2003). Automated Transit Headway Control via Adapative Signal Priority. *Journal of Advanced Transportation*. *vol.* 38 *(1)*, pp. 45-67.

[69] Y. Wadjas, and P. G. Furth. (2003). Transit Signal Priority Along Arterials Using Advanced Detection. *Transportation Research Record: Journal of the Transportation Research Board (1856)*, pp. 220-230.

[70] N. A. Chaudhary, V. G. Kovvali, and S. M. M. Alam. *Guidelines for Selecting Signal Timing Software.* Report No. FHWA/TX-03/0-4020-P2, Texas Transportation Institute, College Station, TX, 2002.

[71] "SYNCHRO 6 User Guide," Trafficware. Sugar Land, TX, 2004.

[72] C. E. Wallace, K. Courage, D. Reaves *et al. TRANSYT-7F user's manual.* University of Florida, Gainesville, Florida, 1984.

[73] E. Chang, and C. J. Messer, "Arterial Signal Timing Optimization Using Passer II-90-Program User's Manual," Texas Transportation Institute, 1991.

[74] X. Zeng, Y. Zhang, K. N. Balke *et al.* (2014). A Real-Time Transit Signal Priority Control Model Considering Stochastic Bus Arrival Time. *Intelligent Transportation Systems, IEEE Transactions on*. *vol.* 15 *(4)*, pp. 1-10.

[75] C.-F. Liao, G. A. Davis, and R. Atherley. (2007). Simulation Study of A Bus Signal Priority Strategy Based on GPS/AVL and Wireless Communications. *Transportation Research Record: Journal of the Transportation Research Board (2034)*, pp. 82-91.

[76] J. Liu, B. Cai, Y. Wang *et al.* (2013). Generating Enhanced Intersection Maps for Lane Level Vehicle Positioning based Applications. *Procedia - Social and Behavioral Sciences*. *vol.* 96, pp. 2395-2403.

[77] N. H. Gartner. (1982). Prescription for Demand-Responsive Urban Traffic Control. *Transportation Research Record: Journal of the Transportation Research Board (881)*, pp. 73-76.

[78] H. Liu, W.-H. Lin, and C.-W. Tan. (2007). Operational Strategy for Advanced Vehicle Location System–Based Transit Signal Priority. *ASCE Journal of Transportation Engineering*. *vol.* 133 *(9)*, pp. 513-522.

[79] C.-W. Tan, S. Park, H. Liu *et al.* (2008). Prediction of Transit Vehicle Arrival Time for Signal Priority Control: Algorithm and Performance. *IEEE Transactions on Intelligent Transportation Systems*. *vol.* 9 *(4)*, pp. 688-696.

[80] F. Dion, H. Rakha, and Y. Zhang. (2004). Evaluation of Potential Transit Signal Priority Benefits along a Fixed-Time Signalized Arterial. *ASCE Journal of Transportation Engineering*. *vol.* 130 *(3)*, pp. 294-303.

[81] "VISSIM 5.40-09 - User Manual," PTV Vision, 2013.

[82] W. Kim, and L. R. Rilett. (2005). Improved Transit Signal Priority System for Networks with Nearside Bus Stops. *Transportation Research Record: Journal of the Transportation Research Board (1925)*, pp. 205-214.

[83] R. L. Bertini, and A. M. El-Geneidy. (2004). Modeling Transit Trip Time Using Archived Bus Dispatch System Data. *ASCE Journal of Transportation Engineering*. *vol.* 130 *(1)*, pp. 56-67.

[84] F. Li, Z. Duan, and D. Yang. (2012). Dwell Time Estimation Models for Bus Rapid Transit Stations. *Journal of Modern Transportation*. *vol.* 20 *(3)*, pp. 168-177.

[85] S. Rashidi, and P. Ranjitkar. (2013). Approximation and Short-Term Prediction of Bus Dwell Time using AVL Data. *Journal of the Eastern Asia Society for Transportation Studies*. *vol.* 10, pp. 1281-1291.

[86] L. Ntaimo, and S. Sen. (2007). A Branch-and-Cut Algorithm for Two-Stage Stochastic Mixed-Binary Programs with Continuous First-Stage Variables. *International Journal of Computational Science and Engineering*. *vol.* 3 *(3)*, pp. 232-241.

[87] Y. Zhang, and K. Tiaprasert. *Enhanced Adaptive Signal Control Using Dedicated Short-Range Communications.* Report No. SWUTC/14/600451, Texas A&M Transportation Institute, College Station, Texas, 2014.

[88] K. Tuerprasert, and C. Aswakul. (2010). Multiclass Cell Transmission Model for Heterogeneous Mobility in General Topology of Road Network. *Journal of Intelligent Transportation Systems*. *vol.* 14 *(2)*, pp. 68-82.

APPENDIX A

BUS SCHEDULE PERFORMANCE ANALYSIS

Table 16: Bus Schedule Performance using TSP_S7 Family Models (WB).

	WB Bus Schedule Delay (second)														
	Seed = 35			Seed = 41			Seed = 47			Seed = 53			Seed = 59		
Bus Veh ID	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay
1	10	-30	-72	6	-30	-68	-2	-22	-74	-2	-34	-74	2	-32	-74
520	10	-22	-68	12	-22	-66	14	-10	-68	0	-18	-72	4	-22	-62
1079	4	-48	-66	0	-50	-74	0	-10	-102	0	-46	-64	2	-10	-66
1589	-2	-12	-8	2	-12	-80	12	-12	-56	14	-48	-76	-2	-12	-76
2165	-2	-34	-56	6	-28	-32	2	-38	-42	0	-38	-88	4	-28	-20
2721	-2	-72	-62	4	-68	10	8	-70	-82	8	-68	-64	4	-70	-72
3267	12	-14	-60	4	-14	-60	4	-12	-58	8	-16	-66	-2	-14	-58
3822	2	-4	-48	0	-12	-50	-2	-10	-54	0	-10	-104	0	-10	-102
4389	-2	-6	-6	-2	-6	-58	-2	-12	-56	-2	-12	-54	12	-12	-46
4922	16	-26	-76	4	-20	-54	0	-26	-82	6	-28	-80	-2	-28	-82
5477	-2	-28	-62	4	-30	-68	2	-30	-78	6	-30	-76	2	-30	-72
6031	-4	-14	-66	-4	-14	-60	-4	-16	-64	6	-14	-66	-2	-16	-68
6587	0	-10	-74	-2	-12	-72	0	-12	-104	0	-4	-74	-2	-10	-100
7125	0	4	-74	2	-12	-70	4	-12	-78	4	-12	-66	16	-12	-52
7675	0	-30	-80	2	-42	-84	66	-24	-54	6	-30	-46	4	-30	-34
8214	6	-28	-72	4	-30	-72	8	-28	-70	-2	-26	-62	8	-30	-56
8760	12	-4	-26	-4	-22	-36	-2	-20	-64	0	-22	-58	0	-22	-68
9293	0	-10	-62	0	-10	-74	0	-14	-52	-2	-6	-66	0	-12	-80
9833	-2	-48	-56	4	-12	-86	4	-12	-76	6	-12	-62	-2	-8	-40
10348	12	-24	-82	0	-30	-40	8	-30	-88	-2	-42	-90	-2	-34	-80
10893	2	-28	-64	-2	-32	-80	10	-30	-80	-2	-72	-72	4	-70	-80
11466	-2	-14	-68	0	-8	-58	4	-12	-66	-2	-12	-70	4	-16	-72
12025	-2	-10	-100	-2	50	-102	0	-10	-102	2	-8	-46	4	-2	-54
12569	-2	-8	-46	0	-12	-44	-2	-4	-6	10	-12	-46	-2	-12	-40
13115	2	-26	-82	2	-26	-58	2	-28	-54	52	-30	-56	0	-32	-88
13670	14	32	-28	10	-28	-94	-2	-30	-62	4	-28	-64	2	-30	-76
14248	4	-14	-10	10	-16	-72	4	-14	-76	4	-14	-60	6	-8	-68
14808	0	-12	-58	-2	-12	-48	2	-10	-58	0	-10	-56	0	-12	-52
15350	2	-12	-56	4	-12	-76	-2	-12	-68	6	-12	-56	16	-12	-76
15899	0	-34	-88	2	-34	-82	2	-26	-78	58	-32	-40	2	-32	-20

* positive values indicate vehicle is late; negative values indicate vehicle is early.
Cell shared as red indicate vehicle being late

Table 17: Bus Schedule Performance using TSP_S7 Family Models (EB).

Bus Veh ID	Seed = 35			Seed = 41			Seed = 47			Seed = 53			Seed = 59		
	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay
2	-22	-36	-48	2	-34	-44	-26	-40	-50	2	-40	-44	0	-34	-44
814	0	-54	-58	-2	-52	-54	-2	0	-56	2	-8	-12	-2	-52	-54
1649	40	-54	-50	0	-52	-38	-2	-56	-58	0	-56	-62	-2	-10	-40
2468	-4	-16	-12	90	-16	24	-4	-16	-14	0	-12	-10	-4	-16	-14
3318	18	-72	-74	6	-28	-72	20	-72	-76	20	-76	-76	20	-76	-78
4138	0	-36	-38	0	-36	-44	0	-36	-34	2	-36	-36	0	-36	-44
4963	-2	-52	10	-2	-48	-62	6	-52	12	4	-50	14	2	-48	6
5813	-4	-54	-52	-2	-52	-60	-4	-54	-56	-22	-52	-66	-2	-52	-66
6632	0	-16	-20	-2	-12	-32	0	-12	-30	0	-16	-30	2	-12	-26
7459	18	-40	-38	18	-38	-40	20	-36	-36	18	-36	-34	16	-36	-70
8254	0	-28	-34	6	-32	-40	-2	-32	-44	-2	-32	-36	-2	-32	-42
9084	0	-8	-10	-2	-46	-42	0	-44	-8	-2	-44	-8	-2	2	-52
9894	-26	-52	-70	-4	-56	-46	-26	-10	-62	-4	-10	-56	0	-48	-50
10753	-2	-12	-12	-2	-16	-10	0	-16	-14	0	-12	-18	-4	-12	-18
11596	16	-28	-76	16	-32	-68	16	-28	-70	22	-70	-68	18	-30	-68
12387	2	-36	-46	0	-36	-44	2	-36	-44	2	-32	-36	0	-36	-42
13190	-2	-50	-8	2	-2	-56	0	-44	-66	0	-46	-50	-2	-50	8
14055	0	-50	-58	-4	-52	-48	0	-54	-58	-4	-56	-66	-2	-52	-64
14888	0	-10	-24	0	-10	-32	-4	-12	-26	0	-12	-28	4	-10	-24
15660	20	-36	-76	18	-38	-40	16	-36	-38	18	-34	-74	20	-32	-38

EB Bus Schedule Delay (second)

* positive values indicate vehicle is late; negative values indicate vehicle is early.
Cell shared as red indicate vehicle being late

Table 18: Bus Schedule Performance using TSP_L7 Family Models (WB).

	WB Bus Schedule Delay (second)														
	Seed = 35			Seed = 41			Seed = 47			Seed = 53			Seed = 59		
Bus Veh ID	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay	Deviation	Lateness	Delay
1	108	60	60	108	62	62	102	56	56	60	108	108	108	60	60
520	22	22	22	22	22	22	22	20	20	24	20	20	22	68	68
1079	32	-52	-52	-24	-16	-16	34	-16	-18	-18	-54	-52	36	82	82
1589	44	44	46	44	44	90	42	40	40	46	42	42	46	44	94
2165	2	2	2	66	4	4	58	2	2	2	2	2	2	4	2
2721	102	66	66	104	102	102	102	66	66	66	108	66	66	66	66
3267	22	24	22	22	24	78	24	24	24	22	22	22	24	24	24
3822	-14	-14	-14	-16	-16	-16	38	-16	56	-16	-16	-16	-18	56	-18
4389	90	84	84	42	46	46	46	46	84	42	46	46	44	46	88
4922	2	2	2	56	52	4	4	4	4	2	4	46	2	2	2
5477	116	64	64	62	62	62	62	62	62	62	62	64	62	62	62
6031	26	26	26	66	24	24	22	22	22	24	22	22	24	66	66
6587	-16	-16	-16	-18	-18	-18	26	-18	-16	-16	26	54	-18	-18	38
7125	42	46	42	42	42	42	44	42	42	44	42	42	44	42	90
7675	-34	2	2	-34	2	2	-34	2	60	-36	4	2	-32	4	4
8214	138	114	64	106	60	60	134	118	62	106	62	112	106	60	60
8760	26	22	74	24	22	24	70	26	78	26	24	24	22	24	26
9293	30	-16	-16	-20	-14	-14	-42	36	-16	-24	80	82	-20	-18	-18
9833	46	44	44	44	100	42	42	42	42	42	42	42	92	44	44
10348	74	2	2	2	48	2	2	2	2	4	2	4	2	2	2
10893	66	64	64	102	114	66	66	102	102	66	66	66	66	66	66
11466	22	22	22	26	70	26	22	22	22	82	24	78	22	82	22
12025	38	-16	-16	-18	-20	-20	-18	58	54	-14	-14	-14	32	58	-14
12569	46	46	96	46	12	108	44	46	108	46	94	84	44	84	84
13115	6	46	6	4	4	6	4	6	4	2	2	2	60	2	2
13670	110	66	66	62	62	62	64	62	62	64	64	64	110	64	64
14248	22	22	24	22	24	22	22	22	22	24	22	66	24	84	80
14808	-18	-18	-18	-18	-16	-18	32	-14	-16	-16	-16	-16	34	-14	-16
15350	42	94	42	42	88	42	44	42	42	42	44	42	44	44	94
15899	-32	0	0	-36	2	2	46	0	0	46	2	56	46	2	4

* positive values indicate vehicle is late; negative values indicate vehicle is early.
Cell shared as red indicate vehicle being late

Table 19: Bus Schedule Performance using TSP_L7 Family Models (EB).

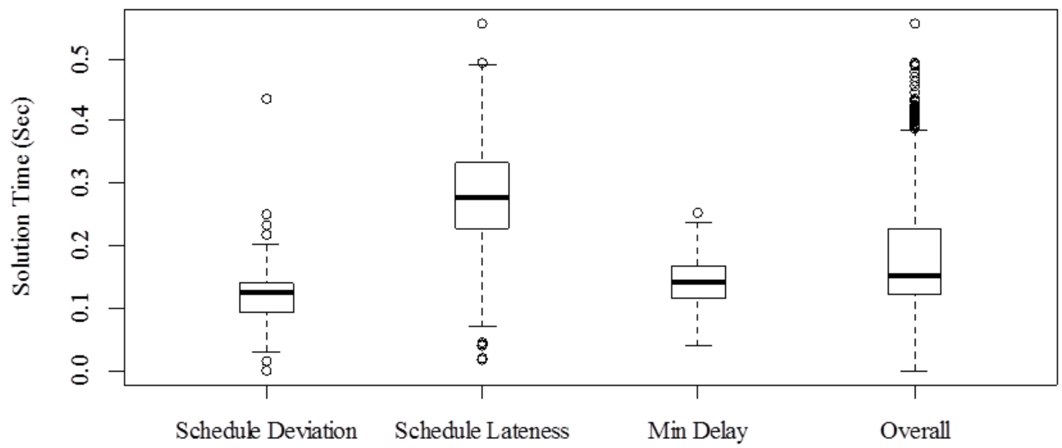
Bus Veh ID	Seed = 35 Deviation	Seed = 35 Lateness	Seed = 35 Delay	Seed = 41 Deviation	Seed = 41 Lateness	Seed = 41 Delay	Seed = 47 Deviation	Seed = 47 Lateness	Seed = 47 Delay	Seed = 53 Deviation	Seed = 53 Lateness	Seed = 53 Delay	Seed = 59 Deviation	Seed = 59 Lateness	Seed = 59 Delay
2	2	-38	-38	58	-32	-36	4	-38	-38	56	-36	-36	4	-38	-38
814	4	6	4	4	2	2	2	2	2	110	0	0	0	0	0
1649	44	-52	-52	40	-10	-52	-10	-52	-52	44	-56	-56	44	-52	-52
2468	-8	-20	-20	-10	-20	-20	-10	-20	-20	-10	-20	-20	-10	-20	-20
3318	24	24	24	24	24	24	24	28	24	24	28	28	24	28	28
4138	6	-32	-26	66	-32	-36	-30	-32	-32	8	-36	-36	-30	-32	-32
4963	10	6	6	10	6	6	12	14	14	6	14	14	10	10	10
5813	44	-56	-56	-10	-52	-52	-10	-56	-52	-44	-56	-56	44	-52	-52
6632	-10	-10	-14	-10	-12	-14	-10	-10	-10	-6	-10	-10	-6	-10	-10
7459	18	18	18	18	18	18	18	22	22	18	18	18	24	22	22
8254	-30	-34	-34	-22	-34	-30	8	-34	-34	-30	-32	-32	-24	-34	-30
9084	0	0	0	4	4	4	0	2	2	0	0	0	2	0	0
9894	-50	-56	-56	40	-54	-56	-50	-56	-56	44	-56	-56	40	-50	-56
10753	-10	-20	-16	-10	-18	-16	-10	-16	-16	-4	-20	-20	-10	-20	-20
11596	26	28	24	26	28	28	24	24	24	28	24	24	24	24	24
12387	-30	-32	-36	8	-30	-32	-24	-32	-32	-24	-30	-30	-24	-32	-36
13190	12	10	6	100	8	98	8	8	8	10	6	6	6	10	10
14055	48	-56	-52	-12	-56	-56	44	-52	-52	-48	-52	-52	48	-52	-56
14888	-10	-14	-10	22	-12	-8	-10	-10	-10	-10	-14	-14	-6	-8	-10
15660	18	18	22	22	18	22	24	18	18	18	18	18	22	22	22

EB Bus Schedule Delay (second)

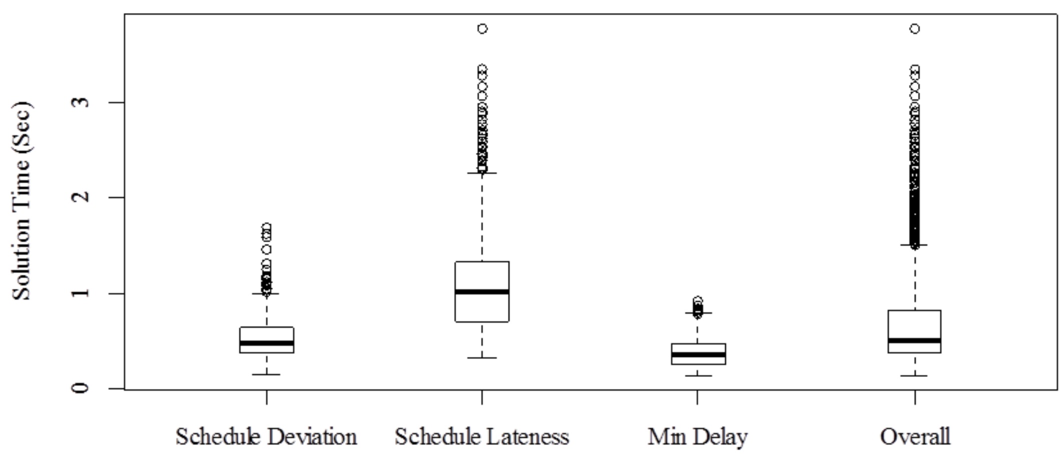
* positive values indicate vehicle is late; negative values indicate vehicle is early.
Cell shared as red indicate vehicle being late

APPENDIX B

SOLUTION TIME

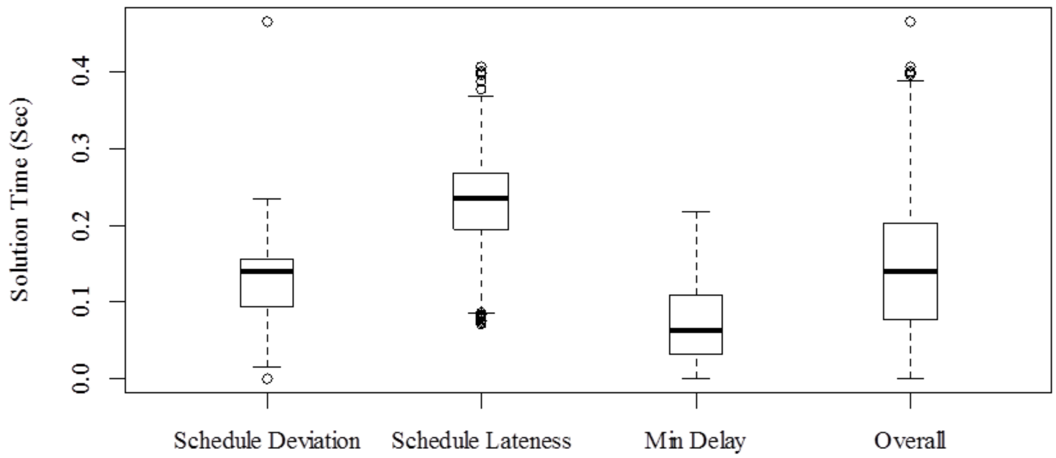


(a) TSP_L7 Model Family

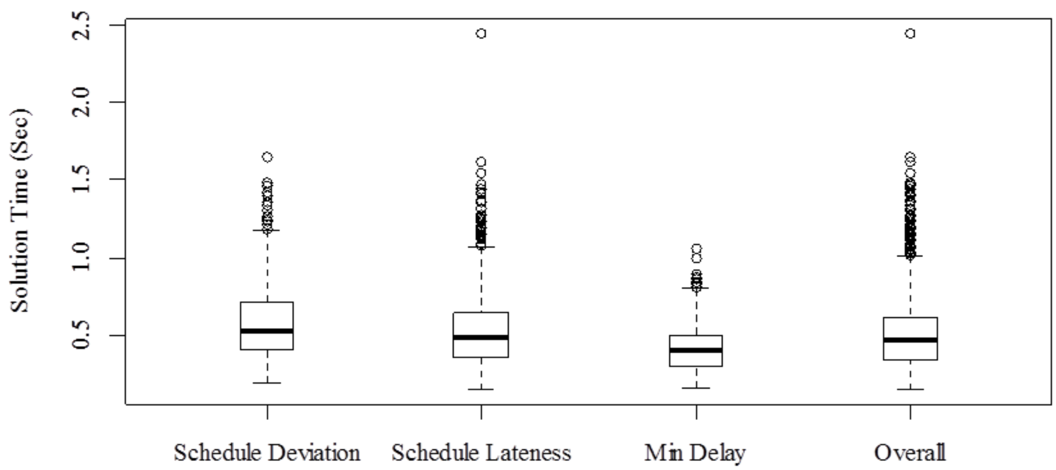


(b) TSP_S7 Model Family

Figure 39: Box Plots of Optimization Solution Time.



(c) TSP_L8 Model Family



(d) TSP_S8 Model Family

Figure 39 Continued.