

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Traffic Signal Optimization with Transit Priority: A Person-based Approach

Permalink

<https://escholarship.org/uc/item/0t4336f3>

Author

Christofa, Eleni

Publication Date

2012

Peer reviewed|Thesis/dissertation

**Traffic Signal Optimization with Transit Priority:
A Person-based Approach**

by

Eleni Christofa

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy
in

Engineering – Civil and Environmental Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Alexander Skabardonis, Chair

Professor Joan Walker

Professor Pravin Varaiya

Spring 2012

Traffic Signal Optimization with Transit Priority:
A Person-based Approach

Copyright © 2012

by

Eleni Christofa

Abstract

Traffic Signal Optimization with Transit Priority:
A Person-based Approach

by

Eleni Christofa

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Alexander Skabardonis, Chair

Traffic responsive signal control with Transit Signal Priority (TSP) is a strategy that is increasingly used to improve transit operations in urban networks. However, none of the existing real-time signal control systems have explicitly incorporated the passenger occupancy of transit vehicles in granting priority or have effectively addressed issues such as the provision of priority to transit vehicles traveling in conflicting directions at signalized intersections. The contribution of this dissertation is the development of a person-based traffic responsive signal control system with TSP that minimizes total person delay in a network by explicitly considering all vehicles' passenger occupancy and transit schedule delay. By using such conditions, the issue of assigning priority to transit vehicles traveling in conflicting directions is also addressed in an efficient way. In addition, the impact of these priority strategies on auto traffic is addressed by minimizing the total person delay in the network under consideration and assigning penalties for interrupting the progression of platoons on arterials. The system is first developed for isolated intersections, and then extended to arterial signalized networks. Evaluation tests for a wide range of traffic and transit operating characteristics show that significant reductions in transit passenger delay can be achieved without substantially increasing auto passenger delay. Furthermore, the system achieves lower vehicle delays compared to signal settings obtained by state-of-the-art signal optimization software. Finally, it utilizes readily deployable technologies, which provide real-time information such as sensors, Automated Vehicle Location and Automated Passenger Counter systems and can be implemented on existing infrastructure in urban multimodal networks.

To my parents, grandparents, and brother,
for their unconditional love and support that have shaped who I am today

Contents

Contents	ii
List of Figures	iv
List of Tables	vi
Acknowledgments	vii
1 Introduction	1
1.1 Motivation	1
1.2 Research Question	2
1.3 Research Contribution	2
1.4 Dissertation Organization	3
2 Literature Review	4
2.1 Transit Signal Priority (TSP)	4
2.2 Signal Coordination with TSP	10
2.3 TSP Implementation for Conflicting Transit Routes	11
2.4 Real-Time Signal Control Systems with TSP	12
2.5 Summary of Literature Review	17
3 Research Approach	19
3.1 Mathematical Program	19
3.2 Data Requirements	22
3.3 Performance Measures	24
3.4 Testing and Evaluation	25
3.5 Summary of Research Approach	27
4 Isolated Intersection	29
4.1 Undersaturated Traffic Conditions	29
4.2 Oversaturated Traffic Conditions	38
4.3 Study Sites	44
4.4 Evaluation	48
4.5 Summary of Findings	59

5	Signalized Arterial	61
5.1	Optimization Procedure	61
5.2	Delay Estimation	63
5.3	Mathematical Program Formulation	75
5.4	Study Site	84
5.5	Evaluation	85
5.6	Extension to Networks	91
5.7	Summary of Findings	93
6	Conclusions	95
6.1	Summary of Research Findings	95
6.2	Contribution	97
6.3	Future Work	98
	Bibliography	100
A	Glossary of Symbols	105

List of Figures

2.1	Offset Adjustment for Transit Priority	6
2.2	Phase Extension and Phase Advance for Transit Priority	8
2.3	Phase Insertion or Phase Rotation for Transit Priority	9
3.1	Emulation-In-the-Loop Simulation Platform	27
4.1	Impact of Changes in Signal Timings on Auto Delays	31
4.2	Queueing Diagram for Lane Group j for Undersaturated Conditions (Auto Delay)	32
4.3	Queueing Diagram for Lane Group j for Undersaturated Conditions (Transit Delay)	35
4.4	Queueing Diagram for Lane Group j for Oversaturated Conditions (Auto Delay)	40
4.5	Queueing Diagram for Lane Group j for Oversaturated Conditions (Transit Delay)	42
4.6	Layout and Bus Routes for the Intersection of Katechaki and Mesogion Avenues	45
4.7	Lane Groups, Phasing, and Green Times for the Intersection of Kate- chaki and Mesogion Avenues	45
4.8	Layout and Bus Routes for the Intersection of University and San Pablo Avenues	47
4.9	Lane Groups, Phasing, and Green Times for the Intersection of Uni- versity and San Pablo Avenues	47
4.10	Percent Change in Person Delay for Different Intersection Flow Ra- tios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of Katechaki and Mesogion)	51
4.11	Percent Change in Person Delay for Different Average Bus to Auto Passenger Occupancy Ratios and $Y = 0.6$ (Test Type I: Intersection of Katechaki and Mesogion)	51
4.12	Percent Change in Person Delay for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of University and San Pablo)	53

4.13	Percent Change in Person Delay for Different Average Bus to Auto Passenger Occupancies and $Y = 0.6$ (Test Type I: Intersection of University and San Pablo)	53
4.14	Intersection Flow Ratios for the 1 Hour Time-Dependent Demand Profile (Test Type II: Intersection of University and San Pablo)	54
4.15	Change in Person Delay for Different Average Bus to Auto Passenger Occupancy Ratios and $Y = 0.6$ (Test Type II: Intersection of Katechaki and Mesogion)	56
4.16	Change in Person Delay for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type III: Intersection of Katechaki and Mesogion)	58
5.1	Pairwise Arterial Signals Optimization	64
5.2	Auto Delay Estimation for Platoon Arrivals	65
5.3	Transit Delay Estimation	72
5.4	San Pablo Avenue Layout (not to scale)	86
5.5	Signal Phasing and Green Times for San Pablo Avenue Segment	86
5.6	Arterial Signal Optimization in a Network	92

List of Tables

3.1	System Technology Requirements and Costs	25
4.1	Person Delays for $Y = 0.80$ and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of Katechaki and Mesogion)	50
4.2	Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type II: Intersection of Katechaki and Mesogion)	55
4.3	Performance Measures for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type III: Katechaki and Mesogion Intersection)	59
5.1	Person Delays on the Arterial Segment for $\bar{o}_b/\bar{o}_a = 40/1.25$ and Five Signal Cycles of Traffic Operations (Test Type I)	87
5.2	Person Delays on the Arterial Segment $\bar{o}_b/\bar{o}_a = 40/1.25$ and 1 Hour of Traffic Operations (Test Type III)	88
5.3	Person Delays per Type of Approach on the Arterial Segment for $\bar{o}_b/\bar{o}_a = 40/1.25$ and 1 Hour of Traffic Operations (Test Type III)	89
5.4	Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$, $\delta_{b,T}^x = 1$ and 1 Hour of Traffic Operations (Test Type III)	90

Acknowledgments

I have enjoyed every moment of my “Berkeley” life because I had the great opportunity to interact with people that have made an impact on me. This journey would have not been the same without the support of those people whom I acknowledge here.

I am greatly indebted to my advisor, mentor, and friend Alexander Skabardonis. I have always been impressed by his ability to find the solution when everything seems to be leading to a dead end. I am grateful to him for teaching me this skill of overcoming obstacles and for helping me grow into an independent researcher. I want to thank him for being a good friend always available to enthusiastically give advice on research and life matters. Our long discussions have been invaluable in shaping the way I think about career and life goals.

I want to thank Joan Walker for her continuous support and invaluable advice on this dissertation and academic life. I am thankful to Pravin Varaiya for his constructive comments on this research and for introducing me to Sensys Networks, through which I had the opportunity to learn more about the operational characteristics of transit signal priority strategies and the technology requirements. I am thankful I had the chance to take classes from and interact with so many distinguished faculty at Berkeley. I would therefore like to extend my thanks to Mike Cassidy, Carlos Daganzo, Mark Hansen, Adib Kanafani, and Samer Madanat. In particular, I thank Samer Madanat for his mentorship and support.

I want to thank my friend and mentor Nikolas Geroliminis from the bottom of my heart. He helped me adjust to the new environment during my first months at Berkeley and has provided invaluable advice about research and career goals since then, but most importantly he has been a great friend. I am grateful to my undergraduate thesis advisor, Matthew Karlaftis, for motivating me to follow a transportation path and apply to graduate school. I also want to thank Ioannis Papamichail and Kostas Aboudolas for all they taught me, for their advice and guidance that helped improve this dissertation research.

My experience in Berkeley would not have been the same if it wasn't for my dear friends and colleagues in the transportation group. I want to thank my office mates and friends, Celeste Chavis, Vikash Gayah, and Karthik Sivakumaran, for their continuous moral support, research recommendations, and for being there to celebrate my successes and help me overcome my failures. I am grateful for meeting my friend Ilgin Güler and sharing the same apartment for two years and I want to thank her for all the good and tough times we went through together. I am thankful that I had the chance to meet, work, and share ideas about research and life with many other colleagues in McLaughlin Hall. I want to thank Juan Argote, Gurkaran Buxi, Robert Campbell, Offer Grembek, Weihua Gu, Josh Pilachowski, and Stella So for their friendship and support throughout all these years.

I will never forget the help I got from my friend Sophia Diamantidou while preparing

my graduate school applications. Her friendship and continuous support are irreplaceable. I also want to thank my friends Tasos Nikoleris and Antonis Papavasiliou for always being available whenever I needed help, for their research advice, and for all the fun we had together.

I am grateful to my parents, Michail and Veta, for their constant love and all the sacrifices they have made to raise and educate me. They have always been supportive of my decisions in life and I could not have been where I am today without their love. I also want to thank my brother Panagiotis for always believing in me and pushing me to aim and reach higher. I am indebted to my grandparents, Charalambos and Eleni for transferring to me their invaluable wisdom, teaching me the value of education from a very young age, and supporting my studies, and to my aunt Stratoula for showing me what it takes to be a great teacher. Finally, I want to acknowledge my extended family for always being there for me.

Last but not least, I want to thank my love and best friend Eric for his endless support. It is only with his patience, guidance, and love that I have managed to reach the end of this journey successfully.

This research was supported by the Gordon F. Newell Memorial Fellowship, the University of California Berkeley Center for Future Urban Transport, a Volvo Center of Excellence, the Eisenhower Graduate Fellowship Program, and the University of California Transportation Center. I would also like to thank the U.S. DOT Exploratory Advanced Research Program for their financial support. I want to wholeheartedly thank the ITS Library staff for all their help with this research and the ITS Payroll Office staff for all of their assistance over the years.

Chapter 1

Introduction

1.1 Motivation

Traffic congestion is one of the biggest problems that urban areas face because it is associated with low mobility and high levels of pollution and fuel consumption. Conflicts among multiple transportation modes that share the same infrastructure further exacerbate this problem. However, multimodal systems are essential for achieving more efficient, sustainable, and equitable transportation operations. Traffic signal control systems, if optimized properly, hold potential to achieve efficient multimodal traffic operations by resolving conflicts for shared space, while mitigating congestion and its negative externalities in urban networks. These systems are traditionally optimized by minimizing total delays for vehicles, thus ignoring the importance of person mobility in networks served by multiple transportation modes. In addition, such vehicle-based optimization can lead to unfair treatment of high occupancy transit vehicles and their passengers.

Transit vehicles contribute less to congestion and pollution per passenger compared to autos, but often their passengers experience higher overall costs than auto users. There is a need to grant priority to transit vehicles at bottlenecks such as signalized intersections, which are responsible for a big portion of their delay. Prioritizing transit vehicles through improvements in facility design (e.g., bus lanes, queue jumper lanes) is not always feasible because of geometric and spatial restrictions. As a result, there is a clear need to optimize signal control systems such that they balance their treatment of transit and auto users by minimizing total person delay in a network.

Transit Signal Priority (TSP) is an operational strategy that facilitates efficient transit operations by providing priority to transit vehicles at signalized intersections. TSP strategies have been implemented in several urban areas in the United States and Europe. Many studies report significant reductions in control delay for transit vehicles and an overall improvement of their operations. However, they are often disruptive to the auto traffic, leading to substantial increases in delay for auto users. Commonly used priority strategies consist of changing signal timings by fixed increments once

a transit vehicle is detected without considering traffic conditions on the rest of the network. In addition, existing systems do not take into account the difference in passenger occupancies between autos and transit vehicles, instead optimizing their signal settings on a per vehicle basis. This also leads to inefficient ways of treating conflicting transit routes, when two or more transit vehicles that are candidates for priority arrive at the same time at an intersection from conflicting directions. Finally, existing traffic signal control systems are based on site-specific implementations, limiting even further their widespread applicability in the real-world.

The remaining sections of this chapter present the research question, identify the contribution of this research, and provide an overview of the structure of the chapters of this dissertation.

1.2 Research Question

The need to manage multimodal transportation systems efficiently and sustainably and to improve person mobility has recently become imperative due to the continuous growth in traffic demand that exceeds network capacities in many cities. Sustainability can be improved by using the existing infrastructure more efficiently. Traffic signal control systems are widely available in urban networks, and they can therefore be used to manage traffic operations more efficiently. Combining traffic signal optimization with TSP strategies is the most cost-effective and widely applicable way to improve the level of service for transit operations and minimize the total person delay in signalized networks (Skabardonis, 2003).

More specifically, the question that motivates this research is: *How should traffic signal control systems be designed so that they provide priority to transit vehicles traveling in conflicting directions, while minimizing the impacts on general traffic in urban networks?*

1.3 Research Contribution

The contribution of this dissertation is the development of a person-based traffic responsive signal control system that can be implemented on isolated intersections and signalized arterials. By minimizing person delay, the system provides priority to transit vehicles at signalized intersections based explicitly on their passenger occupancy. At the same time, the schedule delay that a transit vehicle has when arriving at an intersection is taken into account so that priority is only provided to those vehicles that are late. Therefore, priority is assigned to vehicles traveling in conflicting directions in an efficient way. In addition, this signal control system addresses the impact that these priority strategies have on auto traffic. This is done by minimizing the total person delay in the system under consideration and assigning penalties for interrupting the progression of platoons for the arterial case. Therefore, the system reduces delays for transit vehicles and improves transit schedule adherence. The system is

also flexible because the user can weigh the relative merit of auto and transit delays as desired and therefore allow different trade-offs between them. Finally, its underlying optimization process can be solved quickly to provide optimal signal settings in real-time, thus making it implementable in real-world settings.

Unlike other signal control systems, this person-based traffic responsive signal control system is generic. Therefore, it can be implemented and evaluated on any urban network regardless of the layouts, phasing schemes or traffic and transit characteristics of the intersections. Another advantage is that implementation of the system depends on readily deployable technologies. These include sensing systems that are commonly used in cities (e.g., loop detectors), Automated Vehicle Location (AVL) systems that can track the location of transit vehicles, and Automated Passenger Counter (APC) systems that can provide real-time information on the passenger occupancies of transit vehicles.

Overall, this dissertation contributes to the development of readily implementable strategies that take advantage of existing infrastructure to improve transit and traffic operations in urban multimodal networks. This research is important because it provides the field of transportation with a cost-effective tool that improves person mobility in congested metropolitan areas. This work ultimately supports sustainable transportation systems that will improve quality of life in cities.

1.4 Dissertation Organization

This dissertation is organized as follows. Chapter 2 presents a review of the related literature on traffic signal control systems and TSP strategies. Chapter 3 describes the research approach and the input and associated technology requirements to develop and operate the person-based traffic responsive signal control system. In addition, it presents the evaluation methods and performance measures used by the system. Chapter 4 presents the signal control system that has been developed for an isolated intersection. First, the mathematical program that minimizes total person delay at an intersection is presented. Then the system is evaluated with data from two real-world study sites, and the results from a variety of tests are presented. Chapter 5 extends the system to signalized arterials. First, the mathematical program that minimizes total person delay at two consecutive intersections is described as well as the method of the pairwise optimization used for arterials with multiple intersections. The results from testing the performance of the system with data from a real-world arterial with four intersections are presented. Finally, Chapter 6 includes a summary of the key findings, the dissertation's contribution, and future research directions.

Chapter 2

Literature Review

The literature related to traffic signal control systems and Transit Signal Priority (TSP) strategies is extensive. This section discusses the existing work in these two areas with a focus on signal control systems that have incorporated TSP. First, Section 2.1 describes existing TSP strategies, both active and passive. The impact of active TSP strategies on auto traffic and the disruption of signal coordination are discussed in Section 2.2. Methods used to maintain signal coordination while implementing TSP are also described. Then, Section 2.3 presents the strategies used in signal control systems with TSP to grant priority to transit vehicles traveling in conflicting directions at intersections. Section 2.4 consists of a review of real-time signal control systems with TSP. Finally, Section 2.5 summarizes the limitations of the existing systems that motivated the design of the signal control system developed in this dissertation.

2.1 Transit Signal Priority (TSP)

Transit priority can be achieved both by proper facility design and the use of traffic signal control systems (Skabardonis, 2000). Some examples of facility designs include dedicated bus lanes and queue jumper lanes that allow a bus to bypass the queue and arrive more quickly to the stop line (Baker *et al.*, 2002). In addition to their site-specific character, implementation of such priority schemes requires extra space or reallocation of existing space, which are practices that are often infeasible. Since the goal of this dissertation is to develop priority strategies that take advantage of existing infrastructure, the review of the literature has focused only on TSP strategies.

TSP strategies via traffic control modify normal signal operations to allow transit vehicles to travel through a signalized intersection with reduced delay. Note that this is different than preemption which interrupts normal signal operations in order to serve the transit vehicle with no delay. The objective of TSP is to improve transit efficiency by reducing control delay for transit vehicles (i.e., delay caused by the signals at intersections), and thus maintain schedule adherence and minimize bus bunching, making the system more reliable and attractive to users. Moreover, it results in more

fuel efficient operations both for transit and auto traffic and provides incentives for higher transit ridership (Baker *et al.*, 2002). Existing TSP strategies can be classified in two categories: passive and active, which are described in detail in the following two sections.

2.1.1 Passive Priority Strategies

Passive priority strategies are developed offline based on historical data. They operate continuously without requiring any detection systems and, as a result, regardless of the presence of a transit vehicle (Baker *et al.*, 2002). They mainly include changes in the signal settings such as green times, offsets,¹ and cycle lengths. Passive priority strategies include:

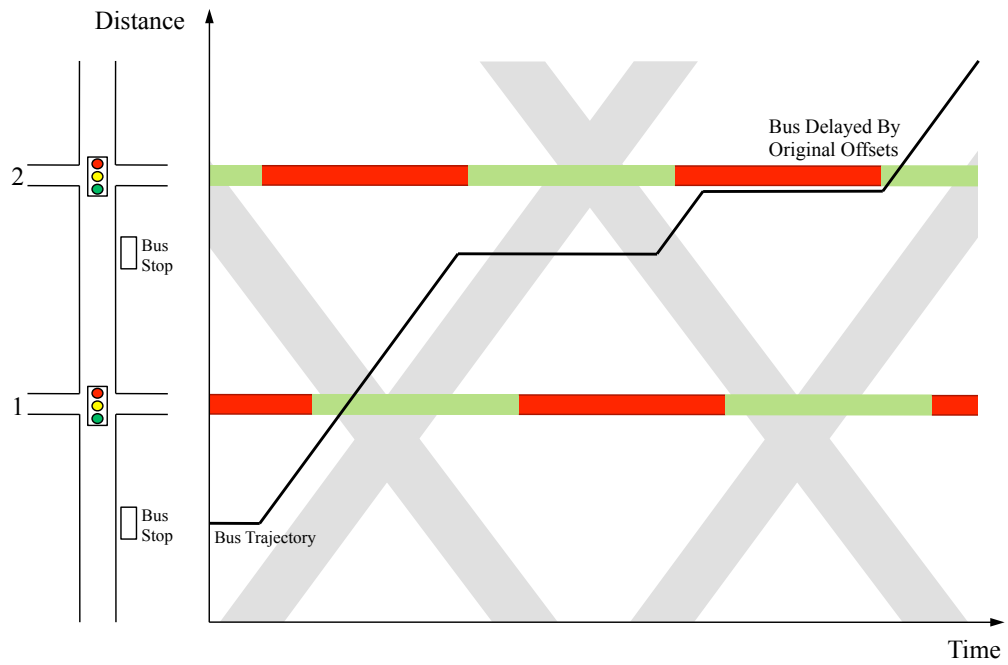
- adjustment of offsets,
- additional green time for the phases² serving transit vehicles, and
- reduction in cycle length.

Figure 2.1 shows time-space diagrams of vehicle trajectories traveling through two consecutive signalized intersections. The trajectories of auto platoons are grouped together and are shown as a grey “band” in each direction, and the trajectory of a bus is shown with a single black line. Vehicles travel in both directions, and the locations of the intersections are denoted on the distance axis. The time axis includes the phase sequence and timings for the signalized intersections. In this example, with no adjustment of offsets the bus would have to stop at the second intersection, if the green was terminated after the passing of the vehicle platoon (Figure 2.1(a)). Transit priority is often provided by adjusting the offsets to account for the lower transit vehicle speeds and dwell times at transit stops (Figure 2.1(b)). The other two passive priority strategies aim at reducing delay for buses by either increasing the green time allocated to phases that serve transit vehicles, thereby reducing the probability of a transit vehicle arriving during a red interval, or by decreasing the length of the cycle and thus increasing the turnover of phases.

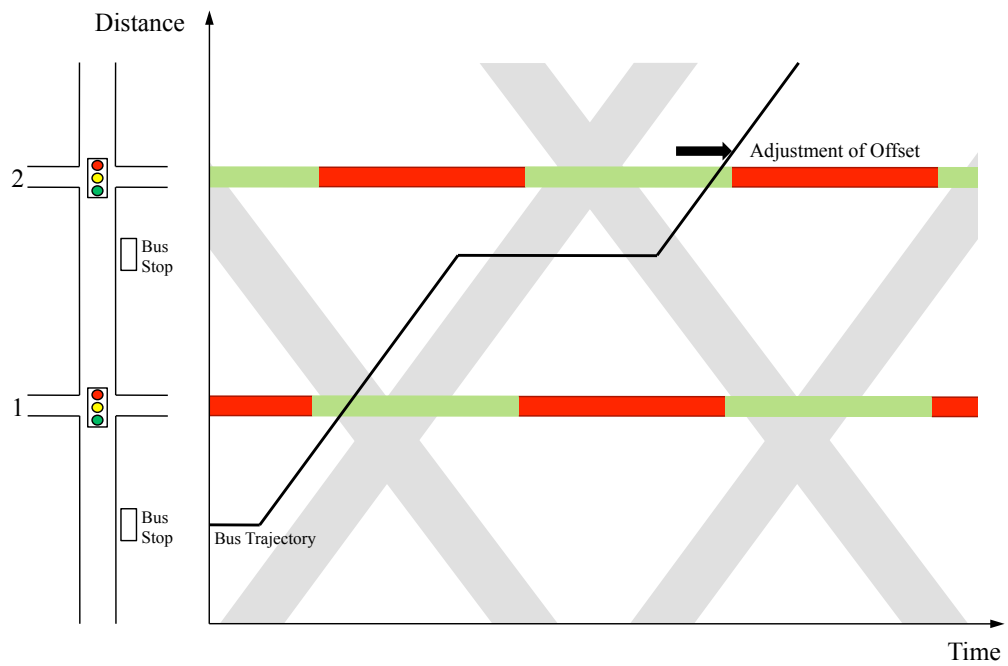
Skabardonis (2000) developed optimal signal timings for bus operations by minimizing a combination of delays and stops offline. Weighting factors for delays and stops that implicitly accounted for passenger loads were included to favor the buses. The optimal signal settings resulted in a 14% decrease in bus delay and a 4% increase in average bus speeds without significant adverse impacts on the rest of the traffic. However, the study also concluded that heavy weighting of buses can lead to modest additional benefits to transit at the cost of excessive delays to the rest of the traffic.

¹*Offset* is the relative time between the defined reference points (e.g., start times) of the coordinated phases at two intersections (Koonce *et al.*, 2008).

²A *phase* is the green (i.e., right-of-way), yellow and red clearance intervals in a cycle, that are assigned to an independent movement or a group of non-conflicting movements (FHWA, 2009).



(a) Initial Signal Settings



(b) Adjusted Offset

Figure 2.1. Offset Adjustment for Transit Priority

Other studies used passive priority to provide progression to buses either by minimizing bus travel times (Estrada *et al.*, 2009) or by changing offsets and the phase sequence at signals (Furth *et al.*, 2010). The results indicate significant improvements to buses and impacts to auto users that vary from small increases to small reductions in their delay.

Passive priority strategies are inexpensive to develop and easy to implement. However, their success depends on the validity of the assumption of low variability of traffic volumes. In addition, such strategies assume that transit vehicles have deterministic dwell times at transit stops (i.e., accurate knowledge of arrival times, so that offsets are adjusted accordingly), which is not realistic for most transit operations.

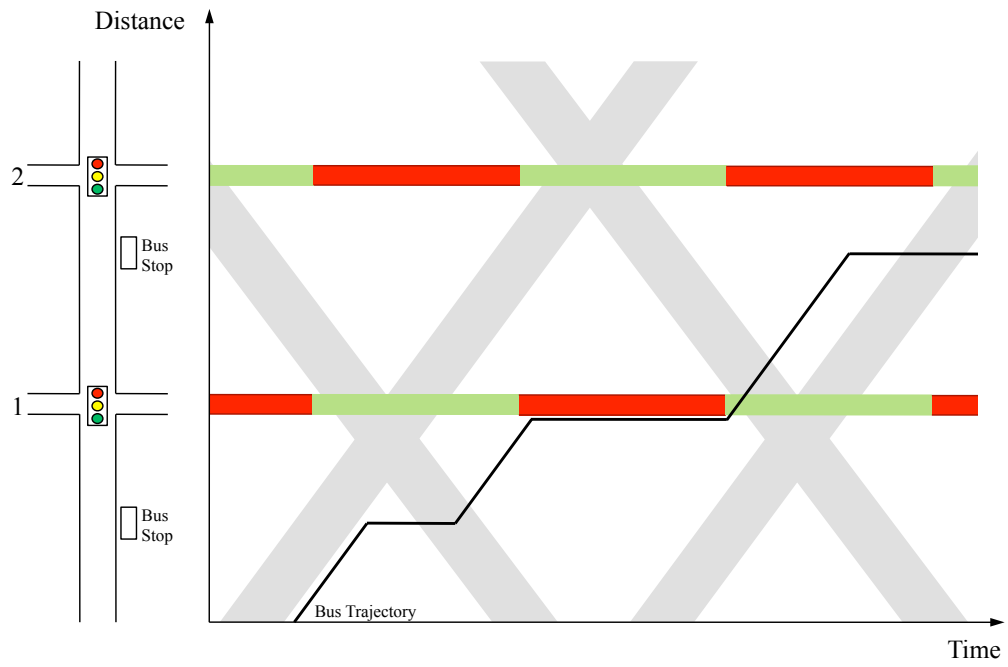
2.1.2 Active Priority Strategies

Active priority strategies are implemented using real-time information on traffic conditions and transit arrivals at the intersection. As a result, they are typically more effective than passive priority strategies. Information on auto and transit operations, which is obtained by sensing technologies is required for the design of such strategies, which consist of:

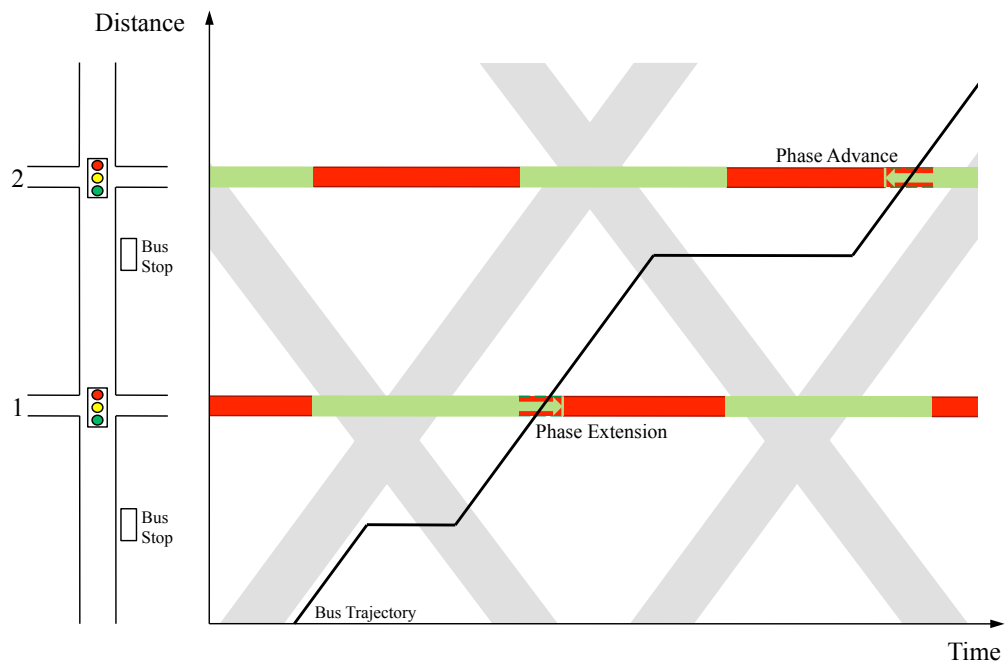
- phase extension,
- phase advance,
- phase insertion, and
- phase rotation.

Figures 2.2 and 2.3 illustrate examples of the active priority strategies mentioned above. As before, the trajectories of auto platoons and a bus are shown. Figure 2.2(a) shows that under the initial signal settings the bus is expected to stop at intersection 1. By extending the green time for that phase at intersection 1, the bus can pass uninterrupted, as shown in Figure 2.2(b). However, it will have to stop at the second intersection, unless the red is truncated (and the next phase is advanced) to allow the bus to pass without delay. Figure 2.3 illustrates provision of priority via phase insertion or phase rotation. While these two strategies are different in practice, their illustration on a one approach time-space diagram appears the same. Figure 2.3(a) shows that under initial signal settings the bus is expected to stop at the second intersection. In this case two options are considered: either a new phase that will serve the bus is inserted, or the phase sequence is changed so that the bus can be served as soon as possible (Figure 2.3(b)). Note that for all the strategies presented here, the progression of the vehicle platoons is maintained for both directions.

Extensive research exists on the design, implementation, and evaluation of active TSP strategies on signalized arterials (Al-Sahili & Taylor, 1996; Balke *et al.*, 2000; Skabardonis, 2000; Kim & Rilett, 2005; Ahn & Rakha, 2006), a few of which are currently operational in several cities around the world (Head, 1998; Baker *et al.*,

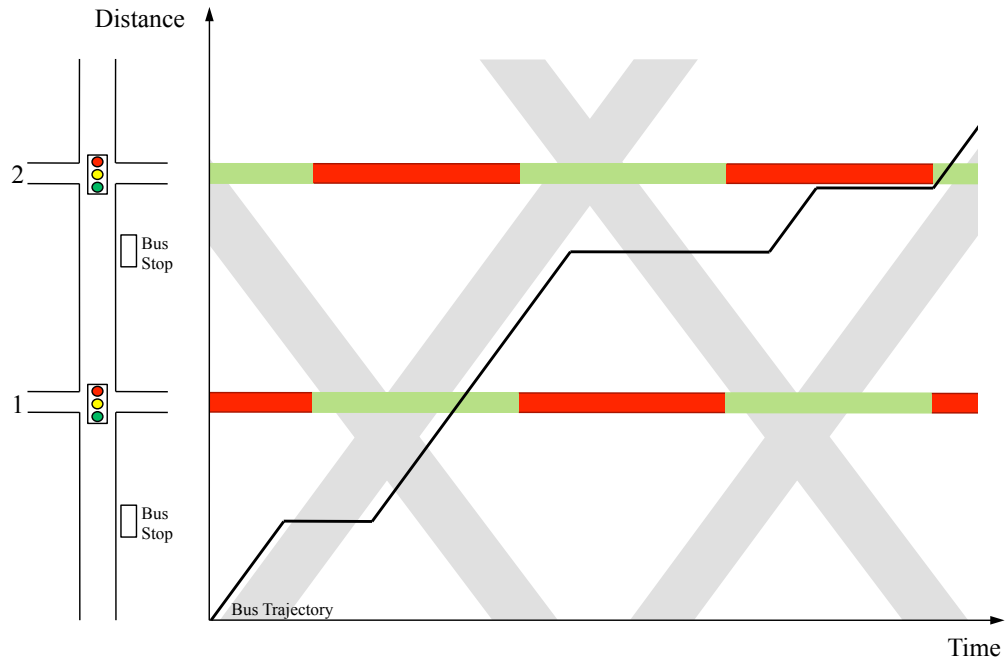


(a) Initial Signal Settings

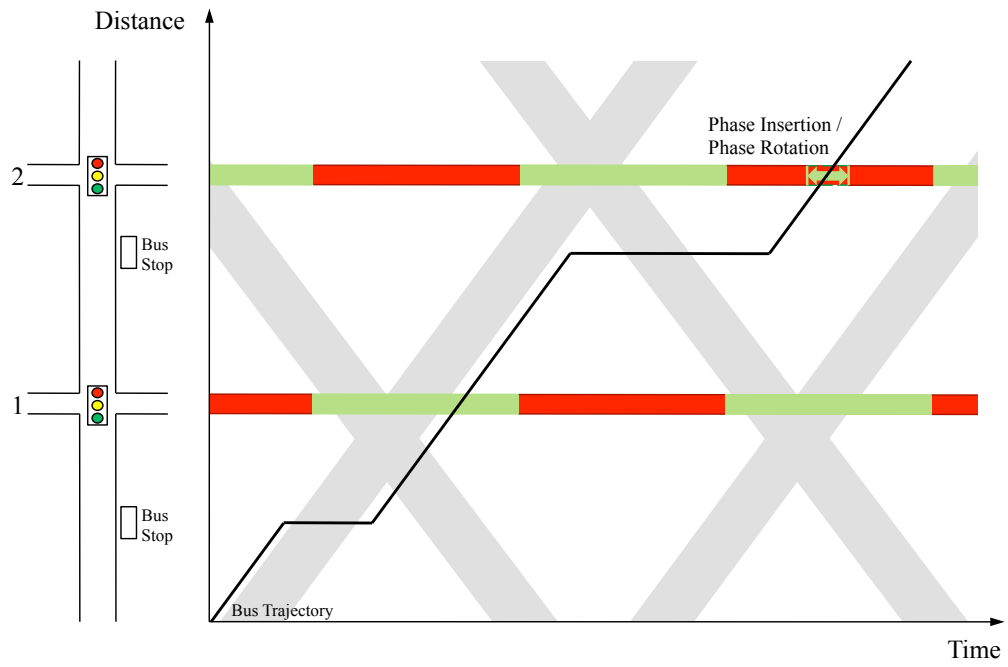


(b) Phase Advance and Phase Extension

Figure 2.2. Phase Extension and Phase Advance for Transit Priority



(a) Initial Signal Settings



(b) Phase Insertion/Phase Rotation

Figure 2.3. Phase Insertion or Phase Rotation for Transit Priority

2002; Nash, 2003). The reported benefits for transit and the whole system vary from modest improvements for the performance of transit vehicles with insignificant impacts on the rest of the traffic (Skabardonis, 2000) to significant reductions in bus travel times and minor increases in the delays for non-transit traffic for moderate demand levels (Balke *et al.*, 2000; Zhou, 2008). Most of the implementations are site-specific and often their success depends on the existence of appropriate facility design for priority (e.g., bus and queue jumper lanes). In addition, the lack of a systematic framework for evaluation of their benefits leads to reported outcomes that result from improvements of the existing signal control systems rather than the active TSP strategies themselves (e.g., benefits are attributed to switching from a fixed-time signal control system to a traffic responsive or adaptive one).

While active priority strategies can be used in real-time and are more effective in improving transit operations than passive priority strategies, they require sensing and communication technologies that increase the cost and complexity of such implementations with no guarantee of success on a network level. Active priority strategies often have detrimental impacts on non-transit traffic (especially cross-street traffic), can cause confusion for motorists, and in many cases are responsible for loss of signal coordination (Chang & Ziliaskopoulos, 2003; Skabardonis, 2000). Finally, existing systems that incorporate active TSP strategies do not have an efficient way of treating the issue of conflicting transit routes due to limited flexibility in granting priority when multiple transit vehicles need to be considered. Some of these critical issues are discussed in the next sections.

2.2 Signal Coordination with TSP

According to the Signal Timing Manual (Koonce *et al.*, 2008), coordination is the ability to synchronize the signals of multiple intersections in order to achieve uninterrupted progression of traffic for one or more directions in a network. Coordination is an important aspect of traffic signal systems because it can be used to reduce delays and stops, which consequently leads to a reduction in fuel consumption and air pollution. During the implementation of TSP strategies, coordination can easily be interrupted due to continuously changing signal settings at the intersections. The recovery period for the transition back to coordination can take several cycles (Balke *et al.*, 2000; Sane & Salonen, 2009), and sometimes this transition is more disruptive than the original interruption (Furth & Muller, 2000). If disruption for transit priority occurs often, the intersection may never be able to recover coordination.

Maintaining coordination on a network while providing priority to transit vehicles is a challenging task in the design and optimization of signal control systems with TSP. Existing literature suggests using a transition period during which phases are skipped or shortened and then restoring progression after the clearance of the queues that might have formed (Baker *et al.*, 2002; Furth & Muller, 2000). A more recent approach proposes to isolate the intersection during provision of transit priority and return to the coordinated mode after the transit vehicle under consideration has passed the

intersection (Sane & Salonen, 2009). The latter, however, can be effective only for low to moderate traffic conditions. Other options for maintaining coordination are to provide priority only to intersections with spare green time, which however limits provision of priority, or increase the system cycle length, which increases delays. Communication between adjacent intersections about platoon arrivals is the most effective way to maintain coordination, especially in cases where cycle lengths are not fixed (Janos & Furth, 2002; Henry & Farges, 1994). This approach of tracking the arrivals of platoons is followed in this dissertation for maintaining coordination on arterials.

2.3 TSP Implementation for Conflicting Transit Routes

The issue of conflicting transit routes occurs when two or more transit vehicles traveling in conflicting directions are expected to arrive at an intersection within the signal control’s optimization interval and they are all candidates for priority. In such cases, the system needs to decide how to grant priority to those vehicles. This is a challenging issue that needs to be addressed when designing signal control systems with TSP in order to ensure that all transit users are treated in an equitable and efficient way and to avoid detrimental impacts on the auto traffic.

This issue has been ignored by many systems that provide priority only to pre-determined routes and specific vehicles (Li *et al.*, 2008; Mauro & Di Taranto, 1989), only to vehicles traveling in non-conflicting directions (Cornwell *et al.*, 1986), or do not provide priority to any of the candidate transit vehicles when such conflicts occur (Ahn & Rakha, 2006). Others have addressed this issue by treating transit vehicles on a first-come, first-served basis (Li *et al.*, 2008), which however can lead to high disruption of traffic operations. In particular, for cases that absolute priority is provided (i.e., when priority is given without reference to some criteria such as schedule delay or passenger occupancy), provision of priority could be wasted. As a result, some systems have incorporated criteria and rules based on transit vehicles’ schedule delay, the time of the priority request relative to the active phase, or functions of queue length and schedule adherence to determine the sequence of priority provision to transit vehicles traveling in conflicting directions (Li *et al.*, 2008; Wadjas & Furth, 2003; Henry & Farges, 1994; Zlatkovic *et al.*, 2012; Li *et al.*, 2005). Other systems have based their decisions on minimization of some transit performance metric, such as total priority delay (Head *et al.*, 2006) or total transit delay weighted by passenger occupancy and schedule deviation (Ma *et al.*, 2011). More recently, Head *et al.* (2006) was extended and heuristics were developed to treat the issue of conflicting transit routes while accounting for the fact that bus arrival times are stochastic (He *et al.*, 2011).

Evaluation of these systems have shown their comparative advantage to first-come, first-served approaches in resolving conflicting requests. However, no system

has been found that has both of the following features: provides priority to transit in an efficient way even when transit vehicles are traveling in conflicting directions and optimizes signal settings for auto and transit users by accounting for their passenger occupancy and schedule delay to minimize total person delay.

2.4 Real-Time Signal Control Systems with TSP

Real-time signal control systems adjust the signal settings based on optimizing some predefined performance measure such as minimizing vehicle delay (Head, 1998). To do this, predictions of the traffic conditions are required as an input to the optimization process. Information needed for these predictions of traffic characteristics is obtained from detectors located at the entrance and/or stop line of the intersection approaches. Adjustments to the signal control settings are then made based on these predictions.

Real-time signal control systems are divided into adaptive and traffic responsive based on how rapidly they respond to variations in traffic flow (Klein *et al.*, 2006). Traffic responsive signal control systems optimize signal settings every minute or two, a time interval that is usually a multiple of the cycle length. The optimization process is called *cyclic optimization* because it maintains the concepts of cycle length, phase green times, and offsets. These are adjusted by the optimization in real-time to accommodate prevailing traffic conditions and achieve certain degrees of saturation or minimize delays, number of stops, or some combination of the two (Conrad *et al.*, 1998).

On the other hand, adaptive signal control systems run on a rolling horizon and do not retain any concept of cycle length, phase green times, or offsets. This is called *acyclic optimization*. An objective function, which is usually a linear combination of several cost elements such as vehicle delay and stops in the system, is minimized over a decision horizon that typically varies between values smaller than 30 seconds to greater than 2 minutes. The optimal signal settings are implemented only for part of the decision horizon (3–5 seconds) and are replaced by new ones every time they are generated (Conrad *et al.*, 1998). As a result, signal settings adapt to prevailing traffic conditions much more quickly than with traffic responsive systems. Note that real-time signal control systems are often collectively called “adaptive” even if they are actually traffic responsive.

Another difference between the two types of real-time signal control systems is that the technological requirements are much more intensive for adaptive signal control systems than for traffic responsive ones. Faster communication speeds and more complex signal controller hardware and software are typically required to make high accuracy predictions that are needed for the operation of adaptive signal control systems with limited time to optimize the signal settings (Gordon & Tighe, 2005). In addition, such systems require twice the detector density of traffic responsive systems (Klein *et al.*, 2006). Adaptive signal control systems require prediction of flows at the individual vehicle level rather than the more macroscopic measures of flow and platoon characteristics that are required for traffic responsive signal control systems.

Traffic responsive and adaptive signal control systems are further divided into subcategories according to the type of control architecture (i.e., fully centralized, hierarchical, and fully decentralized). In a fully centralized control system, all of the calculations and decisions are made by the central controller which determines signal timings for each local controller. On the other hand, fully decentralized systems allow the local controllers to perform all of the necessary calculations independently. Hierarchical systems are a combination of decentralized and centralized control. Such systems optimize objective functions on two or more levels. Hierarchical systems are further characterized as centralized or distributed according to the relative weight assigned to the central and local controllers (Yagar & Dion, 1996).

A number of real-time signal control systems that incorporate transit priority strategies exist in the literature. A description of the most prominent adaptive and traffic responsive signal control systems with TSP follows.

2.4.1 Traffic Responsive Signal Control Systems with TSP

SCOOT (Split, Cycle, and Offset Optimization Technique) and SCATS (Sydney Coordinated Adaptive Traffic System) are the most widely deployed real-time signal control systems. SCOOT is a fully centralized control system which optimizes phase green times, cycle lengths, and offsets in real-time based on saturation level constraints. The objective is to improve traffic progression, thus minimizing vehicle delays and stops (Hunt *et al.*, 1982). Data are obtained by detectors located at the upstream end of each approach. Priority is provided to transit vehicles through phase extension or advance, conditional on schedule-based and headway-based criteria only when traffic conditions are below user-defined levels of saturation (Bretherton *et al.*, 2002). Simulation and field trials have indicated delay savings on the order of 20% with impacts on the auto traffic that vary according to the priority strategy followed (Bretherton, 1996). However, the active priority strategies that are implemented do not explicitly account for the passenger occupancy of vehicles in the optimization process and therefore do not treat the issue of conflicting transit routes in an efficient way.

SCATS is a centralized hierarchical control system that determines phase green times, cycle lengths, and offsets by dividing the network into smaller subnetworks and designing signal settings for each of them independently. The selection of phase green times and cycle lengths is constrained by the degrees of saturation of preselected movements as in SCOOT. The offset decisions are used to achieve coordination within a subnetwork by grouping intersections with compatible cycle lengths. Data are obtained by detectors at the intersection stop line. Absolute transit priority is provided through phase extension or advance. The results from field implementations indicate a significant decrease in transit travel times and their variability, but no significant impact on the auto travel times (Cornwell *et al.*, 1986). However, the results cannot be attributed only to signal priority since the signals were uncoordinated in the "before" case. As a result, it is unclear how much of the benefit of SCATS could

also have been achieved by providing static coordination of signals. As for SCOOT, the TSP logic of SCATS does not account for passenger occupancies of vehicles. In addition, it is restricted to assign priority only to vehicles traveling in non-conflicting directions.

TUC (Traffic-responsive Urban Control) is a traffic responsive signal control system developed at the Technical University of Crete in Greece and is currently being implemented in several cities. The system is designed for heavy traffic conditions, and it optimizes phase green times, cycle lengths, and offsets to avoid spillback queues while taking into account link storage capacity (Diakaki *et al.*, 2003). Priority is provided either by weighting the measurements on approaches that have major transit routes or by adjusting the phase green times at the intersection level to provide absolute priority when a transit vehicle is detected, while at the same time accounting for saturation levels on all other approaches and downstream links. Priority is achieved through phase extension or insertion. Simulation tests on the networks of Tel Aviv and Jerusalem in Israel indicate that TUC with transit priority can achieve improvements in transit vehicles' speed by about 25% in the cases that transit vehicles are served by non-major phases and has an insignificant impact on transit operations when those are served by major phases. However, no field implementation of TUC with transit priority exists up-to-date. The main disadvantage of TUC is that it does not explicitly account for the higher occupancy or schedule delay of transit vehicles in order to provide priority efficiently in the case of conflicting transit routes.

MOTION (Method for the Optimization of Traffic Signals In On-line controlled Networks) is a decentralized and hierarchical signal control system developed in Germany. The system minimizes delays and stops in the network by optimizing phase sequence, phase green times, cycle lengths, and offsets (Busch & Kruse, 2001). Information needed for the optimization consists of volumes, platoons, and occupancies that are obtained from detectors. Priority can be provided to transit vehicles both at the network and the intersection level. At the network level, priority is achieved by determining offsets based on the average travel times of transit vehicles. At the intersection level, green times and phase sequences are adjusted to provide priority to transit vehicles. The level of priority provided depends on the traffic conditions in the network. Results from field implementations or simulation tests have not been reported. As with most existing systems, MOTION does not explicitly incorporate passenger occupancy or schedule delay of transit vehicles in the optimization process and does not present an efficient way of treating priority requests from conflicting directions.

California Partners for Advanced Transportation Technology (PATH) recently developed and implemented an Adaptive Transit Signal Priority System (ATSPS) (Li, 2008). Priority is provided based on a trade-off between bus delay savings and the impact on the rest of the traffic. The phase green times are optimized by minimizing a weighted sum of delays for buses and autos. ATSPS has been tested through hardware-in-the-loop simulation studies as well as through a field operational test on a 2-mile stretch of El Camino Real in San Mateo County, California. Results from

the field test indicate statistically significant bus trip travel time savings on the order of 9–13% without significant increases in auto delay. Despite the benefits achieved, the system has used constant weighting factors for all transit vehicles independent of their direction, passenger occupancy, or schedule delay and it has not been extended to include transit traffic on conflicting routes since it treats priority requests on a first-come, first-served basis.

2.4.2 Adaptive Signal Control Systems with TSP

PRODYN (PROgramme DYNamique) is a fully-decentralized, adaptive signal control system which operates on a rolling horizon using dynamic programming. Transit priority is achieved by including cost elements for the transit vehicles in the objective function that is optimized over the rolling horizon. The cost elements are weighted based on the priority level assigned *a priori* to each transit vehicle and its direction. Coordination is achieved by communicating the forecasts of the traffic streams with the neighboring intersections. Simulation tests on an isolated intersection and a three intersection arterial, revealed different levels of benefit for buses that were correlated with the weighting factors assigned to them. PRODYN also reduced delays for auto users compared to optimal signal settings from TRANSYT-7F for the same level of transit priority (Henry & Farges, 1994).

UTOPIA (Urban Traffic OPTimization by Integrated Automation) system in Turin, Italy is a hierarchical and decentralized system that is capable of providing priority to selected bus routes while simultaneously improving mobility for private vehicles (Mauro & Di Taranto, 1989). UTOPIA consists of closed-loop control strategies that are classified into intersection and area level control. The intersection level control optimizes the signal timings at each intersection, while accounting for traffic conditions at adjacent intersections. The weighting factors for the cost elements of the intersection level objective function are updated and consequently constrained by the area level control decisions. The area level control decisions are made based on an optimization process which minimizes the total travel time spent by private vehicles in the network. The first implementation of UTOPIA took place in a large area in Turin, Italy (Donati *et al.*, 1984). The results from field experiments showed a reduction in travel times for both private and public vehicles on the order of 9–15%. The main limitation of UTOPIA is the provision of priority to preselected vehicles and routes regardless of their passenger occupancy or schedule delay. As a result, it does not provide an efficient way of treating the issue of conflicting transit routes. Furthermore, the system is site-specific and its implementation and fine-tuning are not well documented. These both limit its widespread applicability in the real world.

Research efforts in the 1990s led to the development of SPPORT (Signal Priority Procedure for Optimization in Real-Time) which is a heuristic, fully-distributed, rule-based, adaptive system for optimizing signal timings while assigning priority to transit vehicles (Yagar & Han, 1994; Yagar & Dion, 1996; Conrad *et al.*, 1998; Dion & Hellinga, 2002). A rule is an ordered preference for various types of events such as

the arrival of a platoon on one approach and the existence of a queue of a specific size on another. The system accounts for the impact of stopped transit vehicles on traffic operations, uses person-based performance measures for evaluation, and compensates for lost green times during transit priority. Coordination is maintained indirectly through the introduction of rules that prioritize platoons, clear queues before the arrivals of such platoons, and hold vehicles at upstream intersections to avoid queue spillbacks. Those rules define the ideal request times for changing the signal settings. Simulation tests on an isolated intersection reveal a 21% passenger delay decrease compared to actuated signal control. Despite its advantages compared to previously developed adaptive systems, SPPORT still fails to address the issue of assigning priority to conflicting transit routes in an efficient or system optimal way. This is because priority is based on user-specified priority lists with no consideration of passenger occupancy or schedule delay. As a result, there is no guidance on which lists will give optimal results. In addition, it has been tested only at an isolated intersection through simulation.

The Los Angeles Department of Transportation (LADOT) has developed an adaptive signal control system with TSP that is currently being implemented on a network of 15 corridors for a total length of 450 miles. This control system incorporates centralized TSP with adaptive traffic control to allow provision of priority only to late buses. Advanced surveillance and communication technologies have been used for this implementation and the system has been found to reduce bus delays at the intersections by 33–39% with minimal impact on cross traffic. The issue of providing priority to conflicting transit routes is treated by predetermining which route gets the priority. In addition, the system is site-specific in terms of the signal controller, firmware, and communication technologies used, and as a result it is difficult to implement elsewhere (Li *et al.*, 2008). Finally, the software used for granting priority is proprietary and vendor-specific for each jurisdiction within the corridor where it is implemented, and thus very little is known about how priority is assigned.

Recent research efforts at the University of Arizona have led to the development of PAMSCOD (Platoon-based Arterial Multi-modal Signal Control with Online Data) (He *et al.*, 2012). PAMSCOD is a platoon-based formulation to optimize signal settings on an arterial for both autos and transit vehicles, using information obtained via a vehicle-to-infrastructure communications environment. Assuming availability of such information, the system identifies queues and platoons approaching an intersection and uses a mixed integer linear program to decide on future signal settings. It accounts for progression for both modes and it accommodates transit priority requests. Simulation experiments have shown that the system significantly decreases auto and bus delays for a variety of traffic conditions compared to commonly implemented coordinated-actuated signal control systems. However, the system suffers from high computation times that are prohibitive for real-world applications, and its success depends on the availability of high market penetration of probe vehicles.

The main advantage of traffic responsive and adaptive signal control systems with TSP is their ability to consider traffic demand perturbations in real time and accommodate time-dependent auto and transit vehicle flows, leading to more efficient control. Such signal control systems require good surveillance systems for accurate measurements of traffic flows, location and identification information of transit vehicles as well as accurate algorithms that use measured data to predict future traffic conditions. Especially when transit priority is incorporated, the existence of surveillance systems that can accurately determine transit vehicles' arrival times at the intersection, as well as their schedule adherence and passenger occupancies, is vital for the success of traffic responsive and adaptive signal control systems. Accurate detection systems are usually expensive to install and maintain, and they are hard for practitioners to operate, thus restricting the feasibility of such systems to very few locations. Traffic responsive signal control systems are usually less computationally intensive and have fewer technology requirements, so they are easier to implement in real-world settings compared to adaptive signal control systems. Therefore, the focus of this dissertation is on the development of a traffic responsive signal control system with TSP.

2.5 Summary of Literature Review

The review of the literature shows that several passive and active transit priority strategies exist and have been tested under different conditions worldwide. Lately, the provision of priority to transit vehicles has been incorporated in more advanced real-time signal control systems. Despite the improvements in transit operations that have been reported both through field implementations and simulation tests, existing systems still have a number of shortcomings.

First, none of the existing systems optimize signal settings by explicitly minimizing person delay for all travelers. On the contrary, they usually minimize vehicle delays and provide priority based on rules that are not directly included in the optimization process. Even when this is not the case, there has been no systematic investigation to test how a passenger occupancy-based priority strategy affects transit and traffic operations under a variety of traffic conditions and transit operating characteristics. As a result, existing systems cannot treat the issue of conflicting transit routes in an efficient and optimal way for minimizing person delay. In addition, most systems are limited to minimizing control delay for transit vehicles, thus ignoring the adherence of a transit vehicle to schedule. This often results to providing priority to transit vehicles that are ahead of schedule which could lead to further disruptions in the transit system operations. Even systems that have in some way addressed the above issues are dependent on the availability of high market penetration rates of probe vehicles and highly accurate information which is not currently available in real-

world operations. Finally, their high computation times are prohibitive for real-time operations.

The objective of this dissertation is to develop a person-based traffic responsive signal control system that minimizes person delay by explicitly accounting for the passenger occupancy of autos and transit vehicles. This leads to provision of priority to transit vehicles and introduces an efficient way of assigning priority in cases that transit vehicles travel in conflicting directions and compete for it. At the same time, the system acknowledges the importance of schedule adherence for reliable transit operations. Therefore, it assigns appropriate weights to the delays of transit vehicles in order to avoid prioritizing those that arrive early. This also adds another criterion to decide on priority assignments in cases with multiple conflicting transit routes. Consideration of the impact that priority strategies have on auto traffic and its progression is achieved by assigning the appropriate delays for interrupting the progression of the platoons. Finally, the system is based on the availability of currently deployable technologies, such as detectors, AVL and APC systems, that are often already installed to serve other planning and evaluation purposes as explained in Section 3.2. Such technologies provide the input for accurate predictions of vehicle demand, arrivals, and passenger occupancies.

Chapter 3

Research Approach

A traffic responsive signal control system is developed based on a mathematical program that minimizes person delays for all travelers (i.e., auto and transit users). The system is designed under the assumption that readily deployable technologies are available that allow for real-time accurate estimates of vehicle flows, arrival times, and passenger occupancies. The system is evaluated with both deterministic and stochastic arrival tests for a variety of performance measures that include person delay, vehicle delay, number of stops, and emissions.

This chapter presents the research approach for the person-based traffic responsive signal control system that is developed in this dissertation. Section 3.1 describes the assumptions and the formulation of the mathematical program used to minimize person delays for both transit and autos at an intersection. Section 3.2 lists the data requirements necessary for the operation of the system. Technologies that are needed to obtain these data are also described. Then, Section 3.3 identifies performance measures that are used as part of the evaluation, and Section 3.4 describes the tests used to evaluate the performance of the system. Finally, Section 3.5 summarizes the research approach that is followed in the remaining chapters of this dissertation.

3.1 Mathematical Program

A mathematical program is developed to determine the optimal signal settings for all intersections that minimize total person delay in the system. The model minimizes total person delay by weighting delays for both auto and transit vehicles by their respective passenger occupancies. The issue of conflicting transit routes is addressed by accounting for vehicle occupancies and schedule delay in the objective function for the transit vehicles. Schedule delay is the amount of time that a transit vehicle is behind schedule when arriving at the intersection under consideration. Therefore, a transit vehicle that is further behind schedule or carries more passengers than others receives higher priority.

The mathematical program is formulated based on the assumption that perfect information is available about the vehicle arrivals, traffic demand, passenger occu-

pancies, and lane capacities at intersections. Auto vehicle arrivals are assumed to be deterministic for delay estimation purposes. The cycle length is assumed to be constant for the analysis period and common for all signalized intersections in the network under consideration. In addition, the sequence of the phases as well as the phase design are pre-determined and fixed. It is also assumed that the capacity for each approach at intersections is fixed and not affected by traffic operations, which means that the saturation flow for each of the lane groups¹ is constant. Finally, the model is formulated by treating queues of vehicles as vertical queues at the stop lines and assuming that transit vehicles travel on mixed-use traffic lanes along with autos. However, the formulation of the mathematical model holds even when dedicated lanes for transit vehicles exist. Dedicated transit lanes would actually improve the performance of the signal control system because separating transit vehicles from queues of general traffic facilitates accurate predictions of transit vehicle arrivals at intersections. Consequently, they are expected to improve the system's performance compared to when transit vehicles travel in lanes of mixed traffic.

The mathematical program minimizes total person delay by changing the phase green times at all intersections. The mathematical program is run once for every cycle and its generalized formulation for an intersection r and cycle T is as follows:

$$\begin{aligned} \min \quad & \sum_{a=1}^{A_T^r} o_a d_{a,T}^r + \sum_{b=1}^{B_T^r} o_{b,T}^r (1 + \delta_{b,T}^r) d_{b,T}^r & (3.1a) \\ \text{s.t.} \quad & d_{a,T}^r = d_a^r (g_{i,T}^r) & (3.1b) \\ & d_{b,T}^r = d_b^r (g_{i,T}^r) & (3.1c) \\ & g_{i \min}^r \leq g_{i,T}^r \leq g_{i \max}^r & (3.1d) \\ & \sum_{i=1}^{I^r} g_{i,T}^r + L^r = C & (3.1e) \end{aligned}$$

where:

a : auto vehicle index

b : transit vehicle index

A_T^r : total number of autos present at intersection r during cycle T

B_T^r : total number of transit vehicles present at intersection r during cycle T

o_a : passenger occupancy of auto a $[\frac{\text{pax}}{\text{veh}}]$

$o_{b,T}^r$: passenger occupancy of transit vehicle b for cycle T at intersection r $[\frac{\text{pax}}{\text{veh}}]$

$d_{a,T}^r$: delay for auto a for cycle T at intersection r [sec]

$d_{b,T}^r$: delay for transit vehicle b for cycle T at intersection r [sec]

$\delta_{b,T}^r$: factor for determining the weight for schedule delay of transit vehicle b in cycle T at intersection r

$d_a^r (g_{i,T}^r)$: function relating green times to delays for auto a

¹A *lane group* is defined per the Highway Capacity Manual 2000 (HCM, 2000) as one or more adjacent lanes at each intersection approach that can be served by the same phases.

$d_b^r(g_{i,T}^r)$: function relating green times to the delay for transit vehicle b
 $g_{i,T}^r$: green time allocated to phase i in cycle T at intersection r [sec]
 $g_{i\min}^r$: minimum green time for phase i at intersection r [sec]
 $g_{i\max}^r$: maximum green time for phase i at intersection r [sec]
 I^r : total number of phases in a cycle for intersection r
 L^r : lost time at intersection r [sec]
 C : cycle length [sec]

The objective function consists of the sum of the delay for auto and transit passengers that are present at the intersection during the design cycle T (i.e., cycle currently being optimized). Delays for autos and transit vehicles depend on the green times, $g_{i,T}^r$, which are the decision variables for the mathematical program. In fact, $d_{a,T}^r$ and $d_{b,T}^r$ also depend on the green times of the previous and the next cycles, the yellow times, and numerous other input parameters, which are either pre-specified by the user or collected with the use of surveillance technologies. To simplify the notation in subsequent chapters, the delays for each lane group, transit vehicle, and platoon are included in the objective function as a variable, and this variable is constrained to equal a function as shown in (3.1b) and (3.1c).

The delays of both autos and transit vehicles are weighted by their respective passenger occupancies in the objective function. The delays for transit vehicles are also weighted by a factor $(1 + \delta_{b,T}^r)$ in order to account for the schedule delay that a transit vehicle b has when arriving at intersection r during cycle T . This factor, which is user-specified, can be a linear function of the schedule delay of the transit vehicle as follows:

$$\delta_{b,T}^r = \alpha \Delta l_{b,T}^r \quad (3.2)$$

where $\Delta l_{b,T}^r$ is the schedule delay of the transit vehicle, and α is a user-specified positive parameter that determines the strength of the weighting for the schedule delay in the objective function. In other cases, $\delta_{b,T}^r$ could be a binary variable indicating whether a transit vehicle is ahead or behind schedule as follows:

$$\delta_{b,T}^r = \begin{cases} 1 & \text{if } \Delta l_{b,T}^r \geq \theta \\ 0 & \text{if } \Delta l_{b,T}^r < \theta \end{cases} \quad (3.3)$$

where θ is a user-specified schedule delay threshold to define whether a transit vehicle should be considered late for priority purposes or not (e.g., if a transit vehicle is more than 5 minutes late). In either case, the delay for a transit vehicle that is behind schedule is weighted more than a transit vehicle that is arriving early or on time at the intersection.

Three constraints are introduced for the decision variables. The green times of each phase i and intersection r are constrained by their minimum and maximum green times (constraint 3.1d). Minimum green times, $g_{i\min}^r$, are necessary to ensure safe vehicle and pedestrian crossings. In addition, they ensure that no phase is skipped. Maximum green times, $g_{i\max}^r$ are used to restrict the domain of solutions for the green times of the phases and reduce computation times. The phase green times are also

constrained such that the sum of the green times for all phases at each intersection plus the lost time adds up to the cycle length (constraint 3.1e). The lost time is assumed to be the summation of the yellow times that follow each phase:

$$L^r = \sum_{i=1}^{I^r} y_i^r \quad (3.4)$$

The cycle length is kept constant for every cycle in the analysis period and is common among intersections in the network under consideration. Keeping the cycle length constant is not essential for the formulation. In fact, variable cycle lengths could result in lower person delays at the intersection. However, the constraint of constant and common cycle lengths simplifies the formulation of the mathematical program and facilitates progression of auto traffic at the arterial level. In cases that the cycle length is selected by time of day, it still remains common among all intersections in the network at any time.

The details of the mathematical program, as well as the delay for autos and transit vehicles for both the isolated intersection and the arterial cases are presented in more detail in Chapters 4 and 5, respectively, along with the results of the tests performed for a variety of traffic conditions and transit operating characteristics.

3.2 Data Requirements

The formulation and implementation of this person-based traffic responsive signal control system is based on the availability of real-time information about traffic conditions, arrivals of transit vehicles, and passenger occupancies. The required information can be provided by surveillance and communication technologies that are currently deployable in many urban networks worldwide. Surveillance technologies are used to collect data that are necessary to determine the input for the optimization process. Specifically, these inputs are vehicle demand, vehicle arrival times, and passenger occupancies. Advanced communication technologies that are often part of the sensing systems are used to transmit the information in real time to the controllers so that the system can optimize the signal settings for the next cycle.

Traffic demand can be monitored in real-time by inductive loop detectors placed far enough upstream of the intersection so that the vehicle arrivals are measured under free-flow conditions. For the isolated intersection case, detectors placed upstream provide information for estimating traffic demand in terms of average arrival flows, while for the arterial case they provide information for estimating platoon sizes and arrival times at the intersection. In addition, upstream detectors provide speed information that can be used to estimate average free-flow travel times for the link on which they are located. When located at the downstream end of a link, detectors can also provide information on the turning ratios of the different movements that are necessary to estimate the demands for the different lane groups. Inductive loop detectors can be replaced by other surveillance technologies such as video cameras

and Micro-Electro-Mechanical Sensors (MEMS). All of these technologies record vehicle passage and presence and can provide the same information as inductive loop detectors.

Automated Vehicle Location (AVL) systems incorporate Global Positioning Systems (GPS) and Differential GPS, that are used for tracking transit vehicles continuously, sending the information about their location and speed to the controllers in real time. Depending on the technology used, position estimates can be provided with accuracies of 1–3 meters. A desirable data communication rate for successful implementation of TSP strategies is on the order of once per second but this requires intensive and expensive communication systems. However, even with lower data input rates, accurate prediction models can be used to estimate a vehicle’s arrival time at the stop line as well as its schedule delay. Collection of historical data on transit vehicle positions from GPSs can also be used to estimate dwell times at transit stops which, along with the information on transit vehicle speeds, can lead to accurate estimation of travel times. Dwell time and travel time estimates can then be used along with the real-time position of a transit vehicle to predict its arrival time at the intersection under consideration with improved accuracy. In addition, such systems can provide real-time information about the schedule delay of a transit vehicle arriving at an intersection by comparing the actual arrival time with the schedule.

In the absence of GPSs, other AVL systems such as radio control and loop inductors that recognize transit vehicle signatures can be used to detect transit vehicles, determine their position, and identify their schedule delay. Such sensing systems require transit vehicles to be equipped with a transponder through which they can be identified by the system. Loop inductors are used by Los Angeles Department of Transportation (LADOT) for their centralized TSP implementation and are described in more detail in Li *et al.* (2008).

AVL systems are used extensively by transit agencies mainly for system planning and evaluation purposes. One such example from the San Francisco Bay Area is the San Mateo County Transit District which has equipped all of its buses with GPSs that are used for both visual and voice next-stop announcements on-board and provision of information to transit users off-board (Menczer *et al.*, 2006).

Real-time information about auto passenger occupancies is currently not available and only historical data can provide estimates of average occupancy per auto. Such estimates usually vary slightly from day to day and by time of the day as well as among different intersections or arterials. A typical range of values is 1.2 to 1.5 passengers per vehicle. In the future with the introduction of Connected Vehicle technology, it will be possible to have accurate information about passenger occupancies of individual autos. Transit passenger occupancies are expected to be more variable and since the proposed optimization system depends highly on the number of people on-board, real-time information is highly desirable. Advanced technologies are available that allow for real-time information on the number of passengers boarding and alighting at transit stops. These technologies use mixed infrared sensors, or stereoscopic cameras located near the doors that are able to determine the direction of passenger traffic

and thus estimate how many people are exiting or entering with an accuracy of 95–98%. INFODEV Automatic Passenger Counting, ACOREL Onboard Counter, and EUROTECH Passenger Counter are examples of vendors for Automated Passenger Counter (APC) systems. Such systems can be connected to AVL systems to transmit data in real time.

APC systems have been utilized by agencies such as the OC Transpo in Ottawa, Canada, the Regional Transportation District in Denver, Colorado, the Tri-Met (Tri-county Metropolitan) Transportation District in Portland, Oregon, and in many other cities in the United States and worldwide (Furth *et al.*, 2006). As with the AVL systems, APCs can be used not only for real-time signal control purposes but are often installed to gather information for planning and management decisions (e.g., to determine the optimal frequency of buses in the system). The San Francisco Municipal Transportation Agency (SFMTA) has used APCs to collect data on Muni’s ridership in San Francisco in order to evaluate the efficacy of the system (SFMTA, 2011).

In cases that APC systems are not available, other methods can be used to obtain estimates of transit passenger occupancies to be used as input for the optimization of the traffic signal control system. For example, ticket validation methods can provide information for the number of passengers that are boarding at each transit stop. Combined with historical data (e.g., obtained through observations of alighting passengers), such information can be used to estimate transit vehicle passenger occupancies. Even if no real-time data is available, historical data of passenger occupancies for specific routes and stops can be used to determine the relative level of priority for transit vehicles arriving from conflicting directions.

Estimates of the costs of the required surveillance technologies along with the installation location for each of them and a description of the data input that they provide for the optimization are presented in Table 3.1. Communication technologies have not been reported separately here since it has been assumed that they are part of the surveillance systems.

3.3 Performance Measures

Several performance measures are used to evaluate the impacts of the proposed person-based traffic responsive signal control system on auto and transit passengers at the intersection, arterial, crossstreet, and system levels:

- Time measures (e.g., person delay, vehicle delay)
- Operational measures (e.g., number of stops)
- Environmental measures (e.g., emissions)

These performance measures are used to compare three different strategies. First, the signal settings (phase green times and offsets in the case of signalized arterials) are optimized with TRANSYT-7F (McTrans, 2003). TRANSYT-7F is the most widely-used software package for optimizing signal settings and it does this by minimizing

Table 3.1. System Technology Requirements and Costs

Surveillance Technology	Problem Level	Location on Link	Data Input (for optimization)	Cost ^a (per unit)
Detectors	Isolated Int.	Upstream	Demand Flows	\$500–1,000
	Arterial	Upstream	Platoon Sizes	
	Isolated Int. & Arterial	Downstream	Turning Ratios	
	Arterial	Downstream & Upstream	Travel Times	
AVL Systems	Isolated Int. & Arterial	On-board	Transit Arrival Times	\$600 ^b –8,000
APC Systems	Isolated Int. & Arterial	On-board	Transit Passenger Occupancies	\$10,000–15,000

^a Information was obtained from Schweiger (2003) and RITA (2012b).

^b The lowest price corresponds to systems that report location through cell phones.

delay and stops. Next, the above mentioned performance measures are obtained from testing the vehicle-based optimization. Vehicle-based optimization utilizes the same mathematical program as the proposed person-based optimization but it does not weigh delays of cars and bus by their respective passenger occupancies or schedule delay factors. Therefore, it is essentially a traffic responsive signal control system that optimizes signal settings by minimizing vehicle delay while accounting for changes in traffic demand. Finally, the proposed person-based optimization is tested and evaluated.

3.4 Testing and Evaluation

Evaluation of the proposed person-based traffic responsive signal control system described in Section 3.1 has been performed with three types of tests that vary in their assumptions regarding vehicle arrivals and demand profiles.

Test Type I: Deterministic arrival tests with constant auto demand

Tests of type I are performed under the assumption of deterministic arrivals for constant auto traffic demand that corresponds to undersaturated traffic conditions. Deterministic arrivals represent tests for cases where perfect information is available for the auto traffic demand, the auto and transit vehicle arrivals and passenger occupancies of transit vehicles.

Test Type II: Deterministic arrival tests with time-dependent auto demand

Perfect information about vehicle demands, arrival, and passenger occupancies are also assumed for tests of type II that are performed for a time-dependent auto demand profile that changes once every cycle. While the demand profile is time-dependent, the rate of arrivals within each cycle is still constant. The main difference compared to type I is that some of the cycles operate in oversaturated traffic conditions. As a result, the behavior of the system can be tested for congested traffic conditions.

Test Type III: Stochastic arrival tests with constant auto demand

Tests of type III include stochastic arrival tests that are performed through simulation for undersaturated traffic conditions under the assumption of exponentially distributed vehicle arrivals. The simulation experiments are performed with the microscopic simulation software, AIMSUN (Transport Simulation Systems, 2010). In particular, Emulation-In-the-Loop Simulation (EILS) is used to better represent the behavior of the proposed signal control system in a real-world environment. EILS consists of using an Application Programming Interface (API) that models traffic control so that the signal control system is tested in an environment that emulates its operation in a real-world setting (Stevanovic & Martin, 2007).

The evaluation platform for the simulation tests, shown in Figure 3.1, consists of the AIMSUN model with API and various solvers in MATLAB (The MathWorks, 2009) depending on the type of mathematical program being solved for isolated intersections and arterials. API is used to control the simulation and its code can be written in C++ or Python. For the simulation tests performed in this dissertation the code is written in C++. Since the proposed system is traffic responsive, the optimization of signal settings is performed at the end of each cycle in order to obtain the optimal signal settings for the next cycle. As a result, the API is called at the end of each cycle to read vehicle data from the available sensing systems in the simulated network, such as simulated detectors and AVL. Once the data are collected, the API determines the input for the optimization by estimating the auto traffic demand and passenger occupancies and predicting auto and transit vehicle arrival times. This input is then imported to MATLAB which optimizes the signal settings. After retrieving the optimal signal settings, the API returns them to the traffic controllers in AIMSUN to be implemented during the next cycle.

In a nutshell, EILS is used to test the performance of the developed signal control optimization in more realistic traffic conditions where vehicles do not arrive deterministically and where errors exist in the prediction of the auto and transit vehicle arrivals. The formulation of the mathematical program used for the optimization is still based on the assumption of perfect information for the input. As a result, tests of type III examine how well the system performs when estimates or predictions are used as input to the optimization when perfect information is not available. The simulation has the additional advantage of evaluating the system on the basis of several performance measures as described in Section 3.3.

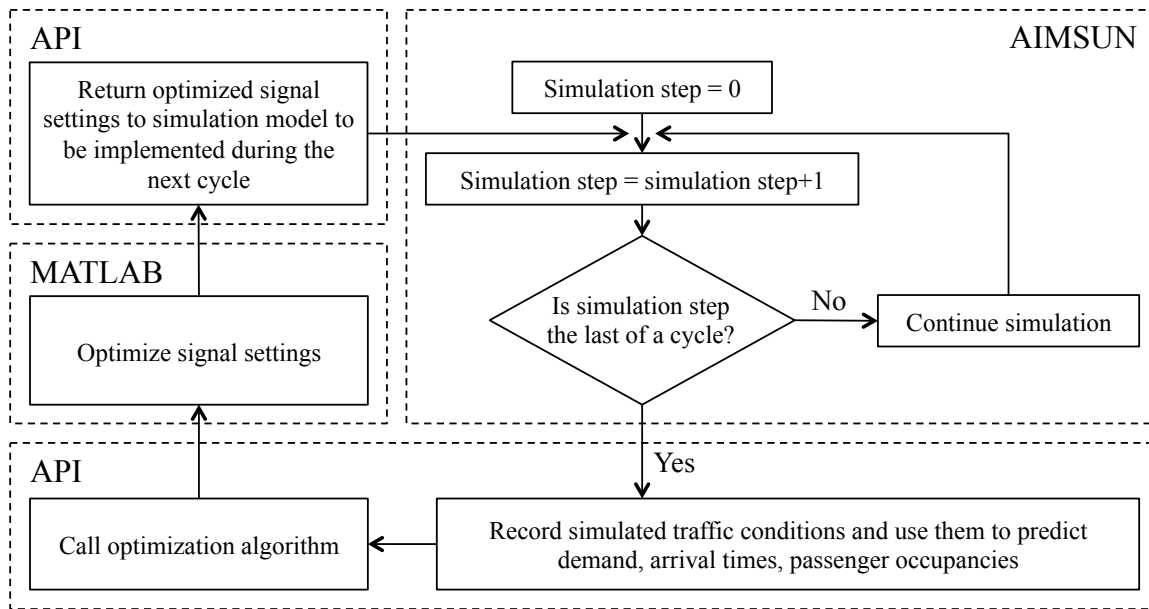


Figure 3.1. Emulation-In-the-Loop Simulation Platform

These three types of tests are performed for the isolated intersection and the arterial cases for a variety of traffic conditions and transit operating characteristics (e.g., auto traffic demand, transit passenger occupancy, etc.) for a one hour period of traffic operations. In this way, one can identify auto demand and passenger occupancy thresholds for which implementation of such a system is beneficial to all users. The outcomes of the tests are presented in Sections 4.4 for isolated intersections and in 5.5 for arterials.

3.5 Summary of Research Approach

A mathematical program is formulated to minimize the total person delay at an intersection accounting for delays of both auto and transit passengers. The mathematical program minimizes total person delay by weighting delays for autos and transit vehicles by their respective passenger occupancies. In addition, transit vehicle delays are weighted by an extra factor that accounts for their schedule delay. In that way, the delays for transit vehicles that are late are weighted by more than those that are early or on time.

Implementation of the proposed system in real-world settings is feasible with the use of existing and deployable surveillance and communication technologies, such as detectors, AVL and APC systems, and their respective communication requirements for real-time operations. Despite the cost of such technologies one should consider that many of these could already be in place to provide data for other planning or

management purposes. For example, AVL and APC systems are useful for transit operations planning and the provision of real-time information at transit stops and online about transit vehicle locations and expected arrivals.

The system is evaluated with the use of performance measures that include time measures (e.g., person and vehicle delays), operational measures (e.g., number of stops), and environmental measures (e.g., emissions) for all users at the intersection, arterial, cross street, and network levels. Several types of tests are performed for this purpose. These include tests with deterministic arrivals, where it is assumed that perfect information is available about the vehicle demand, arrivals, and passenger occupancies and others with stochastic arrivals that are performed through EILS. Simulation tests are used as an alternative to field implementation which is costly and time consuming even if politically and technologically feasible. A variety of traffic conditions and transit operating characteristics are used to test the performance of the system at isolated intersections and arterials.

First, the mathematical program is formulated for an isolated intersection using the same assumptions presented here and supposing that autos arrive at constant rates within a cycle (Chapter 4). Once the performance of the system at an isolated intersection is well understood, it is extended to signalized arterials where autos arrive in platoons at an intersection, influenced by upstream signals. Extensions of the signalized arterial mathematical program to the optimization of signal settings on arterial networks are also discussed (Chapter 5).

Chapter 4

Isolated Intersection

The mathematical program is first formulated and tested for the case of an isolated intersection. The main assumption for the isolated intersection is that the vehicle arrival pattern is not affected by the signal timings of upstream intersections. All assumptions stated in Section 3.1 hold. The mathematical program formulated under these assumptions for the optimization of the signal timings at an isolated intersection is a Mixed Integer Non-Linear Program (MINLP).

This chapter presents in detail the MINLP that is used to minimize total person delay for all auto and transit users at an intersection. This includes a description of how delays for autos and transit vehicles are estimated for the objective function. Section 4.1 presents the mathematical program that is developed first for undersaturated traffic conditions. Then it is extended to handle to oversaturated traffic conditions when demand exceeds available capacity as described in Section 4.2. Section 4.3 describes the two study sites used for the evaluation of the system. Then, Section 4.4 presents the findings from the system evaluation on the study sites under a variety of traffic conditions and transit operating characteristics for the types of tests described in Section 3.4. Finally, Section 4.5 summarizes the findings and insights obtained from the analysis of the test results.

4.1 Undersaturated Traffic Conditions

First, the mathematical program is formulated for an isolated intersection when undersaturated traffic conditions prevail and where the arrival rate of autos is constant for all cycles. Undersaturated traffic conditions correspond to traffic operations where demand does not exceed capacity for any of the lane groups at the intersection under consideration. This means that the following equation holds for any lane group j that can be served by phase i at an intersection and whose vehicles arrive at a constant rate of q_j and can be served at a rate of s_j during a cycle T :

$$\frac{q_j}{s_j} \leq \frac{g_{i,T}}{C} \quad (4.1)$$

where $g_{i,T}$ is the green time allocated to phase i for that cycle, and C is the cycle length. For simplicity the intersection index r has been dropped from the notation in this chapter, because only one intersection is considered. A detailed estimation of auto and transit vehicle delays that comprise the objective function of the mathematical program for undersaturated conditions and the final MINLP used to minimize total person delay are presented next.

4.1.1 Auto Delay

The person delay for the auto passengers included in the objective function for undersaturated traffic conditions consists of the sum of two terms: 1) the person delay that corresponds to the autos that will be served during the design cycle, T , and 2) an estimate of the delay for those that will be served during cycle $T + 1$. The delay estimate is included in the objective function to account for the impact that the design of the signal timings in cycle T will have on the delays of $T + 1$. If the expected measure of delay were not included, the optimized signal timings for the design cycle would provide the minimum green times to all the phases apart from the last one, and this would substantially increase auto delay for the next cycle.

The effect of considering delay estimates in cycle $T + 1$ is illustrated through a simple example. Consider a 2-phase intersection of two one-way streets (Figure 4.1). The eastbound vehicles arrive at a rate of q_1 and are served by phase 1 at a rate of s_1 while the southbound vehicles arrive at a rate of q_2 and are served by phase 2 at a rate of s_2 . The queueing diagrams shown in Figure 4.1 represent the cumulative arrivals and departures of vehicles over time at both approaches for cycle T , which is the cycle currently being optimized, and the next cycle, $T + 1$. The areas of the triangles between the arrival and departure curves in the figure represent the delays for eastbound and southbound traffic under the different signal timing scenarios.

If phase 1 is truncated to give priority to a bus arriving at approach N before the end of its green time (advance of phase 2: case ii), the delays for the vehicles served by phase 1 during the next cycle, $T + 1$, will be increased (triangle ii for the eastbound autos), while the delays for the vehicles at approach N will be decreased for the design cycle T , (triangle ii for the southbound autos). If phase 1 is extended (extension of phase 1: case iii) to provide priority to a transit vehicle arriving after the end of its initial green time, both the delays for that transit vehicle and the delays for the autos that would otherwise be served during phase 1 in the next cycle, $T + 1$, are reduced (triangle iii for the eastbound autos). However, the delays for the autos at approach N that are served by phase 2 in the design cycle, T , are increased as a result of the longer red time interval they experience (triangle iii for the southbound autos). As long as the delay gains during the design cycle, T , are accounted for in the objective function, the losses caused to the vehicles served by the other phase need to also be counted for, whether those vehicles end up being served during the design cycle or the next one. Therefore, the trade-off between the delays for the different approaches is fully captured.

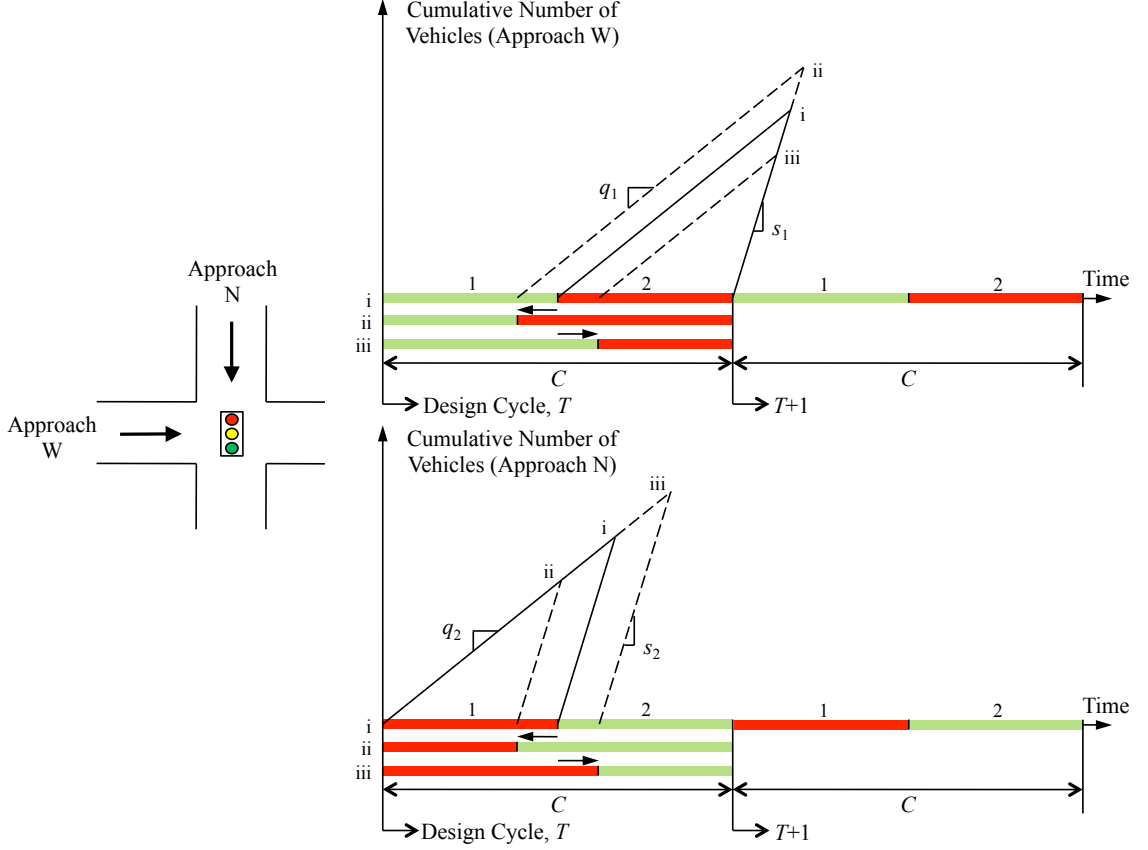


Figure 4.1. Impact of Changes in Signal Timings on Auto Delays

The auto delay is calculated based on the cycle length, green ratio, saturation flow, and arrival rate that is assumed to be constant. The red time interval for a lane group j is $R_j = C - G_j^e$, and the green ratio is $\lambda_j = G_j^e/C$, where G_j^e is the summation of the effective green times for all the phases that can serve lane group j . Further assuming that the vehicles belonging to lane group j arrive at a rate of q_j and are served at a saturation flow of s_j , their total delay, D_j , for one cycle is given by:

$$D_j = \frac{1}{2} \frac{q_j C^2 (1 - \lambda_j)^2}{(1 - \frac{q_j}{s_j})} \quad (4.2)$$

where $C(1 - \lambda_j)$ is the red time allocated to lane group j . Figure 4.2 illustrates the delay for the vehicles of a lane group j , showing the cumulative number of vehicles present at an intersection for cycles $T - 1$, T , and $T + 1$ for that lane group. The cumulative count restarts at the end of the green phase that serves the lane group.

The cycle time for each lane group can be split into three components which are functions of the green times for each phase. The first is the component of the red time from the start of the cycle to the beginning of the green for the subject lane group, $R_j^{(1)}(g_{i,T})$, the second is the duration of the effective green time itself, $G_j^e(g_{i,T})$,

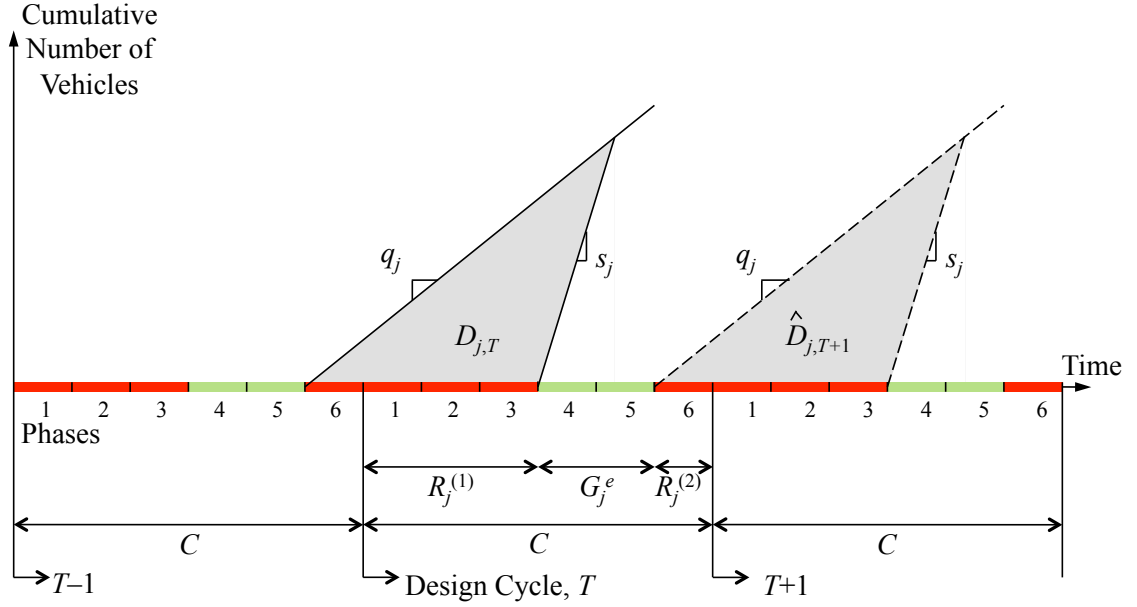


Figure 4.2. Queueing Diagram for Lane Group j for Undersaturated Conditions (Auto Delay)

and the third is the component of red time from the end of the green until the end of the cycle, $R_j^{(2)}(g_{i,T})$. These values are illustrated in Figure 4.2 and are calculated as follows:

$$R_j^{(1)}(g_{i,T}) = \sum_{i=1}^{k_j-1} g_{i,T} + \sum_{i=1}^{k_j-1} y_i \quad (4.3a)$$

$$G_j^e(g_{i,T}) = \sum_{i=k_j}^{l_j} g_{i,T} + \sum_{i=k_j}^{l_j-1} y_i \quad (4.3b)$$

$$R_j^{(2)}(g_{i,T}) = \sum_{i=l_j+1}^I g_{i,T} + \sum_{i=l_j}^I y_i \quad (4.3c)$$

where:

$g_{i,T}$: green time for phase i in cycle T

l_j : the last phase in a cycle that can serve lane group j

k_j : the first phase in a cycle that can serve lane group j .

According to the figure, lane group j can be served by phases 4 and 5, so its effective green time is: $G_j^e = g_4 + g_5$. The shaded area between the solid lines represents the total delay experienced by the autos of lane group j that are served during the design cycle, T , denoted by $D_{j,T}$. The shaded area between the dashed lines represents the estimate of the total delay experienced by the autos of lane group j that are served during the next cycle, $T + 1$, denoted by $\hat{D}_{j,T+1}$. The delay for a

lane group j for cycle T is counted from the end of the green phase that served j in cycle $T - 1$ until the end of the corresponding green phase in T . As a result, the signal timings for the previous cycle, $T - 1$, must be known in order to determine the delays of the vehicles that arrive at the intersection during cycle $T - 1$ but will be served during the design cycle, T . Such queueing diagrams can be drawn for each lane group to estimate the delay for autos under the assumption of First-In-First-Out (FIFO) queueing discipline.

The calculation of auto delays for cycles T and $T + 1$ are presented below along with examples to illustrate how the red times are determined.

Auto delay for cycle T

The total delay for all autos of lane group j that are served during cycle T , $D_{j,T}$, is derived from (4.2) as follows:

$$D_{j,T} = \frac{1}{2} \sum_{j=1}^J \frac{q_j}{1 - \frac{q_j}{s_j}} \left(R_j^{(2)}(g_{i,T-1}) + R_j^{(1)}(g_{i,T}) \right)^2 \quad (4.4)$$

where J is the total number of lane groups at the intersection.

EXAMPLE: Vehicles in lane group j , that can be served by phases 4 and 5, experience red time equal to the sum of the green and yellow time of phase 6 in the previous cycle, $T - 1$, and the green and yellow times of phases 1–3 in the design cycle, T (Figure 4.2).

Auto delay for cycle $T + 1$

The estimate of the total delay for all autos that will be served during cycle $T + 1$, $\hat{D}_{j,T+1}$, is derived from (4.2) as follows:

$$\hat{D}_{j,T+1} = \frac{1}{2} \sum_{j=1}^J \frac{q_j}{1 - \frac{q_j}{s_j}} \left(R_j^{(2)}(g_{i,T}) + R_j^{(1)}(g_{i \text{ next}}) \right)^2 \quad (4.5)$$

where $g_{i \text{ next}}$ is a user-specified estimate for the green time of phase i during cycle $T + 1$. This may be a base case signal timing or an optimal fixed signal timing that has been determined offline.

EXAMPLE: Vehicles in lane group j , that will be served by phases 4 and 5 in the next cycle $T + 1$, experience red time equal to the summation of the green and yellow time of phase 6 in the design cycle T and the green and yellow times of phases 1–3 in cycle $T + 1$ (Figure 4.2).

As a result of these delay components, the first part of the objective function (auto

person delay) for one intersection becomes:

$$\sum_{a=1}^{A_T} o_a d_{a,T} = \bar{o}_a \sum_{j=1}^J \left(D_{j,T} + \hat{D}_{j,T+1} \right) \quad (4.6)$$

where $D_{j,T}$ and $\hat{D}_{j,T+1}$ depend on the decision variables $g_{i,T}$ as shown in (4.4) and (4.5). An average value of passenger occupancy per auto, \bar{o}_a , is used because total vehicle delay is calculated collectively rather than accounting for each vehicle separately. However, the delays experienced by any individual vehicle could be easily estimated with the use of queueing diagrams such as the one in Figure 4.2, given that the arrival time of the vehicle at the back of the queue is known. This is the approach used to estimate transit vehicle delays as shown next.

4.1.2 Transit Delay

The person delay for the transit passengers is the sum of two terms: 1) the person delay that corresponds to the transit vehicles that are served during the design cycle, T , and 2) an estimate of the delay for those that arrive in T but are served in $T + 1$. Under the assumption that information about the arrival times of transit vehicles at the intersection is known only for the design cycle, transit vehicles that arrive at the intersection during cycle $T + 1$ are not taken into account. The exclusion of such vehicles is not expected to significantly affect the performance of the system, because these vehicles will be considered when the signal settings for cycle $T + 1$ are designed.

Transit vehicles travel in mixed traffic lanes with the autos, so the delay of a transit vehicle, b , that arrives in its lane group's queue at some time, t_b , is the same as an auto that arrives at the same time at the back of that lane group's queue. The delay for a transit vehicle that belongs to a lane group j can be calculated by the same queueing diagrams used for auto delay estimation (Figure 4.3). The estimation of the transit delay used in the optimization of each cycle T depends on the actual arrival time of the transit vehicle, t_b , relative to the end of the last phase that can serve its lane group, j , in cycle $T - 1$ and the end of the last phase that can serve j in cycle T . The end of the last phase that can serve j in T can be expressed as:

$$\tau_{j,T} = (T - 1)C + R_j^{(1)}(g_{i,T}) + G_j^e(g_{i,T}). \quad (4.7)$$

The possible cases are summarized next.

Case 1: Arrival before the end of green in T

If a transit vehicle, b , that belongs to lane group j has arrived in the previous cycle, $T - 1$, at some time after the end of the last phase that can serve j ($t_b \geq \tau_{j,T-1}$) or arrives in the design cycle, T , before the end of the phases that can serve it ($t_b \leq \tau_{j,T}$), its delay for cycle T , $d'_{b,T}$, depends on the transit arrival time, t_b , and green times,

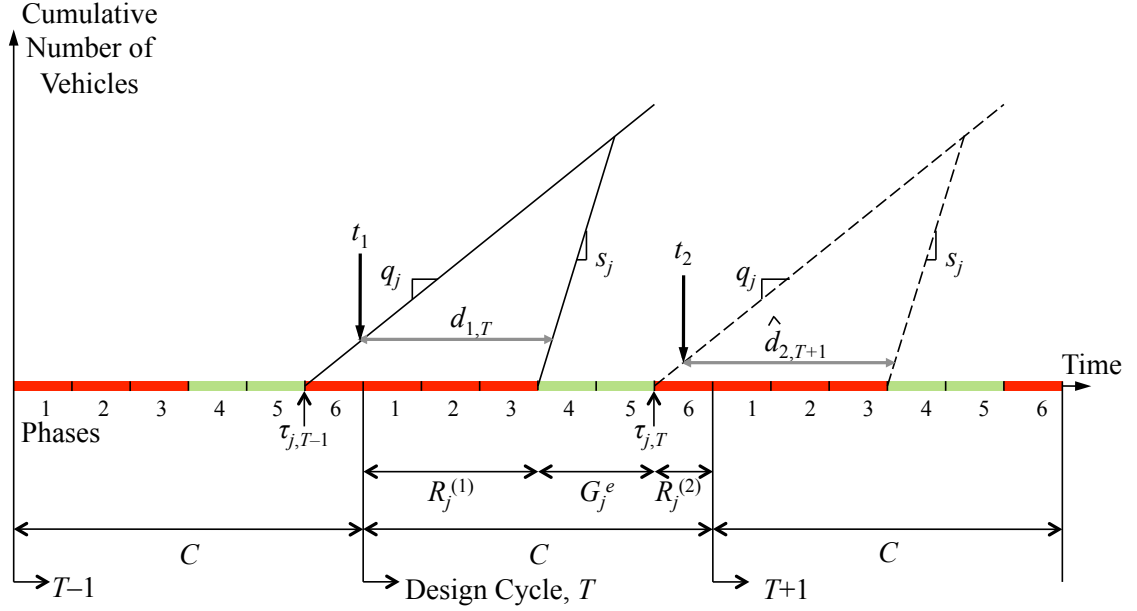


Figure 4.3. Queueing Diagram for Lane Group j for Undersaturated Conditions (Transit Delay)

$g_{i,T}$. This delay is expressed as:

$$d'_{b,T} = (T - 1)C + R_j^{(1)}(g_{i,T}) + \frac{q_j}{s_j} (t_b - \tau_{j,T-1}) - t_b. \quad (4.8)$$

If the transit vehicle arrives before the clearance of its lane group's queue, (4.8) will give a positive delay. Define this as time interval α :

$$\alpha = \{t_b | \tau_{j,T-1} \leq t_b < \tau_{j,T}, d'_{b,T} \geq 0\}. \quad (4.9)$$

However, if the transit vehicle arrives after the clearance of that queue and at a time within the phases that can serve it, (4.8) will give a negative delay, which implies that the true delay for such a transit vehicle will be zero. Define this as time interval β :

$$\beta = \{t_b | \tau_{j,T-1} \leq t_b < \tau_{j,T}, d'_{b,T} < 0\}. \quad (4.10)$$

Therefore, the transit vehicle delay, $d_{b,T}$ for this case is expressed as:

$$d_{b,T} = \max\{d'_{b,T}, 0\}. \quad (4.11)$$

EXAMPLE: If a bus that belongs to lane group j arrives during phase 6 of cycle $T - 1$, (e.g., $t_b = t_1$ in Figure 4.3), or phases 1, 2, or 3 of cycle T , it will be served during cycle T , and its delay is indicated on the queueing diagram as $d_{1,T}$.

Case 2: Arrival after the end of green in T

If a transit vehicle, b , that belongs to lane group j arrives during cycle T after the last phase that can serve its respective lane group ($t_b > \tau_{j,T}$), the transit vehicle will be served during the next cycle, $T + 1$. Define this as time interval γ :

$$\gamma = \{t_b | t_b > \tau_{j,T}\}. \quad (4.12)$$

In this case, the objective function includes an estimate of the delay that the transit vehicle would experience if it arrives after the green that can serve it. In order to estimate this delay, an assumption for the green times for the next cycle, $g_{i \text{ next}}$, must be made. The estimate of the delay for such a transit vehicle, $\hat{d}_{b,T}$, is given by:

$$\hat{d}_{b,T} = TC + R_j^{(1)}(g_{i \text{ next}}) + \frac{q_{j,T}}{s_j} (t_b - \tau_{j,T}) - t_b. \quad (4.13)$$

EXAMPLE: If a bus that belongs to lane group j arrives during phase 6 in cycle T , (e.g., $t_b = t_2$ in Figure 4.3), it can either be served by phase 5 of the design cycle, T , if it is possible to extend the green by a sufficient amount to serve the bus, or it will be served during phase 4 of the next cycle, $T + 1$. The objective function for optimizing cycle T includes an estimate of the delay that the bus would experience if it was to be served during the next cycle, $T + 1$, and its total delay is indicated on the queueing diagram as $\hat{d}_{2,T}$.

4.1.3 Mathematical Program Formulation

As described by the equations above, the mathematical program that minimizes the person delay for auto and transit users at a signalized intersection for one cycle is a Mixed Integer Non-Linear Program (MINLP). The integer variables are introduced due to the different delay formulas that correspond to each of the three time intervals in which a transit vehicle could possibly arrive (α, β, γ). As a result, for each transit vehicle, b , considered in the optimization, there are three binary variables introduced ($w_b^\alpha, w_b^\beta, w_b^\gamma$), where $w_b^f = 1$ if $t_b \in f$, otherwise $w_b^f = 0$, for $f = \{\alpha, \beta, \gamma\}$. A summary of the formulation is shown below:

Objective Function (Auto person delay component):

$$\bar{o}_a \frac{1}{2} \sum_{j=1}^J \frac{q_{j,T}}{1 - \frac{q_{j,T}}{s_j}} \left[\left(R_j^{(2)}(g_{i,T-1}) + R_j^{(1)}(g_{i,T}) \right)^2 + \left(R_j^{(2)}(g_{i,T}) + R_j^{(1)}(g_{i \text{ next}}) \right)^2 \right] \quad (4.14)$$

Objective Function (Transit person delay component):

$$\sum_{b=1}^{B_T} o_{b,T} \left[w_b^\alpha \left((T-1)C + R_j^{(1)}(g_{i,T}) + \frac{q_j}{s_j} (t_b - \tau_{j,T-1}) - t_b \right) + w_b^\gamma \left(TC + R_j^{(1)}(g_{i \text{ next}}) + \frac{q_j}{s_j} (t_b - \tau_{j,T}) - t_b \right) \right] \quad (4.15)$$

Constraints:

$$-(1 - w_b^\alpha)M_1 \leq (T - 1)C + R_j^{(1)}(g_{i,T}) + \frac{q_j}{s_j} (t_b - \tau_{j,T-1}) - t_b \quad \forall b \quad (4.16)$$

$$w_b^\alpha M_1 \geq (T - 1)C + R_j^{(1)}(g_{i,T}) + \frac{q_j}{s_j} (t_b - \tau_{j,T-1}) - t_b \quad \forall b \quad (4.17)$$

$$(1 - w_b^\gamma)t_b \leq \tau_{j,T} \quad \forall b \quad (4.18)$$

$$(1 - w_b^\gamma)M_2 + w_b^\gamma t_b \geq \tau_{j,T} \quad \forall b \quad (4.19)$$

$$G_j^e(g_{i,T}) \geq g_{j \min} \quad \forall j \quad (4.20)$$

$$\sum_{i=1}^I g_{i,T} + \sum_{i=1}^I y_i = C \quad (4.21)$$

$$g_{i,T} \geq g_{i \min} \quad \forall i \quad (4.22)$$

$$g_{i,T} \leq g_{i \max} \quad \forall i \quad (4.23)$$

$$w_b^\alpha + w_b^\beta + w_b^\gamma = 1 \quad \forall b \quad (4.24)$$

$$w_b^\alpha, w_b^\beta, w_b^\gamma \in \{0, 1\} \quad \forall b \quad (4.25)$$

where $g_{i \min}$ is the lower bound and $g_{i \max}$ is the upper bound for the green time of each phase i and M_1 , M_2 are big numbers that are set equal to C and TC , respectively. The constraints are described as follows:

- Constraints (4.16)–(4.19) ensure that the correct delay formula will be added to the objective function for each of the transit vehicles present at the intersection during the design cycle, T .
- Constraint (4.20) refers to the minimum green times for each lane group. Minimum green times are necessary to ensure undersaturated conditions for each lane group; i.e., $g_{j \min} = Cq_{j,T}/s_j$. In addition, they ensure safe pedestrian crossings and guarantee that no phase is skipped.
- Constraint (4.21) ensures that the green times for each phase, which will be the outcome of the optimization plus the lost time, which is essentially the sum of the yellow times, add up to the cycle length.
- Constraints (4.22) and (4.23) set the upper and lower bounds for the decision variables.
- Constraints (4.24) and (4.25) ensure that only one binary variable will be equal to one.

Note that the formulation of the mathematical program above leads to bilinearities (i.e., non-convexity in the objective function) due to the multiplication of the

continuous variables ($g_{i,T}$) with the integer variables ($w_b^\alpha, w_b^\beta, w_b^\gamma$). In order to avoid this problem, three new continuous variables ($g_{i,b}^\alpha, g_{i,b}^\beta, g_{i,b}^\gamma$) are introduced for each phase and for each of the transit vehicles whose delays are included in the objective function (Floudas, 1995). The initial continuous decision variables, $g_{i,T}$, are now defined as:

$$g_{i,T} = g_{i,b}^\alpha + g_{i,b}^\beta + g_{i,b}^\gamma \quad \forall i, b \quad (4.26)$$

where:

$$g_{i,b}^\beta = g_{i,b}^\gamma = 0 \quad \forall i \quad \text{if } t_b \in \alpha, \quad (4.27)$$

$$g_{i,b}^\alpha = g_{i,b}^\gamma = 0 \quad \forall i \quad \text{if } t_b \in \beta, \quad (4.28)$$

$$g_{i,b}^\alpha = g_{i,b}^\beta = 0 \quad \forall i \quad \text{if } t_b \in \gamma. \quad (4.29)$$

As a result, the transit person delay component of the objective function can be rewritten as:

$$\begin{aligned} \sum_{b=1}^{B_T} o_{b,T} \left[w_b^\alpha \left((T-1)C + \sum_{i=1}^{k_j-1} y_i + \frac{q_j}{s_j} (t_b - \tau_{j,T-1}) - t_b \right) \right. \\ \left. + w_b^\gamma \left(TC + R_j^{(1)}(g_{i \text{ next}}) + \frac{q_j}{s_j} \left(t_b - (T-1)C - \sum_{i=1}^{l_j-1} y_i \right) - t_b \right) \right. \\ \left. + \sum_{i=1}^{k_j-1} g_{i,b}^\alpha - \frac{q_{j,T}}{s_j} \sum_{i=1}^{l_j} g_{i,b}^\gamma \right] \quad (4.30) \end{aligned}$$

and constraints (4.22) and (4.23) are replaced by:

$$g_{i,b}^f \geq w_b^f g_{i \text{ min}} \quad \forall i, \forall f \in \{\alpha, \beta, \gamma\} \quad (4.31)$$

$$g_{i,b}^f \leq w_b^f g_{i \text{ max}} \quad \forall i, \forall f \in \{\alpha, \beta, \gamma\}. \quad (4.32)$$

4.2 Oversaturated Traffic Conditions

Oversaturated traffic conditions occur when demand exceeds capacity and residual queues form for one or more of the lane groups at the intersection under consideration. This means that the following equation holds for a lane group j that can be served by phase i at an intersection and whose vehicles arrive at a rate of $q_{j,T}$ and are served at a rate of s_j during a cycle T :

$$\frac{q_{j,T}}{s_j} > \frac{g_{i,T}}{C}. \quad (4.33)$$

The formulation of the mathematical program that minimizes person delay at an isolated intersection for undersaturated conditions is extended to capture delays when oversaturated traffic conditions occur. This formulation is also generalized to allow the arrival rate $q_{j,T}$ to change from cycle to cycle.

4.2.1 Auto Delay

As in Section 4.1.1 the person delay for the auto passengers included in the objective function for oversaturated traffic conditions consists of the sum of two terms: 1) the person delay experienced by autos that are present at the intersection during the design cycle, T , before the end of their respective green times, and 2) an estimate of the person delay experienced by autos that are present at the intersection during the next cycle, $T + 1$, before the end of their respective green times. Both terms include the delays of the vehicles that remain in residual queues. Delays that the auto passengers experience during the next cycle, $T + 1$, are included for the same reason as in the undersaturated case, to account for the impact that the design of the signal timings in the current cycle will have on the delays of the next one. As in Section 4.1.1, the delay for a lane group j is counted from the end of the green phase that serves j in cycle $T - 1$ until the end of the respective green phase in cycle T .

In order to estimate the delay experienced by the autos in oversaturated traffic conditions, the number of autos in the residual queues for each of the lane groups must be estimated. The value $N_{j,T}$ denotes the number of autos in the residual queue of lane group j at time $\tau_{j,T}$, and this is calculated as follows:

$$N_{j,T} = N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T} \left(R_j^{(1)}(g_{i,T}) + G_j^e(g_{i,T}) \right) - G_j^e(g_{i,T})s_j \quad (4.34)$$

where all variables are defined as before.

The delay for autos can be estimated with queueing diagrams as explained in Section 4.1.1. Figure 4.4 represents the cumulative number of vehicles present at an intersection for cycles $T - 1$, T , and $T + 1$ for a lane group j that can be served by phases 4 and 5. The cumulative count restarts at the end of the green phase that serves the lane group and includes the residual queue. The shaded area between the solid lines represents the total delay experienced by the autos that belong to lane group j and are present at the intersection from the end of the green time for j in the previous cycle, $T - 1$, until the end of the green time for j in the design cycle, T , denoted by $D_{j,T}$. The other shaded area between the dashed lines represents an estimate of the delay that the autos of lane group j will experience from the end of the green time for j in the design cycle T , until the end of the green for j in the next cycle $T + 1$, denoted by $\hat{D}_{j,T+1}$.

The calculation of auto delays for cycles T and $T+1$ that can handle oversaturated traffic conditions are presented next.

Auto delay for cycle T

If the autos of lane group j experience oversaturated traffic conditions during cycle T (i.e., $N_{j,T} > 0$), their total delay, $D_{j,T}$, is given by:

$$D_{j,T} = \frac{1}{2} \left[2N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) \right] R_j^{(2)}(g_{i,T-1}) \\ + \frac{1}{2} \left[2N_{j,T-1} + 2q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T}R_j^{(1)}(g_{i,T}) \right] R_j^{(1)}(g_{i,T})$$

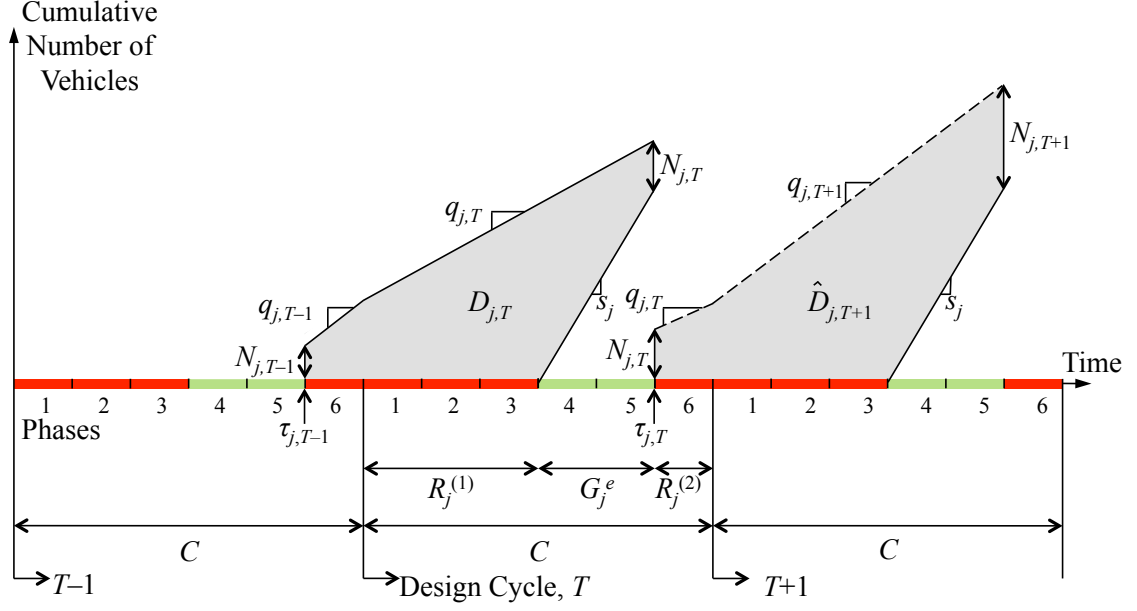


Figure 4.4. Queueing Diagram for Lane Group j for Oversaturated Conditions (Auto Delay)

$$\begin{aligned}
& + \left[N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T}R_j^{(1)}(g_{i,T}) \right] G_j^e(g_{i,T}) \\
& + \frac{1}{2} (q_{j,T} - s_j) G_j^e(g_{i,T})^2.
\end{aligned} \tag{4.35}$$

If the autos of lane group j experience undersaturated traffic conditions during cycle T (i.e., $N_{j,T} \leq 0$), the third and fourth terms of (4.35) change and the total delay for the autos of lane group j , $D_{j,T}$, becomes:

$$\begin{aligned}
D_{j,T} & = \frac{1}{2} \left[2N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) \right] R_j^{(2)}(g_{i,T-1}) \\
& + \frac{1}{2} \left[2N_{j,T-1} + 2q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T}R_j^{(1)}(g_{i,T}) \right] R_j^{(1)}(g_{i,T}) \\
& + \frac{1}{2(s_j - q_{j,T})} \left[N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T}R_j^{(1)}(g_{i,T}) \right]^2.
\end{aligned} \tag{4.36}$$

Auto delay for cycle $T + 1$

If the autos of lane group j experience oversaturated traffic conditions during cycle T (i.e., $N_{j,T} > 0$), the estimate of their total delay for cycle $T + 1$, $\hat{D}_{j,T+1}$ is given by:

$$\begin{aligned}
\hat{D}_{j,T+1} & = \frac{1}{2} \left[2N_{j,T} + q_{j,T}R_j^{(2)}(g_{i,T}) \right] R_j^{(2)}(g_{i,T}) \\
& + \frac{1}{2} \left[2N_{j,T} + 2q_{j,T}R_j^{(2)}(g_{i,T}) + q_{j,T+1}R_j^{(1)}(g_{i \text{ next}}) \right] R_j^{(1)}(g_{i \text{ next}})
\end{aligned}$$

$$\begin{aligned}
& + \left[N_{j,T} + q_{j,T} R_j^{(2)}(g_{i,T}) + q_{j,T+1} R_j^{(1)}(g_{i \text{ next}}) \right] G_j^e(g_{i \text{ next}}) \\
& + \frac{1}{2} (q_{j,T+1} - s_j) G_j^e(g_{i \text{ next}})^2
\end{aligned} \tag{4.37}$$

where $g_{i \text{ next}}$ is specified by the user as explained in Section 4.1.1.

The delay estimate for all autos of lane group j in cycle $T + 1$, $\hat{D}_{j,T+1}$, for undersaturated traffic conditions ($N_{j,T+1} \leq 0$) changes in the same way as (4.36):

$$\begin{aligned}
\hat{D}_{j,T+1} = & \frac{1}{2} \left[2N_{j,T} + q_{j,T} R_j^{(2)}(g_{i,T}) \right] R_j^{(2)}(g_{i,T}) \\
& + \frac{1}{2} \left[2N_{j,T} + 2q_{j,T} R_j^{(2)}(g_{i,T}) + q_{j,T+1} R_j^{(1)}(g_{i \text{ next}}) \right] R_j^{(1)}(g_{i \text{ next}}) \\
& + \frac{1}{2(s_j - q_{j,T+1})} \left[N_{j,T} + q_{j,T} R_j^{(2)}(g_{i,T}) + q_{j,T+1} R_j^{(1)}(g_{i \text{ next}}) \right]^2.
\end{aligned} \tag{4.38}$$

Equations (4.35)–(4.38) are used to calculate the total auto delay for cycle T and the estimate of auto delay for cycle $T + 1$ that participate in (4.6).

4.2.2 Transit Delay

As in Section 4.1.2. the person delay for the transit vehicles for oversaturated traffic conditions consists of the sum of two terms: 1) the person delay experienced by transit vehicles that are present at the intersection during the design cycle T before the end of their respective green times, and 2) the person delay for transit vehicles that arrive before the end of the design cycle T but cannot be served during T . It is assumed that information on the location and arrival times of transit vehicles is available only for the design cycle, and as a result, the transit vehicles that arrive during cycle $T + 1$ are not taken into account.

As stated before, transit vehicles travel in mixed traffic lanes with the autos, so the delay of a transit vehicle that arrives in its lane group's queue at some time t_b is the same as an auto vehicle that arrives at the same time at that lane group's queue. The delay for a transit vehicle that belongs to a lane group j can be calculated by using queueing diagrams as shown in Figure 4.5.

For the case that oversaturated traffic conditions prevail during some cycles, the estimation of the transit delay used in the optimization of each cycle T depends on the actual arrival time of the transit vehicles, t_b , as well as whether or not it is served during cycle T . All of the cases, along with the respective formulas for estimating transit delays, are summarized next.

Transit delay estimate for cycle T

Case 1: Arrival after the end of green in cycle $T - 1$ and before the end of green in cycle T

If a transit vehicle that belongs to lane group j arrives at some time after the end of the last phase that can serve j in the previous cycle, $T - 1$, ($t_b > \tau_{j,T-1}$) and

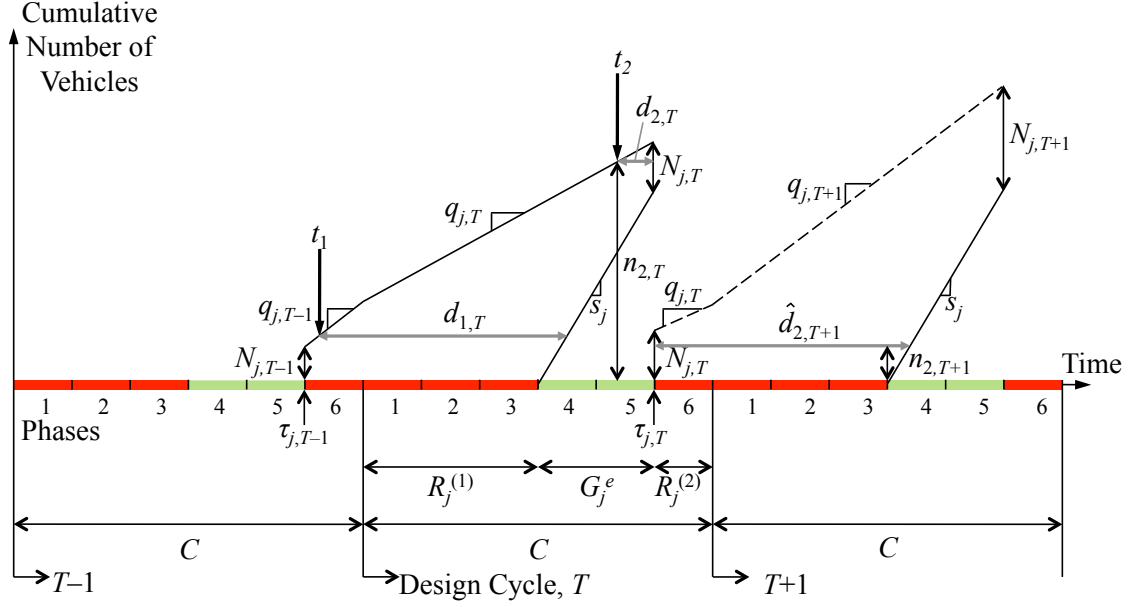


Figure 4.5. Queueing Diagram for Lane Group j for Oversaturated Conditions (Transit Delay)

before the end of the respective phases in the design cycle, T , ($t_b \leq \tau_{j,T}$) (e.g., $t_b = t_1$ in Figure 4.5), its delay during cycle T , $d_{b,T}$, depends on t_b and $g_{i,T}$. This delay is expressed as:

$$d_{b,T} = \begin{cases} \tau_{j,T} - t_b & \text{if } n_{b,T+1} > 0 \\ (T-1)C + R_j^{(1)}(g_{i,T}) + \frac{n_{b,T}}{s_j} - t_b & \text{if } n_{b,T+1} \leq 0 \end{cases} \quad (4.39)$$

where $n_{b,T}$ is the transit vehicle's position in the queue before the end of the phases that serve its lane group j in cycle T (i.e., before $t_{j,T}$). The value of $n_{b,T}$ can be calculated as follows:

$$n_{b,T} = \begin{cases} N_{j,T-1} + q_{j,T-1}(t_b - \tau_{j,T-1}) & \text{if } t_b < (T-1)C \\ N_{j,T-1} + q_{j,T-1}R_j^{(2)}(g_{i,T-1}) + q_{j,T}(t_b - (T-1)C) & \text{if } t_b \geq (T-1)C \end{cases} \quad (4.40)$$

and $n_{b,T+1}$ is its position in the queue after the end of the phases that serve its lane group j in cycle T , (i.e., after $t_{j,T}$), and can be calculated as follows:

$$n_{b,T+1} = n_{b,T} - G_j^e(g_{i,T})s_j. \quad (4.41)$$

A positive position ($n_{b,T+1} > 0$) means that the transit vehicle has not been served in cycle T , and a non-positive position means that it has.

For the transit vehicles that arrive during the green time and are actually served during T ($n_{b,T+1} \leq 0$), it is possible that they are served at the moment they arrive. In this case, (4.39) gives a negative number, but the actual delay for the transit vehicle is 0.

Case 2: Arrival before the end of green in cycle $T - 1$, transit vehicle not served during $T - 1$

If a transit vehicle that belongs to lane group j has arrived at some time before the end of the green time that can serve j in the previous cycle, $T - 1$, and it is still present during cycle T ($t_b \leq \tau_{j,T-1}$) its delay for cycle T , $d_{b,T}$, is as follows:

$$d_{b,T} = \begin{cases} R_j^{(2)}(g_{i,T-1}) + R_j^{(1)}(g_{i,T}) + G_j^e(g_{i,T}) & \text{if } n_{b,T+1} > 0 \\ R_j^{(2)}(g_{i,T-1}) + R_j^{(1)}(g_{i,T}) + \frac{n_{b,T}}{s_j} & \text{if } n_{b,T+1} \leq 0 \end{cases} \quad (4.42)$$

where:

$$n_{b,T} = n_{b,T-1} - G_j^e(g_{i,T-1})s_j. \quad (4.43)$$

For the transit vehicles that arrive after the end of the green that can serve their respective lane groups, their delay during cycle T is assumed to be 0 and an estimate of the delay the vehicle experiences in $T + 1$ is added in the objective function, which is calculated as shown next.

Transit delay estimate for cycle $T + 1$

If a vehicle arrives before the end of the last phase that can serve its lane group in cycle T but cannot be served, or it arrives after the end of that phase and before the end of the cycle, two things could happen: 1) the phases that can serve it will be extended so that the transit vehicle can be served during a following cycle, T , or 2) the transit vehicle will serve during the next cycles. The objective function includes an estimate of the delay that such a transit vehicle would experience if the green time of the phase that can be served it is not extended. As before, the delay estimate for a transit vehicle that cannot be served during cycle T depends on its arrival time, t_b , and whether or not it can be served during the next cycle, $T + 1$. The estimation of the delay experienced during the next cycle $T + 1$ follows one of the two cases below:

Case 1: Arrival before the end of green in T

If a transit vehicle that belongs to lane group j has arrived at some time before the end of the green time that can serve j in T ($t_b \leq \tau_{j,T}$) but cannot be served during that cycle (e.g., $t_b = t_2$ in Figure 4.5), its delay estimate for cycle $T + 1$, $\hat{d}_{b,T+1}$, is as follows:

$$\hat{d}_{b,T+1} = \begin{cases} R_j^{(2)}(g_{i,T}) + R_j^{(1)}(g_{i \text{ next}}) + G_j^e(g_{i \text{ next}}) & \text{if } n_{b,T+2} > 0 \\ R_j^{(2)}(g_{i,T}) + R_j^{(1)}(g_{i \text{ next}}) + \frac{n_{b,T+1}}{s_j} & \text{if } n_{b,T+2} \leq 0 \end{cases} \quad (4.44)$$

where:

$$n_{b,T+2} = n_{b,T+1} - G_j^e(g_{i \text{ next}})s_j. \quad (4.45)$$

Case 2: Arrival after the end of green in T

If a transit vehicle that belongs to lane group j has arrived at some time after the end of the last phase that can serve j in T and that phase is not extended to serve it ($\tau_{j,T} < t_b < TC$), its delay estimate for cycle $T + 1$, $\hat{d}_{b,T+1}$, is as follows:

$$\hat{d}_{b,T+1} = \begin{cases} TC + R_j^{(1)}(g_{i \text{ next}}) + G_j^e(g_{i \text{ next}}) - t_b & \text{if } n_{b,T+2} > 0 \\ TC + R_j^{(1)}(g_{i \text{ next}}) + \frac{n_{b,T+2}}{s_j} - t_b & \text{if } n_{b,T+2} \leq 0 \end{cases} \quad (4.46)$$

where:

$$n_{b,T+2} = N_{j,T} + q_{j,T} (t_b - \tau_{j,T}) - G_j^e(g_{i \text{ next}})s_j. \quad (4.47)$$

Using the delay estimates obtained by the equations presented in Sections 4.2.1 and 4.2.2 (i.e., $D_{j,T}$, $\hat{D}_{j,T+1}$, $d_{b,T}$, $\hat{d}_{b,T+1}$) the components of auto and transit person delay of the objective function are estimated.

4.3 Study Sites

The performance of the person-based traffic responsive signal control system is tested with data from two real-world intersections: the intersection of Mesogion and Katechaki Avenues in Athens, Greece, and the intersection of University and San Pablo Avenues in Berkeley, California, United States. The two study sites have been selected in order to test the performance of the system under different intersection geometries, lane allocations, and phasing schemes, as well as different transit route designs and headways which affect the frequency of transit vehicle conflicts for priority. The selected test sites are described below and the results of the performed tests are presented in Section 4.4.

4.3.1 Intersection of Katechaki and Mesogion Avenues

The intersection of Katechaki and Mesogion Avenues is a busy intersection of two main signalized arterials located in Athens, Greece. This test intersection has been selected because of the high traffic volumes on all approaches, its complicated phasing scheme, and the existence of multiple conflicting bus routes.

The intersection's layout is presented in Figure 4.6. As the figure shows, this is a complex intersection with through and turning traffic in all directions (the main through movement is on Katechaki Avenue). Figure 4.7 presents the lane groups (on the right labeled 1–8r) phasing, and green times for the intersection during the morning peak. The intersection signal operates on a fixed 6-phase cycle. Auto volume data are available at a rate of once per second from loop detectors placed 40 meters upstream of the intersection on each approach. Measured traffic volumes during the morning peak hour (7–8am) are used as a representative demand. These volumes

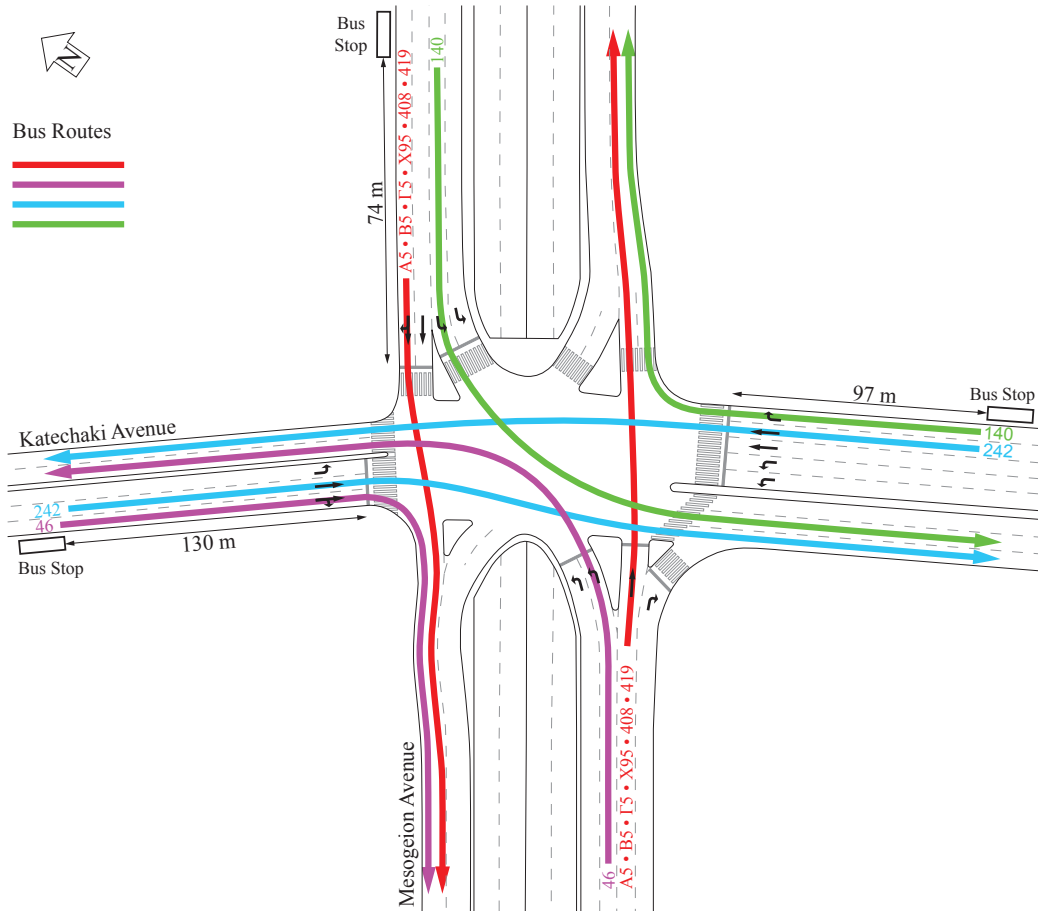


Figure 4.6. Layout and Bus Routes for the Intersection of Katechaki and Mesogion Avenues

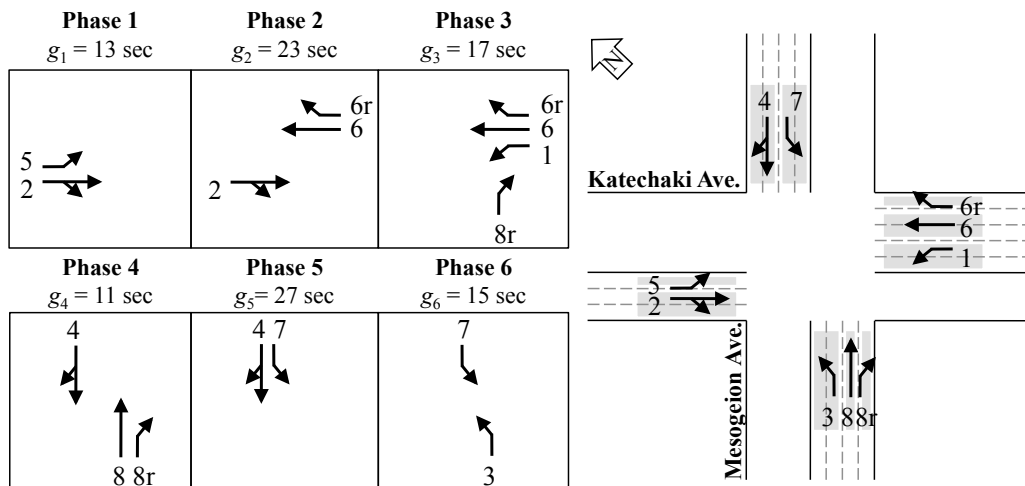


Figure 4.7. Lane Groups, Phasing, and Green Times for the Intersection of Katechaki and Mesogion Avenues

correspond to an intersection flow ratio of $Y = 0.80$.¹ For the signal’s current cycle length of $C = 120$ seconds and lost time of $L = 14$ seconds, this corresponds to a degree of saturation² of $X_c = 0.87$ which indicates nearly saturated traffic conditions.

Nine bus routes travel through the intersection in mixed traffic lanes with headways that vary from 15 to 40 minutes for each route. This corresponds to 43 buses in the morning peak hour. The numbers next to the directional arrows in Figure 4.6 correspond to the different bus routes. The bus routes run in four conflicting directions with 70% traveling on the northeast-southwest approaches (Mesogion Avenue) and the rest on the northwest-southeast approaches (Katechaki Avenue). Their bus stops are located nearside (i.e., upstream of the intersection). The bus stop on the southwest approach is not shown in the figure because of its longer distance from the stop line, which also diminishes its impact on the traffic operations of the intersection. However, the impact of all bus stops on the operation of the intersection is ignored. Information about the bus schedule is available at the Athens Urban Transport Organisation’s website (OASA, 2010).

4.3.2 Intersection of University and San Pablo Avenues

The intersection of University and San Pablo Avenues is selected as the second study site to test the performance of the person-based traffic responsive signal control system on a typical U.S. layout for an intersection of two major arterials. This intersection is also characterized by high traffic volumes on all approaches, but it has a simpler layout and phasing scheme compared to the intersection of Katechaki and Mesogion Avenues. Nevertheless, several bus routes travel through that intersection on multiple conflicting directions.

Figure 4.8 shows the intersection’s layout and Figure 4.9 presents the lane groups, (on the right labeled 1–8r), phasing, and green times for the intersection during the evening peak. The intersection signal operates on a 4-phase cycle. Data about the traffic volumes and signal settings for the evening peak hour have been obtained from previous research studies (Skabardonis *et al.*, 1990) and have been updated based on recent field observations. Evening peak hour traffic volumes (4–5pm) correspond to an intersection flow ratio of $Y = 0.73$. For a cycle length of $C = 80$ seconds and a lost time of $L = 12$ seconds, this corresponds to a degree of saturation of $X_c = 0.89$, also indicating traffic conditions close to saturation.

Six bus routes travel through the intersection in mixed traffic lanes with headways that vary from 10 to 30 minutes on each route, corresponding to 34 buses in the evening peak hour. The bus routes run in three conflicting directions with 60% traveling on the north-south approaches (San Pablo Avenue) and the rest on the east-west approaches (University Avenue). The location of the bus stops varies, with

¹*Intersection flow ratio* is defined as the summation of flow ratios, meaning the ratio of demand to saturation flow, for all critical lane groups at the intersection (HCM, 2000).

²*Degree of saturation* is defined as the ratio of demand volume to the capacity for a subject lane group (also known as volume-to-capacity ratio) (Koonce *et al.*, 2008).

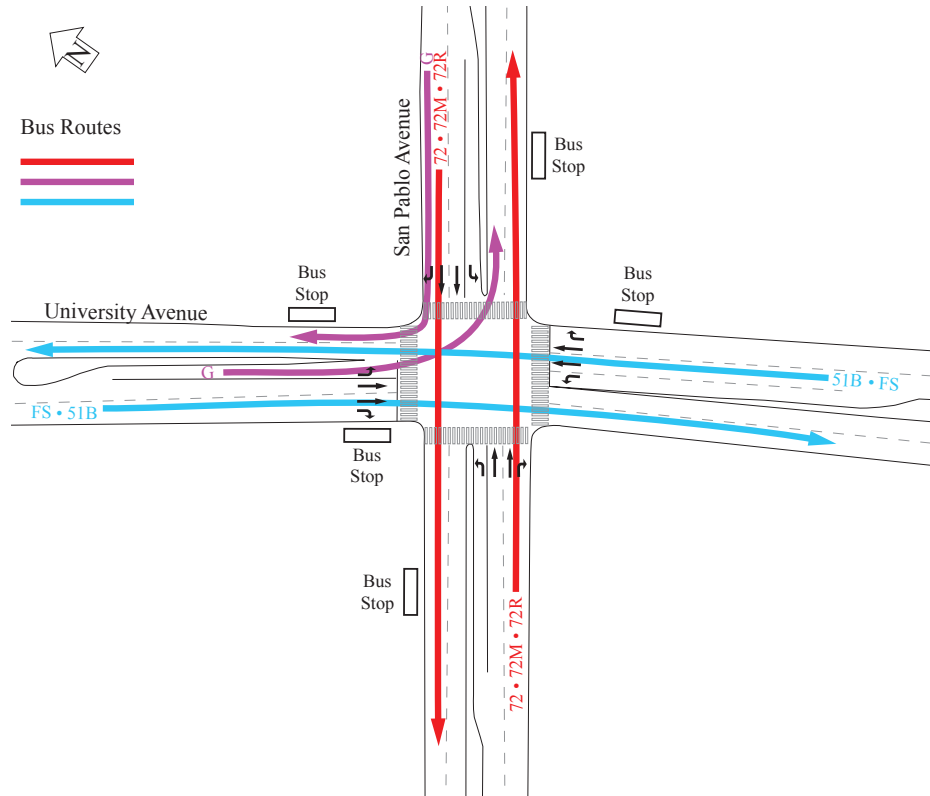


Figure 4.8. Layout and Bus Routes for the Intersection of University and San Pablo Avenues

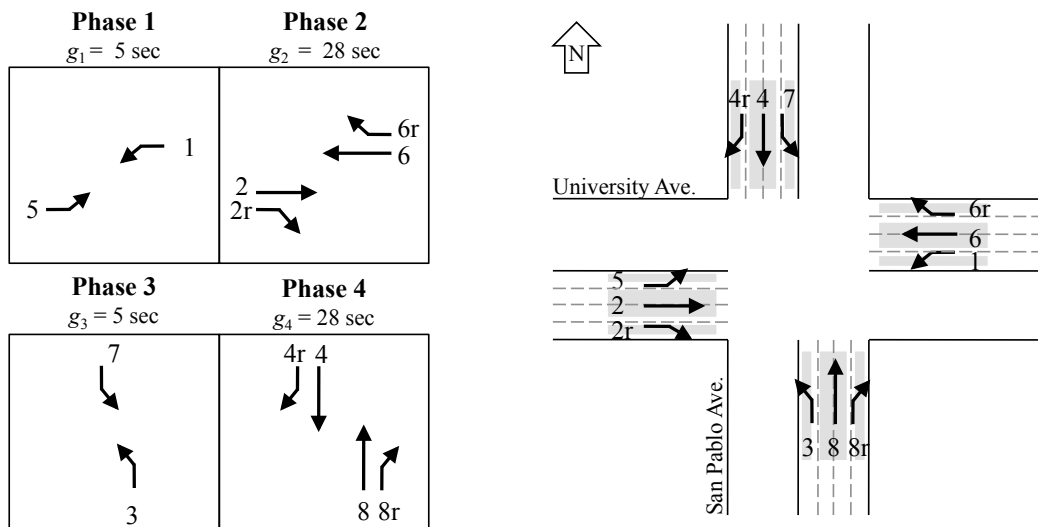


Figure 4.9. Lane Groups, Phasing, and Green Times for the Intersection of University and San Pablo Avenues

some located nearside and others located farside. As in the case of the intersection of Katechaki and Mesogion Avenues, the impact of bus stops on the operation of the intersection is ignored. Information about the bus schedule is available at Alameda-Contra Costa Transit District’s website (AC Transit, 2011).

4.4 Evaluation

The proposed person-based traffic responsive signal control system has been tested using data from the study sites described in Section 4.3. Several tests have been performed as indicated in Section 3.4 for a variety of traffic conditions (undersaturated and oversaturated) and assumptions about the optimization input (availability of perfect information versus predictions for the vehicle arrivals and demands from detectors). These tests include evaluation of two different optimization scenarios for one hour of traffic operations at both study sites: Scenario 1, when only vehicle delay is minimized (i.e., vehicle-based optimization where vehicle delays are not weighted by their respective passenger occupancies), and Scenario 2, when total person delay for both transit and auto passengers is minimized (i.e., person-based optimization where vehicle delays are weighted by their respective passenger occupancies). In addition, tests have been performed that implement the optimal fixed timings from TRANSYT-7F (Scenario 0) for the whole hour. For each scenario, a warm up period equal to one cycle length is used. In addition, each scenario is evaluated ten times in order to account for the effect of variations in bus arrivals at the intersection. The resulting average values of the ten replications are presented. The details of the tests, as well as the performance of the system for a variety of traffic and transit operating characteristics are presented next.

4.4.1 Test Type I: Deterministic Vehicle Arrivals–Constant Auto Traffic Demand

Tests of type I have been performed under the assumption that autos arrive deterministically at a constant rate using data from both study sites. For these tests, it is assumed that perfect information is available about the arrival rates of autos and the arrival times of buses at the intersection.

The auto arrival rates are set as the average flow during the morning peak hour for the intersection of Katechaki-Mesogion Avenues and during the evening peak hour for the intersection of University and San Pablo Avenues. In addition, the average auto occupancy, \bar{o}_a , is assumed to be 1.25 passengers per vehicle at both sites. Bus arrival times at the intersection are simulated based on a shifted normal distribution around their scheduled arrival times since no information is available about the real distribution of their schedule deviation. Furthermore, no schedule delay is considered in these tests. For the buses, the passenger arrivals at the bus stops are assumed to be deterministic and constant because headways are short enough that people do not rely on a published schedule. As a result, the bus occupancy is a function of the time

between the actual arrivals of two consecutive buses of the same route. This means that the buses are assumed to operate as if they arrive empty at the bus stop just upstream of the intersection under consideration, so a larger headway would lead to a greater number of passengers on-board. The passenger occupancy of each bus that arrives at the intersection is given by:

$$o_b = p_m(t_{b,m} - t_{b-1,m}) \quad (4.48)$$

where p_m is the passenger demand for bus route m and $t_{b,m}$ is the actual time that bus b belonging to route m arrives at the back of the queue at the intersection under consideration. Despite the fact that the schedule delay of the buses is not considered directly, it is implicitly taken into account in the optimization process through the higher passenger occupancy expected of late buses. For the initial testing of the signal control system, an average bus occupancy of $\bar{o}_b = 40$ passengers per vehicle is assumed.

The user-specified $g_{i \text{ next}}$ that are used as estimates of the phase green times of the next cycle are set to be the same as the fixed optimal signal timings provided by TRANSYT-7F for the specific traffic conditions. In addition, the upper bounds for the green times of the phases, $g_{i \text{ max}}$, are set equal to $C - \sum_{i=1}^I y_i$. Non-zero lower bounds for the green times of each phase, $g_{i \text{ min}}$, are also introduced to ensure that all phases are allocated some minimum green time. A total minimum green time of 7 seconds is assigned to each of the left-turn phases and 10 seconds to each of the through phases.

The MINLP, described in Section 4.1.3, has a quadratic objective function and linear constraints. As long as the objective function remains convex, the global optimum can be easily found using the Branch and Bound method utilized for solving MINLP problems. Indeed, the Hessian matrix of the objective function is positive-definite for all tested scenarios, so the objective function is convex. The computation time for the optimization of signal settings for one cycle is on the order of 20 seconds for the tests performed, which is sufficiently small to allow for real-world implementations.

Intersection of Katechaki and Mesogion Avenues

Table 4.1 presents the person delay for auto, bus, and total number of users obtained by the three scenarios tested for an intersection flow ratio of $Y = 0.80$. A comparison of the person-based optimization with the vehicle-based one indicates that the former can achieve a reduction in the total person delay at the intersection of 7.1% by reducing the delay of bus users by 35.6% and increasing auto user delay by 3.5%. This translates into a reduction in average bus delay of 13 seconds and an increase in average auto delay of only 1.5 seconds. Comparing the delays obtained by vehicle-based optimization with those delays from implementing the optimized fixed timings obtained by TRANSYT-7F, one observes that the optimal signal settings from vehicle-based optimization outperform those obtained from TRANSYT-7F even in reducing auto delay. As a result, evaluation of the person-based traffic responsive signal control

Table 4.1. Person Delays for $Y = 0.80$ and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of Katechaki and Mesogion)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 0: TRANSYT-7F (Fixed Settings)	55.24	19.49	74.73
Scenario 1: Vehicle-based Optimization	53.22	19.76	72.97
Scenario 2: Person-based Optimization	55.08	19.72	67.80
% Change in person delay between Scenarios 1 & 2	3.50%	-35.60%	-7.08%

system is made by comparing the person delays from person-based optimization with the ones from the vehicle-based optimization, because the latter provides the lowest delays that can be achieved for autos.

The performance of the system has been tested for different intersection flow ratios that vary from 0.4 to 0.8 for average occupancies of 40 passengers per bus and 1.25 passengers per auto. The results are illustrated in Figure 4.10. Specifically, the figure illustrates the percent changes in person delay of auto and bus passengers, as well as total person delay, achieved by the person-based optimization compared to the vehicle-based optimization. The results indicate consistent patterns in the person delay changes for all scenarios. The higher the intersection flow ratio, the lower the benefit for transit users and for all users traveling through the intersection, and as a result, the lower the increase in auto user delay. This is expected due to the fact that higher intersection flow ratios imply higher auto traffic demand and consequently lower flexibility to change the signal timings while maintaining undersaturated conditions. For very high intersection flow ratios, the vehicle-based and person-based optimization of the signal settings result in the same optimal signal timings and the same person delays, since the high auto flow outweighs the higher occupancies of the buses. The figure also shows the 95% confidence intervals of the percent changes for person delays of autos, buses, and all passengers at the intersection. The plotted confidence intervals indicate that the percent changes in person delay for autos, buses, and all travelers of the intersection are significantly different than zero.

Tests have also been performed for different average bus to auto passenger occupancy ratios to investigate how changes in bus ridership affect the provision of priority. The average auto occupancy is kept constant for all the scenarios and equal to 1.25 passengers per vehicle. Figure 4.11 shows the results obtained by comparing the person delays from the person-based optimization with those from vehicle-based optimization for an intersection flow ratio of $Y = 0.6$. The figure indicates that for very low average occupancy ratios, the collective benefits to all passengers diminish as do the benefits to the bus passengers. The converse is also true; the higher the

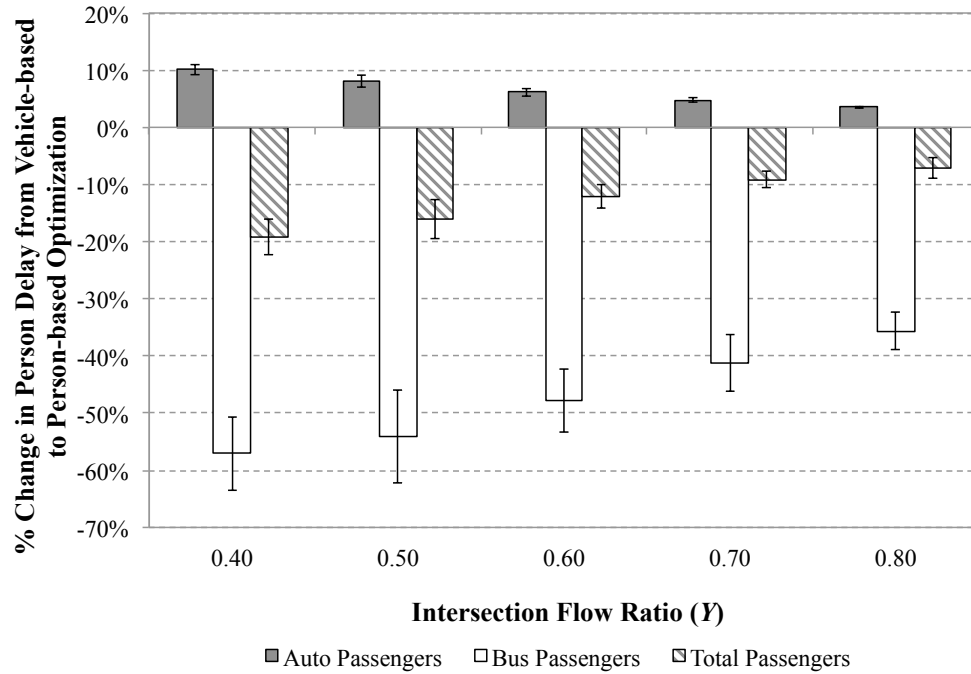


Figure 4.10. Percent Change in Person Delay for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of Katechaki and Mesogion)

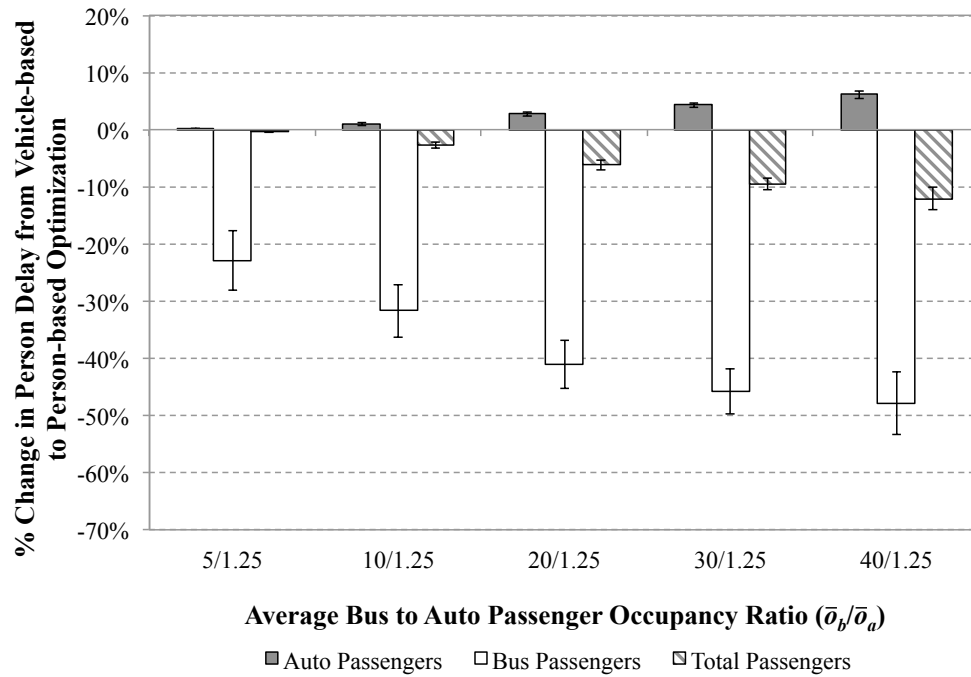


Figure 4.11. Percent Change in Person Delay for Different Average Bus to Auto Passenger Occupancy Ratios and $Y = 0.6$ (Test Type I: Intersection of Katechaki and Mesogion)

passenger occupancy of buses, the higher the savings for their passengers and the higher the delays for auto users compared with the person delays from vehicle-based optimization. This outcome is expected since a higher bus passenger occupancy leads to a larger weight for the bus delays, and given that the intersection is undersaturated, there is spare time to provide more priority to serve the buses. However, for high average bus to auto passenger occupancy ratios, the system eventually converges towards one solution and the benefit to bus passengers levels off once the system has reached the maximum amount of priority that it can provide to transit for the specific traffic conditions.

Similar patterns as the ones observed in Figure 4.11 are observed for all intersection flow ratios tested. However, tests for other intersection flow ratios have shown that the benefits level off at different occupancy ratios for the two optimization scenarios tested. The higher the intersection flow ratio, the lower the spare capacity at the intersection and the lower the occupancy ratio at which benefits for transit users level off.

Intersection of University and San Pablo Avenues

The same tests (Tests Type I) have been performed with data from the evening peak hour at the intersection of University and San Pablo Avenues. Figure 4.12 illustrates the results obtained for intersection flow ratios, Y , that vary from 0.4 to 0.73 and for average occupancies of $\bar{o}_b = 40$ passengers per bus and $\bar{o}_a = 1.25$ passengers per auto. A similar pattern is observed as before regarding the reduction in the benefit obtained by bus users and the significance of the percent changes of person delays for auto, bus, and all users of the intersection as the intersection flow ratio increases. However, the percent benefit for bus users and all travelers and the increase in delay for auto users is much smaller than the respective changes at the intersection of Katechaki and Mesogion Avenues. This can be attributed to the fact that the intersection operates under a shorter cycle that has four phases which reduces the flexibility for providing priority.

A comparison of person delay changes for different average bus to auto passenger occupancy ratios and an intersection flow ratio of $Y = 0.6$ (Figure 4.13) shows that there is no statistically significant improvement of the percent change for bus and auto passenger delay for average bus to auto occupancy ratios greater than 10/1.25. This indicates that the intersection reaches the maximum priority it can provide to buses for the specific traffic conditions even when only 20 passengers are on-board the average bus.

4.4.2 Test Type II: Deterministic Vehicle Arrivals–Time-Dependent Auto Traffic Demand

Using information from the study site of Katechaki and Mesogion Avenues, a time-dependent demand profile that includes signal cycles operating in oversaturated traffic conditions has been constructed to test the performance of the proposed signal control

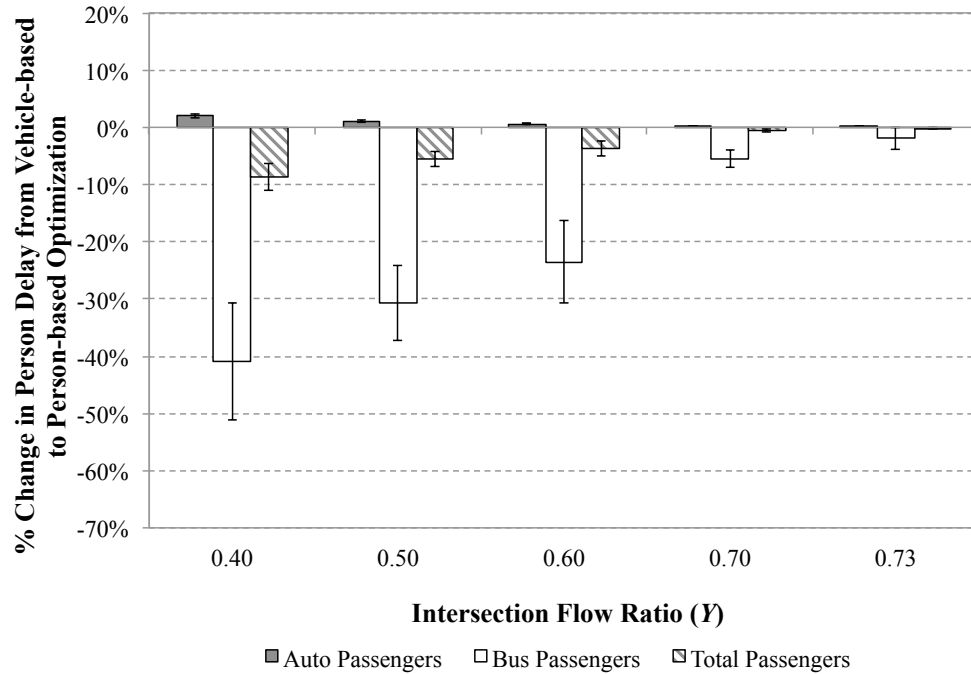


Figure 4.12. Percent Change in Person Delay for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type I: Intersection of University and San Pablo)

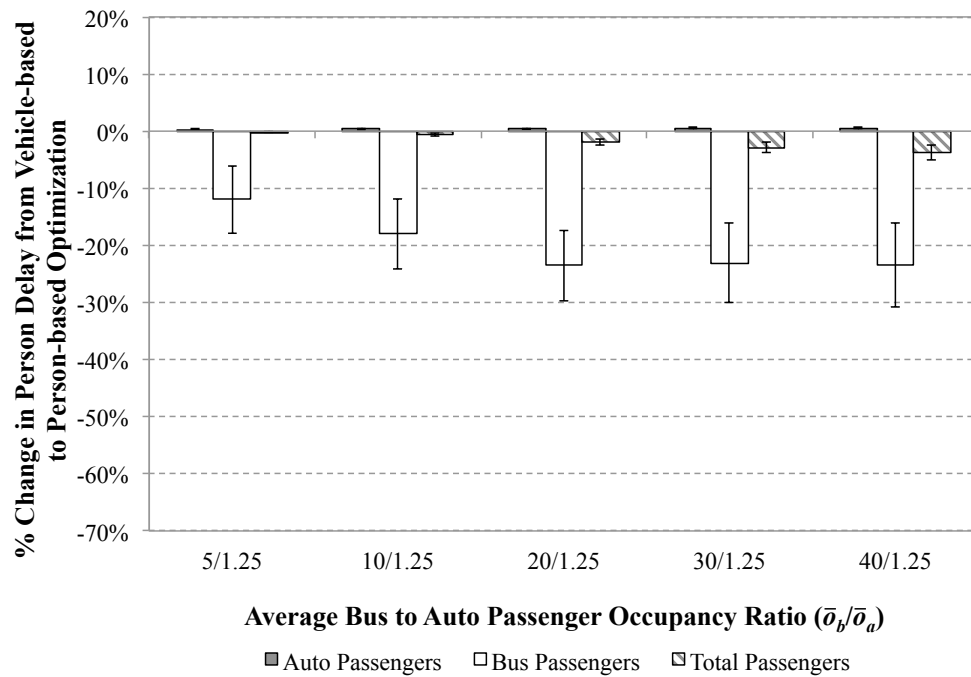


Figure 4.13. Percent Change in Person Delay for Different Average Bus to Auto Passenger Occupancies and $Y = 0.6$ (Test Type I: Intersection of University and San Pablo)

system. The time-dependent demand profile consists of auto traffic demands that correspond to intersection flow ratios from $Y = 0.4$ to 1.17 as shown in Figure 4.14. For a cycle of $C = 120$ seconds and lost time of $L = 14$ seconds, intersection flow ratios higher than 0.883 indicate oversaturated traffic conditions. Even though the system operates at the beginning in oversaturated traffic conditions after the 14th cycle (including the warm-up cycle), demand drops and the system returns to undersaturated traffic conditions by the end of the one hour interval tested. As in tests of type I, perfect information is assumed to be available about the arrival rates of autos and buses at the intersection. The tests have been performed under the same assumptions about passenger occupancies and exclusion of schedule delay as in Section 4.4.1.



Figure 4.14. Intersection Flow Ratios for the 1 Hour Time-Dependent Demand Profile (Test Type II: Intersection of University and San Pablo)

The $g_{i \text{ next}}$ are set equal to the minimum green times, $g_{i \text{ min}}$, for all of the phases except for the last one, which is allocated green time equal to the residual of the cycle length as shown in the following formulas:

$$g_{i \text{ next}} = \begin{cases} g_{i \text{ min}} & \forall i < I \\ C - \sum_{i=1}^{I-1} g_{i \text{ min}} - \sum_{i=1}^I y_i & \forall i = I. \end{cases} \quad (4.49)$$

This is based on the assumption that all lane groups experience similar traffic conditions. As a result, implementing minimum green times over the next cycle would give the minimum delay to all vehicles for cycle $T + 1$, so the estimate of the delay is a

Table 4.2. Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type II: Intersection of Katechaki and Mesogion)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 1: Vehicle-based Optimization	90.07	42.53	132.60
Scenario 2: Person-based Optimization	90.07	27.46	120.06
% Change in person delay between Scenarios 1 & 2	2.81%	-35.45%	-9.46%

lower bound. However, the performance of the signal control system is not sensitive to the values of $g_{i \text{ next}}$ chosen for the next cycle, $T + 1$.

Table 4.2 shows that the person-based optimization reduces total and bus passenger delay compared to the vehicle-based ones even when traffic operates in oversaturated conditions. Total person delay for the intersection is reduced by 9.5%, while bus passenger delay is reduced by 35.5%, indicating the magnitude of improvement in bus operations achieved by providing priority. Auto passenger delays increase by only 2.8%. The percentages translate into an increase in the average auto delay on the order of 4 seconds per vehicle and a decrease in the average bus delay on the order of 16 seconds per vehicle.

The sensitivity of the results to the average bus passenger occupancy is investigated by performing tests with different average bus to auto passenger occupancy ratios. Figure 4.15 illustrates the percent changes in the person delay for auto and bus passengers as well as the percent change in total person delay achieved by the person-based optimization compared to the vehicle-based optimization. For bus occupancies exceeding 10 passengers, there is no statistically significant percent change in bus passenger delays with increasing occupancy. This shows that the system gives similar priority for all five cases regardless of the number of passengers on the buses, which indicates that the results are not very sensitive to the bus passenger occupancy. This is an effect of the saturated traffic conditions that occur for part of the tested time interval and an indication that the maximum possible priority is being given to bus passengers. A similar effect is observed for high flow ratios for tests of type I, because the conditions approach saturation. Just as for tests of type I, the computation time is on the order of 20 seconds for the optimization of the signal settings for one cycle.

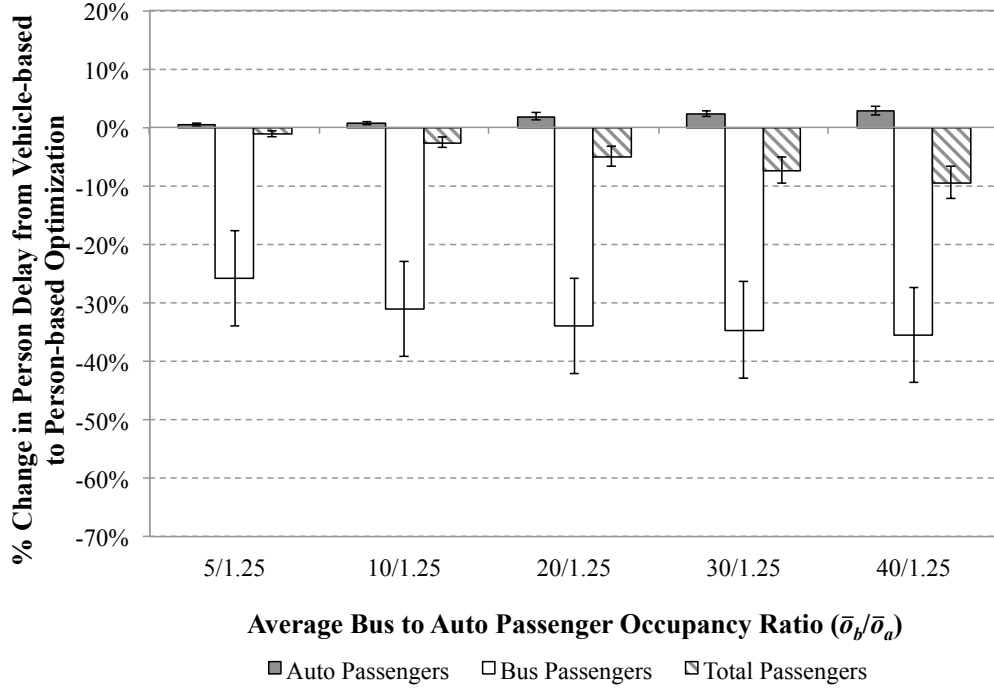


Figure 4.15. Change in Person Delay for Different Average Bus to Auto Passenger Occupancy Ratios and $Y = 0.6$ (Test Type II: Intersection of Katechaki and Mesogion)

4.4.3 Test Type III: Stochastic Vehicle Arrivals—Constant Auto Traffic Demand

The final type of tests for the isolated intersection have been performed with Emulation-In-the-Loop Simulation (EILS) in AIMSUN as described in Section 3.4. These tests are used in order to evaluate the performance of the system in more realistic traffic conditions where vehicles do not arrive deterministically and where errors exist in the estimation of auto demand and the prediction of vehicle arrival times. In addition, EILS allows for the evaluation of the proposed signal control system using additional performance measures that would be hard to assess analytically, such as average speed, number of stops, and emissions. Tests of type III are performed for undersaturated traffic conditions for the same input scenarios as tests of type I for the intersection of Katechaki and Mesogion Avenues.

The auto inter-arrival times are simulated to follow an exponential distribution. In order to estimate the auto demand for each cycle using the same method that would be required in reality, detectors are located approximately 100 meters upstream of the intersection on each approach. Detectors are also located at the exits of each of the approaches in order to measure the exit flow for each lane group and cycle. Exponential smoothing is used on the measured flows of both types of detectors during the previous cycle in order to create an estimate for the demand of the respective lane

group for the next cycle. The estimate of the arrival rate, $\hat{q}_{j,T}$, is a weighted average of the previous estimate and observed value:

$$\hat{q}_{j,T} = eq_{j,T-1} + (1 - e)\hat{q}_{j,T-1} \quad (4.50)$$

where e is a factor between 0 and 1 that determines how much weight is placed on the most recent observation. A value of $e = 0.2$ has been used in the performed tests. The maximum of the two smoothed flows from the two types of detectors is used as an input for the optimization for the next cycle in order to account for cases that signal timings in the previous cycle are not able to serve all of the incoming demand.

The timetable of the bus arrivals at the entry links of the network is fixed and based on the same headways as in the deterministic arrival tests. In order to predict the arrival time of buses at the intersection for the traffic signal optimization, detectors are placed upstream on entry links at distances equivalent to a travel time of one cycle length from the intersection. The prediction of the bus arrivals at the approaches is estimated using an average nominal speed of 45 km/hr. The average passenger occupancy for the autos is assumed to be 1.25 passengers per vehicle, while each transit vehicle is assigned a random number of passengers with an average value of 40 passengers per vehicle. The tests are performed under the assumption that schedule delay is negligible.

The green times for the next cycle, $g_{i \text{ next}}$, and the upper and lower bounds, $g_{i \text{ max}}$ and $g_{i \text{ min}}$, are defined as in the deterministic arrival tests. Ten replications are performed for each of the intersection flow ratios tested before ($Y = \{0.4, 0.5, \dots, 0.8\}$), which allow for variation in the auto and bus arrivals at the intersection. As in the deterministic arrival tests, the Hessian matrix of the objective function is positive-definite for all tested cases, so the problem can be solved using the Branch and Bound method. The computation time for the optimization of signal settings for one cycle remains similar to before, which is promising for implementing the proposed system in real-world settings.

The results from the simulation tests are shown in Figure 4.16. A comparison of the results from the simulation with the ones from the deterministic arrival tests indicates that for the same intersection flow ratio, the percent benefit achieved in person delay for the whole intersection is on the same order of magnitude. The same holds for the magnitude of the percent increase on auto passenger delay. Since no delay at bus stops is considered in the simulation tests, the differences between the results of the test types I and III can be attributed only to the variations in the prediction of auto and bus arrivals, which are not accounted for in tests of type I. This results in a reduction in the benefit that is achieved for bus users and all travelers at the intersection compared to the respective one with perfect information (Test Type I). For example, for an intersection flow ratio of $Y = 0.6$, a 32.2% reduction in person delay for transit users is observed, but the reduction assuming perfect information is 47.8%. Since the optimization relies on estimates of arrival rates for autos and arrival times for transit vehicles that contain errors, it cannot provide the optimal phase durations as it would with perfect information.

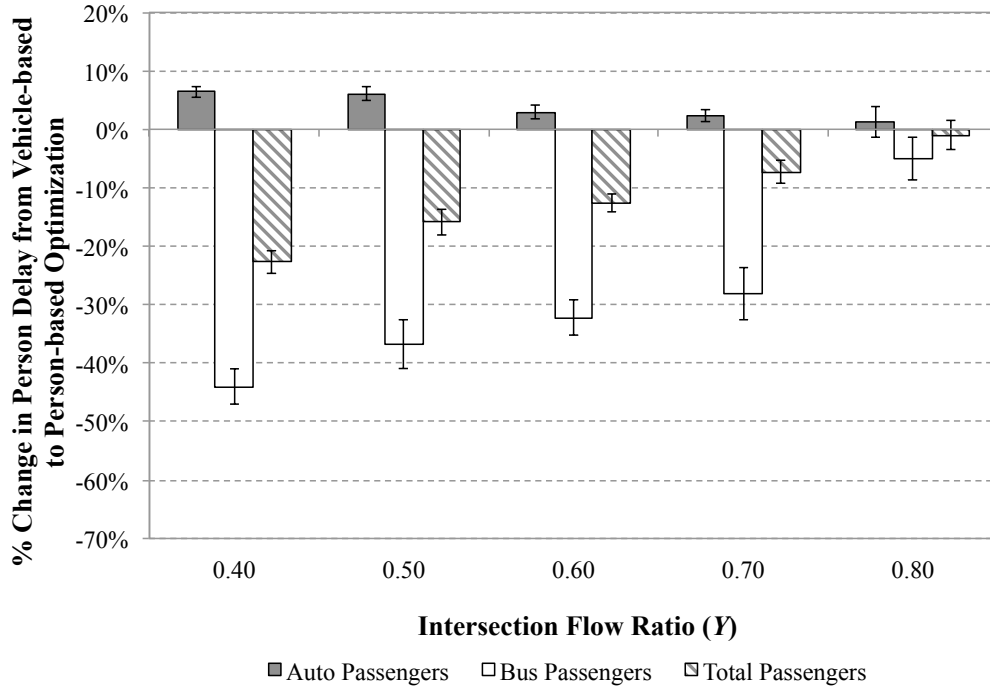


Figure 4.16. Change in Person Delay for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type III: Intersection of Katechaki and Mesogion)

For the case when the intersection flow ratio is $Y = 0.8$, the person-based optimization does not outperform the vehicle-based optimization in terms of reducing the total person delay and the bus delay for all replications. This is another consequence of the error in the arrival estimates for high auto traffic demand. Estimation errors can lead to signal timings that cause oversaturated traffic conditions for some approaches of the intersection. This results in less accurate estimates of auto arrival rates and bus arrival times which further impedes the operation of the signal control system.

Simulation tests also allow for evaluation of the system with several additional performance measures such as the ones shown in Table 4.3. The results shown in the table for an intersection flow ratio of $Y = 0.6$ indicate that there is a decrease in the number of stops for buses on the order of 13.8% and an increase in the average speed for buses of 8.8% when using person-based optimization compared to the vehicle-based optimization. At the same time, the number of stops for autos increases by almost 2.5%, which leads to an increase in pollutant emissions. For example, carbon monoxide (CO) increases by about 0.6%. However, the signal timings from person-based optimization lead to a substantial reduction in the CO emitted by buses of about 8.2%. Overall, the higher the auto traffic demand, the lower the benefit achieved with the proposed system for buses in terms of improving their speed and reducing their stops. In addition, for high auto traffic demand, all of these performance measures

converge to the same values because the two different optimization scenarios result in the same green times.

Table 4.3. Performance Measures for Different Intersection Flow Ratios and $\bar{o}_b/\bar{o}_a = 40/1.25$ (Test Type III: Katechaki and Mesogion Intersection)

Perf. Measure	Veh. Type	Opt. Scenario	Intersection Flow Ratio (Y)				
			0.4	0.5	0.6	0.7	0.8
Speed	Auto	Veh-based	44.40	43.96	43.19	42.41	41.72
		Per-based	43.85	43.42	42.86	42.12	41.53
		% Change	-1.26%	-1.23%	-0.77%	-0.68%	-0.45%
	Bus	Veh-based	36.17	36.29	36.18	35.87	35.21
		Per-based	40.81	40.06	39.38	38.02	35.56
		% Change	12.84%	10.40%	8.83%	5.99%	1.00%
Stops per Vehicle	Auto	Veh-based	0.69	0.71	0.76	0.81	0.86
		Per-based	0.72	0.74	0.78	0.83	0.87
		% Change	3.98%	4.10%	2.40%	2.13%	1.70%
	Bus	Veh-based	0.90	0.90	0.93	0.98	0.98
		Per-based	0.78	0.80	0.80	0.86	0.97
		% Change	-13.40%	-11.44%	-13.75%	-12.23%	-1.18%
CO Emissions (kg)	Auto	Veh-based	33.49	41.87	51.50	62.18	72.00
		Per-based	33.89	42.51	51.80	62.54	72.15
		% Change	1.19%	1.53%	0.59%	0.57%	0.21%
	Bus	Veh-based	0.56	0.57	0.57	0.58	0.60
		Per-based	0.50	0.52	0.53	0.55	0.59
		% Change	-11.20%	-9.16%	-8.22%	-5.54%	-1.39%

4.5 Summary of Findings

This chapter has presented the formulation and testing of the person-based traffic responsive signal control system for an isolated intersection. The mathematical program that has been developed to minimize total person delay at the intersection has been tested for a variety of traffic conditions and transit operating characteristics under the assumption that perfect information about traffic demand and transit arrival times is available. Additional tests have been performed to evaluate the system when such perfect information is not available. In the latter case, performance is based on estimates of traffic demand and transit arrival times. The outcomes of the person-based optimization tests have been compared with the outcomes from vehicle-based optimization, which is commonly used to optimize signal settings in traffic responsive signal control systems.

The results from two real-world study sites indicate that the person-based traffic responsive signal control system can achieve significant reductions in the overall person delay and transit user delay at an intersection for a wide range of auto demands by increasing the auto user delay by only a small amount. For example, under a typical intersection flow ratio of $Y = 0.6$ and an average bus to auto passenger occupancy ratio of $\bar{o}_b/\bar{o}_a = 40/1.25$, the proposed system leads to a reduction in bus passenger delay in simulation of about 32% and an increase in car user delay of about 3%. Higher auto traffic demand results in lower reductions in the overall passenger delay and the bus passenger delay and reduces the negative impact on auto users. For very high auto traffic demand, person-based and vehicle-based optimization lead to the same outcome. Sensitivity analysis with respect to the transit passenger occupancy shows that in general, the higher the passenger occupancy of a transit vehicle the higher the priority provided to it and the higher the benefit for transit users. However, there is a limit to the amount of priority that can be provided for transit vehicles that depends on the traffic conditions at the intersection and the operating characteristics of the transit system.

A comparison of the performance of the system through simulation with the tests performed under the assumption of perfect information has shown that it performs well even without incorporating the prediction errors in auto demand and transit vehicle arrival times. Even though losses in the benefit experienced by transit users are observed due to errors in the predictions, the system still achieves significant reductions in their delay for a variety of auto traffic demands. Accounting for the uncertainty of arrivals in the delay calculation and developing improved prediction algorithms for vehicle arrivals can reduce errors in the system and lead to improved performance for simulation tests or real-world implementations.

Overall, the different tests have shown that the performance of the system depends on the traffic conditions as well as the transit operating characteristics such as passenger occupancy, headways, and the number of routes traveling through the intersection in combination with the intersection's phasing and layout. A major advantage of the system compared to other signal control systems is that the computation time is on the order of 20 seconds per cycle which is short enough to allow for real-world implementations.

Chapter 5

Signalized Arterial

The formulation of the mathematical program presented in Chapter 4 is used as a stepping stone to formulate the mathematical program for the arterial case. This chapter includes a detailed description of the mathematical program that is used to minimize person delay for all auto and transit users on signalized arterials. The optimization procedure is described first in Section 5.1. The assumptions and methodology for estimating delays for autos and transit vehicles are described in Section 5.2, and the final mathematical program is presented in Section 5.3. Then, a signalized arterial with four intersections is described in Section 5.4. The results obtained from the system evaluation on the test arterial are presented in Section 5.5. Section 5.6 identifies ways that the arterial level system can be extended to signalized arterial networks. Finally, Section 5.7 summarizes the findings and insights obtained from the analysis of the test results.

5.1 Optimization Procedure

A mathematical program similar to the one for the isolated intersection is formulated for the optimization of the signal settings for the intersections on an arterial. A signalized arterial network consists of an arterial and its cross streets. The cross streets are considered up to one intersection beyond the main arterial and these external intersections are assumed to operate under a fixed signal timing plan. As a result, vehicles arrive at each intersection in platoons when traveling on the arterial and on the cross-street links.

The main assumption that differentiates the mathematical program for arterials from the mathematical program for isolated intersections is the distribution of vehicle arrivals. On signalized arterials, vehicles arrive in platoons at all approaches since their arrivals are influenced by upstream signals. This simplifies the estimation of auto delay as described in Section 5.2.1. In addition, the mathematical program is formulated under the assumption that there is negligible platoon dispersion. The cycle length is kept constant for the analysis period and common for all intersections along the arterial, because this facilitates vehicle platoon progression. Besides the

cycle length constraints, platoon progression is achieved by optimizing the signals on the arterials for a pair of two intersections at a time and incorporating the delays from interrupting platoon progression in the objective function. Under these assumptions, the mathematical program for the arterial level signal optimization is a Mixed Integer Linear Program (MILP).

The optimization of signal settings for an arterial is based on a pairwise optimization strategy introduced in Newell (1964, 1967). This strategy entails optimization of signal timings for two intersections at a time accounting for the delays of all vehicles arriving at the two intersections during the time period of consideration. Newell (1967) showed that for heavy one-directional traffic near saturation and with no platoon dispersion, the best way to maintain progression and minimize stops is by synchronizing signals for consecutive intersections pairwise in the direction of the heaviest traffic.

The mathematical program that has been developed minimizes the total delay at two consecutive intersections, r and $r + 1$, for all vehicles that are present at the subject intersections during the design cycle, T . This includes the delays that vehicles leaving intersection r and arriving at $r + 1$ experience and also delays for those that leave intersection $r + 1$ and arrive at r . The decision variables are the green times for each phase i , $g_{i,T}^r$, at each intersection r in the design cycle, T , and the choice of green times determines the beginning time of the coordinated phase (i.e., offset). As a result, offsets are effectively optimized through this process only in cases that the coordinated phase is not the first one in the cycle. Otherwise offsets cannot be changed, because the cycle length remains constant. In those cases, in order to maintain progression in a selected direction, however, the offset between an intersection and its adjacent downstream intersection is set equal to the average free flow travel time between the two intersections under consideration.

Before starting the pairwise optimization, the critical intersection on the arterial is identified. The critical intersection is typically defined as the one with the highest intersection flow ratio or the one that has the heaviest transit traffic. Starting with the critical intersection, progression is maintained for the heaviest direction of traffic on the arterial. This means that the phase that serves the heaviest direction is designated as the coordinated one. Once the signal settings for the first two intersections, r and $r + 1$, are optimized, the next pair, $r + 1$ and $r + 2$, will be optimized with the beginning of green for the coordinated phase at $r + 1$ (i.e., offset) constrained by the optimization outcome of r and $r + 1$. This constraint, which is introduced in addition to the constraints presented in Section 3.1, ensures that the beginning of the green for the coordinated phase will be held constant when optimizing the second pair of intersections and can be expressed as:

$$\sum_{i=1}^{c^{r+1}-1} g_{i,T}^{r+1}(2) = \sum_{i=1}^{c^{r+1}-1} g_{i,T}^{r+1}(1) \quad (5.1)$$

where c^{r+1} is the coordinated phase, $g_{i,T}^{r+1}(1)$ are the optimal green times for phase i during cycle T at intersection $r + 1$ obtained from the optimization of the first pair of

intersections, and $g_{i,r}^{r+1}(2)$ are the green times obtained from the optimization of the second pair of intersections in which intersection $r + 1$ belongs. If the coordinated phase is the first in the cycle, then this constraint holds automatically since the optimal offset from the optimization of the first pair of intersections is equal to the predefined offset due to the cycle length constraints. The same constraint is applied to every pair of intersections on the arterial for which coordination should be maintained.

In case it is not clear which direction has the heaviest traffic, the same process can be repeated in the opposing traffic direction, and the signal settings that give the lowest total person delay can be chosen. This is particularly easy to do in practice because the mathematical program can be solved very quickly (as explained in Section 5.3). As a result, both optimizations can be performed fast enough for real-world implementations.

An example of this pairwise optimization procedure is illustrated in Figure 5.1. Assume that the signal settings of a four intersection arterial are being optimized. Furthermore, assume that the critical intersection is intersection 1, and the heaviest traffic direction is from 1 to 4 (Figure 5.1(a)). The signal settings are first optimized for intersections 1 and 2, taking into account the delay of the autos and transit vehicles for the incoming and shared links (i.e., links that connect the two intersections being optimized) as indicated by the arrows in Figure 5.1(b). After the signal settings for intersections 1 and 2 have been optimized, the signal settings for intersection 1 are fixed to the optimal ones. The signal settings for intersections 2 and 3 are then optimized with the offset for intersection 2 constrained by the optimization of 1 and 2 (Figure 5.1(c)). The same process is repeated pairwise until the signal settings of all four intersections are optimized.

5.2 Delay Estimation

For each pair of intersections, the auto delays that contribute to the objective function of the optimization consist of three terms: 1) the delay experienced by vehicles that travel on the incoming links during the design cycle T , 2) the delay experienced by vehicles that travel on the shared links during the design cycle T , and 3) the delay experienced by vehicles that did not get served during the previous cycle which constitute the residual queues at the approaches of the two intersections. This means that a platoon could be experiencing delay while traveling on the incoming link (approaching the first intersection it arrives at) and a portion of that platoon that continues in the subject network could experience delay while traveling on the shared link (approaching the second intersection it arrives at) during cycle T . Similarly, the objective function includes the delays for transit vehicles at the first intersection at which they arrive. For transit vehicles that continue in the network, the delays experienced at the second intersection during cycle T are also included in the objective function.

Since the optimization is conducted using a pairwise approach, as described in Section 5.1, the delays are calculated for each pair of intersections r and $r + 1$ in order

to optimize the signal settings for that pair. There is symmetry in the formulas for the delays of vehicles traveling in the direction of progression (i.e., from intersection 1 to 2 to 3, etc.) and those traveling in the opposing direction (i.e., from 3 to 2 to 1). Suppose that r is the first intersection of the pair being optimized, and the optimization is made from intersection r to $r + 1$ which is the second intersection of the pair. For any platoon, the first intersection at which it arrives is denoted by u , and the second intersection at which a portion of it arrives is denoted by v . This means that for a platoon traveling in the direction of progression $u = r$ and $v = r + 1$, while for a platoon traveling in the opposing direction $u = r + 1$ and $v = r$. The same holds for transit vehicles. This notation is used for the remainder of the chapter because delays are estimated by tracking platoons and transit vehicles. A detailed estimation of auto and transit vehicle delays that constitute the objective function of the mathematical program is presented next.

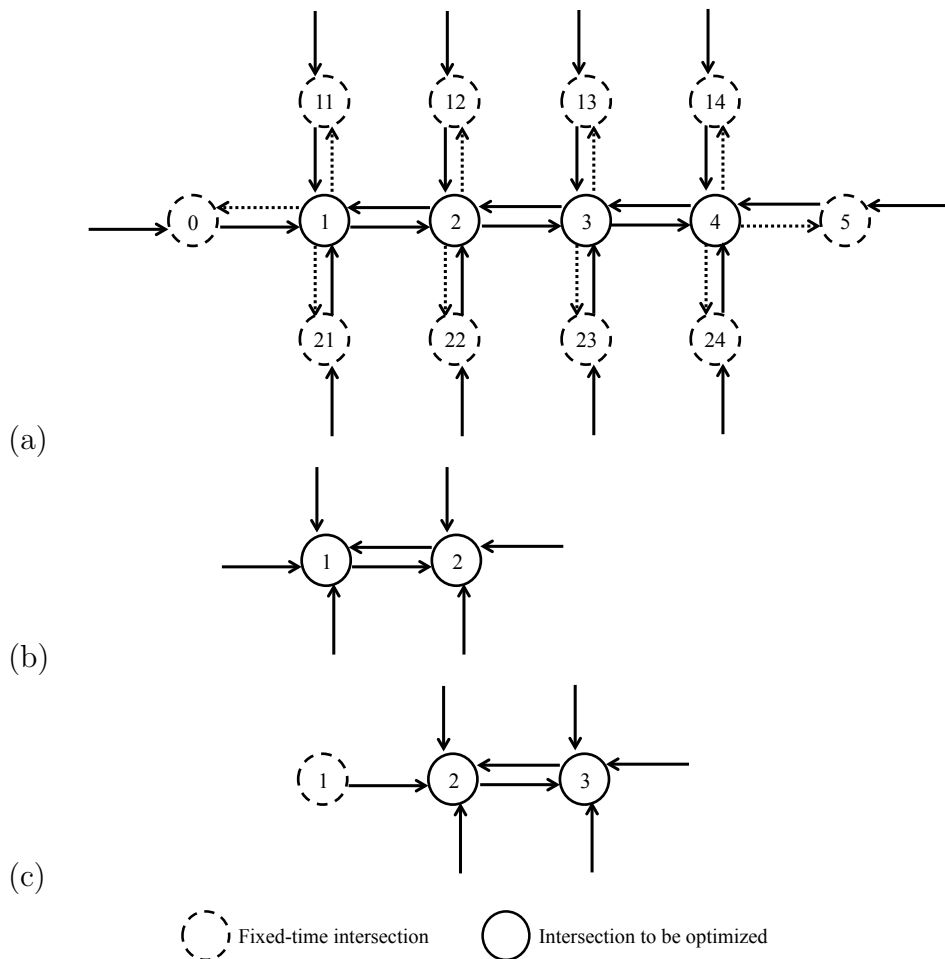


Figure 5.1. Pairwise Arterial Signals Optimization

5.2.1 Auto Delay

Auto delays are estimated assuming that vehicles arrive in platoons with no dispersion. This implies that all vehicles have the same behavior, travel with the same speeds, and maintain the same headways. In addition, since they leave from the upstream intersection at capacity (i.e., at saturation flow), they arrive at the downstream one at capacity. Once the vehicles get the green signal at the intersection, they are also served at capacity. Assuming that Kinematic Wave Theory (Lighthill & Whitham, 1955; Richards, 1956) holds, all vehicle trajectories are parallel at all times, as shown in Figure 5.2. This means that the last vehicle in a platoon that is stopped will experience the same delay as the first vehicle in the platoon that is stopped. This assumption simplifies the estimation of delays. So, the collective delay for all vehicles can be easily estimated knowing only the arrival time of the first vehicle in a platoon at intersection u , $t_{j,T}^u$, the size of that platoon, $P_{j,T}^u$, and the traffic conditions at the approach as expressed by the size of the residual queue of lane group j at the end of the previous cycle $T - 1$, $N_{j,T-1}^u$.

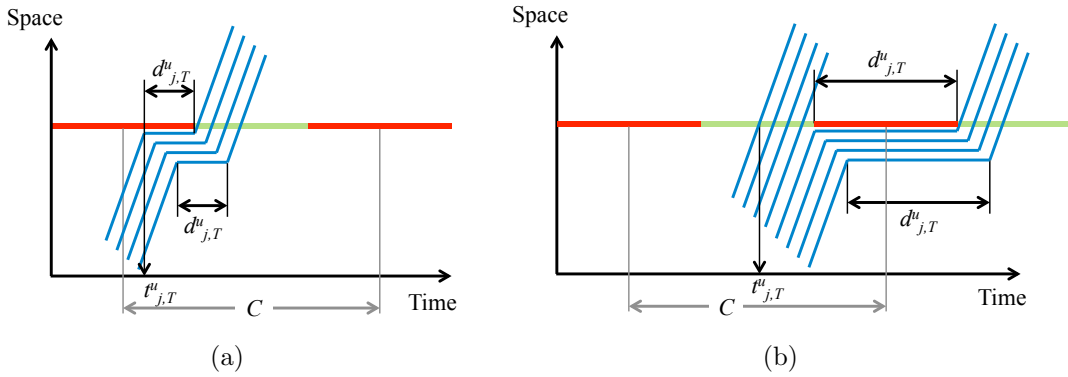


Figure 5.2. Auto Delay Estimation for Platoon Arrivals

In order to account for the impact of the signal control system and transit priority on the auto delays, the delay caused by interrupting the progression of platoons at signalized intersections is taken into account. The delay estimate for autos at both intersections includes delays caused by stopping the head of the platoon (Figure 5.2(a)) and/or the tail of it (Figure 5.2(b)). In addition, the delays experienced by vehicles that are left in the residual queue at the end of the previous cycle are included in the objective function under the same assumption that all of them experience the same amount of delay.

For the platoons that travel on the main arterial, the delay they experience at the second intersection of the pair they arrive at, v , is included in addition to the delay at the first intersection, u . This ensures that the effect of disrupting progression is accounted for in both directions. The same procedure could be followed for all platoons that arrive first at one of the two intersections under consideration (e.g., from cross streets) and their portion that continues downstream. However, this would

add to the computation time of the optimization process. In addition, the formulation of the system has focused only on the platoons traveling on the main arterial because they are usually the biggest and most critical for maintaining progression.

The following sections describe in detail the estimation of auto delays included in the objective function of the mathematical program. Note that for simplicity and reduced computation time, all equations are formulated assuming that there is only one platoon per cycle per lane group. However, the algorithm can easily be extended to include multiple platoons in a cycle for the same lane group as long as the arrival times and sizes of those platoons are known.

Auto Delay for Vehicles in Platoons at u

The delay for the autos arriving in platoons from incoming links at an intersection, u , depends on the actual arrival time of the platoon at the back of its lane group's queue, $t_{j,T}^u$, its size, $P_{j,T}^u$, and the residual queue's length, $N_{j,T-1}^u$, for the subject lane group. Platoon sizes and arrival times can be estimated in advance from the upstream intersection signal timings, while queue length can be estimated with information from sensing technologies or by keeping track of arrivals and departures as shown in (5.5). Based on their arrival times and sizes, as well as the traffic conditions, there are six cases for delay estimation of autos in platoons. For each case, auto platoon delay consists of two components: 1) the delay caused by stopping the head of the platoon, $D_{j,T}^{(H)u}$, which includes any delay incurred before the vehicles start being served and 2) the delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$, which corresponds to the delay experienced by vehicles after the platoon starts being served or after the end of green phase that serves it (whichever comes first) and until the beginning of the green phase that serves it in the next cycle. Note that the delay equations are formulated under the assumption of vertical queues. The six delay estimation cases for auto platoon arrivals, along with the formulas for estimating these delays for an intersection u are summarized next.

Case 1: Arrival before residual queue served, entire platoon served in green

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ before the time that the corresponding residual queue of j from the previous cycle $T - 1$, $N_{j,T-1}^u$, would have finished being served if there was enough green time available. There is enough available green time to serve the residual queue, and spare green time to serve all $P_{j,T}^u$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.2a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.2b)$$

$$P_{j,T}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u - N_{j,T-1}^u. \quad (5.2c)$$

where the beginning of cycle T for intersection u , t_T^u is as follows:

$$t_T^u = t_T + O_T^u \quad (5.3)$$

where $t_T = (T - 1)C$ is the beginning of cycle T at the critical intersection which is the first one to be optimized and O_T^u is the difference between the starting time of cycle T at intersection u and the critical intersection, which is the first to be optimized on the arterial. This quantity is determined *a priori* and does not change because the cycle lengths are common for all intersections.

In this case, all vehicles in the platoon experience delay caused by only stopping the head of the platoon at intersection u , $D_{j,T}^{(H)u}$, but no delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$. These values can be expressed as:

$$D_{j,T}^{(H)u} = P_{j,T}^u \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} - t_{j,T}^u \right) \quad (5.4a)$$

$$D_{j,T}^{(T)u} = 0. \quad (5.4b)$$

The number of vehicles in the residual queue, $N_{j,T-1}^u$, is calculated as:

$$N_{j,T-1}^u = \max \left\{ P_{j,T-1}^u + N_{j,T-2}^u - G_j^{e,u}(g_{i,T-1}^u)s_j^u, 0 \right\}. \quad (5.5)$$

Estimates of queue length can also be obtained with the use of detectors upstream of the stop line.

Case 2: Arrival before residual queue served, insufficient green to serve entire platoon

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ before the time that the corresponding residual queue of j , $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P_{j,T}^u$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.6a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.6b)$$

$$P_{j,T}^u \geq G_j^{e,u}(g_{i,T}^u)s_j^u - N_{j,T-1}^u. \quad (5.6c)$$

All vehicles in the platoon experience delay caused by stopping the head of the platoon $D_{j,T}^{(H)u}$, and a portion of the vehicles experience delay caused by stopping the

tail of the platoon, $D_{j,T}^{(T)u}$. These values can be expressed as:

$$D_{j,T}^{(H)u} = P_{j,T}^u \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} - t_{j,T}^u \right) \quad (5.7a)$$

$$D_{j,T}^{(T)u} = (P_{j,T}^u - G_j^{e,u}(g_{i,T}^u)s_j^u + N_{j,T-1}^u) \left(C - \frac{N_{j,T-1}^u}{s_j^u} \right). \quad (5.7b)$$

The delay estimate caused by stopping the tail of the platoon is equal to one cycle length minus the time it takes to serve the residual queue. This component is subtracted in order to avoid double counting since that delay component has already been captured by (5.7a). However, one can adjust this delay estimate accordingly to change the penalty imposed for stopping the tail of the platoon.

Case 3: Arrival before end of green, insufficient green to serve residual queue

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$, before the end of the green time for j , but there is not enough available green time to serve all $N_{j,T-1}^u$ vehicles in the residual queue. These conditions are summarized as:

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) \quad (5.8a)$$

$$N_{j,T-1}^u \geq G_j^{e,u}(g_{i,T}^u)s_j^u. \quad (5.8b)$$

All vehicles in the platoon experience delay caused by stopping the head of the platoon, $D_{j,T}^{(H)u}$, and by stopping the tail of the platoon, $D_{j,T}^{(T)u}$. These can be expressed as:

$$D_{j,T}^{(H)u} = P_{j,T}^u \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) \quad (5.9a)$$

$$D_{j,T}^{(T)u} = P_{j,T}^u (C - G_j^{e,u}(g_{i,T}^u)s_j^u). \quad (5.9b)$$

Case 4: Arrival after residual queue served, entire platoon served in green

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ after the time that the corresponding residual queue of j , $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, and there is enough spare green time to serve all $P_{j,T}^u$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^u \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.10a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.10b)$$

$$P_{j,T}^u \leq \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u. \quad (5.10c)$$

In this case, vehicles in the platoon do not experience any delay at intersection u . As a result, both the delay caused by stopping the head of that platoon, $D_{j,T}^{(H)u}$, and the delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$, are zero:

$$D_{j,T}^{(H)u} = 0 \quad (5.11a)$$

$$D_{j,T}^{(T)u} = 0. \quad (5.11b)$$

Case 5: Arrival after residual queue served, insufficient green to serve entire platoon

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ after the time that the corresponding residual queue of j , $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P_{j,T}^u$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^u \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.12a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.12b)$$

$$P_{j,T}^u \geq \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u. \quad (5.12c)$$

All vehicles in the platoon experience delay caused by stopping the tail of the platoon at intersection u , $D_{j,T}^{(T)u}$, but no delay caused by stopping the head of the platoon. The delay terms can be expressed as:

$$D_{j,T}^{(H)u} = 0 \quad (5.13a)$$

$$D_{j,T}^{(T)u} = \left[P_{j,T}^u - \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u \right] \times \left(t_{T+1}^u - t_{j,T}^u + R_j^{(1)u}(g_{i_{\text{next}}}^u) \right). \quad (5.13b)$$

Case 6: Arrival after the green

A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ after the end of the phase that can serve it. This case captures all arrivals not satisfying the conditions of cases 1 through 5, and it can also be expressed as:

$$t_{j,T}^u > t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u). \quad (5.14)$$

All vehicles in the platoon experience delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$, but no delay caused by stopping the head of the platoon. The delay

terms can be expressed as:

$$D_{j,T}^{(H)u} = 0 \quad (5.15a)$$

$$D_{j,T}^{(T)u} = P_{j,T}^u \left(t_{T+1}^u - t_{j,T}^u + R_j^{(1)u}(g_{i_{\text{next}}}^u) \right). \quad (5.15b)$$

Auto Delay for Vehicles in Platoons at v

An estimate of the delay that the platoons that travel on the shared links experience at the second intersection of the pair, v , is also included in the objective function. The delay estimation for stopping the head and tail of the platoon at intersection v is based on the same six cases described above. The only differences are the values for the size of the platoon that need to be adjusted based on the portion of the platoon continuing downstream and its arrival time at the second intersection, v , which is a function of its service time at the first intersection u . In addition, some of the estimates of the delays for vehicles that do not get served during cycle T are adjusted.

The size of the platoon in lane group j at the second intersection during cycle T is denoted by $P_{j,T}^v$, and an estimate of this value, $\hat{P}_{j,T}^v$, is used in the optimization. Measurements from detectors located at the upstream end of the shared link between the two intersections determine the auto demand that is expected to arrive at intersection v from the subject link. In order to obtain an estimate of the platoon size of a specific lane group j at v , the measured value from the detectors is multiplied by a factor ψ_j^v , indicating the portion of the incoming demand that will be joining lane group j at intersection v . The platoon size estimate is used instead of a direct calculation of platoon size from the signal settings and arrivals at the upstream intersection, because it reduces the number of bilinearities and trilinearities in the objective function, and as a result, it decreases the computation time of the optimization process.

It is also assumed that the arterial has the same number of lanes all the way through, so autos leave intersection u at saturation flow and arrive at intersection v at the saturation flow as well. When this is not the case, and the number of lanes upstream and downstream are different, an adjustment factor for saturation flow, $S_{j,u}^v$, is used to ensure that the vehicles in the platoon are arriving to v at a rate equal to the saturation flow for lane group j at v . In this case, delay expressions are also adjusted to represent platoon size per lane. Finally, some of the delay estimates have been adjusted to avoid multiplication of more than three decision variables as shown in Section 5.3.

The arrival time for a platoon at the back of its lane group's queue at the second intersection, $t_{j,T}^v$, is estimated based on the arrival case for the first intersection as follows:

$$t_{j,T}^v = \begin{cases} t_T^u + R_j^{(1)u}(g_{i,T}^u) + tt_{j,u}^v & \text{for cases 1, 2, 3, 6} \\ t_{j,T}^u + tt_{j,u}^v & \text{for cases 4, 5} \end{cases} \quad (5.16)$$

where $tt_{j,u}^v$ is the average free flow travel time to traverse the shared links between intersections u and v . For cases 1, 2, 3, and 6 the estimate of the platoon's arrival

time at v is based on the assumption that vehicles from the incoming platoon join the vehicles in the residual queue and travel together as one platoon. As a result, this new platoon is assumed to arrive at the downstream intersection $tt_{j,u}^v$ seconds after the beginning of green when the residual queue starts being served. For cases 4 and 5, the platoons are served by their first intersection as soon as they arrive. This implies that the residual queues are short, and as a result, it has been assumed that the majority of vehicles at the downstream intersection, v , is mainly vehicles from the arriving platoon at u . Therefore, the equation assumes that the arrival time of the platoon depends only on the service time of the platoon at u .

Auto Delay for Vehicles in Residual Queues

The delays for the vehicles that remain in the residual queues of lane groups at both intersections of the pair r and $r + 1$ are estimated based on the size of the residual queue and whether or not it can be entirely served during cycle T . Each case, along with the respective formulas for estimating delays for the vehicles in residual queues, is summarized below:

Case 1: Residual queue served in green

The residual queue of a lane group j that contains the vehicles that were not served by the end of cycle $T - 1$, $N_{j,T-1}^r$, can be entirely served during cycle T :

$$N_{j,T-1}^r \leq G_j^{e,r}(g_{i,T}^r)s_j^r. \quad (5.17)$$

So, the total delay experienced by all vehicles in the residual queue, $D_{j,T}^{(Q)r}$, can be expressed as:

$$D_{j,T}^{(Q)r} = N_{j,T-1}^r R_j^{(1)r}(g_{i,T}^r). \quad (5.18)$$

Case 2: Insufficient green to serve residual queue

The residual queue of a lane group j that contains the vehicles that were not served by the end of cycle $T - 1$, $N_{j,T-1}^r$, cannot be entirely served during cycle T :

$$N_{j,T-1}^r \geq G_j^{e,r}(g_{i,T}^r)s_j^r. \quad (5.19)$$

So, the total delay experienced by all vehicles in the residual queue, $D_{j,T}^{(Q)r}$, can be expressed as:

$$D_{j,T}^{(Q)r} = N_{j,T-1}^r R_j^{(1)r}(g_{i,T}^r) + (N_{j,T-1}^r - G_j^{e,r}(g_{i,T}^r)s_j^r) C \quad (5.20)$$

because the vehicles that do not get served will have to wait for an extra cycle before they start being served.

5.2.2 Transit Delay

In addition to auto delay, the objective function includes the delay for transit vehicles present at the two intersections during cycle T , which consists of two terms: 1) the delay transit vehicles experience at the first intersection they arrive, u and 2) the delay they experience at the downstream intersection, v , for those that travel on the main arterial to another intersection that belongs to the pair being optimized. In addition, transit vehicles that do not get served during the cycle in which they arrive experience an extra component of delay equal to $R_j^{(1)u}(g_{i_{\text{next}}}^u)$. This delay estimate is included as a penalty in the optimization in order to force the system to advance the phase and serve the transit vehicle earlier. The estimate is assumed to be equal to the delay that a transit vehicle would experience if it was the first one in the queue to be served and the next cycle was operating under some user-specified signal timings, $g_{i_{\text{next}}}^u$. The choice of the signal timings for the next cycle can change according to how much priority a transit vehicle should be given.

Perfect information on the arrival times of transit vehicles is assumed for the design cycle T just as for the isolated intersection case. Transit vehicles travel in mixed lanes with the autos, so the delay of a transit vehicle b that arrives at the back of its lane group's queue at intersection u at some time $t_{b,T}^u$ is the same as a platoon of size one that arrives at the same time at the queue (see Figure 5.3).

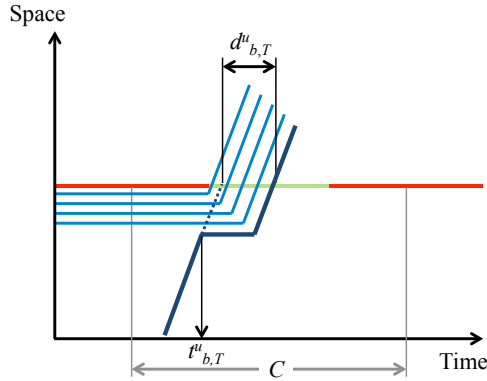


Figure 5.3. Transit Delay Estimation

Transit Delay at u

The estimation of the transit delay used in the optimization of each cycle T depends on the actual arrival time of the transit vehicle at intersection u , $t_{b,T}^u$, as well as whether the vehicle is served during cycle T or not, which also depends on traffic conditions on the subject approach (e.g., residual queue length). Note that the delay equations are formulated based on the assumption that a transit vehicle arrives at the back of the queue before or after the arrival of the platoon due to its dwell time at bus stops. This means that when arriving at the intersection, the bus observes

only the residual queue in front of it. The four delay estimation cases for transit arrivals, along with the formulas for estimating these delays for an intersection u are summarized next.

Case 1: Arrival before residual queue served, transit vehicle served in green

A transit vehicle that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{b,T}^u$ before the time the corresponding residual queue, $N_{j,T-1}^u$, would have finished being served and there is enough available green time to serve the residual queue in front of it. These conditions are summarized as:

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.21a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u. \quad (5.21b)$$

In this case, the transit vehicle will be served by cycle T and its delay, $d_{b,T}^u$, can be expressed as:

$$d_{b,T}^u = t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} - t_{b,T}^u. \quad (5.22)$$

Case 2: Arrival before residual queue served, transit vehicle not served in green

A transit vehicle that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{b,T}^u$ before the time the corresponding residual queue, $N_{j,T-1}^u$, would have finished being served, but there is not enough available green time to serve the residual queue in front of it. These conditions are summarized as:

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.23a)$$

$$N_{j,T-1}^u \geq G_j^{e,u}(g_{i,T}^u)s_j^u. \quad (5.23b)$$

In this case, the transit vehicle is not served during cycle T and it experiences delay, $d_{b,T}^u$, that can be expressed as:

$$d_{b,T}^u = t_{T+1}^u - t_{b,T}^u + R_j^{(1)u}(g_{i_{\text{next}}}^u). \quad (5.24)$$

As explained before this delay estimate includes the delay actually being experienced by the transit vehicle during cycle T but also an estimate of the delay experienced until the beginning of the green time in the next cycle, $T + 1$.

Case 3: Arrival after residual queue served and before the end of green

A transit vehicle that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{b,T}^u$ after the time the corresponding residual queue, $N_{j,T-1}^u$, would have finished being served and before the end of the green time for the phase that can serve j . There is enough available green time to serve the residual queue in front of it. These conditions are summarized as follows:

$$t_{b,T}^u \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.25a)$$

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) \quad (5.25b)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u. \quad (5.25c)$$

In this case, the transit vehicle is served as soon as it arrives at the intersection and as a result its delay, $d_{b,T}^u$, is zero:

$$d_{b,T}^u = 0. \quad (5.26)$$

Case 4: Arrival after the end of green

A transit vehicle that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{b,T}^u$ after the end of the green time for the phase that can serve j , which is expressed as follows:

$$t_{b,T}^u > t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u). \quad (5.27)$$

In this case, the transit vehicle is not served during cycle T , and it experiences delay, $d_{b,T}^u$, which can be expressed as:

$$d_{b,T}^u = t_{T+1}^u - t_{b,T}^u + R_j^{(1)u}(g_{i_{\text{next}}}^u). \quad (5.28)$$

Transit Delay at v

The delay for a transit vehicle that arrives at its second intersection, v , after it is served by the first intersection, u , is also included in the objective function to account for the impact that the signal timings at one intersection have on the other. This means that some level of progression for the transit vehicles between adjacent intersections is taken into account as well.

As before, the delay for these transit vehicles is based on the arrival time of the vehicles at the downstream intersection, v , and whether or not they can be served during cycle T , which depends on the traffic conditions. Their delays can be estimated based on one of the four cases presented above for intersection u . To estimate the delay, the arrival time for the transit vehicle at the its lane group's queue at second

intersection, v , denoted by $t_{b,T}^v$, depends on the arrival case at the first intersection, u . The two different cases are summarized as:

$$t_{b,T}^v = \begin{cases} t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + tt_{b,u}^v & \text{for case 1} \\ t_{b,T}^u + tt_{b,u}^v & \text{for case 3} \end{cases} \quad (5.29)$$

where $tt_{b,u}^v$ is the expected travel time for the shared link between intersections u and v for a transit vehicle b , and it includes the lost time due to transit stops. For cases 2 and 4, the transit vehicle is not served during cycle T . As a result, there is no delay at the downstream intersection, v , included in the objective function for cycle T .

5.3 Mathematical Program Formulation

In the preceding sections of this chapter, the estimation of auto and transit delay for a cycle and a pair of intersections has been presented. Overall, the total delay for the autos consists of three terms: 1) the delay experienced by any platoon at the first intersection of the pair at which it arrives, u , $D_{j,T}^u$, 2) the delay experienced by the two platoons that travel on the arterial on both directions at the second intersection at which they arrive, v , $D_{j,T}^v$, and 3) the delay experienced by the vehicles that are already in residual queues at both intersections in the pair being optimized, when the cycle under consideration starts, $D_{j,T}^{(Q)r}$.

The formulas presented in Section 5.2 are used to calculate the delays for auto passenger component of the objective function. The person delay component for autos in the objective function for the vehicles arriving in a platoon from an incoming link at intersection u is as follows:

$$\sum_{j \in J_{IN}^u} \bar{o}_a D_{j,T}^u = \bar{o}_a \sum_{j \in J_{IN}^u} \left(D_{j,T}^{(H)u} + D_{j,T}^{(T)u} \right) \quad (5.30)$$

where J_{IN}^u is the set of lane groups for the incoming links at intersection u and the sum of the delay components is given by:

$$\begin{aligned} D_{j,T}^{(H)u} + D_{j,T}^{(T)u} &= (h_{j,T}^1 + h_{j,T}^2) P_{j,T}^u \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} - t_{j,T}^u \right) \\ &+ h_{j,T}^2 \left(P_{j,T}^u - G_j^{e,u}(g_{i,T}^u) s_j^u + N_{j,T-1}^u \right) \left(C - \frac{N_{j,T-1}^u}{s_j^u} \right) \\ &+ h_{j,T}^3 P_{j,T}^u \left[\left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) + \left(C - G_j^{e,u}(g_{i_{\text{next}}}^u) \right) \right] \\ &+ \left[h_{j,T}^5 \left(P_{j,T}^u - \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u \right) + h_{j,T}^6 P_{j,T}^u \right] \\ &\times \left(t_{T+1}^u - t_{j,T}^u + R_j^{(1)u}(g_{i_{\text{next}}}^u) \right) \end{aligned} \quad (5.31)$$

Similarly, the person delay component for autos in the objective function for the two platoons traveling on the arterial (i.e., on shared links) and arriving at the second

intersection, v , that also belongs to the pair currently being optimized, is as follows:

$$\sum_{j \in J_{SH}^v} \bar{o}_a D_{j,T}^v = \bar{o}_a \sum_{j \in J_{SH}^v} \left(D_{j,T}^{(H)v} + D_{j,T}^{(T)v} \right) \quad (5.32)$$

where J_{SH}^v is the set of lane groups for the shared link at intersection v and the sum of the delay components is given by:

$$\begin{aligned} D_{j,T}^{(H)v} + D_{j,T}^{(T)v} &= (z_{j,T}^1 + z_{j,T}^2) \hat{P}_{j,T}^v \left[t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} - t_{j,T}^v \right] \\ &+ z_{j,T}^2 \left(\hat{P}_{j,T}^v - G_j^{e,v}(g_{i,T}^v) s_j^v + N_{j,T-1}^v \right) \left(C - \frac{N_{j,T-1}^v}{s_j^v} \right) \\ &+ z_{j,T}^3 \hat{P}_{j,T}^v \left[\left(t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) - t_{j,T}^v \right) + (C - G_j^{e,v}(g_{i_{\text{next}}^v})) \right] \\ &+ z_{j,T}^5 \left[\hat{P}_{j,T}^v - \left(t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) - t_{j,T}^v \right) s_j^v \right] \\ &\times (C - G_j^{e,v}(g_{i_{\text{next}}^v})) \\ &+ z_{j,T}^6 \hat{P}_{j,T}^v (C - G_j^{e,v}(g_{i_{\text{next}}^v})). \end{aligned} \quad (5.33)$$

In order to estimate this component of the objective function, estimates of the arrival times at the back of the queue of the corresponding lane group at the downstream intersection are needed, $t_{j,T}^v$, which are based on the cases shown in (5.16) can be expressed as follows:

$$t_{j,T}^v = (h_{j,T}^1 + h_{j,T}^2 + h_{j,T}^3 + h_{j,T}^6) \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + t_{j,u}^v \right) + (h_{j,T}^4 + h_{j,T}^5) (t_{j,T}^u + t_{j,u}^v) \quad (5.34)$$

Finally, the person delay for the autos that are already in the residual queue of an intersection u is estimated as follows:

$$\bar{o}_a \sum_{j=1}^{J^r} D_{j,T}^{(Q)r} = \bar{o}_a (x_{j,T}^1 + x_{j,T}^2) N_{j,T-1}^r R_j^{(1)r}(g_{i,T}^r) + \bar{o}_a x_{j,T}^2 (N_{j,T-1}^r - G_j^{e,r}(g_{i,T}^r) s_j^r) C \quad (5.35)$$

where J^r is the total number of lane groups at intersection r .

The total delay for the transit vehicles consists of two terms: 1) the delay experienced by the transit vehicles at the first intersection of the pair they arrive, u , $d_{b,T}^u$, and 2) the delay experienced by the transit vehicles that travel on the arterial at the second intersection they arrive, v , $d_{b,T}^v$. The person delay for transit vehicles can be estimated based on the equations presented in Section 5.2.2. The transit person delay component of the objective function for the transit vehicles at the first intersection at which they arrive during cycle T , $d_{b,T}^u$, is expressed as follows:

$$\begin{aligned} \sum_{b=1}^{B_T^u} o_{b,T}^u (1 + \delta_{b,T}^u) d_{b,T}^u &= \zeta_{b,T}^1 \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} - t_{b,T}^u \right) \\ &+ (\zeta_{b,T}^2 + \zeta_{b,T}^4) \left(t_{T+1}^u - t_{b,T}^u + R_j^{(1)u}(g_{i_{\text{next}}^u}) \right) \end{aligned} \quad (5.36)$$

and the person delay for those that continue in the network to the other intersection of the pair being optimized and arrive at their second intersection, v , during cycle T , experience delay, $d_{b,T}^v$, which is expressed as follows:

$$\sum_{b=1}^{B_T^v} o_{b,T}^v (1 + \delta_{b,T}^v) d_{b,T}^v = \eta_{b,T}^1 \left(t_T^v + R_j^{(1)v} (g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} - t_{b,T}^v \right) \quad (5.37)$$

$$+ (\eta_{b,T}^2 + \eta_{b,T}^4) \left(t_{T+1}^v - t_{b,T}^v + R_j^{(1)v} (g_{i_{\text{next}}}^v) \right). \quad (5.38)$$

In order to estimate this component of the objective function, estimates of the arrival times at the back of the queue of the corresponding lane group at the downstream intersection, $t_{b,T}^v$, are needed. These estimates are based on the cases shown in (5.29) and can be expressed as follows:

$$t_{b,T}^v = \zeta_{b,T}^1 \left(t_T^u + R_j^{(1)u} (g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + t_{b,u}^v \right) + \zeta_{b,T}^2 (t_{b,T}^u + t_{b,u}^v). \quad (5.39)$$

So, the objective function of the mathematical program that minimizes person delays for two intersections for cycle T , is as follows:

$$\begin{aligned} & \sum_{u=1}^2 \left(\sum_{j \in J_{IN}^u} \bar{o}_a D_{j,T}^u + \sum_{b=1}^{B_T^u} o_{b,T}^u (1 + \delta_{b,T}^u) d_{b,T}^u \right) \\ & + \sum_{v=1}^2 \left(\sum_{j \in J_{SH}^v} \bar{o}_a D_{j,T}^v + \sum_{b=1}^{B_T^v} o_{b,T}^v (1 + \delta_{b,T}^v) d_{b,T}^v \right) + \sum_{u=1}^2 \sum_{j=1}^{J^r} \bar{o}_a D_{j,T}^{(Q)r}. \end{aligned} \quad (5.40)$$

Constraints for Autos in Platoons at u

This section presents all of the constraints that are used to identify the appropriate case and determine the corresponding delay estimate portion of (5.31) to be included in the objective function of the mathematical program. A big value constant M is used with the binary variables to determine which constraints to activate for the relevant cases. For this mathematical program, M is set equal to TC for a cycle indexed by T with length C .

Case 1

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u} (g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + (1 - h_{j,T}^1) M \quad (5.41a)$$

$$N_{j,T-1}^u \leq G_j^{e,u} (g_{i,T}^u) s_j^u + (1 - h_{j,T}^1) M \quad (5.41b)$$

$$P_{j,T}^u \leq G_j^{e,u} (g_{i,T}^u) s_j^u - N_{j,T-1}^u + (1 - h_{j,T}^1) M \quad (5.41c)$$

Case 2

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + (1 - h_{j,T}^2)M \quad (5.42a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - h_{j,T}^2)M \quad (5.42b)$$

$$P_{j,T}^u + (1 - h_{j,T}^2)M \geq G_j^{e,u}(g_{i,T}^u)s_j^u - N_{j,T-1}^u \quad (5.42c)$$

Case 3

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) + (1 - h_{j,T}^3)M \quad (5.43a)$$

$$N_{j,T-1}^u + (1 - h_{j,T}^3)M \geq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.43b)$$

Case 4

$$t_{j,T}^u + (1 - h_{j,T}^4)M \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.44a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - h_{j,T}^4)M \quad (5.44b)$$

$$P_{j,T}^u \leq \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u + (1 - h_{j,T}^4)M \quad (5.44c)$$

Case 5

$$t_{j,T}^u + (1 - h_{j,T}^5)M \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.45a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - h_{j,T}^5)M \quad (5.45b)$$

$$P_{j,T}^u + (1 - h_{j,T}^5)M \geq \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u \right) s_j^u \quad (5.45c)$$

Case 6

$$t_{j,T}^u + (1 - h_{j,T}^6)M \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) \quad (5.46)$$

Constraints for Autos in Platoons at v

Similarly, the constraints that identify the appropriate case and the corresponding delay estimate portion of (5.33) for the objective function for the autos in platoons arriving at their second intersection, v , are as follows:

Case 1

$$t_{j,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} + (1 - z_{j,T}^1)M$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - z_{j,T}^1)M \quad (5.47a)$$

$$\hat{P}_{j,T}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v - N_{j,T-1}^v + (1 - z_{j,T}^1)M \quad (5.47b)$$

Case 2

$$t_{j,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} + (1 - z_{j,T}^2)M \quad (5.48a)$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - z_{j,T}^2)M \quad (5.48b)$$

$$\hat{P}_{j,T}^v + (1 - z_{j,T}^2)M \geq G_j^{e,v}(g_{i,T}^v)s_j^v - N_{j,T-1}^v \quad (5.48c)$$

Case 3

$$t_{j,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) + (1 - z_{j,T}^3)M \quad (5.49a)$$

$$N_{j,T-1}^v + (1 - z_{j,T}^3)M \geq G_j^{e,v}(g_{i,T}^v)s_j^v \quad (5.49b)$$

Case 4

$$t_{j,T}^v + (1 - z_{j,T}^4)M \geq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} \quad (5.50a)$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - z_{j,T}^4)M \quad (5.50b)$$

$$\hat{P}_{j,T}^v \leq \left(t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) - t_{j,T}^v \right) s_j^v + (1 - z_{j,T}^4)M \quad (5.50c)$$

Case 5

$$t_{j,T}^v + (1 - z_{j,T}^5)M \geq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} \quad (5.51a)$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - z_{j,T}^5)M \quad (5.51b)$$

$$\hat{P}_{j,T}^v + (1 - z_{j,T}^5)M \geq \left(t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) - t_{j,T}^v \right) s_j^v \quad (5.51c)$$

Case 6

$$t_{j,T}^v + (1 - z_{j,T}^6)M \geq t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) \quad (5.52)$$

Constraints for Autos in Residual Queues

The delay estimate portion of (5.35) for the objective function for the autos that are in residual queues is determined accordingly with the use of the following constraints:

Case 1

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - x_{j,T}^1)M \quad (5.53)$$

Case 2

$$N_{j,T-1}^u + (1 - x_{j,T}^2)M \geq G_j^{e,u}(g_{i,T}^u)s_j^u. \quad (5.54)$$

Constraints for Transit Vehicles at u

The constraints that identify the case in which a transit vehicle falls and the corresponding delay estimate portion of (5.36) to be included in the objective function for the transit vehicles arriving at the first intersection, u , are as follows:

Case 1

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + (1 - \zeta_{b,T}^1)M \quad (5.55a)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - \zeta_{b,T}^1)M \quad (5.55b)$$

Case 2

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} + (1 - \zeta_{b,T}^2)M \quad (5.56a)$$

$$N_{j,T-1}^u + (1 - \zeta_{b,T}^2)M \geq G_j^{e,u}(g_{i,T}^u)s_j^u \quad (5.56b)$$

Case 3

$$t_{b,T}^u + (1 - \zeta_{b,T}^3)M \geq t_T^u + R_j^{(1)u}(g_{i,T}^u)s_j^u + \frac{N_{j,T-1}^u}{s_j^u} \quad (5.57a)$$

$$t_{b,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) + (1 - \zeta_{b,T}^3)M \quad (5.57b)$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u + (1 - \zeta_{b,T}^3)M \quad (5.57c)$$

Case 4

$$t_{b,T}^u + (1 - \zeta_{b,T}^4)M \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) \quad (5.58)$$

Constraints for Transit Vehicles at v

Similarly, the constraints that identify the corresponding delay estimate portion of (5.31) to be included in the objective function for transit vehicles arriving at the second intersection, v , are as follows:

Case 1

$$t_{b,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^u}{s_j^v} + (1 - \eta_{b,T}^1)M \quad (5.59a)$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - \eta_{b,T}^1)M \quad (5.59b)$$

Case 2

$$t_{b,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} + (1 - \eta_{b,T}^2)M \quad (5.60a)$$

$$N_{j,T-1}^v + (1 - \eta_{b,T}^2)M \geq G_j^{e,v}(g_{i,T}^v)s_j^v \quad (5.60b)$$

Case 3

$$t_{b,T}^v + (1 - \eta_{b,T}^3)M \geq t_T^v + R_j^{(1)v}(g_{i,T}^v) + \frac{N_{j,T-1}^v}{s_j^v} \quad (5.61a)$$

$$t_{b,T}^v \leq t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) + (1 - \eta_{b,T}^3)M \quad (5.61b)$$

$$N_{j,T-1}^v \leq G_j^{e,v}(g_{i,T}^v)s_j^v + (1 - \eta_{b,T}^3)M \quad (5.61c)$$

Case 4

$$t_{b,T}^v + (1 - \eta_{b,T}^4)M \geq t_T^v + R_j^{(1)v}(g_{i,T}^v) + G_j^{e,v}(g_{i,T}^v) \quad (5.62)$$

Other Constraints

For platoons of lane group j at intersection u :

$$h_{j,T}^1 + h_{j,T}^2 + h_{j,T}^3 + h_{j,T}^4 + h_{j,T}^5 + h_{j,T}^6 = 1 \quad (5.63)$$

$$h_{j,T}^1, h_{j,T}^2, h_{j,T}^3, h_{j,T}^4, h_{j,T}^5, h_{j,T}^6 \in \{0, 1\} \quad (5.64)$$

For platoons of lane group j at intersection v :

$$z_{j,T}^1 + z_{j,T}^2 + z_{j,T}^3 + z_{j,T}^4 + z_{j,T}^5 + z_{j,T}^6 = 1 \quad (5.65)$$

$$z_{j,T}^1, z_{j,T}^2, z_{j,T}^3, z_{j,T}^4, z_{j,T}^5, z_{j,T}^6 \in \{0, 1\} \quad (5.66)$$

For autos in residual queues of lane group j :

$$x_{j,T}^1 + x_{j,T}^2 = 1 \quad \forall j \in J^r \quad (5.67)$$

$$x_{j,T}^1, x_{j,T}^2 \in \{0, 1\} \quad \forall j \in J^r \quad (5.68)$$

For a transit vehicle b at intersection u :

$$\zeta_{b,T}^1 + \zeta_{b,T}^2 + \zeta_{b,T}^3 + \zeta_{b,T}^4 = 1 \quad \forall b \quad (5.69)$$

$$\zeta_{b,T}^1, \zeta_{b,T}^2, \zeta_{b,T}^3, \zeta_{b,T}^4 \in \{0, 1\} \quad \forall b \quad (5.70)$$

For a transit vehicle b at intersection v :

$$\eta_{b,T}^1 + \eta_{b,T}^2 + \eta_{b,T}^3 + \eta_{b,T}^4 = 1 \quad \forall b \quad (5.71)$$

$$\eta_{b,T}^1, \eta_{b,T}^2, \eta_{b,T}^3, \eta_{b,T}^4 \in \{0, 1\} \quad \forall b \quad (5.72)$$

The following constraints correspond to the initial constraints of the mathematical program (3.1d) and (3.1e):

$$\sum_{i=1}^{I^r} g_{i,T}^r + \sum_{i=1}^{I^r} y_i^r = C \quad \forall r \quad (5.73)$$

$$g_{i,T}^r \geq g_{i \min}^r \quad \forall i, r \quad (5.74)$$

$$g_{i,T}^r \leq g_{i \max}^r \quad \forall i, r \quad (5.75)$$

All of the constraints above are described as follows:

- Constraints (5.41)–(5.46) ensure that the correct delay formula is added to the objective function for each of the autos arriving in platoons at their first intersection, u .
- Constraints (5.47)–(5.52) determine the respective formula for the two platoons arriving at their second intersection, v .
- Constraints (5.53)–(5.54) determine the delay formula for autos in residual queues at both intersections.
- Constraints (5.55)–(5.58) ensure that the equivalent delay formulas for transit vehicles arriving at their first intersection, u , and constraints (5.59)–(5.62) at their second intersection, v , is added to the objective function.
- Constraints (5.63)–(5.72) make sure that only one binary variable will be equal to one.

- Constraint (5.73) ensures that the green times for each phase at each intersection (i.e., the outcome of the optimization) plus the lost time, which is essentially the sum of the yellow intervals, add up to the common cycle length.
- Constraints (5.74)–(5.75) set the upper and lower bounds for the continuous decision variables.

Note that the mathematical program formulation as presented has bilinearities and trilinearities caused by multiplication between the continuous variables (i.e., green times for the two intersections that are considered for each pair, $g_{i,T}^r$) and binary decision variables (i.e., variables that determine which delay formula to be used for each of the cases presented in Section 5.2 for platoons, residual queues, and transit vehicles). The existence of bilinearities and trilinearities introduces non-linearities in the objective function and constraints. In order to avoid this problem, convex relaxations for bilinearities and trilinearities as described in Meyer & Floudas (2004) are used. Examples of such convex relaxations are presented next.

Define a variable μ to be equal to the bilinearity under consideration, for example:

$$\mu = \xi \sum_{i=1}^{l_j} g_{i,T}^r \quad (5.76)$$

where $\xi \in \{0,1\}$ is a binary variable and the continuous variables are $g_{i,T}^r \in [g_{i \min}^r, g_{i \max}^r]$. Then, μ replaces the bilinearity in the objective function and the following four constraints are added to the mathematical program:

$$\mu \geq \xi \sum_{i=1}^{l_j} g_{i \min}^r \quad (5.77)$$

$$\mu \geq \sum_{i=1}^{l_j} g_{i,T}^r + \xi \sum_{i=1}^{l_j} g_{i \max}^r - \sum_{i=1}^{l_j} g_{i \max}^r \quad (5.78)$$

$$\mu \leq \xi \sum_{i=1}^{l_j} g_{i \max}^r \quad (5.79)$$

$$\mu \leq \sum_{i=1}^{l_j} g_{i,T}^r + \xi \sum_{i=1}^{l_j} g_{i \min}^r - \sum_{i=1}^{l_j} g_{i \min}^r \quad (5.80)$$

Define ρ as equal to the trilinearity under consideration, for example:

$$\rho = \xi \chi \sum_{i=1}^{l_j} g_{i,T}^r \quad (5.81)$$

where $\xi, \chi \in \{0,1\}$ are binary variables and the continuous variables are $g_{i,T}^r \in [g_{i \min}^r, g_{i \max}^r]$. Then, ρ replaces the trilinearity in the objective function and the fol-

lowing seven constraints are added to the mathematical program:

$$\rho \geq \sum_{i=1}^{l_j} g_{i,T}^r + \chi \sum_{i=1}^{l_j} g_{i,\max}^r + \xi \sum_{i=1}^{l_j} g_{i,\max}^r - 2 \sum_{i=1}^{l_j} g_{i,\max}^r \quad (5.82)$$

$$\rho \geq \chi \sum_{i=1}^{l_j} g_{i,\min}^r + \xi \sum_{i=1}^{l_j} g_{i,\min}^r - \sum_{i=1}^{l_j} g_{i,\min}^r \quad (5.83)$$

$$\rho \geq 0 \quad (5.84)$$

$$\rho \leq \xi \sum_{i=1}^{l_j} g_{i,\max}^r \quad (5.85)$$

$$\rho \leq \chi \sum_{i=1}^{l_j} g_{i,\max}^r \quad (5.86)$$

$$\rho \leq \sum_{i=1}^{l_j} g_{i,T}^r + \chi \sum_{i=1}^{l_j} g_{i,\min}^r - \sum_{i=1}^{l_j} g_{i,\min}^r \quad (5.87)$$

$$\rho \leq \sum_{i=1}^{l_j} g_{i,T}^r + \xi \sum_{i=1}^{l_j} g_{i,\min}^r - \sum_{i=1}^{l_j} g_{i,\min}^r \quad (5.88)$$

After performing the convex relaxations of bilinearities and trilinearities, the mathematical program consists of an objective function that is linear in its continuous and binary variables and has linear constraints. As a result, the final mathematical program is a Mixed Integer Linear Program (MILP) that can be solved very quickly, with computation times on the order of 5 to 10 seconds for a signalized arterial with four intersections as shown from the tests performed in the Section 5.5.

5.4 Study Site

The performance of the person-based traffic responsive signal control system is tested at four signalized intersections of a real-world arterial. In particular, the study site consists of the intersections along San Pablo Avenue at Ashby Avenue, Heinz Avenue, Grayson Street, and Dwight Way located in Berkeley, California. This segment of San Pablo Avenue has been selected due to the variety in phasing schemes utilized on these four intersections and the existence of conflicting bus routes at two out of the four intersections.

Figure 5.4 shows the segment of San Pablo Avenue along with the bus lines that travel through the four intersections. The link lengths between the intersections of the selected segment vary from 220 to 550 meters and the existing signal control is fixed-time coordinated. Figure 5.5 presents the phasing and green times for all intersections during the evening peak. As indicated in the figure, the intersections consist of a variety of signal phasing schemes that cover all of the basic possible phasing schemes. All intersections operate under a common cycle length of 80 seconds and

the demand used for the tests corresponds to the evening peak hour (4–5pm). During that time period, all four intersections operate in undersaturated traffic conditions with intersection flow ratios varying from 0.3 to 0.6. The intersection of Ashby and San Pablo Avenues is the critical one.

Five bus routes travel through the segment under consideration in mixed traffic lanes with headways that vary from 12 to 30 minutes on each route. This corresponds to an average of 24 buses per hour for the analysis period. The numbers next to the directional arrows in Figure 5.4 correspond to the different bus routes. Of the buses that travel in the corridor, 60% travel on San Pablo Avenue and 40% on the two cross streets: Ashby Avenue and Dwight Way. At these cross streets, the bus routes travel in two conflicting directions. The location of the bus stops varies with some of them being located nearside and some others farside (Figure 5.4). As in the case of the intersection of University and San Pablo Avenues, the bus schedule is available at the Alameda-Contra Costa Transit District’s website (AC Transit, 2011).

5.5 Evaluation

The arterial person-based traffic responsive signal control system has been tested using data from the study site described in Section 5.4. First, tests have been performed for a few cycles assuming that perfect information exists on the platoon sizes and the arrival times of platoons and transit vehicles at the intersections. These give an idea of the maximum benefit that could be achieved by the proposed system. Next, tests have been performed with Emulation-In-the-Loop Simulation (EILS) to evaluate the system when perfect information is not available and predictions of inputs are based on measured quantities from the simulated network as would be done in reality. For these scenarios, the warm-up period has been set equal to the common cycle length of all intersections in the study segment of San Pablo Avenue. Each scenario has been evaluated ten times and the average values of these ten replications are presented in this section.

5.5.1 Test Type I: Deterministic Vehicle Arrivals

Deterministic arrival tests have been performed for the four-intersection segment for five cycles and the total, auto, and bus passenger delays of all intersections are collectively reported in Table 5.1 for the three optimization scenarios tested (i.e., TRANSYT-7F, vehicle-based optimization, and person-based optimization). A comparison of the person-based optimization with the vehicle-based optimization indicates that the person-based traffic responsive signal control system can achieve a reduction in the total person delay by 2.7% by reducing the delay of the bus users by 9.7% and increasing auto user delay by only 1.5%. A comparison of the delays obtained by vehicle-based optimization with the ones obtained with optimal fixed signal settings from TRANSYT-7F shows that the optimal signal settings of vehicle-based optimization outperform those from TRANSYT-7F. This is the same result as

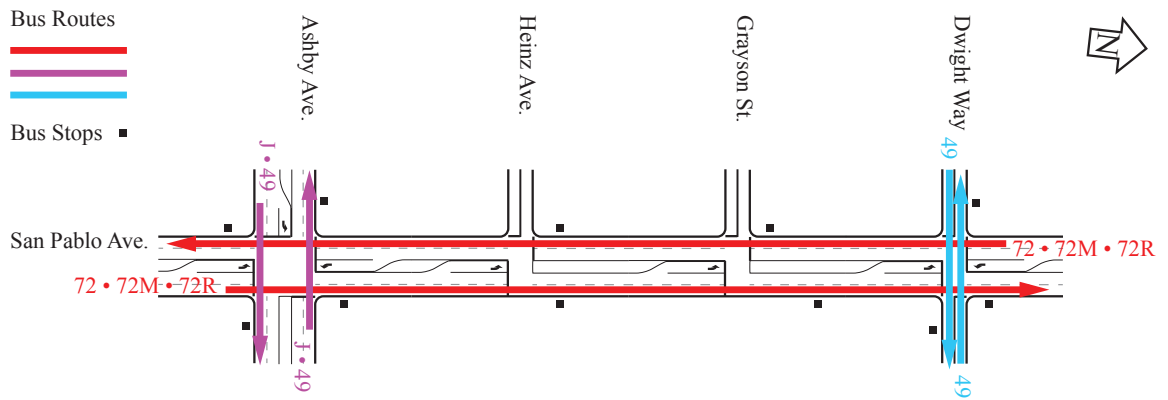


Figure 5.4. San Pablo Avenue Layout (not to scale)

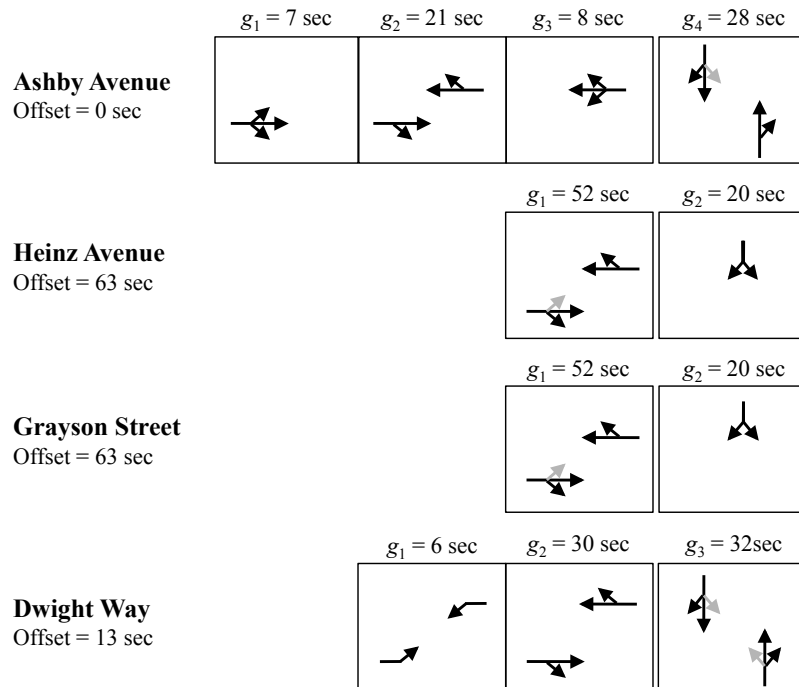


Figure 5.5. Signal Phasing and Green Times for San Pablo Avenue Segment

the one observed for the isolated intersection. Therefore, evaluation of the person-based traffic responsive signal control system for the simulation tests that follow is performed by comparing person delays from person-based optimization with the ones from the vehicle-based optimization, because the vehicle-based optimization provides the lowest delays that can be achieved for autos.

5.5.2 Test Type III: Stochastic Vehicle Arrivals

EILS tests have been performed with AIMSUN for one hour of operations, and the outcomes from the two optimization methods, vehicle-based (Scenario 1) and person-based (Scenario 2), are compared. The evening peak average flows are used as the input for auto demand in the simulation tests. The auto inter-arrival times on the incoming links are simulated to follow an exponential distribution. However, these vehicles stop at upstream fixed-time signalized intersections before they arrive at the intersections on the arterial. This ensures that vehicles arrive in platoons at the intersections being optimized. As a result, the auto demand input for an approach is estimated based on measurements from detectors located at the upstream end of each of the incoming links under consideration. Exponential smoothing is used on the measured counts during the previous cycle, as in (4.50) in order to estimate the demand of the respective lane group for the next cycle. Predictions of auto arrival times at the intersections are based on an average free flow speed of 45 km/hr. The average auto occupancy, \bar{o}_a , is assumed to be 1.25 passengers per vehicle.

The timetable of the bus arrivals at the entry links of the network is fixed and based on headways obtained from the actual schedule posted on the Alameda-Contra Costa Transit District website (AC Transit, 2011). The arrival time of the buses at the intersections is predicted based on their location on a link at the end of the previous cycle. Information on the location of vehicles and bus stops determine the estimated arrival time of a bus at the intersection. For simplicity, dwell times at all bus stops and for all buses are set to 30 seconds. For each bus that stops, an

Table 5.1. Person Delays on the Arterial Segment for $\bar{o}_b/\bar{o}_a = 40/1.25$ and Five Signal Cycles of Traffic Operations (Test Type I)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 0: TRANSYT-7F (Fixed Settings)	5.91	2.70	8.61
Scenario 1: Vehicle-based Optimization	4.14	2.45	6.60
Scenario 2: Person-based Optimization	4.20	2.22	6.42
% Change in person delay between Scenarios 1 & 2	1.45%	-9.73%	-2.71%

Table 5.2. Person Delays on the Arterial Segment $\bar{o}_b/\bar{o}_a = 40/1.25$ and 1 Hour of Traffic Operations (Test Type III)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 1: Vehicle-based Optimization	137.61	59.29	196.90
Scenario 2: Person-based Optimization	135.07	57.96	193.03
% Change in person delay between Scenarios 1 & 2	-1.84%	-2.25%	-1.96%

additional 6 seconds are added to its estimated travel time to reach the intersection in order to account for lost time due to acceleration and deceleration. The average speed assumed for buses is 36 km/hr which is slower than the free-flow speed for autos. Bus passenger occupancies, o_b , are fixed to 40 passengers per vehicle on all buses.

The green times for the next cycle, $g_{i \text{ next}}^r$, are set to be the same as the fixed optimal signal timings provided by TRANSYT-7F for the specific traffic conditions under evaluation. In addition, the upper bounds for the green times of the phases, $g_{i \text{ max}}^r$, are set equal to $C - \sum_{i=1}^{I^r} y_i^r$ at each intersection r . Non-zero lower bounds for the green times of each phase, $g_{i \text{ min}}^r$, are also introduced for the green times for each intersection r to ensure that all phases are allocated some minimum green time. A total minimum green time of 7 seconds is assigned to each of the left-turn phases and 12 seconds to each of the through phases. The resulting mathematical program is an MILP, which is solved with CPLEX (IBM, 2011). CPLEX is very efficient in solving this type of problem. As a result, optimization of signal settings for four intersections can be performed in less than 10 seconds which is sufficiently short time for real-world implementations.

Table 5.2 presents the person delay for auto users, bus users, and all travelers obtained from the two scenarios tested. A comparison of the outcome of the person-based optimization with the one obtained from the vehicle-based optimization indicates that the proposed signal control system achieves a reduction in total person delay by 1.8% for the arterial segment. This translates to a 2.3% reduction of bus passenger delay and a 2.0% reduction of auto user delay.

Despite the expectations that person-based optimization would result in higher delays for auto users, the test presented here shows that person-based optimization could lead to lower delays for all users compared to vehicle-based optimization. This can be attributed to three reasons:

1. Autos that are traveling on the same direction as transit vehicles benefit from the provision of priority, and as a result, their delays are reduced along with those of transit. In particular, the higher the auto demand in the direction of

Table 5.3. Person Delays per Type of Approach on the Arterial Segment for $\bar{o}_b/\bar{o}_a = 40/1.25$ and 1 Hour of Traffic Operations (Test Type III)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
<i>Main Arterial Northbound</i>			
Scenario 1: Vehicle-based Optimization	48.42	29.18	77.60
Scenario 2: Person-based Optimization	48.18	28.59	76.77
% Change in person delay between 1 & 2	-0.50%	-2.03%	-1.07%
<i>Main Arterial Southbound</i>			
Scenario 1: Vehicle-based Optimization	52.84	22.82	75.66
Scenario 2: Person-based Optimization	52.09	22.49	74.58
% Change in person delay between 1 & 2	-1.42%	-1.42%	-1.42%
<i>Cross Streets</i>			
Scenario 1: Vehicle-based Optimization	36.34	7.29	43.63
Scenario 2: Person-based Optimization	34.80	6.87	41.67
% Change in person delay between 1 & 2	-4.24%	-5.76%	-4.49%

transit vehicles, the higher the benefit in person delay reduction for auto users. This is shown with the breakdown of passenger delays for cross streets and the arterial per direction (Table 5.3).

2. Since the proposed system is used to optimize two intersections at a time and minimize total person delay for one cycle, it does not guarantee global optimality for the whole hour and arterial segment. This explains the fact that person-based optimization outperforms vehicle-based optimization for auto delays.
3. Performance of the two types of optimization through simulation is highly dependent on the accuracy of auto demand estimates and the auto and transit vehicle arrival predictions at the intersections. As a result, it is possible that person-based optimization does not operate as expected when inaccuracies exist in the arrival estimates, because the mathematical program formulation does not account for such inaccuracies.

Table 5.3 shows the auto, bus, and total passenger delays for each arterial direction and for the cross streets. Although the southbound direction has lower auto traffic demand than the northbound direction, both directions have similar bus flows. Note that both directions are served by the same phases at all intersections, except Ashby Avenue, and as a result, they experience similar green times. The observed difference between the results of the person-based and vehicle-based optimization are similar for both directions.

Table 5.4. Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$, $\delta_{b,T}^r = 1$ and 1 Hour of Traffic Operations (Test Type III)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 1: Vehicle-based Optimization	137.61	59.29	196.90
Scenario 2: Person-based Optimization	135.16	57.57	192.74
% Change in person delay between Scenarios 1 & 2	-1.77%	-2.90%	-2.13%

These results show a reduction in delays for bus users by 5.8% and for auto users by 4.2% for cross streets, compared to the vehicle-based optimization. On cross streets, there are fewer buses and less auto traffic traveling compared to the main arterial. As a result, provision of priority to buses on cross streets leads to longer green time for them which also substantially benefits auto users. This outweighs the delay increase to cross-street auto traffic caused by priority provision to the high frequency transit vehicles traveling on the main arterial.

Tests have also been performed under the assumption that all transit vehicles arrive late at the intersections in the network. Accounting for their schedule delay translates into weighting the delay of transit vehicles by a factor of 2, in this case $\delta_{b,T}^r = 1$. A comparison of vehicle-based and person-based optimization indicates that when schedule delay is considered, the benefit to transit users improves to a 3% reduction (Table 5.4). At the same time, autos benefit by the extra green provided to serve buses faster. The ratio of average passenger occupancy of buses over autos is also expected to affect the level of priority provided.

Overall, the simulation tests indicate that the person-based traffic responsive signal control system for arterials is promising for achieving lower total person delays in the system and providing priority to transit vehicles without imposing extra delays on autos, while in some cases it even reduces auto user delays. The level of benefit obtained depends on the layouts, phasing schemes, as well as on the traffic and transit operating characteristics of the intersections. For example, a phasing scheme where the coordinated phase is not the first one would lead to more flexibility in providing priority because one has the advantage of being able to adjust the beginning of the coordinated phase. The results have also demonstrated the need for better demand and arrival prediction algorithms as well as incorporating input inaccuracies into the delay estimation. These would improve the accuracy of the optimization input and would lead to a more robust signal control system. This is critical for the success of the system in real-world settings.

5.6 Extension to Networks

The person-based traffic responsive signal control system that has been developed for an arterial can be readily extended to networks. The signal control system is formulated based on the assumption that vehicles from cross streets arrive in platoons from upstream signalized intersections that operate under fixed-time control. This structure facilitates extension to signalized arterial networks. The intersection from which the pairwise optimization is initiated and the direction in which it progresses are both specified by the user. As a result, certain rules can be used to choose which intersection, arterial, and direction should be optimized first in a network. Once signal settings are obtained on this first arterial, the timings can be fixed, and an intersection along the arterial can be used as a starting point for optimizing another crossing arterial. By repeating this process, an entire network of arterials may be optimized, with progression on some arterials being prioritized over others.

Consider the grid street network shown in Figure 5.6. The first intersection and arterial for optimization would be selected by some criteria, for example choosing the critical intersection with the highest intersection flow ratio and the arterial with the greatest passenger flow in cars and buses. This first arterial may be optimized using the procedure described and evaluated in this chapter. Figure 5.6(a) illustrates the selected arterial starting with intersection 1 while being optimized in the direction of increasing intersection number (i.e., 1, 2, 3, ...). After the signal settings for all of these intersections are optimized, then they are considered fixed, and the pairwise optimization may be performed starting from the first arterial and moving outward. For example, if intersection 1 has a busy cross street, the pairwise optimization may be implemented advancing in both directions away from the first arterial as shown in Figure 5.6(b). Ultimately, this results in a branching pattern that can cover at least the busiest arterials in a network.

Further work is required to identify the most effective way to implement arterial signal optimization on a two-dimensional network. An obvious challenge compared to the single arterial problem is that optimization in a branching pattern will result in many instances where pairwise optimization may approach an intersection from two directions. For example, in Figure 5.6, it is not clear whether the next optimization should be of the arterial from intersection 11 to 12 to 13 or from intersection 2 to 12 to 32. In the interest of minimizing person delays, one strategy may be to optimize arterials in the order of passenger flows prioritizing busier streets first. Alternatively, there may be particular patterns that yield consistently well-performing results network-wide.

Although there is no guarantee that these methods will achieve a global optimum solution for minimizing person delay in a network, there are many cases when this can be expected to outperform static optimization methods that are currently used in practice. A straightforward implementation of the arterial level signal control system is the case of major one-way paired arterials. A related extension is the optimization of signals for traffic leaving a central location such as in an evening peak or following

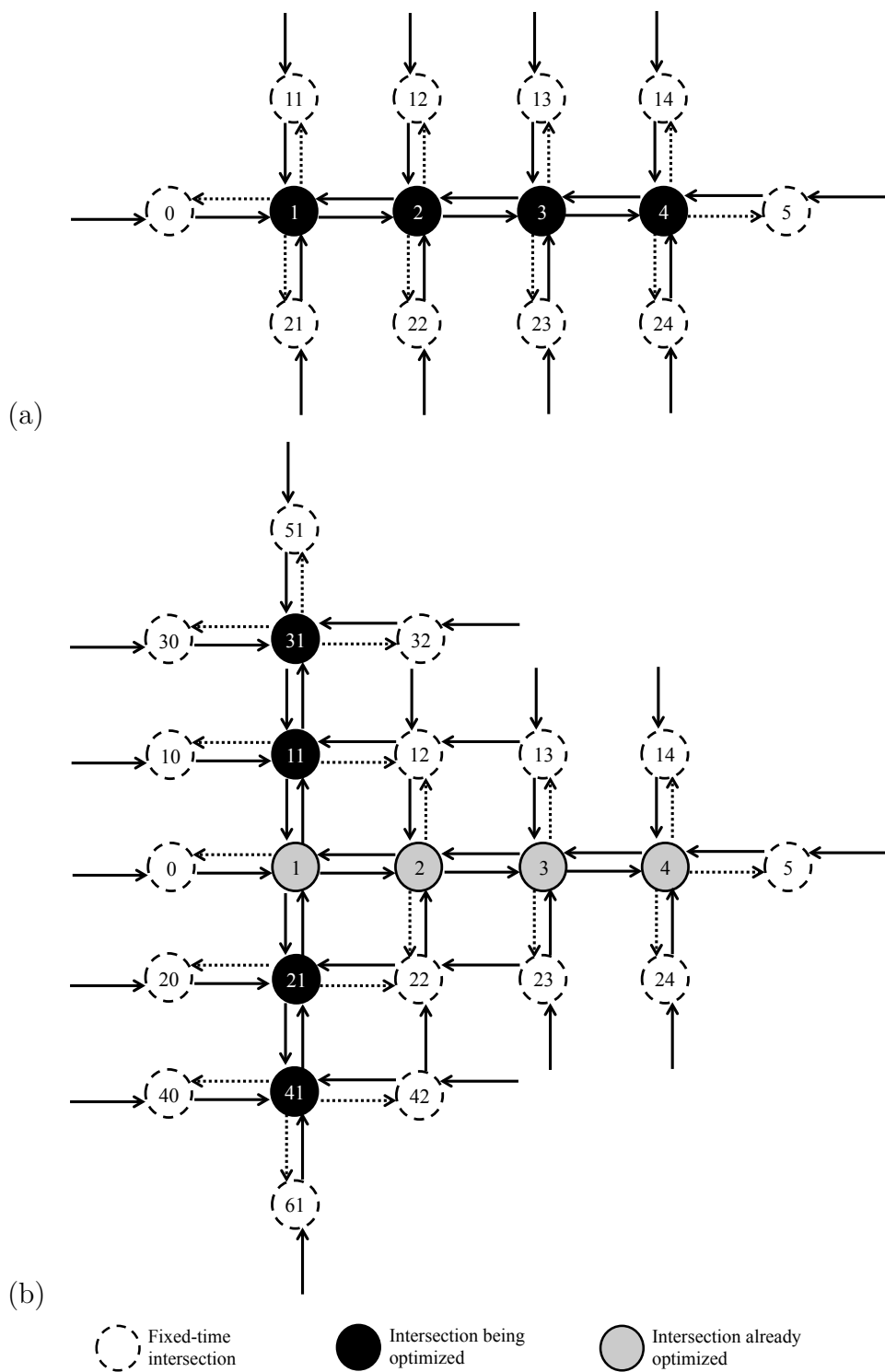


Figure 5.6. Arterial Signal Optimization in a Network

a special event. Such traffic is likely to travel in diverging directions, so the pairwise optimization of arterials can easily be implemented to follow the same directions as the dominant demand. Finally, a direct extension is the case of two major crossing arterials. The signal control system will first optimize the signal settings on the one arterial, keep the common intersection fixed, and then optimize the signal settings on the cross arterial.

5.7 Summary of Findings

This chapter has presented the formulation and testing of the arterial level person-based traffic responsive signal control system. The system has been evaluated with deterministic arrival tests to show that under perfect information about inputs for demand and arrival times, it outperforms static signal settings provided by TRANSYT-7F, even for auto delays. In addition, the system has been evaluated through simulation to test its performance under more realistic traffic and transit operations.

The person-based traffic responsive signal control system has been shown to reduce the delay for all travelers and bus users at an intersection by up to 5% for the selected study site and in most cases to reduce auto user delay as well. This is due to the fact that autos traveling on links with buses also benefit by the provision of transit priority. In addition, since the proposed system does not guarantee global optimality for the whole arterial, it is likely that the two optimization scenarios result in different traffic conditions that could make the same replication not be exactly comparable for the two scenarios. Inaccuracies in the input estimates deteriorate the performance of both types of optimization that could contribute to the decrease in auto user delay when the results are compared between these two scenarios. Furthermore, the tests have shown that buses traveling on cross streets with low auto demand experience very high benefits when transit priority is provided due to the much higher weight cross streets have with person-based optimization compared to vehicle-based optimization. Finally, accounting for schedule delay provides additional benefit to transit users without negatively impacting auto users.

The results from the evaluation of the system through simulation are promising for reducing person delay for an arterial segment and providing priority to transit vehicles. In addition to its performance, the low computation times and requirement only of readily deployable technologies contribute to the feasibility and economic viability of its implementation in real-world settings. The main limitations for testing the system through simulation or in real-world arterials are the input prediction algorithms, the fact that the formulation does not account for the estimation inaccuracies, and the assumption about negligible platoon dispersion. Prediction algorithms need to be designed and calibrated carefully to be able to provide the optimization with accurate input estimates for all levels of traffic conditions. In addition, the delay equations need to be adjusted to account for inaccuracies in the input estimates. The assumption that there is no platoon dispersion can be relaxed by splitting the platoons for each approach into smaller platoons and adjusting the mathematical program

formulation to capture delays appropriately for more than one platoons per lane group. Note, however, that more decision variables would lead to higher computation times, so one should consider the benefit of accounting for platoon dispersion versus the additional complexity of the optimization. Finally, the system can be expanded to arterial signalized networks by using the proposed pairwise optimization method along multiple arterials as outlined in Section 5.6.

Chapter 6

Conclusions

The need for efficient and sustainable management of multimodal transportation systems is steadily increasing due to growing demand in urban networks that are already reaching capacity. Increases in traffic congestion caused by growth in population and car ownership threaten mobility in cities around the world. The challenge is made more difficult, because the problem must be addressed with limited funds, so it is imperative that existing infrastructure systems be used efficiently. With traffic signal control systems already widely deployed in urban street networks, one of the most cost-effective ways to improve efficiency and sustainability of urban transportation systems is to develop signal control strategies that enhance person mobility. In order to achieve this goal, these strategies must resolve conflicts between travel modes and provide priority to higher occupancy transit vehicles while considering all users of the road network in their design.

This dissertation addresses this problem by answering the following question: *How should traffic signal control systems be designed so that they provide priority to transit vehicles traveling in conflicting directions, while minimizing the impacts on general traffic in urban networks?*

A solution for this problem has been made possible through the development of a person-based traffic responsive signal control system. The remaining sections of this chapter include a summary of the key research findings, the contribution of the dissertation, and future research directions.

6.1 Summary of Research Findings

A person-based traffic responsive signal control system has been developed, tested and evaluated on the isolated intersection and signalized arterial level. The system provides priority to transit vehicles with a minimum impact to auto traffic. Evaluation tests have shown that the proposed system outperforms static optimized signal settings obtained from state-of-the-art software such as TRANSYT-7F, even for auto delays.

In particular, for the case of an isolated intersection, the main findings from the evaluation tests are as follows:

1. The person-based traffic responsive signal control system can achieve substantial reductions in total passenger delay for a variety of intersection traffic conditions and transit operating characteristics by greatly reducing transit delay and imposing very low additional delay on auto traffic compared to the vehicle-based optimization scenario.
2. The negative impacts on autos diminish with higher auto traffic demand. For congested traffic conditions, person-based and vehicle-based optimization converge to the same solution because high auto traffic outweighs the higher occupancies of transit vehicles.
3. Higher transit passenger occupancies result in higher person delay reductions until the system starts operating close to saturation at which point the optimization converges to the same solution as with lower passenger occupancies. This happens because there is no flexibility to provide additional priority.
4. The system is robust since uncertainty in auto demand and transit vehicle arrivals does not negate the benefits.

The major findings from the evaluation tests of the system on a signalized arterial are as follows:

1. The person-based traffic responsive signal control system can reduce total passenger delay for all intersections by decreasing transit passengers' delay without substantially impacting auto users' delay compared to the vehicle-based optimization scenario.
2. Autos traveling in the same direction as high frequency transit traffic also benefit from the provision of priority because their approach is weighted more than in the vehicle-based optimization.
3. Transit vehicles traveling on cross streets with low auto traffic experience very high benefits when priority is provided. Low traffic demand leads to low weighting factors and as a result, transit vehicles on the cross streets suffer high delays with vehicle-based optimization. However, cross-street delays have a much higher weight with person-based optimization. This outweighs the delay increase to cross-street auto traffic caused by priority provision to the high frequency transit vehicles of the main arterial.
4. Accounting for schedule delay in providing priority increases the benefit for transit users without negatively affecting auto traffic. Incorporating schedule delay increases the weight imposed on transit users' delay, and as a result, it leads to higher level of priority for transit vehicles.

6.2 Contribution

The main contribution of this dissertation is the development of a person-based traffic responsive signal control system that accounts for passenger occupancy and schedule delay to provide priority to transit vehicles while maintaining progression for auto traffic. By minimizing person delays instead of using more common strategies that minimize vehicle delays, priority is provided to higher occupancy transit vehicles. At the same time, the schedule adherence of transit vehicles is taken into account in order to provide priority only to the vehicles that are late. This allows for a more systematic and efficient treatment of the issue of conflicting transit routes. By incorporating delay penalties for interrupting a platoon's progression, the system also minimizes the negative impact that provision of priority would have on the auto traffic. This system is particularly advantageous because its underlying optimization process can be solved quickly. Therefore, it provides optimal signal settings in sufficiently short time to allow for real-time operations in contrast with other signal control systems.

This dissertation also presents a systematic evaluation of the performance of such a person-based system and its sensitivity in transit passenger occupancy and auto traffic demand on isolated intersections and signalized arterials. The evaluation provides insights on ranges of traffic conditions and transit operating characteristics for which such a system is beneficial for all users.

Unlike other signal control systems, the proposed system is generic because it can be implemented on any intersection layout and signal phasing scheme. It is also flexible because the user can weigh the relative merit of auto and transit delays as desired and even replace the weighting factors in the objective function to minimize any type of metric, such as total travel cost, by using values of time as the weights in the objective function. Another advantage is that it can be implemented with readily deployable technologies, that are often installed to serve other purposes in the system (e.g., AVL systems are used to evaluate the on-time performance of a bus route). Finally, the formulation of this system facilitates its expansion to signalized arterial networks.

Implementation of the proposed system is expected to reduce bus bunching and improve overall schedule adherence which would lead to more reliable transit operations and potentially increase ridership. A consequence of this would be less auto demand which would mitigate congestion and its externalities and would assist the efforts to improve sustainability in cities. The broader impact of this work is that it provides the field of transportation with a cost-effective tool that will allow for more efficient traffic operations and improved person mobility in urban environments. This work ultimately supports sustainable transportation systems that will improve quality of life.

6.3 Future Work

There are several ways that the performance of this person-based traffic responsive signal control system can be improved by for example incorporating the uncertainty of the input into the formulation of the mathematical program and developing more advanced prediction algorithms for the input used or accounting for platoon dispersion. More specifically, some areas for future related research are:

1. *Accounting for input inaccuracies in prediction algorithms and delay estimation.* In this dissertation, the mathematical program formulation is based on the assumption of accurate inputs. At the same time, predictions of vehicle arrivals have been estimated based on a nominal speed under the assumption of zero variability. There is a need to design improved prediction algorithms with the use of real-time data from new data sources such as the Connected Vehicle technology (RITA, 2012a) and adjust the delay estimation formulas to account for inaccuracies in the input. Adjusting the prediction algorithms and delay calculations would lead to a more robust traffic signal control system.
2. *Incorporating platoon dispersion in the signal control system.* This dissertation has developed the system based on the assumption of negligible platoon dispersion. This is valid for traffic operations close to saturation and short to medium signal spacings. In order for the system to be more robust, platoon dispersion needs to be accounted for in the input used in the optimization. This can be done with information that will become available with the Connected Vehicle technology. Such detailed real-time information about the vehicle trajectories can facilitate grouping of vehicles into smaller platoons that are homogeneous in their characteristics. In addition, minor changes to the formulation need to be made to account for the existence of multiple platoons of the same lane group.
3. *Determining intersection capacity in real-time.* This dissertation has assumed that a link's saturation flow (i.e., capacity) remains constant and unaffected by the interactions of multiple travel modes. It is essential to investigate how link capacities change in the presence of transit, freight, and in general multimodal operations (e.g., bus stops, freight deliveries, on-street parking, bicycle traffic). Models that predict how such multimodal operations affect capacity can be used to more accurately determine saturation flows for the different approaches and improve the performance of real-time signal control systems.
4. *Extension to grid networks.* The person-based traffic responsive signal control system that has been developed in this dissertation has been tested and evaluated only on isolated intersections and signalized arterials. However, the mathematical model formulation for the arterial level signal control system facilitates its extension to signalized grid networks. Cases of direct implementations include one-way paired arterials and special events. In the latter case, traffic leaves a central location in diverging directions so arterial signal optimization

can be performed in the directions of heaviest traffic. Finally, the system can be directly extended to optimize signal settings on two major crossing arterials. Application of this signal control system in networks is important because this is the context in which arterials operate in real cities.

Bibliography

2000. *Highway Capacity Manual, 3rd Edition*. TRB Special Report 209. National Research Council, Washington, DC.
- AC Transit. 2011. *Alameda-Contra Costa Transit District*. <http://www.actransit.org>.
- Ahn, K., & Rakha, H. 2006. System-wide impacts of green extension transit signal priority. *IEEE Intelligent Transportation Systems Conference*, 91–96.
- Al-Sahili, K.A., & Taylor, W.C. 1996. Evaluation of bus priority signal strategies in Ann Arbor, Michigan. *Transportation Research Record: Journal of the Transportation Research Board*, **1554**, 74–79.
- Baker, R.J., Collura, J., Dale, J.J., Head, L., Hemily, B., Ivanovic, M., Jarzab, J.T., McCormick, D., Obenberger, J., Smith, L., & Stoppenhagen, G.R. 2002. *An Overview of Transit Signal Priority*. Technical Report. ITS America.
- Balke, K.N., Dudek, C.L., & Urbanik II, T. 2000. Development and evaluation of intelligent bus priority concept. *Transportation Research Record: Journal of the Transportation Research Board*, **1727**, 12–19.
- Bretherton, D. 1996. Current developments in SCOOT: Version 3. *Transportation Research Record: Journal of the Transportation Research Board*, **1554**, 48–52.
- Bretherton, D., Bowen, G., & Wood, K. 2002. Effective urban traffic management and control: SCOOT Version 4.4. *European Transport Conference*.
- Busch, F., & Kruse, G. 2001. MOTION for SITRAFFIC – A modern approach to urban traffic control. *IEEE Intelligent Transportation Systems Conference*, 61–64.
- Chang, E., & Ziliaskopoulos, A. 2003. Data challenges in development of a regional assignment: Simulation model to evaluate transit signal priority in Chicago. *Transportation Research Record: Journal of the Transportation Research Board*, **1841**, 12–22.
- Conrad, M., Dion, F., & Yagar, S. 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**, 100–109.
- Cornwell, P.R., Luk, J.Y.K., & Negus, B.J. 1986. Tram priority in SCATS. *Traffic Engineering and Control*, **27**(11), 561–565.
- Diakaki, C., Dinopoulou, V., Aboudolas, K., Papageorgiou, M., Ben-Shabat, E., Seider, E., & Leibov, A. 2003. Extensions and new applications of the traffic-responsive urban control strategy: Coordinated signal control for urban networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1856**, 202–211.
- Dion, F., & Hellenga, B. 2002. A rule-based real-time traffic responsive signal control system with transit priority: Application to an isolated intersection. *Transportation Research Part B*, **36**(4), 325–343.
- Donati, F., Mauro, V., Roncolini, G., & Vallauri, M. 1984. A hierarchical decentralised traffic light control system. The First Realisation: Progetto Torino. *9th World Congress of the International Federation of Automatic Control*, **2**.
- Estrada, M., Trapote, C., Roca-Riu, M., & Robuste, F. 2009. Improving bus travel times with passive traffic signal coordination. *Transportation Research Record: Journal of the Transportation Research Board*, **2111**, 68–75.
- FHWA. 2009. *Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD)*. Technical Report. U.S. Department of Transportation, Federal Highway Administration.
- Floudas, C.A. 1995. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press, USA.
- Furth, P.G., & Muller, T.H.J. 2000. Conditional bus priority at signalized intersections: Better service with less traffic disruption. *Transportation Research Record: Journal of the Transportation Research Board*, **1731**, 23–30.
- Furth, P.G., Hemily, B., Muller, T.H.J., & Strathman, J.G. 2006. *Using Archived AVL-APC Data to Improve Transit Performance and Management*. TCRP Report 113. Transportation Research Board.
- Furth, P.G., Cesme, B., & Rima, T. 2010. Signal priority near major bus terminal: A case study of Ruggles Station, Boston, Massachusetts. *Transportation Research Record: Journal of the Transportation Research Board*, **2192**, 89–96.
- Gordon, R.L., & Tighe, W. 2005. *Traffic Control Systems Handbook*. Technical Report FHWA-HOP-06-006. U.S. Department of Transportation, Federal Highway Administration.

- He, Q., Head, K.L., & Ding, J. 2011. A heuristic algorithm for priority traffic signal control. *In: 90th Annual Meeting of the Transportation Research Board*. Transportation Research Board.
- He, Q., Head, K.L., & Ding, J. 2012. PAMSCOD: Platoon-based arterial multi-modal signal control with online data. *Transportation Research Part C: Emerging Technologies*, **20**(1), 164–184.
- Head, L. 1998. *Improved Traffic Signal Priority for Transit*. TCRP Project A-16. Interim Report. Transportation Research Board, National Research Council.
- Head, L., Gettman, D., & Wei, Z. 2006. Decision model for priority control of traffic signals. *Transportation Research Record: Journal of the Transportation Research Board*, **1978**, 169–177.
- Henry, J.J., & Farges, J.L. 1994. P.T. priority and Prodyn. *Proceedings of the 1st World Congress on Application of Transport Telematics and Intelligent Vehicle-Highway Systems*, **6**, 3086–3093.
- Hunt, P.B., Bretherton, R.D., Robertson, D.I., & Royal, M.C. 1982. SCOOT on-line traffic signal optimisation technique. *Traffic Engineering and Control*, **23**, 190–192.
- IBM. 2011. *IBM ILOG CPLEX, Version 12.1: High Performance Mathematical Programming Engine*. <http://www.ilog.com/products/cplex>.
- Janos, M., & Furth, P.G. 2002. Bus priority with highly interruptible traffic signal control: Simulation of San Juan’s Avenida Ponce de Leon. *Transportation Research Record: Journal of the Transportation Research Board*, **1811**, 157–165.
- Kim, W., & Rilett, L.R. 2005. Improved transit signal priority system for networks with nearside bus stops. *Transportation Research Record: Journal of the Transportation Research Board*, **1925**, 205–214.
- Klein, L.A., Mills, M.K., & Gibson, D.R.P. 2006. *Traffic Detector Handbook: Third Edition*. Technical Report FHWA-HRT-06-108, FHWA-HRT-06-139. U.S. Department of Transportation, Federal Highway Administration.
- Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., & Urbanik, T. 2008. *Traffic Signal Timing Manual*. Technical Report FHWA-HOP-08-024. U.S. Department of Transportation, Federal Highway Administration.
- Li, L., Lin, W.H., & Liu, H. 2005. Traffic signal priority/preemption control with colored petri nets. *IEEE Intelligent Transportation Systems Conference*, 694–699.
- Li, M. 2008. *Toward Deployment of Adaptive Transit Signal Priority Systems*. PATH Research Report UCB-ITS-PRR-2008-24. California Partners for Advanced Transit and Highways, University of California, Berkeley.

- Li, Y., Koonce, P., Li, M., Zhou, K., Li, Y., Beard, S., Zhang, W.B., Hegen, L., Hu, K., Skabardonis, A., *et al.* 2008. *Transit Signal Priority Research Tools*. PATH Research Report UCB-ITS-PRR-2008-4. California Partners for Advanced Transit and Highways, University of California, Berkeley.
- Lighthill, M.J., & Whitham, G.B. 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **229**(1178), 317–345.
- Ma, Q., Head, K.L., & Ding, J. 2011. A dynamic programming model for bus signal priority with multiple requests. *In: 90th Annual Meeting of the Transportation Research Board*. Transportation Research Board.
- Mauro, V., & Di Taranto, C. 1989. UTOPIA. *Proceedings of the 6th IFAC-IFIP-IFORS Symposium on Control, Computers, and Communications in Transportation*, 245–252.
- McTrans. 2003. *TRANSYT-7F User's Manual*. University of Florida.
- Menczer, W.B., Zatarain, K., Beal, D., Requa, J., McLemore, C., Costello, T., & Giorgis, J.D. 2006. *SMART: Opportunities for Improving Ridership*. Technical Report. Federal Transit Administration.
- Meyer, C.A., & Floudas, C.A. 2004. Trilinear monomials with mixed sign domains: Facets of the convex and concave envelopes. *Journal of Global Optimization*, **29**(2), 125–155.
- Nash, A. 2003. Implementing Zurich's transit priority program. *Transportation Research Record: Journal of the Transportation Research Board*, **1835**, 59–65.
- Newell, G.F. 1964. Synchronization of traffic lights for high flow. *Quarterly of Applied Mathematics*, **21**(4), 315–324.
- Newell, G.F. 1967. *Traffic Signal Synchronization for High Flows on a Two-Way Street*. Research Report. Institute of Transportation and Traffic Engineering, University of California.
- OASA. 2010. *Search Route*. www.oasa.gr.
- Richards, P.I. 1956. Shock waves on the highway. *Operations Research*, **4**(1), 42–51.
- RITA. 2012a. *Intelligent Transportation Systems Joint Program Office: Connected Vehicle Research*.
http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm.
- RITA. 2012b. *Intelligent Transportation Systems Joint Program Office: Knowledge Resources – Cost*. <http://www.itscosts.its.dot.gov/its/benecost.nsf/CostHome>.

- Sane, K., & Salonen, M. 2009. SYVARI – A new idea for using public transport priority at coordinated traffic signals. *Proceedings of the 16th ITS World Congress*.
- Schweiger, Carol L. 2003. *Real-Time Bus Arrival Information Systems: A Synthesis of Transit Practice*. TCRP Synthesis 48. Transportation Research Board, Washington, D.C.
- SFMTA. 2011. *Transit Effectiveness Project (TEP) Data*.
<http://www.sfmta.com/cms/rtep/tepdataindx.htm>.
- Skabardonis, A. 2000. Control strategies for transit priority. *Transportation Research Record: Journal of the Transportation Research Board*, **1727**, 20–26.
- Skabardonis, A. 2003. Benefits of advanced traffic signal systems. *In: Gillen, D. (ed), Measuring the Performance of ITS in Transportation Services*. Kluwer Academic Publishers.
- Skabardonis, A., Deakin, E., Harvey, G., & Stevens, A. 1990. *A Study of Arterial Operational Improvements*. Technical Report. Metropolitan Transportation Commission.
- Stevanovic, A., & Martin, P.T. 2007 (May). *Integration of SCOOT and SCATS in VISSIM Environment*. Presented at the PTV Users Group Meeting.
- The MathWorks. 2009. *Matlab User's Manual*.
- Transport Simulation Systems. 2010. *Aimsun Users Manual v6.1*.
- Wadjas, Y., & Furth, P.G. 2003. Transit signal priority along arterials using advanced detection. *Transportation Research Record: Journal of the Transportation Research Board*, **1856**, 220–230.
- Yagar, S., & Dion, F. 1996. Distributed approach to real-time control of complex signalized networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1554**, 1–8.
- Yagar, S., & Han, B. 1994. A procedure for real-time signal control that considers transit interference and priority. *Transportation Research Part B Methodological*, **28**, 315–315.
- Zhou, K. 2008. *Field Evaluation of San Pablo Corridor Transit Signal Priority (TSP) System*. PATH Working Paper UCB-ITS-PWP-2008-7. California Partners for Advanced Transit and Highways, University of California, Berkeley.
- Zlatkovic, M., Stevanovic, A., & Martin, P.T. 2012. Development and evaluation of an algorithm for resolving conflicting transit signal priority calls. *In: 91st Annual Meeting of the Transportation Research Board*. Transportation Research Board.

Appendix A

Glossary of Symbols

Chapter 3

a	=	auto vehicle index
A_T^r	=	total number of autos present at intersection r during cycle T
α	=	user-specified positive parameter which determines the weight of the schedule delay in the objective function $[\frac{1}{sec}]$
b	=	transit vehicle index
B_T^r	=	total number of transit vehicles present at intersection r during cycle T
C	=	cycle length [sec]
$d_a^r(g_{i,T}^r)$	=	function relating green times to the delay for auto a
$d_{a,T}^r$	=	delay for auto a for cycle T at intersection r [veh-sec]
$d_b^r(g_{i,T}^r)$	=	function relating green times to the delay for transit vehicle b
$d_{b,T}^r$	=	delay for transit vehicle b for cycle T at intersection r [veh-sec]
$\delta_{b,T}^r$	=	factor for determining the weight for schedule delay of transit vehicle b in cycle T at intersection r
$\Delta_{b,T}^r$	=	schedule delay of transit vehicle b arriving at intersection r during cycle T [sec]
$g_{i,T}^r$	=	green time allocated to phase i at intersection r during cycle T [sec]
$g_{i \max}^r$	=	maximum green time for phase i at intersection r [sec]
$g_{i \min}^r$	=	minimum green time for phase i at intersection r [sec]
i	=	phase index
I^r	=	total number of phases in a cycle for intersection r
L^r	=	intersection lost time [sec]
o_a	=	passenger occupancy of auto a $[\frac{pax}{veh}]$
$o_{b,T}^r$	=	passenger occupancy of transit vehicle b for cycle T at intersection r $[\frac{pax}{veh}]$
r	=	intersection index
T	=	cycle index

- θ = user-specified schedule delay threshold to define whether a transit vehicle should be considered late for priority purposes or not [sec]
- y_i^r = yellow time for phase i at intersection r [sec]

Chapter 4

- a = auto vehicle index
- A_T = total number of autos present at an intersection during cycle T
- α, β, γ = indeces for the time interval a transit vehicle could possibly arrive in the case of undersaturated traffic conditions.
- b = transit vehicle index
- B_T = total number of transit vehicles present at an intersection during cycle T
- C = cycle length [sec]
- $d_{a,T}$ = delay for auto a and cycle T at the intersection [veh-sec]
- $d_{b,T}$ = delay for transit vehicle b and cycle T at the intersection [veh-sec]
- $d'_{b,T}$ = delay calculation for transit vehicle b that determines whether the transit vehicle arrives in time interval α or β at the intersection [veh-sec]
- $\hat{d}_{b,T+1}$ = estimate of the delay transit vehicle b experiences in cycle $T + 1$ [veh-sec]
- D_j = total delay for vehicles in lane group j during a cycle [veh-sec]
- $D_{j,T}$ = total delay for vehicles in lane group j experienced during cycle T (defined as the interval from the end of green time in cycle $T - 1$ until the end of green time in cycle T) [veh-sec]
- $\hat{D}_{j,T+1}$ = delay estimate for vehicles in lane group j experienced during cycle $T + 1$ (defined as the interval from the end of green time in cycle T until the end of green time in cycle $T + 1$) [veh-sec]
- e = weighting factor for the most recent observation in exponential smoothing
- f = set of indeces that determine the time interval in which a transit vehicle arrives in the case of undersaturated traffic conditions
- $g_{i,b}^f$ = continuous variable for green times that is determined based on the time interval in which a transit vehicle could possibly arrive in the case of undersaturated traffic conditions ($f = \alpha, \beta, \text{ or } \gamma$) [sec]
- $g_{i \max}^r$ = maximum green time for phase i at an intersection [sec]
- $g_{i \min}^r$ = minimum green time for phase i at an intersection [sec]
- $g_{j \min}$ = minimum green time that needs to be allocated to lane group j to ensure undersaturated traffic conditions [sec]

$g_{i \text{ next}}$	=	user-specified value of the green time for phase i in cycle $T + 1$ [sec]
$g_{i,T}$	=	green time for phase i in cycle T at an intersection [sec]
G_j^e	=	summation of the effective green times for all phases that serve lane group j [sec]
i	=	phase index
I	=	total number of phases in a cycle
j	=	lane group index
J	=	total number of lane groups at an intersection
k_j	=	the first phase in a cycle that can serve lane group j
l_j	=	the last phase in a cycle that can serve lane group j
L	=	intersection lost time [sec]
λ_j	=	green ratio for lane group j
m	=	bus route index
M_1, M_2	=	big numbers
$N_{j,T}$	=	number of autos in the residual queue of lane group j at the end of the last phase that serves that lane group in cycle T
$n_{b,T}$	=	position of transit vehicle b in the queue after the end of the phases that serve its lane group j in cycle T
o_a	=	passenger occupancy of auto a $[\frac{\text{pax}}{\text{veh}}]$
\bar{o}_a	=	average passenger occupancy for auto a $[\frac{\text{pax}}{\text{veh}}]$
o_b	=	passenger occupancy of transit vehicle b $[\frac{\text{pax}}{\text{veh}}]$
$o_{b,T}$	=	passenger occupancy of transit vehicle b for cycle T $[\frac{\text{pax}}{\text{veh}}]$
\bar{o}_b	=	average passenger occupancy for transit vehicle b $[\frac{\text{pax}}{\text{veh}}]$
p_m	=	passenger demand for bus route m $[\frac{\text{pax}}{\text{sec}}]$
q_j	=	arrival rate of autos in lane group j $[\frac{\text{veh}}{\text{sec}}]$
$q_{j,T}$	=	arrival rate of autos in lane group j in cycle T $[\frac{\text{veh}}{\text{sec}}]$
$\hat{q}_{j,T}$	=	arrival rate estimate of autos in lane group j in cycle T $[\frac{\text{veh}}{\text{sec}}]$
r	=	intersection index
R_j	=	red time interval for lane group j
$R_j^{(1)}$	=	component of red interval for lane group j before the green that serves its lane group [sec]
$R_j^{(2)}$	=	component of red interval for lane group j after the green that serves its lane group [sec]
s_j	=	saturation flow for vehicles in lane group j $[\frac{\text{veh}}{\text{sec}}]$
t_b	=	arrival time of transit vehicle b at the back of its lane group's queue [sec]
$t_{b,m}$	=	actual arrival time of bus b that belongs to route m at the back of its lane group's queue [sec]
T	=	cycle index
$\tau_{j,T}$	=	end of the green phases in cycle T that serve lane group j [sec]
X_c	=	intersection degree of saturation
w_b^f	=	binary variable that determines the time interval in which a

		transit vehicle could possibly arrive in the case of undersaturated traffic conditions ($f = \alpha, \beta,$ or γ)
y_i	=	yellow time for phase i [sec]
Y	=	intersection flow ratio

Chapter 5

b	=	transit vehicle index
B_T^u	=	total number of transit vehicles present at intersection u during cycle T
c^{r+1}	=	coordinated phase for intersection $r + 1$
C	=	cycle length [sec]
$d_{b,T}^u$	=	delay for transit vehicle b and cycle T at intersection u [veh-sec]
$d_{j,T}^u$	=	delay for an auto in a platoon of lane group j at intersection u for cycle T [veh-sec]
$\delta_{b,T}^u$	=	factor for determining the weight for schedule delay of transit vehicle b in cycle T at intersection u
$D_{j,T}^u$	=	total delay for vehicles in a platoon that belongs to lane group j for cycle T arriving at intersection u from an incoming link caused by both stopping the head and the tail of the platoon [veh-sec]
$D_{j,T}^{(H)u}$	=	total delay for vehicles in a platoon that belongs to lane group j for cycle T at intersection u caused by stopping the head of the platoon [veh-sec]
$D_{j,T}^{(T)u}$	=	total delay for vehicles in a platoon that belongs to lane group j for cycle T at intersection u caused by stopping the tail of the platoon [veh-sec]
$D_{j,T}^{(Q)r}$	=	total delay for vehicles in a residual queue of lane group j for cycle T at intersection r [veh-sec]
$\eta_{b,T}$	=	binary variable for determining arrival case for transit vehicle b in cycle T at intersection v
$g_{i \max}^r$	=	maximum green time for phase i at intersection r [sec]
$g_{i \min}^r$	=	minimum green time for phase i at intersection r [sec]
$g_{i \text{ next}}^u$	=	user-specified value of the green time for phase i in cycle $T + 1$ at intersection u [sec]
$g_{i,T}^r$	=	green time for phase i in cycle T at intersection r [sec]
$g_{i,T}^{r+1}(1)$	=	green time for phase i in cycle T at intersection v as optimized from the first pair of intersections [sec]
$g_{i,T}^{r+1}(2)$	=	green time for phase i in cycle T at intersection v as optimized from the second pair of intersections [sec]
$g_{i,T}^r$	=	green time for phase i in cycle T at intersection r [sec]
$g_{i,T}^u$	=	green time for phase i in cycle T at intersection u [sec]

$G_j^{e,r}$	=	summation of the effective green times for all phases that serve lane group j at intersection r [sec]
$G_j^{e,u}$	=	summation of the effective green times for all phases that serve lane group j at intersection u [sec]
$h_{j,T}$	=	binary variable for determining arrival case for a platoon that belongs to lane group j in cycle T at intersection u
i	=	phase index
I^r	=	total number of phases in a cycle at intersection r
j	=	lane group index
J^u	=	total number of lane groups at intersection u
J_{IN}^u	=	set of lane groups for the incoming links at intersection u
J_{SH}^v	=	set of lane groups for the shared link at intersection v
l_j	=	the last phase in a cycle that can serve lane group j
μ	=	substituted variable for bilinearities
M	=	big number
$N_{j,T-1}^r$	=	the number of vehicles in the residual queue of lane group j of intersection r at the end of cycle $T - 1$
$N_{j,T-1}^u$	=	the number of vehicles in the residual queue of lane group j of intersection u at the end of cycle $T - 1$
\bar{o}_a	=	average passenger occupancy for auto a
$o_{b,T}^u$	=	passenger occupancy of transit vehicle b for cycle T at intersection u [$\frac{\text{pax}}{\text{veh}}$]
O_T^u	=	difference between the starting time of cycle T at intersection u and the critical intersection [sec]
$P_{j,T}^u$	=	size of platoon of lane group j in cycle T at intersection u [veh]
$\hat{P}_{j,T}^v$	=	estimate of platoon size that belongs to lane group j arriving at intersection v during cycle T
ψ_j^v	=	factor that indicates the portion of the incoming demand to intersection v that will be joining lane group j at that intersection
r	=	intersection index
ρ	=	substituted variable for trilinearities
$R_j^{(1)r}$	=	component of red interval for lane group j of intersection r before the green that serves its lane group at intersection r [sec]
$R_j^{(1)u}$	=	component of red interval for lane group j of intersection u before the green that serves its lane group at intersection u [sec]
s_j^r	=	saturation flow for lane group j at intersection r
s_j^u	=	saturation flow for lane group j at intersection u
$S_{j,u}^v$	=	saturation flow adjustment factor to account for changes in number of lanes along a link from intersection u to lane group j at v
t_T	=	starting time of cycle T at the critical intersection

t_T^u	=	starting time of cycle T at intersection u [sec]
$t_{j,T}^r$	=	arrival time of the first vehicle of a platoon that belongs to lane group j at the back of its lane group's queue at intersection r during cycle T [sec]
$t_{j,T}^u$	=	arrival time of the first vehicle of a platoon that belongs to lane group j at the back of its lane group's queue at intersection u during cycle T [sec]
$t_{b,T}^u$	=	arrival time of transit vehicle b at the back of its lane group's queue at intersection u during cycle T [sec]
$tt_{b,u}^v$	=	expected travel time for transit vehicle b to traverse the shared link between intersections u and v (includes lost time due to transit stops)
$tt_{j,u}^v$	=	average free flow travel time to traverse the shared link between intersections u and v
T	=	cycle index
u	=	index for the intersection that is the first that a platoon arrives at
v	=	index for the intersection that is the second that a platoon arrives at if it travels on the arterial
$x_{j,T}$	=	binary variable for determining arrival case for autos in the residual queue of lane group j
χ	=	binary variable participating in a bilinearity or trilinearity
ξ	=	binary variable participating in a bilinearity or trilinearity
y_i^r	=	yellow time for phase i at intersection r [sec]
$\zeta_{b,T}$	=	binary variable for determining arrival case for transit vehicle b at intersection u
$z_{j,T}$	=	binary variable for determining arrival case for a platoon that belongs to lane group j in cycle T at intersection v