STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

ADA Notice

For individuals with sensory disabilities, this document is available in alternate formats. For information call (916) 654-6410 or TDD (916) 654-3880 or write Records and Forms Management, 1120 N Street, MS-89, Sacramento, CA 95814.

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA13-2157-A		

4. TITLE AND SUBTITLE	5. REPORT DATE
Advanced Traffic Signal Control Algorithms	September 2013
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR	8. PERFORMING ORGANIZATION REPORT NO.
Dr. Alex Skabardonis	UCB-ITS-PRR-2013-XX-A

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
California PATH 1357 S.46th Street, Building. 452 Richmond, CA 94804-4648	11. CONTRACT OR GRANT NUMBER 65A0376 DTFH61-10-H-00002

12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Research, Innovation and System Information 1227 O Street Sacramento CA 95814	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This work was performed as part of the California PATH program of the University of California, in cooperation with the State of California, Transportation Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

16. ABSTRACT

The goal of this project is to develop advanced signal control strategies for mobility, safety, and environment based on connected vehicle data, i.e., real-time information on vehicles' location, speed and characteristics as well as communication to the signal control infrastructure. The final report for the project describes the findings from the literature review on existing adaptive signal control systems, their features and operational experiences from their implementation. Several performance measures are proposed for evaluating signal control algorithms, and procedures for estimating the performance measures from connected vehicle data were developed using statistical techniques and kinematic wave theory. This project developed and tested through simulation a number of control strategies to improve mobility: queue spillback avoidance, control for congested networks, and dynamic lane allocation. The results indicated that the proposed strategies improve the traffic performance. This project also developed strategies for improving intersection safety with emphasis on avoidance of red light running (RLR) related collisions. A prototype in-vehicle driver speed advisory system for minimum fuel consumption and emissions was also developed. Field tests show that the prototype system achieved significant fuel savings.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Traffic signals, adaptive control, connected vehicles, mathematical models	No Restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	125	

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in Braille, large print, audiocassette, or compact disk. To obtain a copy of this document in one of these alternate formats, please contact: the Division of Research and Innovation, MS-83, California Department of Transportation, P.O. Box 942873, Sacramento, CA 94273-0001.

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Advanced Traffic Signal Control Algorithms

Alexander Skabardonis
Steven Shladover
Wei-Bin Zhang
Liping Zhang
Jing-Quan Li
Kun Zhou
California PATH, UC Berkeley

Juan Argote
Eleni Christofa
Yiguang Xuan
Transportation Engineering, UC Berkeley

Matt Barth
Kanok Boriboonsomsin
Haitao Xia
CE-CERT, UC Riverside

Andreas Winckler
Darren Liccardo
BMW Group Technology Office

California PATH Research Report
UCB-ITS-PRR-2013-XX

This work was performed as part of the California PATH program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Final Report for Interagency Agreement #65A0376
FHWA Exploratory Advanced Research Program
Cooperative Agreement DTFH61-10-H-00002

September 2013

ISNN xxxx-xxx

ABSTRACT

The goal of this project is to develop advanced signal control strategies for mobility, safety, and environment based on connected vehicle data, i.e., real-time information on vehicles' location, speed and characteristics as well as communication to the signal control infrastructure. This document is the final report for the project. It describes the findings from the literature review on existing adaptive signal control systems, their features and operational experiences from their implementation. Several performance measures are proposed for evaluating signal control algorithms, and procedures for estimating the performance measures from connected vehicle data were developed using statistical techniques and kinematic wave theory. We developed and tested through simulation a number of control strategies to improve mobility: queue spillback avoidance, control for congested networks, and dynamic lane allocation. The results indicate that the proposed strategies improve the traffic performance. We also developed strategies for improving intersection safety with emphasis on avoidance of red light running (RLR) related collisions. A prototype in-vehicle driver speed advisory system for minimum fuel consumption and emissions was also developed. Field tests show that the prototype system achieved significant fuel savings.

Keywords: *Traffic signals, adaptive control, connected vehicles, mathematical models*

ACKNOWLEDGEMENTS

This work was performed by the California PATH Program at the University of California at Berkeley, in cooperation with the California Department of Transportation (Caltrans), Division of Research and Innovation (DRI) under Cooperative Agreement DTFH61-10-H-00002 with the Federal Highway Administration (FHWA) of the US Department of Transportation, under the Exploratory Advanced Research (EAR) program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein.

The authors thank Peter Huang, Joe Bared and Dr. David Yang of FHWA R&D for their support and advice during the project. We also thank Asfand Siddiqui, Pete Hansra and Greg Larson of Caltrans DRISI for their support, and Kai Leung of Caltrans Traffic Operations Headquarters for modifications of the 170E signal controller software. We thank the project technical expert panel members Professor Darcy Bullock of Purdue University, Edward Fok of the FHWA Resource Center, San Francisco, Professor Larry Head of University of Arizona, Peter Koonce of the city of Portland, Oregon, and Professor Michael Zhang of UC Davis for their comments and suggestions. Professors Carlos Daganzo and Pravin Varaiya of UC Berkeley provided comments and suggestions throughout the study.

EXECUTIVE SUMMARY

The goal of this project is to develop advanced signal control strategies based on connected vehicle (CV) data, i.e., real-time information on vehicles' location, speed and characteristics as well as communication to the signal control infrastructure. A comprehensive literature review was performed with emphasis on existing adaptive traffic signal control systems and CV data and traffic control. We developed algorithms for the estimation of performance measures (MOEs) based on CV data, and developed and tested several control concepts to improve mobility and safety. A prototype system was also developed and field tested to provide real-time speed advisories to drivers to minimize fuel consumption and emissions.

We developed procedures for the estimation of MOEs from CV data based on statistical techniques and kinematic wave theory, and investigated alternative sampling strategies for CV data. We also developed a data fusion method to estimate queue lengths from loop and CV data, based on a weighted combination of the queue length from the two data sources estimated separately using Kalman filtering techniques.

- Testing of the proposed estimation methods show that for undersaturated conditions penetration rates of 50% are needed for accurate estimation of average travel time, delay, and stops, and about 80% of the vehicles need to be equipped to obtain accurate estimates of queue lengths. However the required penetration rates are only 10% for oversaturated conditions. This is promising because improved control strategies can be designed for oversaturated conditions even at the early stages of CV implementation.

- The comparison of estimates of MOEs using the sampled data according to several sampling strategies (i.e., uniform sampling, speed based sampling, acceleration based sampling, J2735 sampling) with the baseline truth values show that there is not a clear "break point" with respect to market penetration where the probe sampling results become significantly better. None of the analyzed probe sampling strategies showed a clear superiority to the others. The variability in the sampled data can be reduced by aggregating data from multiple signal cycles, but this introduces additional latency in the response of the traffic control system. This represents an additional trade-off in considering how to use CV data to implement real-time adaptive signal control.

We developed and tested control strategies to improve mobility based on CV data including queue spillback avoidance, control of congested grid networks and dynamic lane allocation.

- The queue spillback avoidance strategy consists of a queue spillback detection method and a signal control action to prevent spillback occurrence in the next signal cycle. The results from testing the proposed strategy through simulation on a real-world arterial, San Pablo Avenue in Berkeley, indicate correct detection of queue spillbacks in more than 80% of the signal cycles for a range of penetration rates. The proposed signal control prevented the occurrence of spillbacks by maintaining shorter queue than the critical link length and by redistributing the delays at further upstream links that have more storage space for residual queues. More importantly, the proposed strategy reduced the variation of the maximum queue length on the critical link.

- Control strategies for congestion avoidance in grid networks were developed and tested in a real-life network Post-Oak in Houston, Texas. The first strategy consisted of a perimeter control approach, were vehicular entrances on the perimeter of the model were limited by reducing the green times on the network entry links. The second strategy consisted of signal phase changes to the upstream and downstream intersections in a network link that spillbacks are detected. The final strategy consisted of reducing the system cycle length and providing additional green time to those network links that spillback first detected. The third strategy produced the highest benefits. In general, the level of improvements in MOEs that can be obtained from the proposed strategies depend on the characteristics of the particular network under study.

- We developed a dynamic lane grouping (DLG) control strategy that changes the lane utilization at the signalized intersection approach to accommodate spatial variations in traffic demands based on real-time origin-destination information from CV data. The strategy was evaluated through numerical analysis and microscopic simulation for several scenarios consisting of number of lanes, traffic levels, and signal settings. The results show that the proposed strategy can achieve significantly better performance in terms of maximum flow ratio and average delay. For example, for the three-lane approaches and spatial variation of demand of 40%, DLG reduces the maximum flow ratio by 41% and the average delay by 35% compared to fixed lane allocation.

We developed and tested control strategies to improve safety using CV data focusing on how to minimize the occurrence of red-light-running (RLR), a major safety issue at signalized intersections.

- Dynamic all-red extension (DARE) inserts an all-red interval only when there is a high probability of RLR related collision. We developed and calibrated a prediction algorithm for DARE using detailed trajectory data collected with micro-radar sensors at a real world intersection in the Bay Area. The evaluation of the proposed algorithm on the San Pablo Ave test site with continuous detection (which is similar to CV data), shows that the proposed algorithm achieved a correct detection/activation rate over 95% at a false alarm rate of less than 5%.
- A significant factor in RLR frequency is the arrival flow during the yellow change interval. We developed control strategies based on segmentation of traffic platoons based on CV trajectory data. Three control strategies including green extension, phase termination, and double cycle, were proposed for platoon-based control. The evaluation results indicate that the proposed strategies are effective in reducing the flow during the yellow interval in coordinated arterials.

A prototype system was developed to provide in-vehicle speed advisories to drivers for minimum fuel consumption. The prototype system consists of a research vehicle, a cloud server and wireless communication devices to traffic signals. An algorithm provides drivers the optimal speed profile based on inputs from traffic signal's status and vehicle's current state. Also, a control strategy for signal priority to individual vehicles (APIV) was tested alone as well as in combination with the speed advisory system. APIV adjusts the signal settings to provide additional green time to favor individual vehicles approaching a signalized intersection. The system was field tested in a controlled environment, a fixed-time traffic signal at the Richmond Field Station campus of UC Berkeley, and in real-world arterial El Camino Real with three coordinated actuated signals.

- The results from the Richmond Field Station field test show that the speed advisory system achieved fuel savings of 13.6% compared to the uninformed driving scenario. For the APIV approach without the speed advisory fuel savings of 19.1% were achieved. The combination of both approaches showed fuel savings of 28.4%. Further analysis indicates that fuel savings for the informed driving scenario are mainly from the drivers' early slowing down and cruising through the intersection without having to come to a complete stop. The fuel savings due to APIV are mainly contributed by an increased proportion of driving through the green interval.
- A new speed advisory algorithm was developed for actuated coordinated traffic signals and the presence of other vehicles. Due to limited time and resources, it was not possible to do any meaningful measurements on the El Camino Real field tests. Nevertheless important observations were made during the test runs performed. It was observed that is extremely difficult to follow recommended speeds that are continuously changing over time, while driving with other vehicles. This indicates that besides the speed advisory algorithm itself, the method of displaying the speed recommendation is crucial for the success of similar systems in the future. The algorithms will be revised based on the lessons learned and field tested in a follow-up research effort.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Problem Statement

Highway congestion is in large part caused by the inefficiency with which the roadway network capacity is managed. Traffic signals are the mechanism for managing arterial network capacity, yet the control of traffic signals has not significantly changed over the past several decades, despite rapid and profound changes in electronics, sensors and communication technologies, and software. The main impediment to improvements in traffic signal control systems has been the limited ability of available surveillance systems (fixed-point detectors) to measure the true state of the traffic network and its response to signal control changes. Vehicle to infrastructure (V2I) and vehicle-to-vehicle (V2V) cooperation could provide this comprehensive real-time information on the movements and interactions of vehicles in the entire road network and allow a transformational change on traffic control methods.

Although a majority of the large cities in the United States employ pre-timed traffic control, it is estimated that almost 80 percent of all traffic signals in the nation use traffic detectors, mainly for local intersection control (traffic actuation) and a small number of traffic responsive systems. Regardless of the principle of operation, these systems essentially rely on "point" detection and therefore provide only binary information on the presence or absence of vehicles. The traffic signal control algorithms based on this type of vehicle information were developed in the late 1970s and are, with minor modifications, still in operation today.

There is a significant gap between the theoretical advances in traffic control methodologies plus the developments in hardware and software, and the real-world practice in traffic control:

- Existing models inadequately describe traffic operations at intersections, arterials and networks, especially at near-capacity or oversaturated conditions;
- Computational complexity for signal optimization is too high for practical application on typical size signal controlled street networks
- Control actions in real-time have until now are based on flow and occupancy observations at fixed and limited locations which may not be appropriate for the entire range of traffic conditions

1.2 Project Goal and Objectives

The goal of this project is to develop advanced signal control strategies based on the impending widespread availability of probe data from connected vehicles, leading to a transformational change in how we control traffic along urban arterials and networks.

Vision of real-time information to transform traffic signal operations: The advent and progressive deployment of vehicle-infrastructure cooperation through connected vehicles offers the potential for significant improvements is in the efficiency of traffic signal control. Ideally, knowledge of the location, speed, routing and destination of every vehicle in the network in real-time allows for optimal signal settings and significant improvements in traffic performance compared to existing control strategies. However, traditional infrastructure-based traffic detectors only provide point specific information on the presence of vehicles. Ultimately, with full market penetration of traffic data probes, it will be possible to know the complete current state of the traffic network and to predict how it will evolve. With this knowledge, meaningful measures of performance can be estimated accurately for the network, and signal control strategies can be defined and adjusted to optimize those measures. Examples include:

- Knowledge of the vehicle trajectories provides improved prediction of traffic arrivals which permits development of control strategies for i) delay minimization at individual intersections, and ii) coordinated signal operation along arterials or networks to minimize travel time and/or number of stops for the entire traffic stream or for specific origin-destination pairs;
- Knowledge of the vehicle trajectories permits the development and implementation of optimized signal settings that explicitly minimize fuel consumption and air pollutant emissions;
- Knowledge of trajectories of vehicles approaching the signalized intersection approach permits implementation of safety related control measures to reduce red light running, such as adaptive all-red-extension, and offset adjustments to reduce vehicle arrivals during the yellow clearance interval without compromising the intersection traffic performance.

Vision of real-time information to transform the driving experience: In conjunction with the new traffic control strategies, Connected Vehicle (CV) systems can also bring traffic signal status and alert/advisory information into the vehicles, allowing drivers and the in-vehicle systems themselves to better cooperate with the traffic signal control systems. The benefits of this improved cooperation could include more efficient and safe operation at the intersections and traffic signal based alerts to encourage driving habits for less fuel consumption. At this level, efficient traffic management is not only the responsibility of the roadside infrastructure, but it can be shared with the vehicles, so that they can work together as a well-integrated system.

The project has the following objectives:

- Develop comprehensive set of performance measures for evaluating probe based signal control systems
- Develop procedures for estimating the selected performance measures based on CV data
- Develop and evaluate new control strategies enabled by the CV data

1.3 Organization of the Report

This document is the final report for the project. Chapter 2 summarizes the findings from the literature review on existing adaptive signal control systems, their features and operational experiences from their implementation. Chapter 3 presents the selected performance measures for development and evaluation of signal control algorithms, and the methodology and findings of estimating the selected performance measures from CV data. Chapters 4 through 6 describe the formulation and testing of control strategies to improve mobility: queue spillback avoidance, control for congested networks, and dynamic lane allocation. Chapters 7 and 8 describe the development and testing of strategies for improving safety: dynamic all red extension, and minimization of vehicle arrivals during the yellow change interval. Chapter 9 describes the development and testing of a driver speed advisory system to minimize fuel consumption and emissions. The last Chapter summarizes the study findings and discusses future research directions.

Appendix A includes the final report prepared by BMW with detailed description of the prototype system for providing in-vehicle speed advisories. Appendix B includes the assembled bibliography. Appendix C includes the publications produced based on the research effort.

CHAPTER 2

LITERATURE REVIEW

This Chapter describes the findings from the review of the state-of-the-art and state-of practice of traffic signal control systems and strategies. The review of the literature placed emphasis on algorithms, systems and strategies developed over the last ten years, with particular attention to the application of probe vehicle data for traffic control on urban street networks.

2.1 Sources of Information

The literature search was primarily based on the following ongoing and recently completed major research studies:

- PATH TO 6322 "Measure and Field Test the Effectiveness of Adaptive Signal Control" [1]
- NCHRP Synthesis Project 20-05 "Adaptive Control Systems" [2]
- NCHRP 3-90 "Control Strategies for Oversaturated Conditions," [3]
- NCHRP 3-97 "Traffic Signal Control Strategies for Varying Demands and Capacities" [4]
- NCHRP Project 3-66: "Traffic Signal State Transition Logic Using Enhanced Sensor Information" [5]

The literature search was supplemented by online searches at the Transportation Research Board (TRB) TRIS database, and the Institute of Transportation Studies, University of California, Berkeley catalog. We also reviewed the documents and presentations posted at the TRB Signal Systems Committee website, particularly those relate to adaptive signal control:

http://www.signalsystems.org.vt.edu/documents.html

The materials posted provide detailed information on the characteristics of several adaptive traffic signal systems including principles of the strategy, system architecture, data requirements, communications requirements, local controller and central hardware requirements, plus installation, operation and maintenance costs. Additional references obtained and reviewed were from the recent IEEE ITSC Conference (September 2010), and papers presented at the 90th TRB Annual Meeting (January 2011). In addition, we requested from the FHWA Contract Manager unpublished documents on ongoing FHWA research studies, and he provided to us the documents available, namely documentation on the ongoing Pooled Fund study at the University of Virginia.

Over 300 references were identified. The reference list is included in Appendix B. The reference list is organized by a) general references, b) adaptive traffic signal control strategies (ACS-Lite, BALANCE, MOTION, LADOT ATCS, OPAC, PRODYN, RHODES, SCATS, SCOOT, TUC, UTOPIA, and Other Online Strategies), c) methods and algorithms for the optimization of signal settings, d) performance measurement, and e) Connected Vehicle data and traffic control.

Connected Vehicle Traffic Signal Information Exchange Meeting

We attended a two-day Connected Vehicle traffic signal information exchange meeting at Arizona State University in Tempe Arizona on December 14 and 15, 2010 organized by Edward Fok of the FHWA Resource Center, San Francisco, and Greg Krueger of USDOT. "The objective of the meeting was to use the input from the meeting to build a three to five year vehicle-to-infrastructure (V2I) technologies traffic

signal research program plan". Meeting participants received copies of the progress notes for this project provided by Edward Fok of FHWA Resource Center with permission from the project AOTR.

The meeting was very well attended and all participants were actively involved in discussions on how connected vehicles can make a significant difference in traffic signal operations and control. Detailed meeting minutes [6] were provided by the meeting organizers. The following is a short summary of the discussion topics:

- Connected Vehicle data specific Issues
  Sampling standard J2735
  Location of the RSE/Antenna
  Vehicle persistence test
  # of RSEs in a corridor
  Processing power of signal controller (# vehicles broadcasting)
  Controller cabinet physical requirements
  Alternative data sources (blue tooth)
- Performance measurement
  System delay
  Rate of delay
  Travel time
  Queue length
  # vehicles in dilemma zone
- Traffic signal operations and control strategies
  Real time control
  Incident detection
  Queue management
  Traffic signal priority
  Platoon flow
  New optimization models that utilize O-D information
  Work zone control

2.2 Findings

Traffic signals along arterials and networks operate as coordinated to provide progression to the major through movements. Different timing plans (cycle length, green times and offsets) are implemented to account for the variability of traffic throughout the day. Most of the existing signal systems use fixed time timing plans prepared off-line based on historical data (often called "first generation strategies"). These plans are typically implemented by time of day (TOD), e.g., am, midday and pm peak periods. This is illustrated in Figure 2.1A. Fixed-time plans, however, cannot deal with the variability of traffic patterns throughout the day, and they become outdated because of the traffic growth and changes in traffic patterns.

An increasing number of first generation control systems use traffic actuated controllers operating in coordination with a common background cycle length. These systems provide improved through progression by utilizing the spare green time in the signal cycle from the "early" termination of actuated phases [7]**.** At the same time, they may reduce the total intersection delay by responding to the cycle-by-cycle fluctuations in traffic volumes. Simulation results and field studies have shown that coordinated actuated signals significantly improved the performance on the arterial through traffic at the expense of the cross-streets.

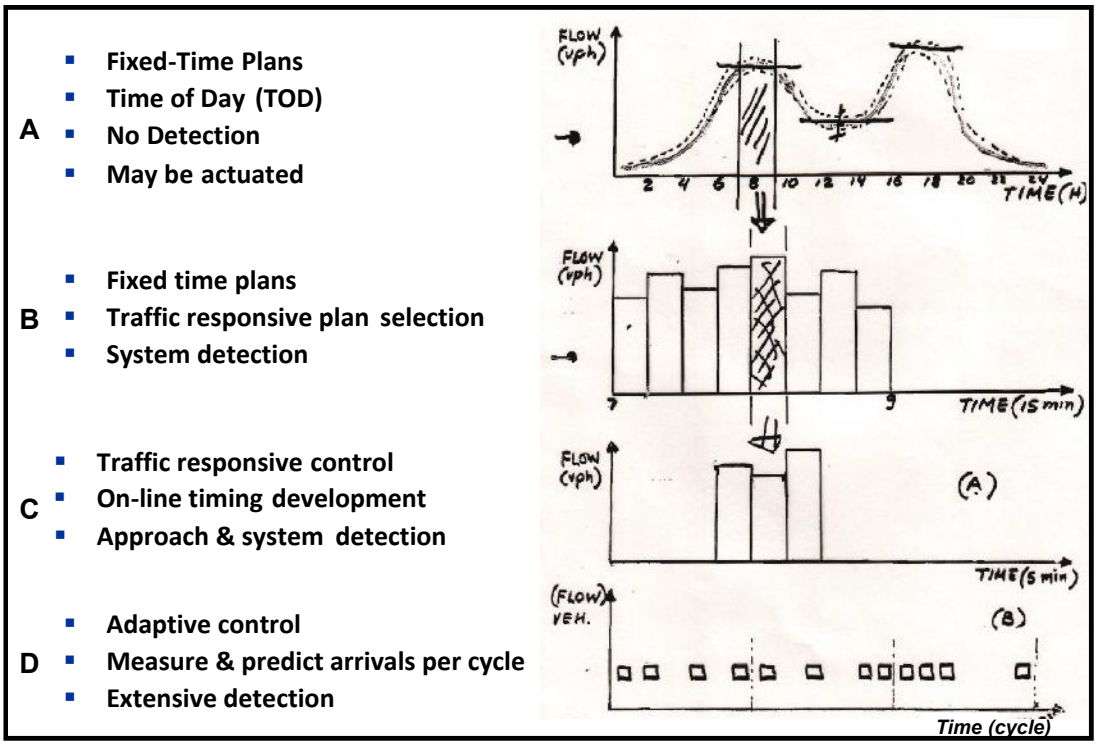


Figure 2.1 Traffic Signal Control vs. Variability of Traffic Flow

A number of signal systems employ several fixed-time timing plans to account for the variability of traffic flow. These timing plans are typically selected based on volume and occupancy data collected from system detectors located in key locations of the network (Figure 2.1B). The system operator may also override the timings based on real-time surveillance data.

"On-line" control systems update the timing plans in real-time based on data from detectors located on each intersection approach. Such strategies fall into two major categories: i) traffic responsive or timing plan update (e.g., "UTCS Second Generation", SCATS and SCOOT) that adjust the signal settings while maintaining a common cycle length (Figure 2.1C), and ii) adaptive control policies (OPAC, RHODES, PRODYN) that continually optimize the timings at each intersection over a short time interval (rolling horizon) as shown in Figure 2.1D. Note that often all on-line systems are called adaptive.

2.2.1 Existing Adaptive Traffic Signal Control Systems

The characteristics of the existing adaptive signal control systems that have been implemented in the field are summarized in Table 2.1.

SCOOT and SCATS are the most widely deployed systems. The SCATS (Sydney Coordinated Adaptive Traffic System) control system uses detector data at the intersection stopline to measure the degree of saturation (volume/capacity ratio), as function of the utilization of the phase green time. It then adjusts on-line the background fixed-time plans. The Splits-Cycle-Offsets-Optimization-Technique (SCOOT) method originally developed in England uses data from detectors located at the upstream end of each approach to estimate the size and shape of traffic platoons for each signal cycle, and then adjusts the timings such as the platoon to clear the intersection during the green time thus minimizing vehicle delays and stops.

Table 2.1 Existing Adaptive Signal Control Systems

System	Installa-tions	Architecture	Detection	Controller	Communications
SCOOT	Over 200 worldwide	Centralized	Exit loops	NEMA (EPAC) or special	Once per second for hold, force-off omit and detector data
SCATS	Over 50 worldwide	Hierarchical (plan based)	Stop bar loops	2070 or special	Strategic control from central and local tactical control
OPAC	2	Decentralized	Exit loops	NEMA (with VS-PLUS firmware) and VME co-processor)	Once per cycle
RHODES	4	Decentralized	Fully actuated design	2070 (with NextPhase firmware and VME co-processor)	Peer-to-peer over IP, event based on upstream detections
BALANCE /MOTION* (Germany)	5	Central	Near Stop Bar	European	Once per second
INSYNC		Decentralized	Near Stop Bar	Existing(Insync Software)	Ethernet
ACS Lite	4	Decentralized	Stop bar loops Upstream	NEMA 2070	Serial or Ethernet
ATCS (Los Angeles)	1	Centralized	Fully actuated with system detectors for VOS	2070 with LADOT firmware	Once per second
TUC *(Chania, Greece)*	5	Central	System loops for VOS	European	Once per cycle
UTOPIA *(Torino, Italy)*	1	Distributed	Fully actuated design	European	

** BALANCE and MOTION are different systems with many similarities*

The SCOOT and SCATS systems have been field evaluated in a number of studies. Most of the field evaluation studies use "before" and "after" measurements of travel times, intersection delays, and the number of stopped vehicles to determine the improvements in the performance measures following implementation of the new control system. Evaluation of SCOOT deployments has reported an average reduction of 12% in intersection delay over fixed time plans. Evaluation of SCATS deployment in Oakland County Michigan reported between 7 to 32% improvements in travel times during different times of the day. In the majority of these evaluations the existing baseline fixed-tine plans were not recently optimized.

The TUC control strategy is using detector occupancy measurements at midblock locations to balance and minimize the queues at each traffic signal, taking into account link storage capacity. This formulation is potentially particularly useful in dealing with oversaturated conditions. Recent field tests in Southampton, UK showed that TUC produced similar performance with the SCOOT strategy.

Adaptive systems continually optimize the signal settings over a short time interval ranging from 15 to 30 seconds (rolling horizon), without necessarily maintaining a common cycle length in the network. Most of these approaches evolved from experimental control of isolated intersections (MOVA, OPAC, RHODES and PRODYN). The Federal Highway Administration (FHWA) sponsored the development of

three adaptive control system (ACS) prototypes - OPAC, RHODES, and RTACL that were field tested in Reston, Virginia; Seattle, Washington, and Chicago, Ill, respectively. Issues related to installation and operation of detection systems, communications, and controllers were significant challenges in each of the field tests. RTACL performed below expectations. OPAC did not improve network conditions and in some cases delay and travel time actually increased. Simulation tests of OPAC using the CORSIM simulation model were consistent with the field test results. RHODES showed no significant difference in arterial travel times, but the cycle times were significantly reduced in Seattle. In a recent simulation study RHODES outperformed the optimal fixed-time plans along a major arterial in Southern California (Figure 2.2). The UTOPIA system in Turin provides absolute priority to public transit vehicles and at the same time optimizes the signal settings for the rest of the traffic stream. Reported benefits include a 20 percent increase in the average bus speeds without disbenefits to the auto traffic.

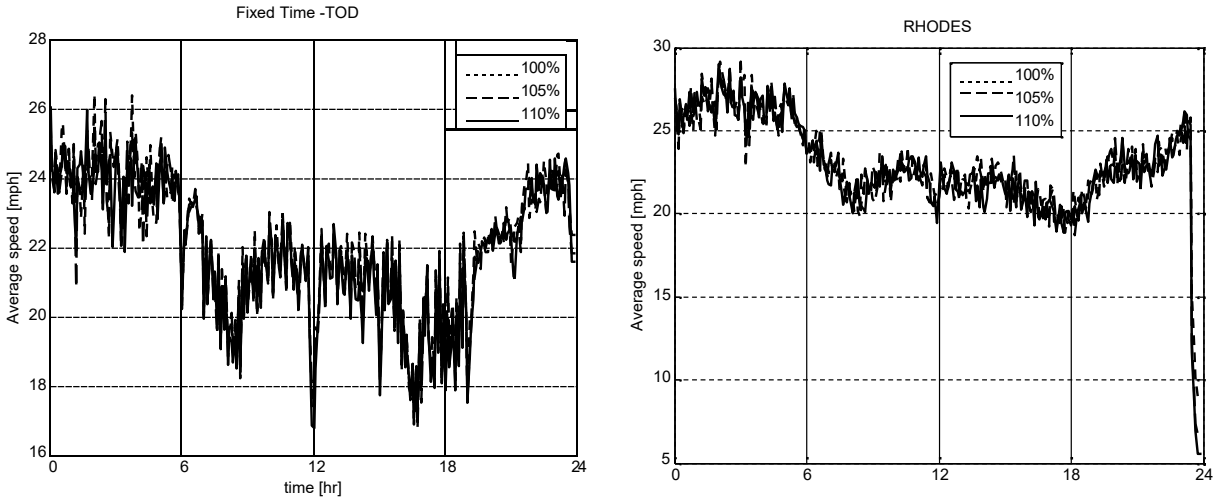


Figure 2.2 Fixed-Time vs. Adaptive Control (RHODES) –Lomita Avenue, CA

Major impediments for the deployment of adaptive signal control systems are the complexity of the systems' logic (underlying algorithms and software implementation), high installation, operation and maintenance costs and limited evidence on the benefits. Most of the cited benefits are based on limited data and do not relate to the geometric, traffic and control characteristics of the specific project areas. More important there is no evidence that adaptive systems are effective in dealing with oversaturated conditions. It has been observed that adaptive systems delay the onset of congestion, but there is no evidence on their effectiveness during congestion. Also, the high costs for detectors and communications (about $65,000 per traffic signal) is a major factor for the limited deployment of adaptive systems. Infrastructure maintenance is also a major issue, especially the hardware related to the specific adaptive system in place. Detection placement and malfunctions often are critical factors for the effectiveness of the chosen traffic control strategy.

Recently FHWA in cooperation with signal controller manufactures and Universities developed the Adaptive Control Software *Lite* (ACS Lite) strategy with the goal of providing a "widely deployable" system that automates monitoring of traffic signal performance and adjustment of signal timing. The *Lite* designation reflects a focus on reducing traditionally high installation and operations costs, which have been the primary impediment limiting the deployment of adaptive systems in the U.S. ACS Lite is designed for the widely deployed closed-loop arterial systems, essentially using system detector data to fine-tune the background fixed-time plans. Field evaluations in four sites showed that the signal timing adjustments by ACS Lite provided substantial reductions in vehicle delay, arterial travel time, vehicle stops, and fuel consumption.

2.2.2 Connected Vehicle Data and Traffic Control

The use of connected vehicle probe data in the development of signal control strategies has been limited. Most of the work has been i) on the potential and limitations of probe data to describe operating conditions in a road network and its relationship to signal control [8], and ii) the potential of using probe data for providing real-time information to drivers, e.g., pilot implementation of speed advisories on head-up displays in Germany [9]. Ongoing work at the University of Virginia as part of the FHWA Pooled Fund study has proposed the use of probe data for better queue management and clustering of vehicle platoons [10]. Also, in a California PATH study aiming to develop a dynamic all-red extension strategy to reduce red-light-running (RLR) collisions, it was found that using vehicle trajectory data approaching the intersection can improve the prediction of RLR occurrences by up to 40%, compared with conventional point detection from loops [11].

A number of studies are concerned with the estimation of performance measures at traffic signals from probe data. It was found that probe data could provide satisfactory estimates of arterial travel times and midblock flow rates even at low market penetration, but queue length estimation errors, however, were high even at 100% market penetration [12]. As part of the VII California project, the use of probe vehicle data to represent arterial traffic was investigated based on a VISSIM micro-simulation of the El Camino Real corridor with 20 signalized intersections in the San Francisco Bay Area [13]. The simulated vehicle trajectories at 1 second resolution were treated as the ground truth for this corridor and were then sampled by a post-processor program that implements the probe vehicle sampling rules prescribed in the draft SAE J2735 standard. Different sampling strategies were tested by adjusting parameters in the post-processor to determine how the fidelity of the probe data would be affected. This research showed that the current probe vehicle sampling rules impose serious constraints on the usefulness of the data for real-time adaptive signal control.

CHAPTER 3

ESTIMATION OF TRAFFIC PERFORMANCE MEASURES WITH CONNECTED VEHICLE DATA

3.1 Introduction

The purpose of this Chapter is to present the proposed measures of effectiveness (MOEs) for the development and evaluation of connected vehicle based traffic signal control strategies.

Several measures of performance have been proposed and used for evaluating traffic operations at highway facilities controlled by traffic signals. Table 3.1 shows typical performance measures and their usage in practice.

Table 3.1 MOEs and Their Usage in Practice [14]

MOE	Typical Usage
Travel Time	Used in long range planning studies at regional or corridor level to evaluate traveler benefits of alternative improvements. Used to evaluate traveler benefits of signal timing improvements for individual facility.
Speed	Used to evaluate alternatives in long range planning studies at regional or corridor level. Used to evaluate benefits of signal timing improvements for individual facility. Used to estimate fuel consumption and air quality impacts.
Delay	Used to evaluate alternatives in long range planning studies at regional or corridor level. Used to evaluate benefits of signal timing improvements for individual intersection or facility. Used to determine the LOS at signalized intersections per HCM Used to estimate fuel consumption and air quality impacts.
Queue	Used to identify hot spots, operations problems at points of facility (left turn bays, blockages, safety)
Stops	Used to evaluate quality of signal timing plans along arterials/networks Used to determine estimates of fuel and air pollution emissions
Travel Time Variance	Used to evaluate benefits of traffic operations improvements that reduce variability but not mean travel time or delay.

In addition the following indicators of performance are commonly used to assess the quality of traffic operations at highway facilities:

- Level of Service (LOS) per the Highway Capacity Manual (HCM) [15]
- Volume/Capacity (v/c)

Level of service (LOS) is a quality measure describing operational conditions within a traffic stream, generally in terms of such service measures as speed (travel time), freedom to maneuver, traffic interruptions, and comfort & convenience. Six LOS are defined for each type of facility that has analysis procedures available; LOS A represents the best operating conditions, and LOS F the worst. The average delay (sec/veh) is used to characterize the LOS at signalized intersections.

Volume/capacity ratios greater than 1.00 indicate that traffic demand is higher than capacity which results in oversaturated conditions with long delays sand excessive queue lengths.

3.2 Proposed Performance Measures

The proposed MOEs for the development and evaluation of traffic signal control algorithms are shown in Table 3.2. These performance measures cover a wide range of operating environments (isolated intersections, arterials, grid networks) traffic patterns (undersaturated vs. oversaturated conditions) and

objectives/constraints (mobility, safety, environment). The shaded entries are the MOEs that can be directly computed from Connected Vehicle data.

Table 3.2 Proposed Performance Measures for Signal Systems

I. MOBILITY

Operating Environment	Operating Conditions	Performance Measure	Units	Comments
Intersection	Undersaturated	Average control delay	sec/veh	Difference free-flow travel time and actual travel time
		Max back of queue	#veh (ft/l)	Average and 95th % of the max extend of queue throughout the cycle
		Cycle failure	%	Proportion of cycles that queue failed to clear during green
		Green time utilization	%	Proportion of green utilized by traffic demand served by the phase
	Oversaturated	Throughput	#	# vehicles served at the intersection per time interval
Arterial/ Grid Network	Undersaturated	Average travel time	(min)	Average travel time for movements served by coordinated signal phases
		Average travel speed	(mph)	Average travel speed for movements served by coordinated signal phases
		Travel time variability	(min)	st deviation, 80th or 95th percentile of travel times served by coordinated phases
		# of stops/stop rate	#/(%)	Average # of stops (fraction of veh stopped) for movements served by coordinated phases
		Total delay	veh-hr	Delay of all vehicles served in the system
		% vehicles in the green	%	Proportion of platoon arriving during the green time per signal cycle
		Bandwidth efficiency	%	Proportion of the green through bandwidth to the signal cycle
		Attainability	%	Proportion of green bandwith to the min green time for the through phase
		Transit delay[1]	sec/bus	average delay to transit vehicles at traffic signals
		Acceleration noise	ft/sec^2	Standard deviation of veh accelerations
	Oversaturated	Throughput	#	# veh served
		Extend of queue	#/mi	Distance or # of street segments with queue spillback
		Congestion duration	hr	Duration of oversaturated conditions
II. SAFETY				
Intersection/ Arterial/ Grid Network	Undersaturated/ Oversaturated	# accidents per type	#/yr	# of accidents by severity and/or traffic movement (e.g., # left turn related accidents)
		Encroachment time (ET)	# conflicts	Surrogate conflict measure
		# RLR	#	# of red light running violators
		# vehicles in yellow	#/cycle	# vehicles in platoon arrive in the yellow clearance interval per signal cycle
III. ENVIRONMENTAL				
Intersection/ Arterial/ Grid Network	Undersaturated/ Oversaturated	Fuel Consumption	gal	Excess fuel consumption due to delay & stops
		HC/CO/NOx/CO2/PM	/gr/m,	Air pollutant emissions / concentrations
		Noise	[db]	Inceased noise level due to congestion

Several studies have been conducted in estimating MOEs from probe data. Most of these studies have focused on the estimation of queue length which is a commonly used measure for designing and evaluating control strategies and is difficult to estimate with static sensors. Comert and Cetin [16] studied

the distribution of average queue lengths assuming known penetration rate of probe vehicles and investigated the impact of the penetration rate of the equipped vehicles on the estimation accuracy. Venkatanarayana et al. [10] estimated queue lengths by using the position of the last equipped probe vehicle in the queue. Hao and Ban [17] and Cheng et al. [18] used Kinematic Wave Theory (KWT) to estimate queue lengths. Other studies have focused on estimating travel time and identifying the number of probe vehicles or penetration rate necessary for reliable travel time estimation on urban arterials investigated the impact of penetration rate on the estimate accuracy [19,20]. CV data were also used by Vasudevan [12] for the estimation of mid-block flow rates and speeds. However, none of the above referenced studies provided a comprehensive description of the estimation process and identified the penetration rate thresholds that ensure accurate estimates of MOEs. This is the first study that systematically determines minimum penetration rates that allow accurate estimates for a variety of MOEs essential for real-time traffic management applications.

The following sections 3.3, 3.4 and 3.5 present the methodology and findings for estimating arterial and intersection MOEs from CV data for traffic signal control. Detailed description of the work performed and the results are included in Appendix C.[12]

3.3 Test Sites

The estimation of MOEs and the impact of the CV technology penetration rate on the accuracy of the MOEs was investigated using actual vehicle trajectories collected as part of the FHWA Next Generation Simulation (NGSIM) Program [21] as the ground truth data source, and simulated vehicle trajectories from a real-world arterial (El Camino Real) in the San Francisco Bay Area [13]. The NGSIM dataset was used to test the proposed estimation procedures for undersaturated traffic conditions, the conditions where the queues are completely dissipated at within each signal cycle. Since no real data was available for oversaturated traffic conditions, when not all the demand can be served within the cycle and residual queues form, the El Camino simulated data was used. A more detailed description of the two datasets follows.

3.3.1 NGSIM Data: Peachtree Street, Atlanta, GA

NGSIM data collected at Peachtree Street, Atlanta, GA for four signalized intersections have been chosen for this study. The NGSIM dataset consists of vehicle trajectories; the format includes vehicle ID, time, position, lane, speed, and acceleration. All data are available at a resolution of $\Delta T = 0.1$ seconds. In addition, vehicle type, vehicle length, and their origins and destinations are available. The traffic signals' settings (cycle length, green, yellow, and offsets) are also known. The data used in this study were collected from 12:45 to 13:00 pm and present only undersaturated conditions. A total of 1,115 vehicles were processed. Since the MOEs estimated in this paper describe the traffic conditions for a given direction, only the trajectories of vehicles traveling northbound of the main arterial were considered. This resulted in a total of 228 vehicle trajectories processed. The effect of penetration rate on the resulting known vehicle trajectories is shown in Figure 3.1 for the NB direction.

[1] Argote, J., E. Christofa, Y. Xuan, and A. Skabardonis, **"**Estimation of Measures of Effectiveness based on Connected Vehicle Data," 14th IEEE ITSC Conference, Washington DC, October 2011.

[2] Argote, J., E. Christofa, Y. Xuan, and A. Skabardonis, "Estimation of Arterial Measures of Effectiveness with Connected Vehicle Data,' 91st TRB Annual Meeting, Washington DC, January 2012.

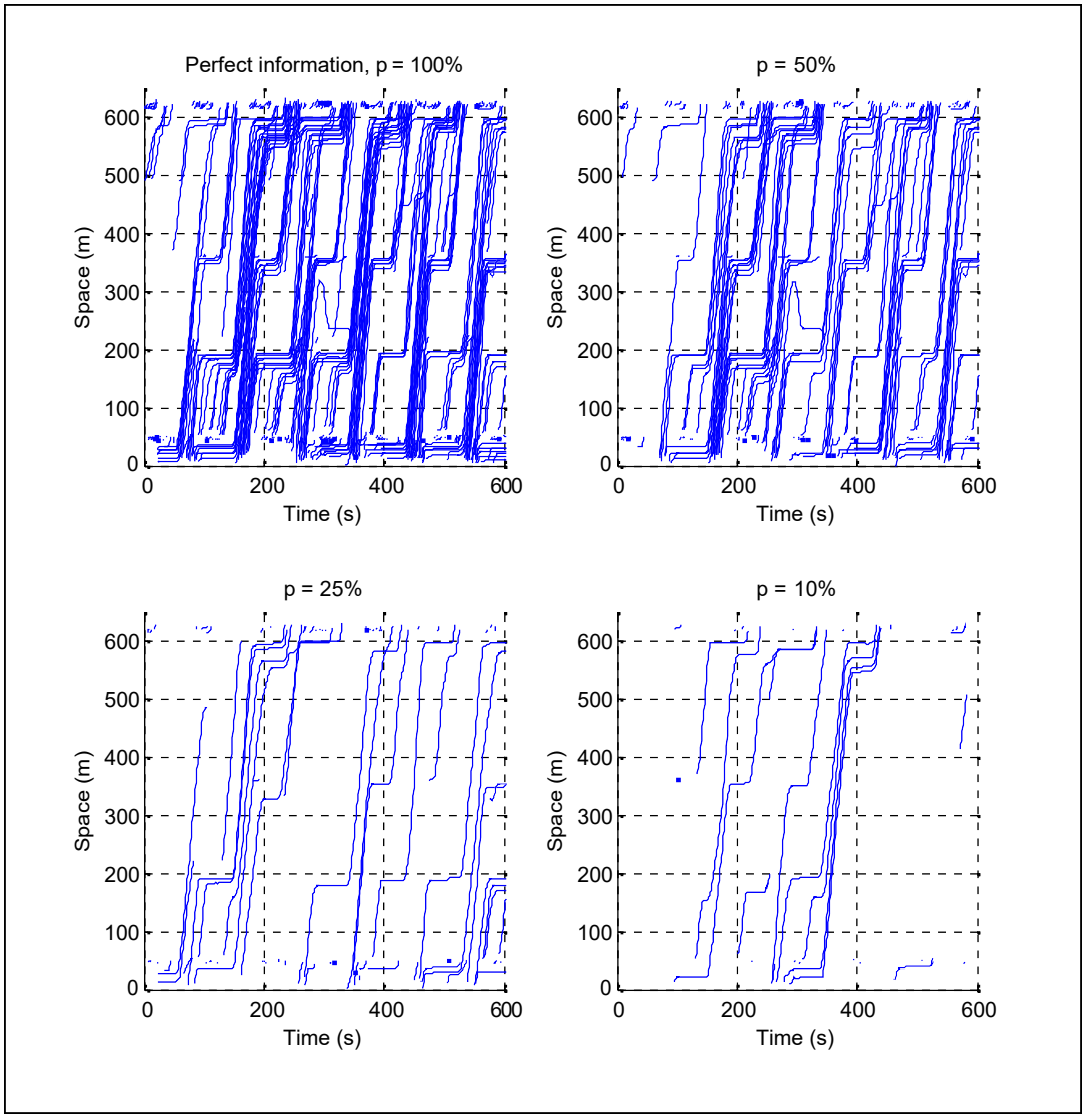


Figure 3.1 Trajectory Plots for Different Penetration Rates—Peachtree Blvd NB

3.3.2 Simulated Data: El Camino Real, San Francisco Bay Area, CA

A section of El Camino Real arterial in Palo Alto/Mountain View, San Francisco Bay Area was simulated using the VISSIM microscopic simulation model [13]. El Camino is a major arterial with substantial commercial, retail and residential/commuter traffic. The study section consists of 18 signalized intersections for a total length of 6.3 miles (10 km). There are three through lanes plus left turn lanes on each signalized intersection approach. Traffic signals are coordinated actuated with a common cycle length of 130 sec. The test site location is shown in the maps of Figure 3.2.

The simulation provides vehicle trajectories as the baseline "truth model" to be sampled by probe vehicles. The trajectories were exported with the same fields and resolution as the NGSIM dataset. A total of 1,332 vehicle trajectories were processed.

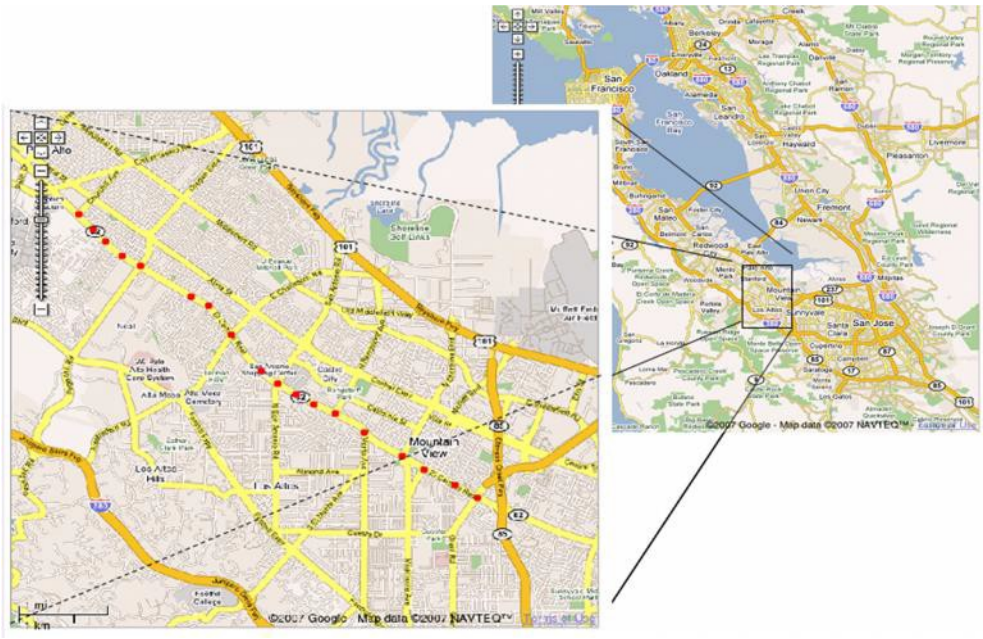


Figure 3.2 El Camino Real Study Location

3.4 Estimation of Arterial Performance Measures

This section presents the estimation of arterial-based MOEs for different market penetration rates *p*, for the CV equipped vehicles. For each of the *p* values considered, 10,000 samples were obtained where each vehicle trajectory was sampled following a Bernoulli trial with probability *p*. Box plots were used to reveal the arterial MOE estimate variation and identify the minimum *p* required to accurately estimate a given MOE for both undersaturated and oversaturated conditions. On each box, the central mark represents the median of the estimates, while the edges of the box show the 25th and 75th percentiles (see for example Figure 3.3). These plots also include a whisker range that extends 1.5 times the distance between the 25th and 75th percentiles at both extremes of the box. A penetration rate was considered acceptable if the whisker range, which corresponds to ± 2.7 the standard deviation of the estimate, laid within a 10% of the average value for the ground truth. If the estimation errors are assumed to be normally distributed, this criterion ensures that the MOE estimate would lay within a 10% of the ground truth with a probability of 0.9965.

3.4.1 Average Speed

The average speed for all lanes in the observed direction for a single sample was obtained by using Edie's generalized average speed definition [22], including all the vehicles in the sample:

$$\bar{v} = \frac{\sum_{i=1}^{S} l_i}{\sum_{i=1}^{S} t_i}.$$

Where:
S is the total number of vehicle trajectories sampled,
l_i is the total distance traveled by vehicle i,
t_i is the total time spent by that vehicle to traverse the distance l_i.

Figure 3.3 shows the boxplot of the average speeds for different penetration rates for both the NGSIM and the simulated data. Based on the threshold of $\pm 2.7\sigma$ defined above, it is shown that penetration rates higher than 35% are necessary to provide accurate speed estimates for an arterial for undersaturated traffic conditions and higher than 5% for ovesaturated traffic conditions.

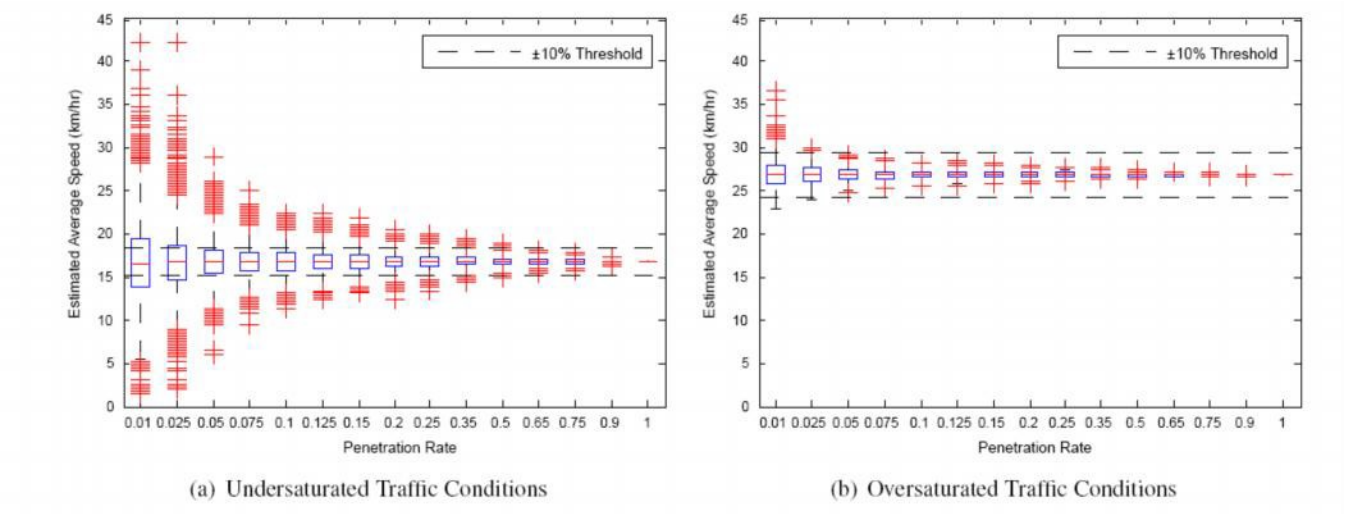


Figure 3.3 Box Plots of Average Speed Estimates vs. Penetration Rate

3.4.2 Average Delay per Unit Distance

Delay is defined as the difference between the actual travel time to traverse an arterial section and the travel time under free flow conditions. The average delay per unit distance for a given sample was obtained by the following expression:

$$\bar{D} = \frac{1}{S}\sum_{i=1}^{S}\frac{1}{l_i}\left(t_i - \frac{l_i}{v_f}\right).$$

Where:
D is the average delay for a single sample of trajectories,
v_f is the free flow speed on the arterial corridor.

The free flow speed used for Peachtree street and El Camino Real was 40 km/hr (25 mph) and 64 km/hr (40 mph) respectively. This MOE can also be interpreted as the difference between the sampled average pace and the free flow pace. Again, the box plot in Figure 3.4(a) indicates that samples obtained under low penetration rates present a significant number of outliers, which could lead to a misinterpretation of the traffic conditions. For this case, penetration rates higher than 75% would be necessary for accurate estimates in undersaturated conditions. On the other hand, Figure 3.4(b) shows that only a 15% penetration rate would be required in oversaturated conditions.

Alternatively, we can simply compute the average total delay, without considering a distance normalization. In that case the algorithm formulation would be:

$$\bar{D} = \frac{1}{S}\sum_{i=1}^{S}\left(t_i - \frac{l_i}{v_f}\right).$$

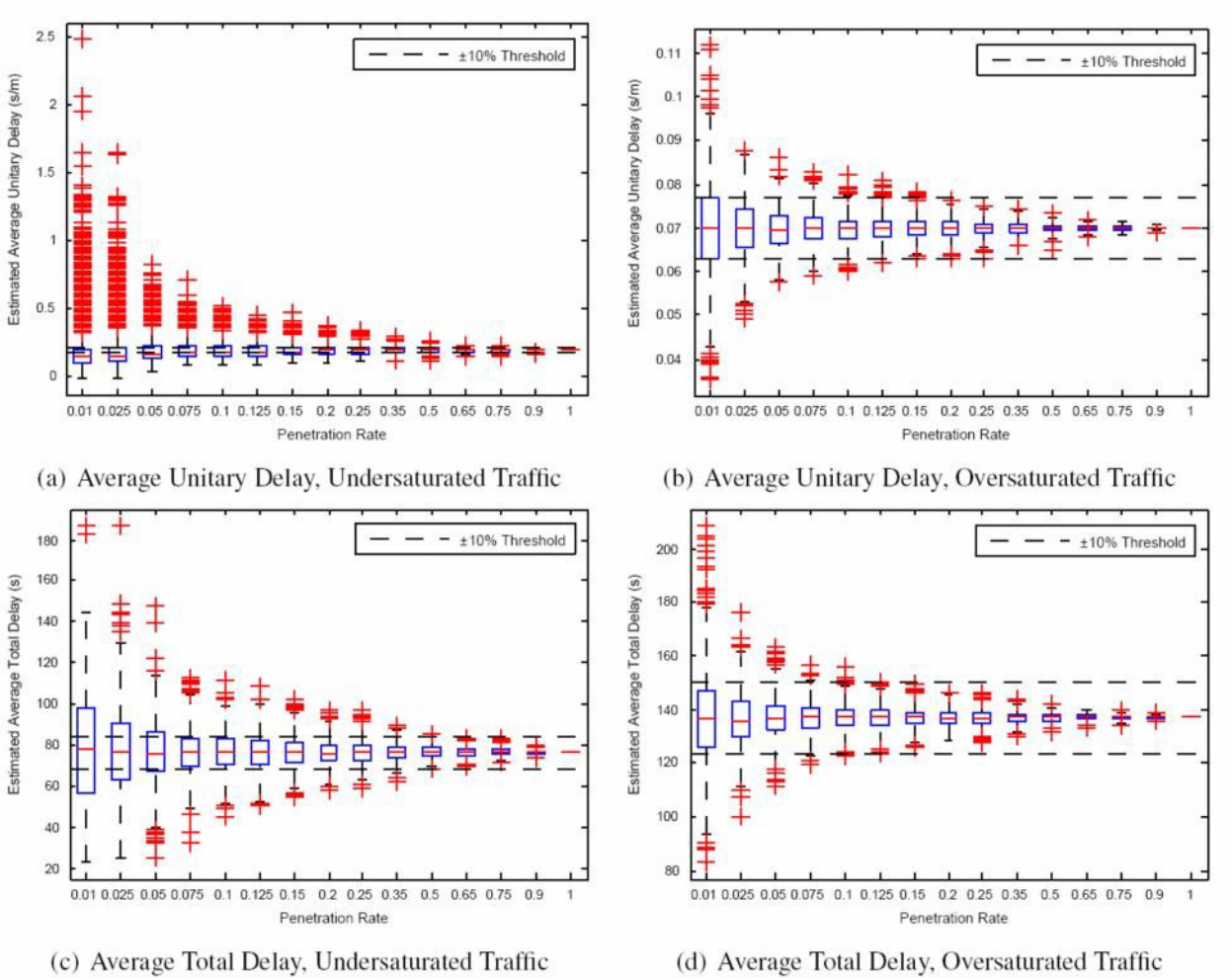
The results presented in Figure 3.4(c) and Figure 3.4(d) show that in this case penetration rates higher than 50% would be necessary for accurate estimates in undersaturated conditions and 10% in oversaturated conditions.

3.4.3 Average Number of Stops

The average number of stops was obtained as the mean number of stops of all vehicles within the sample that were traveling through the arterial on the direction under consideration. A stopped vehicle was defined as a vehicle with speed less than 5 km/h (3.1 mph). There is a high variability in the estimates for the average number of stops for low p. In this case, p values higher than 50% and 20% would be necessary to obtain results that would lie close to the ground truth value for the undersaturated and oversaturated case respectively.

3.4.4 Average Acceleration Noise

Acceleration noise is defined as the standard deviation of a vehicle's acceleration and is used as a measure of the smoothness of traffic flow along signalized arterials. The average acceleration noise was estimated as the mean of the vehicle accelerations' standard deviation [23]. The results show that penetration rates on the order of 10% and above should be sufficient to obtain reliable standard deviations for the acceleration for undersaturated conditions but only 1% is necessary for oversaturated conditions.



(a) Average Unitary Delay, Undersaturated Traffic

(b) Average Unitary Delay, Oversaturated Traffic

(c) Average Total Delay, Undersaturated Traffic

(d) Average Total Delay, Oversaturated Traffic

Figure 3.4 Box Plots of Delay Estimates vs. Penetration Rate

3.4.5 Summary

Testing of the proposed estimation methods in two different datasets concludes that in general lower penetration of CV is needed in order to obtain accurate estimate when in oversaturated traffic conditions in comparison with undersaturated ones (Table 3.3). This result is logical since in oversaturated conditions we have a greater amount of vehicles, which increases the number of CV sampled. This implies that for congested networks, in which real-time control strategies are more critical, accurate estimation of MOEs can be done with lower penetration rates.

Table 3.3 Minimum Penetration Rate for MOE Estimates within 10% of the Ground Truth

MOE	Undersaturated Required p	Oversaturated Required p
Average Speed (km/hr)	35%	5%
Average Total Delay (s/m)	50%	10%
Average Number of Stops	50%	20%
Acceleration Noise (m/s^2)	10%	1%

3.5 Estimation of Intersection Performance Measure: Queue Length

Three methods were developed and tested for estimating queue length and the most accurate one is identified.

3.5.1 Discretization of Time and Space

First, the time-space of interest is discretized in different parallelogram regions accounting for the different links in the arterial *l = {1, 2,. . .},* the different signal cycle times *k= {1, 2, . . }* and the backward wave speed of the stopped and accelerating vehicles (Figure 3.5). This discretization allows us to identify several cells noted by the pairs *(k,l)* with their corresponding queues. Such unconventional discretization of time and space allows to capture the cases where queue propagates so long that cars have to still decelerate and stop even after the onset of the green time for the next cycle. This is particularly useful for oversaturated traffic conditions. The backward wave speed was estimated from the El Camino Real dataset as 19.44 km/h (12.08 mph).

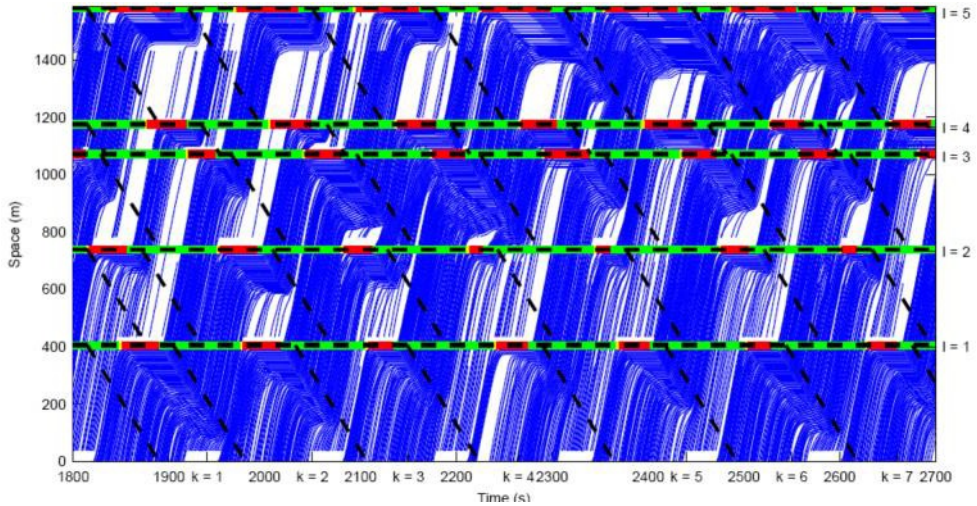


Figure 3.5 Discretization of Space and Time

3.5.2 Identification of Deceleration and Acceleration Points within a Queue

Let us now focus on a specific cell *(k,l)* and drop the notation *(k,l)* to facilitate expression. Next, all the deceleration points corresponding to that cell have to be identified so that the positions of the stopped vehicles can be determined. The time corresponding to the deceleration point of vehicle *i (i=1, . . . ,s)* within a cell t^d_i is determined by selecting the time that satisfies the following condition:

$$t_i^d = min\{t | v_i(t) > v_c \;\&\; v_i(t+\Delta T) \le v_c \;\&\; (t, y_i(t)) \in \text{cell}(k,l)\}$$

where *vi(t)* is the speed of vehicle *i* at time *t*, v_c is a constant speed threshold (e.g., 5 km/hr (3 mph)), i.e., the vehicle is practically stopped, *DT* is the sampling interval and $y_i(t)$ is the position of vehicle *i* at time *t*. These conditions identify the points in time and space within the cell where vehicle i joins the queue. A vehicle is considered to be in queue when it starts decelerating at a lower speed than the specified threshold. Therefore, it is necessary to consider the minimum of the times (or the earliest deceleration point) that satisfies these conditions, so that non-desirable deceleration points are not taken into account.

All the times that satisfy (4), the set $\{t_i^d\}$, can now be rearranged defining a new set of times $U = \{t^d_{(i)}\}$ so that $y_{(1)}(t^d_{(1)}) \ge y_{(2)}(t^d_{(2)}) \ge \ldots \ge y_{(s)}(t^d_{(s)})$. Note that all the times included in this unfiltered set U do not necessarily correspond to deceleration points within the queue. As shown in Figure 4.4, some of the points included in U may correspond to vehicles which are not in the queue, e.g., stopping within the link in order to park. Therefore, we need to impose additional constraints to eliminate these undesired points.

Figure 3.6 indicates that the undesired deceleration points are far from the deceleration points within the queue. Thus, the distance between such points could be used as a filtering criterion. If vehicles within the same queue are sampled independently with probability *p*, given by the CV market penetration rate, the number of vehicles between adjacent sampled vehicles is a random variable satisfying a geometric distribution with parameter p. The *100(1-e)th* percentile of a geometric distribution with parameter p is *ln(e)=ln(1-p)* (we used the 90th percentile in this study, which corresponds to *ln(0.1)=ln(1-p)).* Therefore, the filtering process is defined by the next condition, with NL being the number of lanes on the approach and Kj being the jam density per lane:

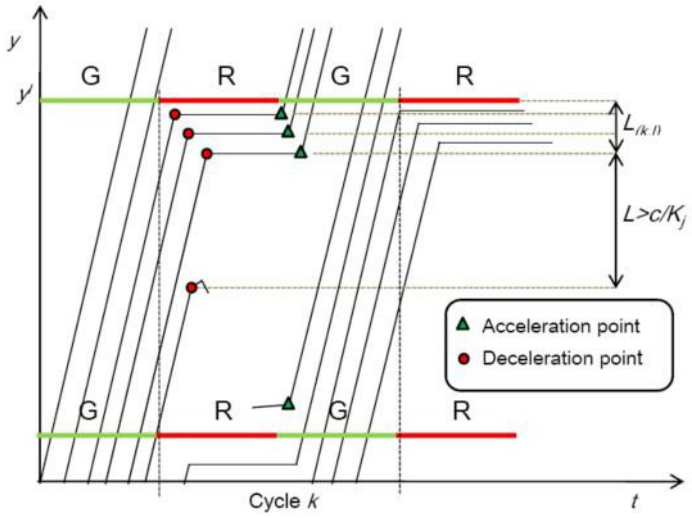


Figure 3.6 Identification of Deceleration and Acceleration Points within the Queue

$$R = \left\{ (i) \middle| d_{(i)} \triangleq |y_{(i)}(t^d_{(i)}) - y_{(i+1)}(t^d_{(i+1)})| > \frac{c}{K_j} = \max\left\{ \frac{\ln(\varepsilon)/\ln(1-p)}{N_L K_j}, 1/K_j \right\} \right\} \quad (5)$$

After applying (5) to trajectories $(i) \in \{(1), \ldots, (s-1)\}$, the set of vehicle indices to be considered in the queue length estimation is given by:

$$A = \begin{cases} \{(1)\} & \text{if } s = 1 \\ \{(1), \ldots, (s)\} & \text{if } R = \{\emptyset\} \\ \{(1), \ldots, \min\{R\}\} & \text{otherwise} \end{cases} \quad (6)$$

Finally, F defines the set of all deceleration points that belong to vehicles within the queue of interest.

$$F = \{t^d_{(i)} | (i) \in A\} \quad (7)$$

Note that F is a subset of U, and if two deceleration points are too far apart, we would suspect that the most upstream one is not actually in the queue. A similar method can be applied to identify the acceleration points, which are denoted similarly as $t_a(i)$ and are always paired with deceleration points. Figure 3.7 shows the deceleration (red asterisks) and acceleration (green asterisks) points at intersection 2 of Peachtree Street after filtering.

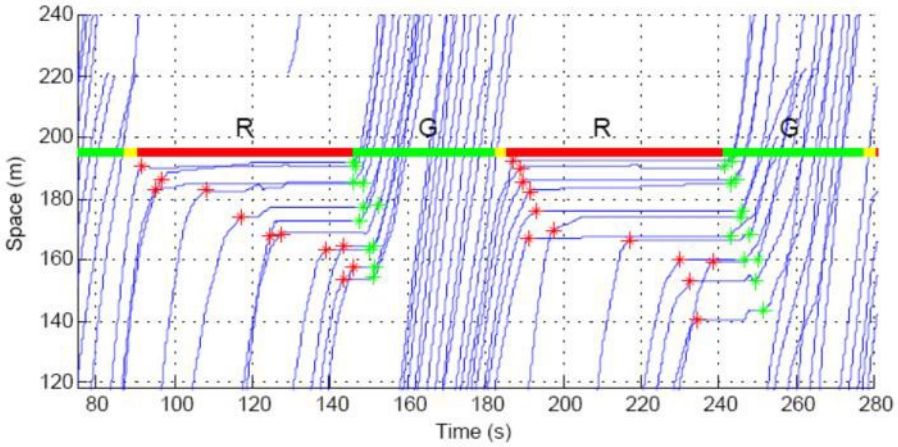


Figure 3.7 Filtered Deceleration and Acceleration Points

3.5.3 Queue Length Estimation Methods

Queue length can be easily estimated if the penetration rate of CVs is 100%. But with only a fraction of vehicles sampled, queue length estimation is a challenge. We analyzed three possible methods to estimate queue length; for each method 2,000 samples are obtained where each vehicle trajectory is sampled following a Bernoulli trial with probability *p*.

Maximum Likelihood (ML) Queue Length Estimation

Assuming that the position of the vehicles in queue can be characterized by a discrete uniform distribution, we can use Maximum Likelihood (ML) estimation. In this case, the estimator of the queue length, ˆLML will be the relative position of the vehicle located further apart from the intersection, i.e., the maximum order statistic for the relative position of the vehicles within the queue with respect to the intersection (Figure 4.8). From equation (6), if we define the number of elements in A (or its cardinality) to be n_A we have:

$$\hat{L}_{ML} = |y^l - y_{(n_A)}(t^d_{(n_A)})|$$

where y^l is the location of the downstream intersection of link l. If this algorithm is applied assuming 100% penetration rate, the estimate obtained gives the ground truth queue length. However, for penetration rates other than 100% this estimator is biased, i.e., the average value given by the estimator will not be equal to the real value of the queue length. Instead, this estimator tends to underestimate the actual queue length. For low penetration rates it is very unlikely that the sampled vehicles will include the last vehicle in the queue and therefore the queue length will be underestimated.

Method of Moments (MM) Queue Length Estimation

The method of moments is also used under the assumption that the vehicles in queue are characterized by a discrete uniform distribution. In this case, the queue length estimator, ˆLMM is equal to two times the mean of the sample relative positions:

$$\hat{L}_{MM} = 2\left|y^l - \frac{1}{n_A}\sum_{i=1}^{n_A} y_{(i)}(t^d_{(i)})\right|$$

This method is theoretically unbiased. However, in reality the cars in a queue will not be uniformly spaced, and even with a 100% penetration rate we would not observe the actual queue length, which means that this method yields worse estimates than the previous one.

Queue Length Estimation based on Kinematic Wave Theory (KWT)

Queue length estimation based on KWT requires at least two measurements of deceleration-acceleration points from all the vehicle trajectories in a given cell *(k,l)*. Figure 3.8 shows how the shockwaves (the threshold between the free flowing and queued states) predicted by KWT can be estimated. The intersection point of the two shockwaves determines the maximum possible queue length, $\hat{L}_{KW}$:

$$\hat{L}_{KW} = \left| y^l - \frac{\frac{y_{(1)}(t^d_{(1)})}{w^d} - \frac{y_{(1)}(t^a_{(1)})}{w^a} + t^a_{(1)} - t^d_{(1)}}{\frac{1}{w^d} - \frac{1}{w^a}} \right| \quad (10)$$

where:

$$\frac{1}{w^d} = \frac{t^d_{(n_A)} - t^d_{(1)}}{y_{(n_A)}(t^d_{(n_A)}) - y_{(1)}(t^d_{(1)})} \quad (11)$$

$$\frac{1}{w^a} = \frac{t^a_{(n_A)} - t^a_{(1)}}{y_{(n_A)}(t^a_{(n_A)}) - y_{(1)}(t^a_{(1)})} \quad (12)$$

In undersaturated traffic conditions the beginning of the first shockwave does not necessarily match the beginning of the red phase in the signal cycle. So it is necessary to sample at least two vehicle trajectories in at least one of the travel lanes within a given cycle, in order to be able to estimate the maximum potential queue length. Thus, before we apply the proposed estimation algorithm we need to know which the minimum acceptable penetration rate *p* is.

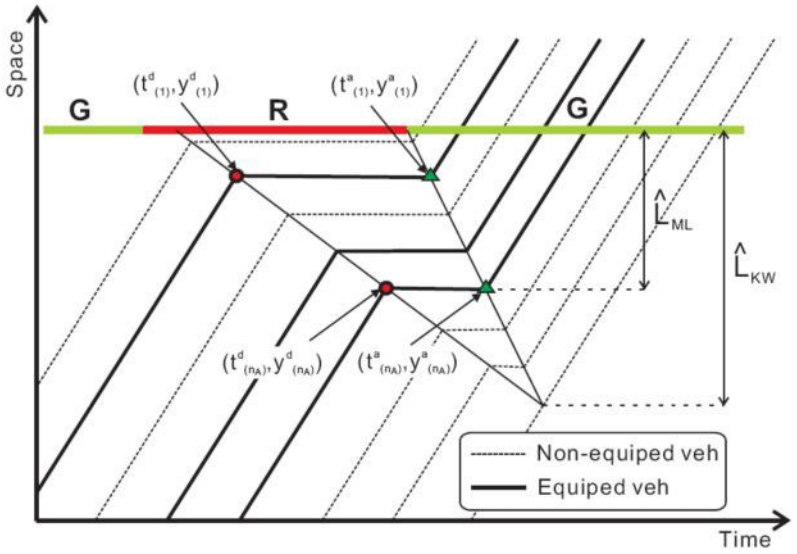


Figure 3.8 ML and KWT Queue Length Estimation

A possible metric to determine p would consist of calculating the value that guarantees that at least 90% of the time two vehicle trajectories on the same lane will be sampled. Let us assume that the number of vehicles traveling on one lane can be modeled as a discrete Uniform *(N_{min}, N_{max})* distribution. Then, for a given number of cars traveling on one lane, approaching an intersection, the probability of selecting at least two of them, based on a binomial distribution, is:

$$P(s_L \geq 2|N_L = m) = 1 - P(s_L < 2|N_L = m) = 1 - \sum_{j=0}^{1} \binom{m}{j} p^j (1-p)^{m-j} \quad (13)$$

where s_L is the number of vehicles sampled in lane L and N_L is the total number of vehicles at that lane for a given cycle. This formula in combination with the assumption that N_L can be modeled as a discrete Uniform distribution, allows us to compute the probability of sampling at least two vehicles traveling on lane L in a given cycle as:

$$P(s_L \geq 2) = 1 - \sum_{m=N_{min}}^{N_{max}} P(s_L < 2|N_L = m)P(N_L = m) =$$

$$1 - \frac{1}{N_{max} - N_{min} + 1} \sum_{m=N_{min}}^{N_{max}} \left(\sum_{j=0}^{1} \binom{m}{j} p^j (1-p)^{m-j} \right) \quad (14)$$

To illustrate this, let us assume that the number of vehicles arriving at an intersection in a specific lane ranges from 1 to 10 vehicles Figure 3.9 shows the probability $P(s_L \geq 2)$ for different penetration rates. It can be seen that penetration rates higher than 50% ensure sampling of at least two vehicles that travel on the same lane within a cell more than 75% of the time.

Figure 3.10 compares the average absolute relative errors obtained by the three queue length estimation methods. Average absolute relative errors are used here instead of box plots because all three estimators are biased and the results from different intersections are combined. Unfortunately, the undersaturated

conditions in the NGSIM dataset are not appropriate for the use of the estimator based on the KWT. The low number of vehicles and the non-uniformity of their arrivals at the intersections result in the overestimation of the queue length. For oversaturated conditions, the KWT estimator still has the highest absolute error, which however is much smaller than the one for undersaturated conditions.

The maximum likelihood estimation method clearly provides the best results, although it would require a penetration rate of at least 80% to guarantee that the average absolute relative error of the estimation does not exceed 10% for undersaturated traffic conditions. However, for oversaturated conditions, a penetration rate of only 20% is sufficient to provide an absolute relative error lower than 10%.

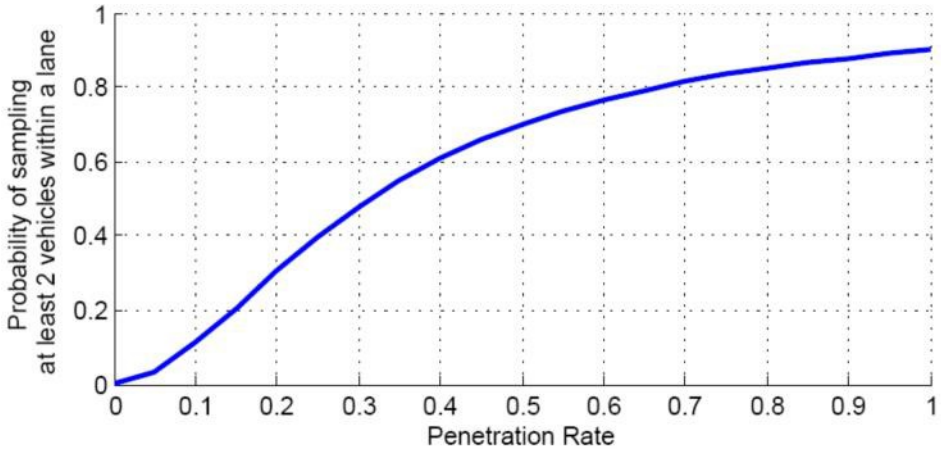


Figure 3.9 Probability of Sampling at Least Two Vehicles for a Given Penetration Rate, (N_{min} = 1, N_{max} = 10)

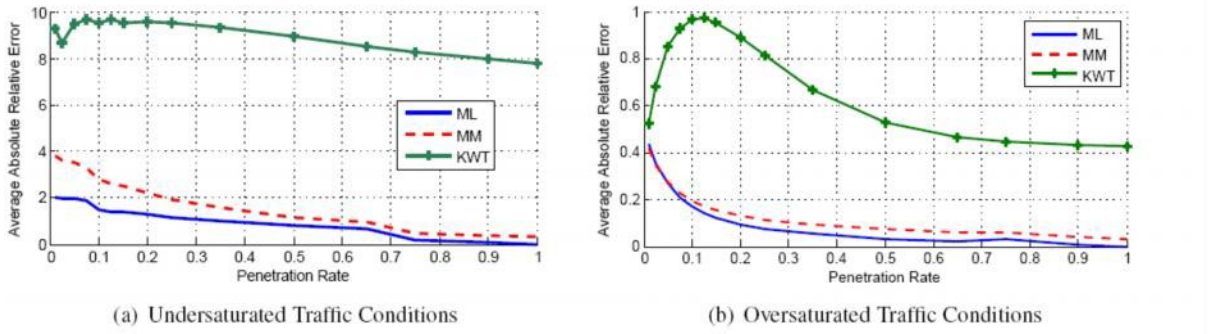


Figure 3.10 Queue Length Estimation: Average Absolute Relative Error vs. Penetration Rate

In summary the finding from testing the three queue length estimation methods, Maximum Likelihood, Method of Moments, and one based on Kinematic Wave Theory show that 80% of the vehicles need to be equipped to obtain accurate estimates for undersaturated conditions but only 10% for oversaturated conditions. This is promising in the sense that accurate estimates of queue length can be obtained for low penetration rates and as a result, improved control strategies can be designed for oversaturated conditions even at the first stages of the CV implementation.

3.6 Evaluation of Probe Vehicle Sampling Strategies for Traffic Signal Control

This section describes the methodology and findings from the evaluation of sampling strategies of probe vehicle data for estimating MOEs for traffic signal control. Detailed description of the work performed and the results are included in Appendix C.[3]

We first run the VISSIM a microscopic simulation model to generate the trajectories of all vehicles. The simulated vehicle trajectories at 1 sec resolution were treated as the ground truth for the corridor. We post-processed the simulated trajectories using custom written software to emulate alternative sampling strategies and penetration rates. We applied the sampling strategies and market penetration rates to the trajectory data to generate snapshots to be sent to roadside equipment (RSE) within wireless communication range. Finally, the MOEs are estimated from the snapshots received at each RSE. We focused on the critical intersections of the arterial as the starting point for evaluation (California Avenue and Page Mill Road). The simulation covers a two-hour peak period, and with a signal cycle length of 130 seconds, we encounter 55 complete signal cycles during the analysis period.

3.6.1 Candidate Sampling Strategies

Uniform Sampling

This is the simplest sampling strategy at fixed time intervals. Uniform sampling at intervals of 5 sec, 10 sec and 20 sec was studied to provide a starting point for comparison with more complicated strategies that aim to improve sampling efficiency.

The uniform sampling strategy is not particularly efficient because it may include multiple redundant samples during times when nothing is changing (vehicles are stopped or cruising at constant speeds). When it is necessary to make most efficient use of the wireless communication channel to avoid communication problems the sampling rules can become more complicated in attempts to reduce redundancy of sampling.

SAE J2735 Messaging Standard

The Society of Automotive Engineers (SAE) invested considerable effort in definition of a probe sampling strategy as part of its J2735 messaging standard for DSRC. The full description of the SAE J2735 sampling protocol requires ten pages of text and does not have to be repeated here. Prior research at PATH and UMTRI has shown it to have some significant limitations, including biased speed estimates based on under-sampling of stopped vehicles, blind spots associated with privacy-motivated interruptions in probe snapshot sequences, and long latencies associated with the rules for when snapshots are uploaded from vehicles. Nevertheless, the key features of the J2735 probe sampling rules are:

- Snapshots generated periodically, by events and for stops and starts
  - 6 s intervals for speeds < 20 mph
  - 20 s intervals for speeds > 60 mph
  - Linear interpolation of sampling interval between 20 – 60 mph
  - Special events – vehicle status changes (TCS, ABS activation)
  - Stop (no movement for 5 s, no other stops within 15 s)
  - Start (after speed exceeds 10 mph)
  - Snapshot buffer size of 32 (default minimum)
- Priority for snapshot transmission: events, stops and starts, then periodic

[3] Shladover, S.E, and J-Q Li, "Evaluation of Probe Vehicle Sampling Strategies for Traffic Signal Control," 14th IEEE ITSC Conference, Washington DC, October 2011.

- FIFO priority for buffer, then overwritten
- May delete alternate snapshots to reduce buffer overflows
- Association of snapshots within a common Probe Status Number (PSN) is severely restricted (120 s and 1 km limit en route, no snapshots in first 500 m)
- Transmission to Roadside Equipment (RSE) only once, when first coming into range (no subsequent transmission until reaching next RSE)

Modified J2735 Strategy

A variation of J2735 was then studied. We first studied the case in which the Probe Status Number switching was turned off, and probe status numbers were maintained throughout the life of the vehicle. The original PSN switching provision was incorporated to protect driver privacy when the VII program was based on the concept of all raw probe data snapshots being uploaded to the "cloud", where they could potentially be mined for patterns that would reveal an individual's trip. That concept is no longer valid, and in future real implementations we should be able to assume that the raw probe snapshots will be aggregated locally at or near the intersections, so that the raw data will never become available for somebody to mine inappropriately. Then, we uploaded probe snapshots more frequently than the original J2735, reducing the number of snapshots accumulated per upload from four to two, in order to reduce the latency in availability of the probe data.

Speed Based Sampling Strategy

A purely speed-based sampling strategy was also studied, without the other complications of the J2735 protocol. In this sampling strategy, when vehicles travel above 20 mph they are sampled relatively infrequently (30 s intervals). As the speed declines to zero, the sampling interval is reduced to 4 s. When the speed is between 0 and 20 mph, a linear interpolation determines the sampling interval. This strategy is expected to better capture data about highly congested situations. After two snapshots are accumulated on the vehicle, the snapshots are uploaded to the nearest intersection RSE.

Speed-Acceleration Based Sampling Strategy

We also changed the sampling protocol more dramatically so that it focuses on changes in speed. In this way we would hope to minimize wasted snapshots but still respond as fast as possible when changes in vehicle motions occur. An acceleration–based sampling strategy was also studied, in which a low sampling rate was used when speed was not changing, but the sampling rate was increased in proportion to the rate of acceleration in order to capture the important dynamic conditions associated with vehicles starting and stopping at a higher fidelity. In more detail, if the absolute value of the current acceleration rate is no more than 0.2 m/s^2, the sampling interval is set as 10 seconds. Otherwise, the sampling interval is set as 1.0/(0.001 + |acceleration rate|), with the results bounded between 2 and 30 seconds to eliminate extremely large and small sampling intervals.

3.6.2 Estimation of Intersection MOEs: Delays, Number of Stops, Queue Lengths

The measures of effectiveness (MOEs) we are estimating from the sampled data for each signal cycle at an individual intersection (El Camino Real at California Avenue) include:

- the total delay for vehicles traversing the intersection along El Camino, measured as the excess of the actual travel time over the free-flow time for vehicles to go from upstream mid-block to downstream mid-block
- the percentage of vehicles that need to stop at the intersection on El Camino.
- the maximum queue length along El Camino

These MOEs can only be estimated in a meaningful way for a complete signal cycle because they vary too dramatically within any individual signal cycle. If we were to try to sample them more frequently the variations within the signal cycle (differences between green onset and stale green or between red onset and end of red) would dominate everything else and it would not be possible to make sensible control decisions. The downside of estimating these MOEs over a complete signal cycle is that it is not possible to make feedback control changes within the cycle, but only to use the data to make adjustments on subsequent signal cycles, introducing a control response latency of at least one signal cycle length.

If it were possible to take snapshots of the state of every vehicle at very frequent update intervals (of the order of one second) this would produce a very accurate representation of the motions of the vehicles and of the complete state of the network. However, such frequent sample updates would overwhelm the wireless network, even using DSRC technology, and would not be practical. The challenge is in finding a sampling frequency and sampling strategy that can produce "accurate enough" outputs at sampling rates that the wireless system can accommodate. The analysis here begins with the simplest of strategies, sampling at uniform update intervals, and then proceeds to more complicated strategies that seek to produce the needed description of the state of the network more efficiently (with fewer total samples and less burden on the wireless channel).

The performance of any probe sampling strategy is heavily dependent on the density of probe vehicles, which is the product of traffic density and the market penetration of probe vehicles. When traffic density is low, it is particularly difficult to collect probe data if the market penetration of probe vehicles is also low because the probe vehicles will only rarely show up on any given road segment. Early in the development of a connected vehicle system, the market penetration will inevitably be low and it could take quite a few years for the market penetration to rise above 50% of the vehicle population. When traffic density is high and there is a high market penetration of probe vehicles, these vehicles may overwhelm the wireless channel and data system unless their sampling frequency is reduced. Therefore it is important to evaluate the probe sampling strategies across a full range of market penetrations.

The uniform sampling strategies have been evaluated at update intervals of 5, 10 and 20 seconds to give an indication of the sensitivity of performance to sampling rates that can vary over a factor of four. These establish a baseline for comparison with the other strategies, in which the sampling rates can vary based on vehicle speed or acceleration.

Delay

The delay is defined as the difference between the actual time needed to traverse the intersection and the ideal free-flow time. Since the delay encountered by an individual vehicle is heavily dependent on when it arrives during the signal cycle, this variable also has to be estimated for entire signal cycles in order to balance the effects of vehicles that arrive on different parts of the cycle. In order to account for the full sequence of deceleration, stop and acceleration back to cruise by the vehicles that need to stop at a signal, it is important to measure the delay between locations on opposite sides of the intersection and far enough away to be outside the locations where queues form. The aggregate measure of delay is the number of vehicle-seconds by which the actual travel time exceeds the free-flow time during each signal cycle.

Figure 3.11 shows an example of the accumulation of delay over time at the example intersection at California Ave. through several signal cycles. The blue diamonds labeled "Base" represent the true values from the VISSIM model simulation, while the other plots represent the samples at different market penetrations.

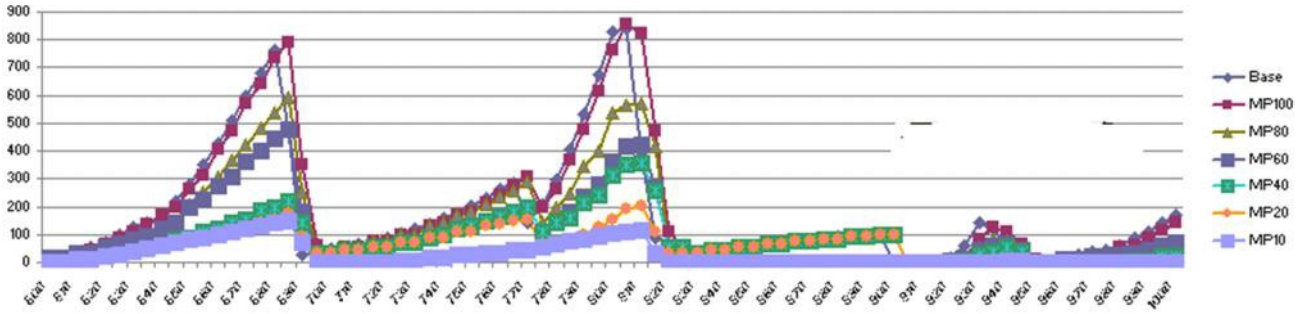


Figure 3.11 Delay Accumulation vs. Time--California Avenue Intersection

It can be seen from Table 3.4 that the mean delay estimates are unbiased for the 10 sec and 5 sec sampling intervals, but at 20 sec sampling intervals the delays are systematically under-estimated. At the low market penetration levels the standard deviation is at least half the mean value. In contrast, at the full market penetration level, there is a factor of at least five between the standard deviations and mean values of the ratios, providing for much better estimates of the actual delays.

Table 3.4 Ratios of Sampled to True Intersection Delays

	Market Penetration (%)					
	10	**20**	**40**	**60**	**80**	**100**
20 s sample updates:						
Mean ratio	0.862	0.958	0.973	0.963	0.938	0.941
Standard deviation of ratio	0.546	0.491	0.360	0.282	0.259	0.215
10 s sample updates:						
Mean ratio	0.978	1.061	1.071	1.044	1.039	1.032
Standard deviation of ratio	0.418	0.356	0.293	0.265	0.257	0.264
5 s sample updates:						
Mean ratio	0.985	1.025	1.040	1.023	1.010	1.000
Standard deviation of ratio	0.446	0.349	0.254	0.184	0.168	0.152

Figure 3.12 shows the coefficients of variation for the delay estimates at the California Avenue intersection. The unmodified J2735 and uniform sampling at 20 sec update interval are noticeably worse than the other strategies and a relatively smooth trend of improvement with respect to market penetration can be seen. The unmodified J2735 sampling is problematic here because the privacy-motivated changes of probe status number cause the loss of many samples, making it impossible to track the times that vehicles need to traverse the intersection.

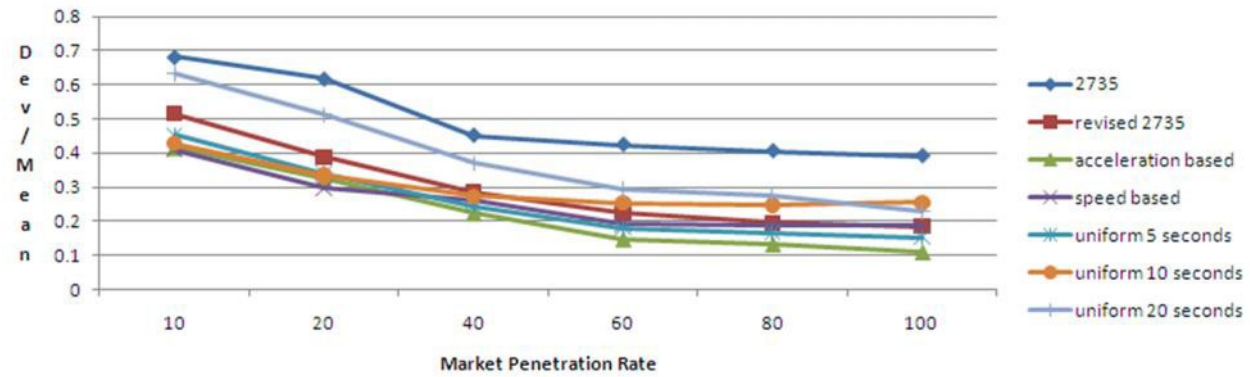


Figure 3.12 Coefficient of Variation—Delay Estimates vs. Alternative Sampling Strategies

These results indicate that in order to use the probe sampled data for signal control, it will probably be necessary to aggregate data from more than one signal cycle for market penetrations less than 60%. This aggregation increases the latency in the traffic control system, meaning that adjustments would have to be made after the passage of multiple signal cycles rather than just one cycle.

Maximum Queue Length

Figure 3.13 shows the sampling of the queue length (blue diamonds) compared with the underlying VISSIM simulation with the actual queue length (red squares) at the California Ave. intersection. The individual signal cycles are clearly visible in this plot of about 15 minutes of simulated time. Even with this high market penetration, the estimated queue length is noticeably shorter than the actual queue length since there were always unequipped vehicles in the queue, which could not be counted. This is a systematic bias that is impossible to avoid, but we should consider the possibility of compensating for it if we have a priori knowledge of the effective local market penetration of equipped vehicles.

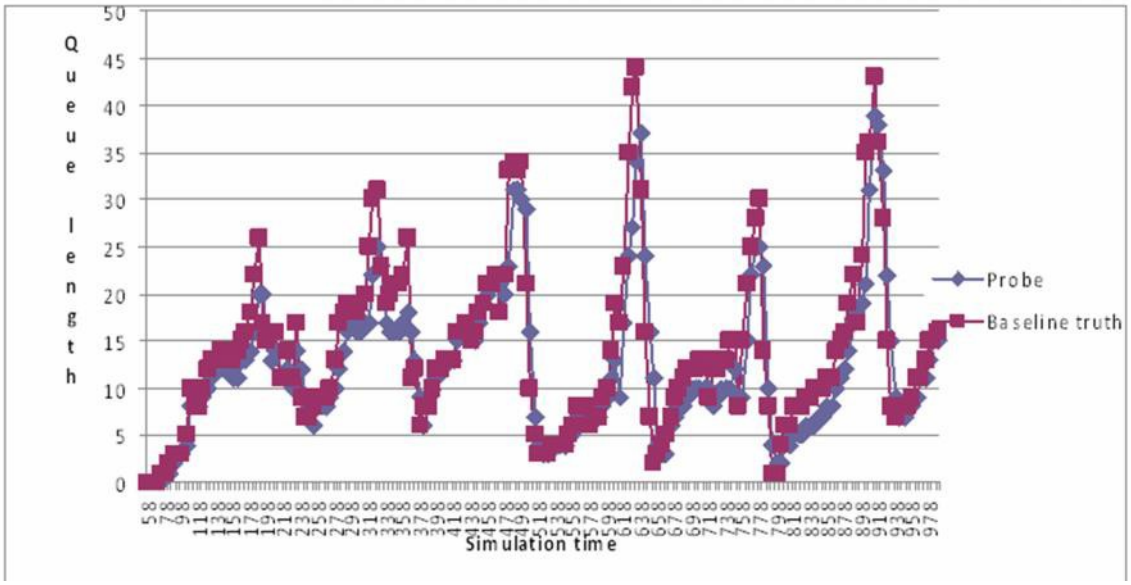


Figure 3.13 Queue length vs. Time—90% Market Penetration and 5 sec Sample Updates

The mean values and standard deviations of these ratios are summarized in Table 3.5 as a function of market penetration and sampling rate.

Table 3.5 Ratios of Sampled to True Queue Lengths

Sample Updates	Market Penetration (%)					
	10	**20**	**40**	**60**	**80**	**100**
20 s sample updates:						
Mean ratio	0.06	0.13	0.26	0.37	0.48	0.61
Standard deviation of ratio	0.040	0.063	0.091	0.126	0.134	0.141
10 s sample updates:						
Mean ratio	0.09	0.18	0.36	0.54	0.70	0.88
Standard deviation of ratio	0.051	0.073	0.087	0.106	0.114	0.096
5 s sample updates:						
Mean ratio	0.09	0.17	0.31	0.46	0.59	0.73
Standard deviation of ratio	0.049	0.059	0.065	0.071	0.077	0.073

These results imply that for any given sampling rate and an estimated market penetration, the sampled queue length estimate could be scaled up by the inverse of the mean ratio shown in the table to produce a more realistic estimate of the actual queue length. The limitation in this approach is that there is still too much variability in the data. At the lower market penetrations, the standard deviation is about half the mean value, which shows that the estimate would still be rather inaccurate. At the full market penetration, the standard deviation is only about one tenth of the mean value, so that estimate would be much more accurate.

Figure 3.14 plots the ratios of the sampled maximum queue length estimates at California Ave. for each simulated signal cycle to the actual queue lengths in the simulation outputs for a variety of market penetration levels. Each market penetration level is represented by 55 data points, corresponding to the 55 signal cycles that occurred during the two hours of simulated traffic.

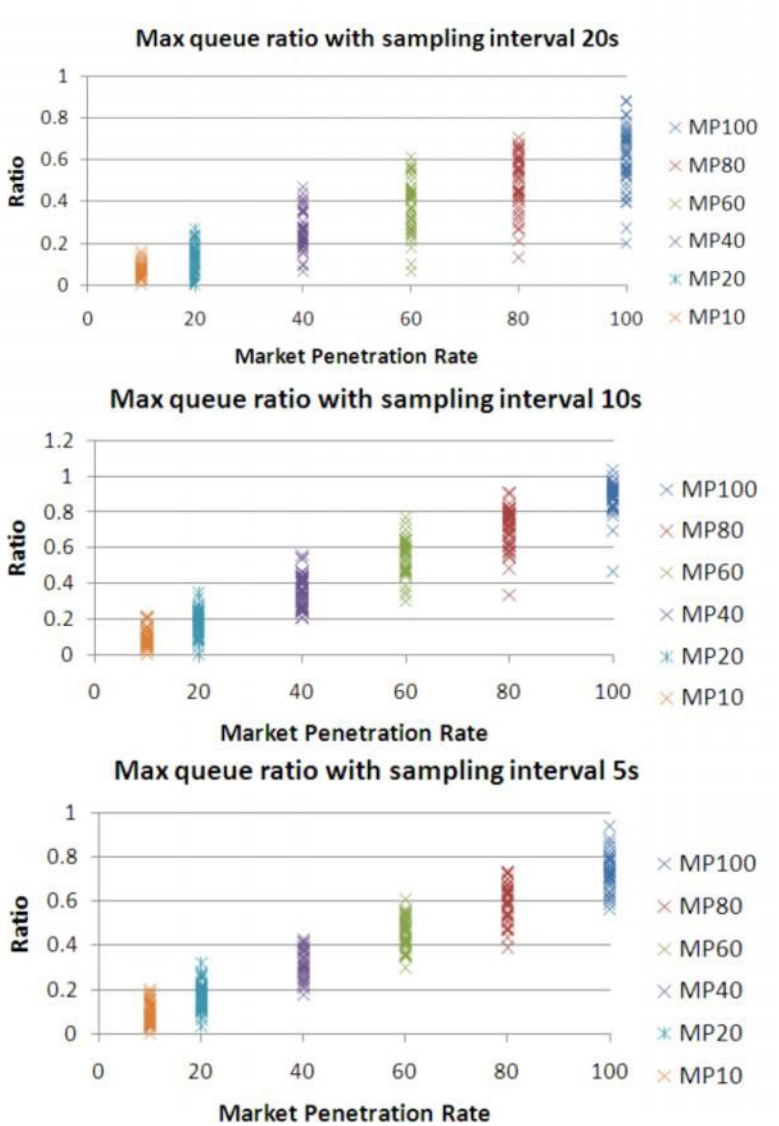


Figure 3.14 Ratios of Sampled to Actual Maximum Queue Lengths per Signal Cycle

Figure 3.15 shows the results for the estimates of queue length at the California Ave. intersection. The three uniform sampling strategies provide a good starting point for comparison, with the slowest update rate (20 s interval) clearly producing the poorest results. For this variable, the uniform sampling at 5 s update intervals and all the more sophisticated strategies are producing very similar results. The importance of market penetration rate is clearly visible here, showing the value of increased market penetration of probe vehicles in producing results with less random variation.

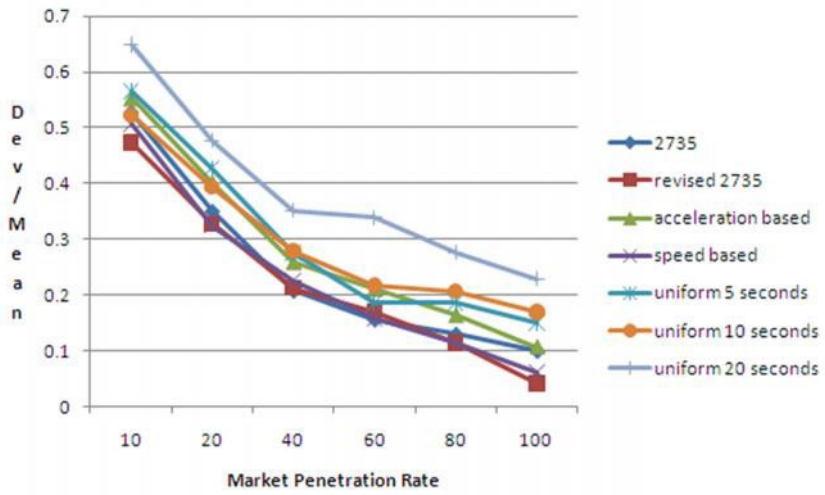


Figure 3.15 Coefficient of Variation--Maximum Queue Length vs. Alternative Sampling Strategies

Number of Stopped Vehicles

Similar to the approach for queue length estimation, we identify the number of probe vehicles that report a stopping event compared to the total number of vehicles that stop in the VISSIM simulation. The ratio gives an indication of the effectiveness of the probe sampling, and these results are reported for the fixed sample intervals of 5, 10 and 20 seconds.

The ratios of the stopping percentage for each sampled case compared to the baseline truth are shown in Table 3.6. These results show unbiased estimates for the sampling intervals of 10 sec and 5 sec, but at the 20 sec sampling intervals the stop ratios are seriously over-estimated because of the loss of samples from many of the vehicles that did not have to stop (they traversed the intersection in less than 20 s, so they were not sampled). These results also show that the standard deviations at the lower market penetrations are quite large, limiting the usefulness of the individual point estimates in these cases for real-time control.

Table 3.6 Ratios of Sampled to True % Stops vs. Market Penetration and Sampling Rate

Sample Updates	**Market Penetration**					
	10	**20**	**40**	**60**	**80**	**100**
20 s sample updates:						
Mean ratio	1.356	1.420	1.549	1.531	1.514	1.513
Standard deviation of ratio	0.751	0.472	0.446	0.421	0.487	0.456
10 s sample updates:						
Mean ratio	1.071	1.089	1.115	1.098	1.091	1.105
Standard deviation of ratio	0.587	0.380	0.249	0.202	0.222	0.223
5 s sample updates:						
Mean ratio	0.914	0.924	0.947	0.938	0.935	0.946
Standard deviation of ratio	0.528	0.325	0.196	0.133	0.127	0.113

3.6.3 Sampling Average Travel Speeds

The approach to sampling travel speeds is substantially different from the approach to sampling the other variables. Speed is a continuous variable associated with each vehicle, and each vehicle should be generating multiple speed samples as it traverses an intersection or a roadway segment. Because it is a continuous variable, it can be sampled more frequently than the other variables, while still remaining meaningful as a descriptor of traffic conditions. This more frequent sampling opens the possibility of getting updates on conditions at shorter intervals than a full signal cycle if there are enough equipped vehicles to produce sample snapshots.

The geographic scope for aggregating travel speed samples should be smaller than a full block or a full approach to a traffic signal, because speed can vary significantly within such a length of roadway. As illustrated in Figure 3.16, we subdivided the long block from Page Mill to California along El Camino Real into four segments to see how much difference we would find. If we had subdivided it into a larger number of segments the number of vehicles per segment would have been reduced proportionately, introducing a different problem with more frequent voids in the sampled data set.

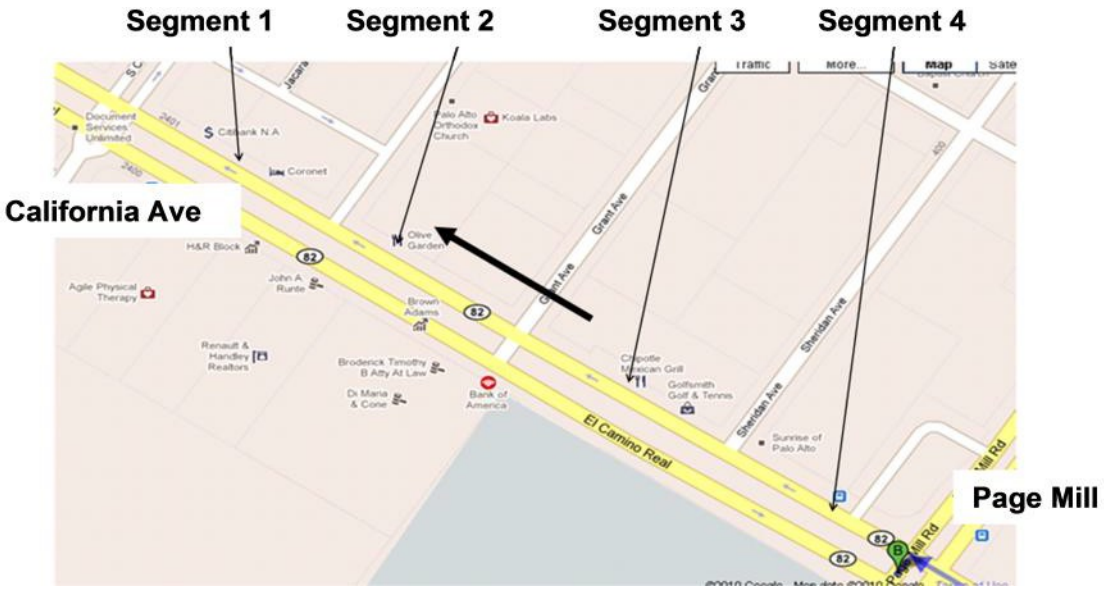


Figure 3.16 Probe Speed Sampling Segments from Page Mill to California Avenue

The relevant statistics for estimating speed within each segment are the mean and standard deviation of the speed estimates. When dealing with the non-uniform sampling intervals, it is important to weight the samples by the duration of time preceding the snapshot so that we do not under-represent the steady-state or slowly-changing conditions in the summary statistics. Figure 3.17 shows two examples of mean speed estimates from probe sampling for Segment 1, the closest segment to the California Ave. intersection on the northbound approach along El Camino Real from Page Mill Rd. Since this segment is closest to the intersection at California Ave., the impact of that signal's settings on the mean speed can be clearly seen.

The mean speed estimates from all the sampling strategies were generally quite accurate and unbiased, at around 17 m/sec (38 mph). The important differences that were identified in comparing the different probe sampling strategies were in the standard deviations of these speed estimates. Table 3.7 shows the RMS errors of the speed estimates for the four segments on the approach from Page Mill to California as a function of sampling interval and market penetration. These indicate the accuracy with which the probe samples can be used to estimate the actual speed in that section. Segment 1, the most important one

because it is closest to the traffic signal, has by far the largest RMS errors, even at the fastest sampling update rate. These errors are large because the underlying speed variations in this segment are large. This is where we have a mixture of vehicles queued while stopped on the red signal and vehicles cruising through at constant speed on the green signal.

Segment 3 has the smallest root mean square (RMS) speed errors because vehicles in this segment are most likely to be cruising at close to constant speed while traveling between intersections, while Segment 2 may still contain some vehicles queued at the intersection waiting for a green signal. Since the speeds in Segment 3 are almost constant, virtually any probe samples can represent the speed of the segment quite accurately, regardless of how infrequent they are. Segment 4 has slightly larger RMS speed errors because it has more speed variability than Segment 3, accommodating the accelerations from a stop of the vehicles that had to stop at the upstream intersection at Page Mill as well as the vehicles that were able to cruise through that intersection without stopping.

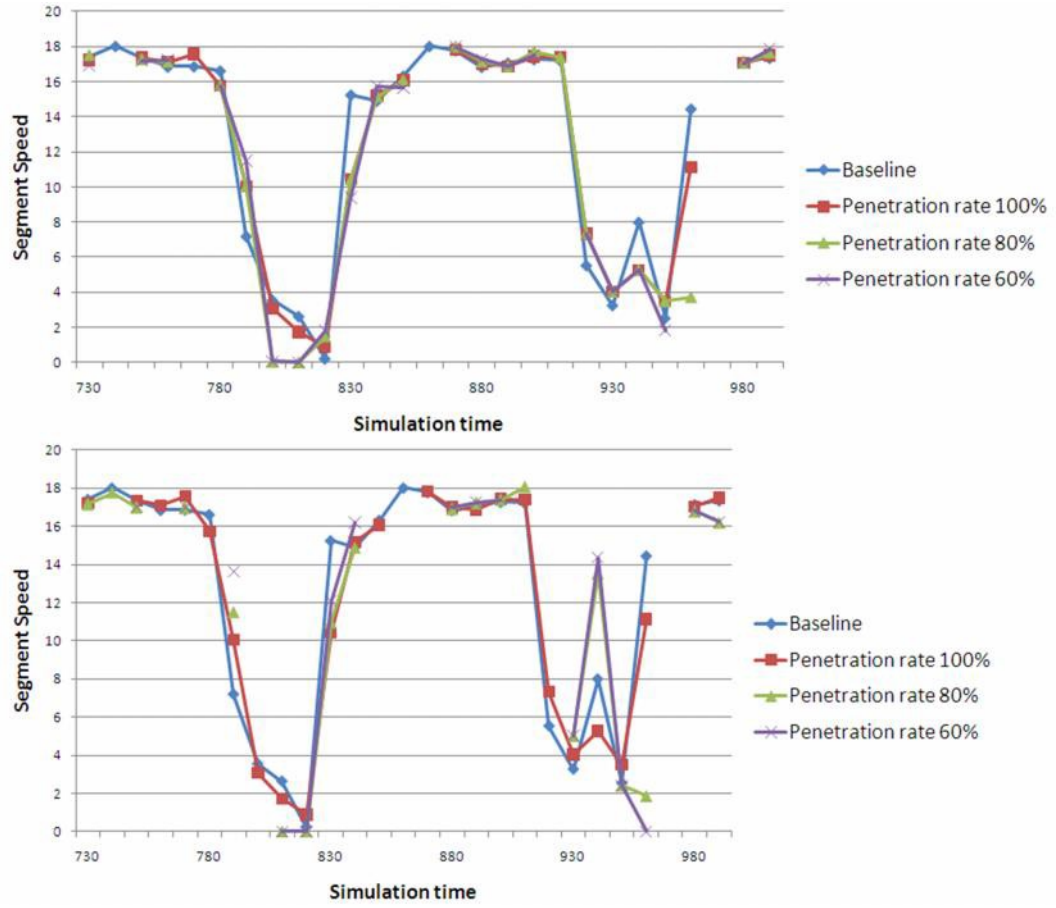


Figure 3.17 Mean Speed vs. Time on Segment 1 (California to Page Mill) at 5 sec and 10 sec Sample Intervals

Table 3.7 RMS Error of Speed Estimates at El Camino Real Segments (Page Mill to California Ave)

Segment #	Sample Interval (sec)	Market Penetration (%)					
		10	20	40	60	80	100
1	5	6.64	6.10	4.96	4.31	3.86	2.91
	10	7.04	6.57	5.80	5.10	4.70	3.85
	20	7.82	7.57	7.61	6.74	6.34	6.03
2	5	1.52	1.70	1.19	1.06	0.85	0.74
	10	1.81	1.67	1.51	1.63	1.43	1.36
	20	1.92	1.68	1.58	1.58	1.55	1.56
3	5	1.00	0.93	0.77	0.70	0.68	0.62
	10	0.91	1.13	1.08	0.89	0.85	0.78
	20	0.97	1.04	1.07	1.04	1.03	1.00
4	5	1.12	1.08	0.97	0.92	0.79	0.72
	10	1.26	1.22	1.14	1.10	1.02	0.90
	20	1.21	1.24	1.11	1.18	1.15	1.17

3.6.4 Communication Costs

These probe sampling strategies produced substantially different numbers of snapshots, and the comparison among strategies should factor in their efficiency, accounting for the burden that these place on the wireless channel. For the California Ave. intersection on El Camino Real, the number of snapshots and uploaded data packets generated by the northbound approaching vehicles during a two-hour simulated peak period at 100% vehicle market penetration are tabulated below (Table 3.8). This helps explain the poorer statistics for the 20 sec update interval and the similarities of many of the other strategies, which use similar numbers of snapshots. This comparison implies limited efficiency advantages for the more complicated sampling strategies, and points out the importance of the total number of probe snapshots in producing better estimates of traffic conditions.

Table 3.8 Communication Burden Imposed by Each Probe Sampling Strategy on One Major Approach During Two Hours at 100% Market Penetration

Sampling Strategy	Data Packets	Snapshots
Uniform at 20 s update interval	1809	2521
SAE J2735	2426	5425
Uniform at 10 s update interval	3796	6872
Modified SAE J2735	3943	7514
Acceleration based	4204	7725
Speed based	5744	12,461
Uniform at 5 s update interval	6015	14,252

Since the values shown in Table 3.8 apply to only the northbound traffic on El Camino Real, it would be necessary to double them to account for both directions of traffic. If we considered all directions of traffic at the busiest intersection on the corridor (El Camino at Page Mill) it would probably be necessary to double these values again. Even multiplying the values in Table 3.8 by a factor of four, it is important to note that the communication burdens associated with these data rates are well within the capabilities of DSRC technology. Even at the 5 sec update interval, the probe vehicles would be generating an average of about three data packets per second at the intersection, which is well within the capacity of the DSRC channel.

3.6.5 Summary of Testing Sampling Strategies

The results of this simulation-based probe sampling study have shown the challenges to effective use of probe sampling in real-time traffic signal control. These results show that there is not a clear "break point" with respect to market penetration where the probe sampling results become significantly better. None of the probe sampling strategies that were analyzed showed a clear superiority to the others, and none was able to provide high-fidelity estimates of the traffic conditions at the intersection at low to moderate market penetrations. Rather, there is a close to linear relationship between the market penetration and decreasing variability in the estimates of the MOEs. This means that we need to think carefully about how much variability in the estimates of the MOEs will be acceptable for each signal control strategy in order to determine what minimum probe vehicle market penetration will be needed to support that strategy.

The variability in the sampled data can be reduced by aggregating data from multiple signal cycles, but this introduces additional latency in the response of the traffic control system. This represents an additional trade-off in considering how to use probe data to implement real-time adaptive signal control.

3.7 Data Fusion

In this section, we propose and test an algorithm to improve the estimation accuracy of MOEs, specifically queue length from CV data by fusing the probe trajectory data and loop detector data. Detailed description of the work performed and the results are included in Appendix C.[4]

Data fusion of loop and probe vehicle data has been used to estimate travel times on freeways and arterial streets and other ITS applications [24].

3.7.1 Approach

First, we estimate the queue length using Kalman filtering technique on the probe vehicle data and loop detector data separately. Second, we solve the Kalman filtering formulations and set the weights as the reciprocal of the covariance in the Kalman filtering. Finally, the weighted combination of the queue length from the two data sources is the estimated queue length.

The weights are obtained from the covariance of the Kalman filtering methods that are based on the queuing shock wave theory. The queuing shock wave is generated when a vehicle joins the queue end, and the vehicle state is changed from arrival to jam. The weights are updated dynamically. The queue length from loop detector data is estimated based on an input-output model.

Kalman Filtering Approach for Probe Vehicle Data

The formal definition is as follows: L_n is a system state at step n that indicates the queue length to the stop bar (m); s_n is a system state at step n that indicates the queuing shockwave speed (m/s); d_n is the measured queue length to the stop bar obtained from the probe vehicle (m); ΔT_{n-1} is the time interval between steps *n-1* and *n* (s); and a_n,b_n, e_n, and f_n are noises with normal distributions. The state equations and measurement equations are as follows.

[4] Li Jing-Quan, K Zhou, S Shladover, and A. Skabardonis, "Estimating Queue Distance under the Connected Vehicle Technology: Using Probe Vehicle, Loop Detector, and Fused Data," paper 13-2274, 92nd TRB Annual Meeting, Washington DC, January 2013 (forthcoming Transportation Research Record)

State equations:

$$L_n = L_{n-1} + s_{n-1}\Delta T_{n-1} + a_{n-1} \quad (1)$$

$$s_n = s_{n-1} + b_{n-1} \quad (2)$$

Measurement equations:

$$d_n = L_n + e_n \quad (3)$$

$$\frac{d_n - d_{n-1}}{\Delta T_{n-1}} = s_n + f_n \quad (4)$$

Note that equations (1) and (3) are based on the relationship between the distance to the stop bar and shock wave speed; in equation (2), we assume that the shock wave at step *n* is roughly same as step *n-1*. Such an assumption is reasonable since there is generally no a significant change in the system. Equation (4) is used to determine the current shock wave speed.

Kalman Filtering Approach for Loop Detector Data

The formal definition is as follows: s_n is a system state at step *n* that indicates the queuing shockwave speed (m/s); q_n is the measured flows obtained from the advanced loop detector (vehicles/sec); k_n is the measured density (vehicles/meter); k_j is the jammed density (vehicles/meter); and g_n and h_n are sensor noises with normal distributions.

State equations:

$$s_n = s_{n-1} + g_{n-1} \quad (5)$$

Measurement equations:

$$\frac{q_n}{k_n - k_j} = s_n + h_n \quad (6)$$

Note that flow q_n can be estimated as the average vehicle flows passing the loop detector. Density k_n can be estimated using the flow and space mean speed. In this study, we tested our method using the data generated from the VISSIM simulation. We then used the free flow speed instead, since it is unclear if the simulation data can accurately represent the loop detector occupancy.

The Kalman filtering equations for the probe vehicles (1 to 4) and equations for the loop detectors (5 and 6) can be solved using the classical iterative process. Covariance can be obtained during the iteration. The matrix operation library Eigen 3.1 is used to solve the Kalman filtering formulations.

3.7.2 Application

We tested the proposed approach on the El Camino Real simulation testbed described in the previous sections. The intersection approach at El Camino Real and Grant Rd intersection was selected. A one-hour simulation was run and second-by-second vehicle trajectory data were obtained. The ground truth queue length was obtained from the second-by-second vehicle trajectory data under 100% market penetration rate. Equipped vehicles were randomly selected from the simulated trajectories.

In order to evaluate the estimation accuracy under different penetration rates of CV vehicles, we first determine the maximum queue length within each cycle. Then, we compute the Mean Absolute Percentage Error (MAPE) from the estimated queue length and the ground-truth queue length:

$$\frac{1}{n}\sum_{n}\left|\frac{estimated\ value - ground\ truth}{ground\ truth}\right|$$

where n is the number of cycles studied. Figure 3.18 shows the computed MAPE under different penetration rates. As expected, MAPE increases with the decrease in the penetration rate. However, our event-based estimation algorithm which uses probe data and signal timing data performs much better than our approach in section 3.6.2 that simply scaled the queue size using the penetration rate. For example, in our case studies, when the penetration rate is 50%, MAPE is less than 18%, which is far less than 50%. MAPE is around 60% even with a very low penetration rate, 10%. The estimation improvement is mainly due to exploiting the advantage of probe trajectory data: even if there is only one probe vehicle, its stopped location provides a lower bound on the real queue length. If this probe vehicle happens to be in the end of the queue, the estimation error is small.

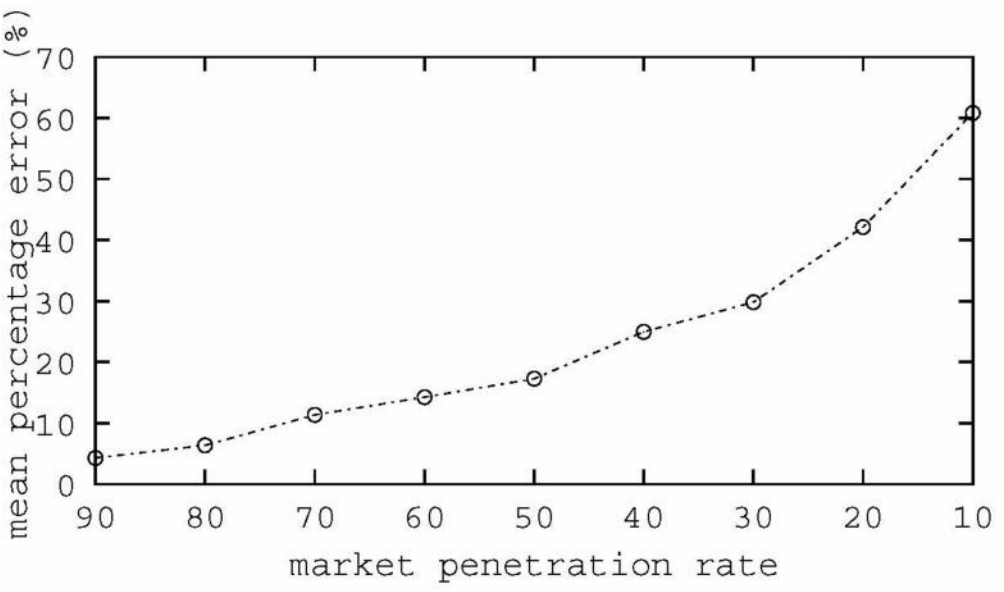


FIGURE 3.18 MAPE of Maximum Queue Length vs. Penetration Rate

We then estimated the queue length using the input-output model with loop detector data. The MAPE with loop detector data is 9.1% in our case studies. Such error rate falls between the rates with CV penetration rates 70% and 80%. Note however that simulated loop detector data represent ideal conditions; real-world loop detector data often suffer from data losses and errors that degrade the estimation accuracy of queue lengths under real-world conditions.

The data fusion method is based on weighted combination of loop and probe trajectory data. If the performance of one method is much better than the performance of the other one, it is unexpected that the fusion algorithm can outperform the better method. For example, if the penetration rate is close to 90%, MAPE is less than 5%, which is already very good. Fusing loop detector data may not be needed. When the penetration rate is close to 10%, it is desirable to only use the input-output model with loop detector data. However, when the penetration rate is 70%, the input-output model with loop data and event-based method with probe trajectory data result in similar performance. We run the data fusion algorithm and the resulting MAPE is 5.96%, which is slightly improved from 9.1% with the input-output model and is fairly improved from 11.35% with 70% penetration rate. Figure 3.19 shows the maximum queue length estimated with different methods within certain cycles. We can see that the data fusion algorithm balances

the underestimate and overestimate from the input-output method and event-based method, thereby improving the estimation accuracy.

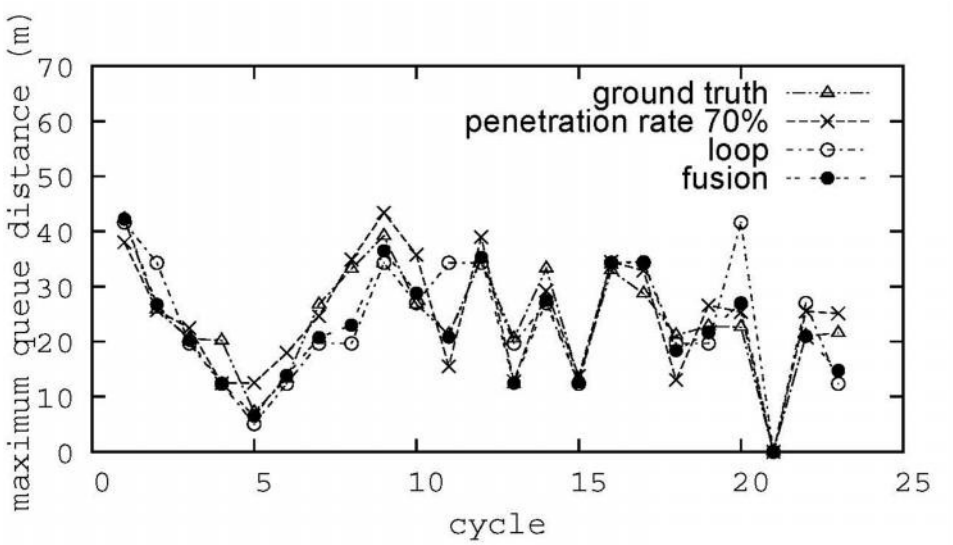


FIGURE 3.19 Estimated Maximum Queue Length with Different Data Sources

The traditional data fusion methods may not work well when either the method with probe trajectory data or the method with loop detector data is superior to the other one. In such situation, it is desirable to use the data source with the superior performance as the primary source, and utilize the other source to support the primary data source.

CHAPTER 4

CONTROL STRATEGIES FOR MOBILITY: QUEUE SPILLBACK AVOIDANCE

4.1 Introduction

When traffic demand exceeds the capacity at signalized intersections, results in growing queues. If these saturated traffic conditions persist for a long time, the queues will eventually spillback, blocking upstream intersections. Depending on the demand pattern and signal settings of the intersections, the impact of queue spillbacks could vary from small reductions in the upstream intersection's capacity to a complete blockage of any vehicle discharge during the green time interval, a situation also known as *de facto red* [25]. Consequently, queue spillbacks can lead to excessive delays and gridlock on signalized networks. A previous study observed that under the occurrence of spillbacks delay can increase by 50-100% for short links [26].

This Chapter describes the work performed to detect queue spillbacks at signalized intersections and develop signal control strategies to mitigate their occurrence, based on Connected Vehicle (CV) data. Detailed description of the work performed and the results are included in Appendix C.[5]

While CV data have been recently used in queue length estimation, very few studies have focused on queue spillback detection and development of signal control strategies to manage such conditions. Venkatanarayana et al. [10] used *CV* data to develop real-time traffic signal control strategies for oversaturated traffic conditions. The queue length estimation was based on the position of the last equipped vehicle in queue and the signal control strategies consisted of adjustments of the green time and/or offsets at the upstream intersection every time the end of the estimated queue reached a certain threshold. The optimal end of queue thresholds used, were determined through simulation for the different signal control strategies that were tested [27]. These algorithms were based on a fixed end of queue threshold, essentially operating as when information is provided from fixed location detectors and were not adjusted for different penetration rates of the equipped vehicles. He et al. [28] developed a platoon-based adaptive signal control system that uses information from vehicle-to-infrastructure communications to adjust the signal settings in arterial signalized networks in real time considering different types of vehicles. At each intersection, the system accounts for the capacity of the downstream links in order to determine signal settings that prevent the occurrence of spillbacks. The system is complex which leads to high computation times that are prohibitive for real-world applications, especially when oversaturated traffic conditions prevail.

4.2 Approach

The detection of the potential occurrence of a spillback on a link for a signal cycle is divided into three steps: 1) obtaining the stopping position of the last *CV*-equipped vehicle that joined the queue in that cycle, 2) determining the critical queue length that indicates the possibility of a queue spillback within a cycle (a value that we will refer to as ideal queue length threshold), and 3) combining the results of steps 1 and 2 with other sources of information (i.e., the *CV* market penetration rate and signal settings) to determine if a spillback could potentially occur within the next cycle. The initial queue length estimation on the subject link is performed with the Maximum Likelihood Estimation (MLE) method. As described in Section 4.2 the MLE method provides more accurate queue length estimates than other more complicated estimation methods. While step 1 is straightforward, step 2 requires considering different

[5] Christofa E, J.Argote, and A. Skabardonis, "Arterial Queue Spillback Detection and Signal Control based on Connected Vehicle Technology," paper 13-4903, 92nd TRB Annual Meeting, Washington DC, January 2013 (forthcoming Transportation Research Record)

inputs such as the link geometry and the traffic demand. The methodology to determine the ideal queue length threshold is presented in the next subsection. After that, two algorithms are developed that combine the results of steps 1 and 2 to detect potential queue spillbacks.

4.2.1 Ideal Queue Threshold Estimation

The ideal queue threshold, L_{lim}, is the maximum acceptable queue length that ensures that a queue spillback will not occur within the next cycle. This parameter can be estimated based on the expected number of vehicles to accumulate in the critical link during the next cycle. The critical link is defined as the link directly upstream of the critical intersection. This is the link that is more likely to suffer from queue spillbacks in the arterial of interest, and therefore the focus of our analysis. The ideal queue threshold for that link can be obtained as:

$$L_{\text{lim}} = L - max\left\{F, \frac{Q}{NK_j}\right\} \quad (1)$$

where L is the link length (in m), F is a safety factor (in m) as described in (*23*), Q is the expected number of vehicles to accumulate in the critical link during the next cycle, N is the number of lanes of the critical link (we assume that the vehicles will distribute themselves homogeneously across lanes), and K_j is the jam density per lane (in veh/m) on the link, typically corresponding to the inverse of the spacing of stopped vehicles (e.g., 7 m/veh). The parameter F guarantees that the link will always have extra space. This safety factor tries to compensate for possible measurement errors and the stochasticity in arrivals. For simplicity, we have assumed a constant value of F equal to three times the spacing of stopped vehicles per lane ($3/K_j$). The term Q can be obtained through a simple input-output analysis:

$$Q = q_{CV}^{u} C/p - \sum_{i=1}^{I_{ca}^{d}} s_i^d G_i^d, \quad (2)$$

where the superscripts $\{u, d\}$ denote upstream and downstream respectively. q_{CV}^{u} is the total flow of *CV*-equipped vehicles in the direction of interest entering the critical link from the upstream intersection (in veh/hr, obtained using a moving average over the last five cycles), C is the cycle length (in hr), which is assumed to be common among all intersections in the arterial segment of interest, p is the *CV* market penetration, which is assumed to be known in this study, but can also be estimated based on *CV* measurements as in (*19*), s_i^d is the downstream saturation flow for movement i (in veh/hr) and G_i^d (in hr) the effective green time for phase i serving movement i at the critical approach, and I_{ca}^{d} is the set of phases at the critical approach.

4.2.2 Gap-based Potential Spillback Detection

When the *CV* market penetration, p, is lower than 100%, the last *CV*-equipped vehicle that is stopped and identified as the back of the queue by the MLE algorithm may not be the last vehicle in that queue. Assuming that the ideal queue threshold, which is also the maximum acceptable queue length, L_{lim}, is calculated as previously explained, we need to determine if the real queue spills back to that location. The gap-based spillback detection method is based on the estimation of the distance between the last *CV*-equipped stopped vehicle that is detected and the actual last vehicle in queue, which we term gap length, X. Figure 4.1 provides a graphic illustration of the ideal threshold value, L_{lim}, and gap length, X, for a case where $X = 4/K_j$.

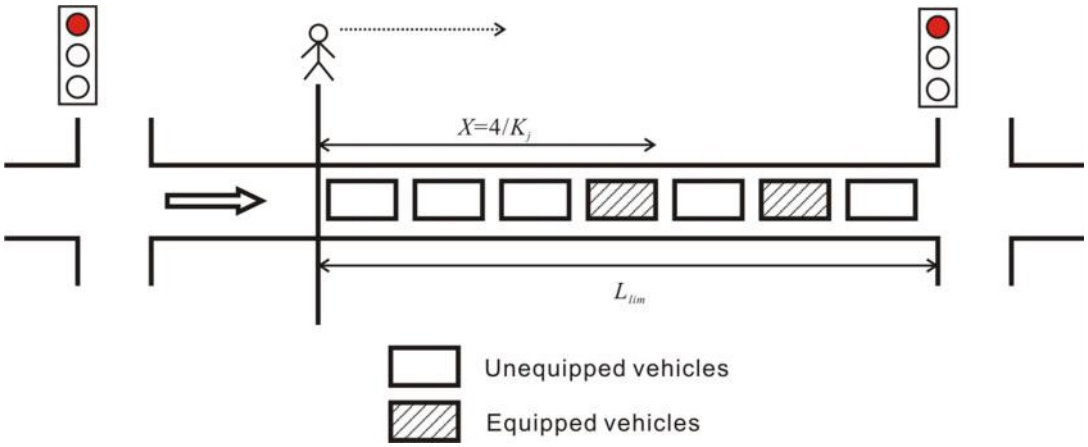


Figure 4.1 Illustration of the Ideal Threshold Value, L_{lim}, and Gap Length, X

X is a random variable that models the number of non-equipped vehicles that join the queue immediately after the last *CV*-equipped in queue for a given cycle. This variable follows a truncated geometric distribution for which the probability of a success corresponds to the penetration rate of the *CV* technology (the probability of sampling a *CV*-equipped vehicle). Based on that variable, we seek the gap length value $X(p) = \min_x \{x \mid \Pr(X \le x \mid L_{\text{lim}}) \ge 1-\alpha\}$, given a false negative rate α. Note that the term negative refers to the situation where a queue does not grow past the threshold, while positive corresponds to the opposite situation. Therefore, if the last *CV*-equipped stopped vehicle detected on a given cycle is within X from the ideal queue threshold, L_{lim}, this method will trigger a potential spillback detection alert. For a given ideal queue threshold, L_{lim}, the distribution of X follows the probability mass function below:

$$\Pr(X = x|L_{\text{lim}}) = \begin{cases} p(1-p)^{NK_j x} & \text{when } 0 \le x < L_{\text{lim}} \\ (1-p)^{NK_j L_{\text{lim}}} & \text{when } x = L_{\text{lim}}. \end{cases} \quad (3)$$

Note that the second case corresponds to cycles for which no *CV*-equipped vehicle is sampled within a queue of length L_{lim}. If we ignore the incidence of the cycles where no *CV*-equipped vehicle is detected, the gap length, X, can be estimated as follows:

$$\Pr(X \le x|L_{\text{lim}}) = 1-(1-p)^{(NK_j x)} \ge 1-\alpha, \quad (4)$$

which results in:

$$NK_j x \ge \frac{\ln(\alpha)}{\ln(1-p)}, \quad (5)$$

and therefore, the minimum gap length $X^*(p)$, is given by the following expression:

$$X^*(p) = \min\left\{L_{\text{lim}}, \frac{1}{NK_j}\left\lceil \frac{\ln(\alpha)}{(\ln(1-p))}\right\rceil\right\} = \min\{L_{\text{lim}}, X(p)\}, \quad (6)$$

where $\lceil x \rceil$ is the smallest integer that is greater than or equal to $x \in \mathbb{R}$. In cases with very low *CV* penetration rates, it is possible to have multiple consecutive cycles in which no *CV*-equipped vehicle is detected. In such cases the minimum gap lengths, $X^*(p)$, are equal to L_{lim}, since the gap lengths $X(p)$ are greater or equal to L_{lim}, and the presence of a single stopped *CV*-equipped vehicle within the link would trigger a spillback detection alert. Therefore, ignoring the number of cycles that have passed since the last *CV*-equipped vehicle was detected, denoted by $n \in \{1,2,...\}$, could lead to an overestimation of potential spillbacks. To avoid this issue, the minimum gap length is adjusted dynamically to account for every

cycle that passes during which no *CV*-equipped vehicle has been detected. For each of those cycles the gap length, $X(p)$, is reduced by an amount equal to the length in queue of the number of vehicles that are served by the downstream intersection in one cycle. Thus, in this case, the dynamically adjusted minimum gap length used by the algorithm is:

$$X^*(n,p) = \text{mid}\left\{0, X(p) - (n-1)\frac{\sum_{i=1}^{I_{ca}^d} s_i^d G_i^d}{NK_j}, L_{\text{lim}}\right\} \quad (7)$$

Figure 4.2 depicts how the minimum gap is dynamically adjusted and shows the position of the first *CV*-equipped stopped vehicle that would trigger a spillback detection alert. The dynamically adjusted minimum gap length $X^*(n, p)$ may vary in every cycle. If a vehicle was detected in the previous cycle, $X^*(n, p)$ will be equal to $X^*(p)$, which is given by equation (6). On the other hand, its value could decrease for every successive cycle in which no *CV* is detected, until $X(p) - (n-1)\sum_{i=1}^{I^{ca}} s^d_i G^d_i / NK_j < 0$. In that case, $X^*(n, p) = 0$ and the detection of a single *CV*-equipped vehicle that is stopped in the critical link downstream of the ideal threshold will trigger the detection of a potential spillback.

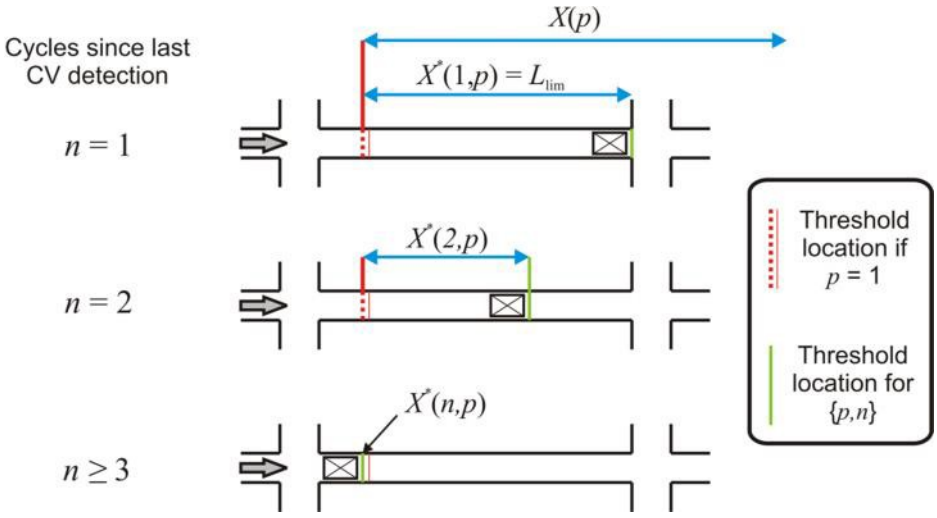


Figure 4.2 Illustration of the Effect of *n* on the value of the Min Gap Length, $X^*(n,p)$, for a link for which $X(p) > L_{\text{lim}}$, $0 < X(p) - \frac{\sum_{i=1}^{I_{ca}^d} s_i^d G_i^d}{NK_j} < L_{\text{lim}}$, and $X(p) - 2\frac{\sum_{i=1}^{I_{ca}^d} s_i^d G_i^d}{NK_j} < 0$

4.2.3 Shockwave-based Potential Spillback Detection

This method takes into account additional information on the signal settings of the upstream intersection in detecting a potential queue spillback. If the beginning of the red phase for the critical direction at the upstream signal is known, it is possible to use the position of the last *CV*-equipped vehicle in queue to determine the maximum potential queue length using kinematic wave theory. If the upstream signal has only two phases, the position of the last vehicle in queue for a given cycle can be estimated as shown in Figure 4.3.

The detection algorithm formulation will vary depending on whether the initial phase at the upstream signal is the through or turn phase, denoted respectively with the superscripts $\{th,tu\}$. However, in both cases the logic of the algorithm is the same. The stopping time and location of the last *CV*-equipped vehicle served by the upstream intersection (t_{cv},y_{cv}) is used to project the traffic shockwaves that arise due

to vehicles joining the queue. This projection is then used to estimate where the last potential vehicle in the queue could stop.

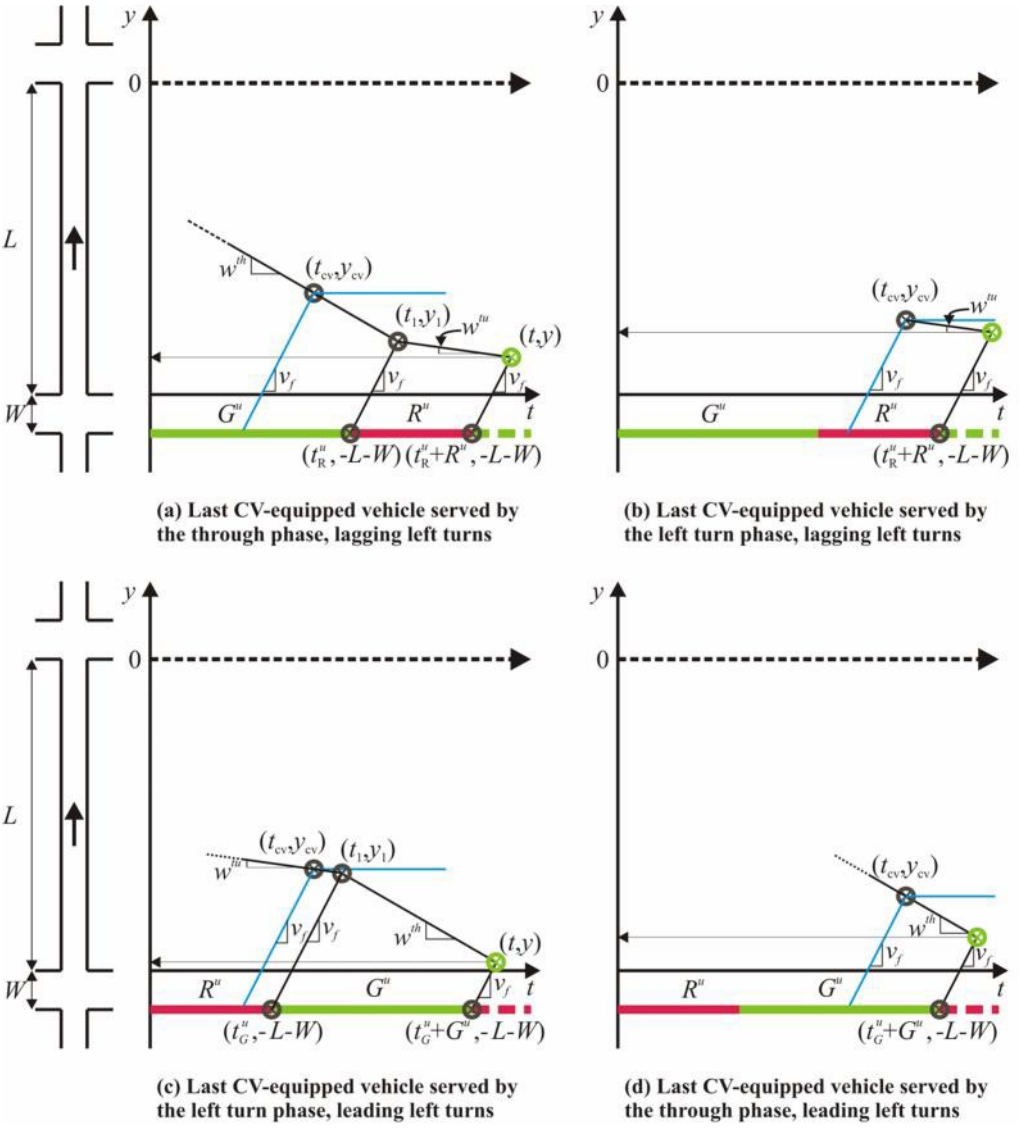


Figure 4.3 Shockwave-based Potential Spillback Detection Method

For the sake of brevity, let us assume in this section that the upstream intersection in the critical link presents a lagging left-turn phase. Then, the first step of the algorithm consists of determining whether the last *CV*-equipped vehicle was served by the upstream through or turn phase. This can be easily achieved since the *CV* technology allows us to follow the trajectory of that vehicle. Thus, it is possible to know when that vehicle crossed the upstream intersection with a maximum error equal to the time interval between consecutive position measurements, which are assumed to be obtained every second.

If the last *CV*-equipped vehicle that is detected in the cycle enters the link during the upstream through phase then the time and position where the last vehicle is projected to stop, (t, y), can be calculated using kinematic wave theory as follows:

$$(t, y(t)) = \left(\frac{y_1 + (L+W) + v_f(t_R^u + R^u) - w^{tu} t_1}{v_f - w^{tu}}, y_1 + w^{tu}(t - t_1) \right), \quad (8)$$

where:

$$(t_1, y_1(t_1)) = \left(\frac{y_{cv} + (L+W) + v_f t_R^u - w^{th} t_{cv}}{v_f - w^{th}}, y_{cv} + w^{th}(t_1 - t_{cv}) \right). \quad (9)$$

W is the intersection width, w^{th} and w^{tu} are the shockwave speeds for the through and turn phases (note that w^{th} should also account for the cross-street vehicles that turn right in red), v_f is the arterial free-flow speed, t_R^u is the beginning of red time for the through phase at the upstream intersection. Note that if the last *CV*-equipped vehicle enters during the through phase we need to account for cars that could enter during the remaining phase time and then the turning cars.

In the case that the last detected vehicle enters during the turn phase, obtaining (t, y) is simpler since we only need to account for the vehicles turning onto the critical link during the remaining green time for the turn phase, and is calculated as follows:

$$(t, y(t)) = \left(\frac{y_{cv} + (L+W) - v_f(t_R^u + R^u) - w^{tu} t_{cv}}{v_f - w^{tu}}, y_{cv} + w^{tu}(t - t_{cv}) \right). \quad (10)$$

Assuming that the link traffic can be described by a triangular fundamental diagram, the values of w^{th} and w^{tu} can be obtained based on the *CV* through and turning flows, q_{CV}^{th} and q_{CV}^{tu}, respectively. The sum of these two flows is equal to the total flow from the upstream intersection, q_{CV}^u, which was previously used in the ideal queue threshold calculation. Note that the arrivals of through vehicles occur during a phase of duration G^u and the turning vehicles join the queue during a phase of duration R^u. However, these flow values (q_{CV}^{th} and q_{CV}^{tu}) are obtained considering that through and turning vehicles join the queue continuously throughout the signal cycle. Thus, we need to consider correction factors to get the appropriate w^{th} and w^{tu} values:

$$w^{th} = \frac{q_{CV}^{th} \frac{C}{pG^u}}{(q_{CV}^{th} \frac{C}{pG^u})/v_f - NK_j}, \quad (11a)$$

$$w^{tu} = \frac{q_{CV}^{tu} \frac{C}{pR^u}}{(q_{CV}^{tu} \frac{C}{pR^u})/v_f - NK_j}. \quad (11b)$$

These two parameters can be used to solve equations (8), (9), and (10). Finally, the algorithm just compares the position of the maximum predicted queue length y with the ideal threshold L_{lim}. If $|y| > L_{\text{lim}}$, a potential spillback alert is triggered. The shockwave-based method does not depend on the offsets of the signals between the two intersections, because it focuses on the position of the back of the queue as estimated based on the last detected stopped vehicle and not on the queue dissipation mechanism.

4.3 Proposed Signal Control Strategy

Once a potential queue spillback is detected in an arterial signalized network, an alternative signal control strategy is activated to limit the growth of the target queue and prevent queue spillbacks during the next cycles. The proposed signal control strategy is based on the concept of metering vehicles at the

intersection immediately upstream of the critical one. We assume that the critical intersection has already been optimized based on some pre-defined criteria, and no changes in signal settings (splits, offsets, cycle length) can be applied to that critical intersection. The algorithm adjusts the signal timings at the upstream intersection taking into consideration the arrivals of all through, left, and right turning vehicles at both the critical and upstream intersections.

The level of metering (i.e., the reduction in green time for the critical phases, which are the upstream intersection signal phases contributing traffic to the critical link and being subject to green time reduction) is determined by the number of vehicles served during the cycle under consideration at the critical intersection. This means that the green time for the critical phases are reduced to allow into the link as many vehicles as the number that can be served by the downstream intersection during that cycle. In mathematical terms, the total reduction in green time for the upstream critical phases can be estimated using the following expression:

$$\sum_{i \in I_c} G_i^u s_i^u = \sum_{i=1}^{I_{ca}^d} G_i^d s_i^d, \tag{12}$$

where G_i^u is the green time for the phase that serves movement i and s_i^u is the saturation flow for movement i at the upstream intersection, I_c is the set of critical movements at the upstream intersection, and G_i^d, s_i^d, and I_{ca}^d are the same parameters as defined in equation (2).

In cases that the flows of interest at the upstream intersections (through from the main arterial, and left and right from the two cross streets) are unbalanced in their demands, the green time is reduced for the heaviest traffic movement, which is now considered to be the only critical movement. If there is significant amount of traffic from the other movements joining the critical link queue, then the green time for the heaviest movement should be further reduced. In this way, the critical link will be able to accommodate traffic from the non-critical movements. Assuming that the extra green time allocated to the non-critical phases does not significantly affect the amount of vehicles entering the critical link, the green time allocated to the critical phase, $G_{i_c}^u$, can now be expressed as follows:

$$G_{i_c}^u s_{i_c}^u = \sum_{i=1}^{I_{ca}} G_i^d s_i^d - \sum_{i=1}^{I_{nc}^u} C q_{CV,i}^u / p, \tag{13}$$

where $s_{i_c}^u$ is the saturation flow for the heaviest upstream movement, $I_{i_{nc}}^u$ is the set of upstream non-critical movements which contribute flow to the critical link but whose green time is increased and $q_{CV,i}^u$ is the average *CV*-equipped vehicle flow for each movement i coming from the upstream intersection.

The proposed strategy ensures that the queue will not increase over the next cycle preventing imminent queue spillbacks. However, if this strategy is only applied at a single link and there is not sufficient queue storage capacity at upstream links, the spillback problem could transfer to another part of the network, causing long queues at intersections located further upstream. When these long queues are likely to cause spillbacks, the proposed signal control strategy can be extended to their respective upstream intersections so that queues are stored on longer links and do not cause intersection blockages at other parts of the network.

4.4 Application

The proposed spillback detection methods and metering signal control strategy have been tested through simulation on a four-signal segment of San Pablo Avenue, a major arterial in Berkeley, California shown in Figure 4.4.

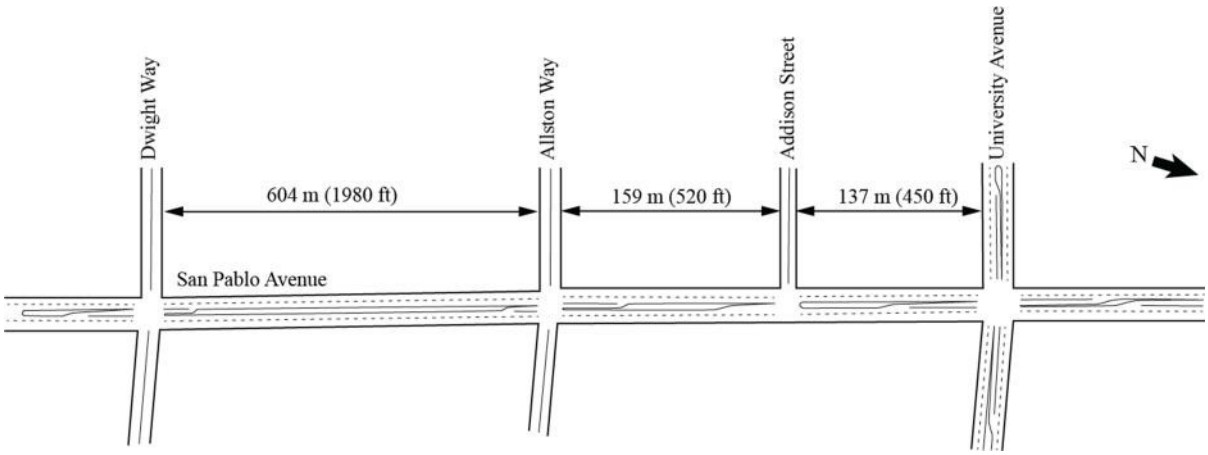


Figure 4.4 San Pablo Avenue Test Site

The heaviest traffic direction for the evening peak period is the northbound one and that was chosen as the focus for the tests performed in this study. All signals operate as fixed-time coordinated with a common cycle length of 80 seconds. Figure 4.5 shows the phasing and signal settings for all intersections during the evening peak period. The intersection of University and San Pablo Avenues is the critical intersection.

The signal settings and demands used in the simulation are based on field data collected by the research team. However, the tests have been performed with entry demand values for the northbound direction that are 3, 2, and 1.5 times higher than the original demand for the first, second, and third 15-minute interval and equal to the original demand for the last 15-minute time interval. This allows the queues in the uncontrolled scenario to grow until creating spillbacks and then dissipate by the end of the simulation period. The rest of the entry volumes have been kept constant throughout the whole hour.

The simulation tests were performed using the AIMSUN microscopic simulation model [29] through emulation-in-the-loop simulation. This technique consists of connecting the algorithm for the proposed strategy with the simulator through its Application Programming Interface (API).

The simulation tests were run for one hour with a warm-up period of 10 minutes and a time-dependent demand profile that led to oversaturated traffic conditions and spillback formation upstream of the critical intersection. The following *CV* market penetration levels were tested, $p \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75, 1.00\}$, in order to determine the accuracy and effectiveness of the spillback detection methods and the alternative signal control strategy. Each penetration rate scenario and method tested was run 10 times and the results presented are the averages of those 10 replications.

The performance measures to evaluate the proposed strategies include average delay, number of stops, and maximum queue length for the critical link, its upstream main arterial link and the cross-street link at the upstream intersection. Travel time and average number of stops are also analyzed for the entire four signal arterial segment in both the northbound and southbound directions.

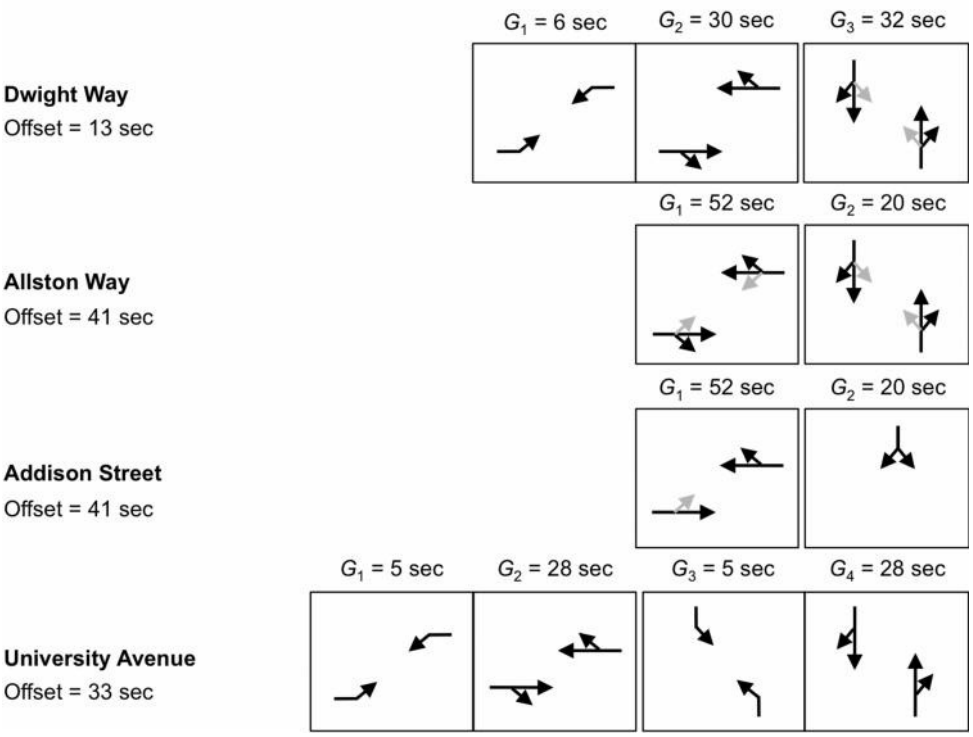


Figure 4.5 San Pablo Avenue: Signal Phasing and Timing

4.4.1 Potential Spillback Detection Results

The two queue spillback detection methods described in Sections 4.3.2 and 4.3.3 were evaluated through simulation based on the percentage of cycles that a potential queue spillback was correctly detected. In order to achieve a complete evaluation of the detection algorithms' effectiveness for different penetration rates, the results are presented as percentages of cycles that fall under one of the following categories:

- correct potential spillback detection,
- correct non-potential spillback detection,
- false positive or false alarm (type I error): potential spillback situation detected when it was not true,
- false negative (type II error): potential spillback present but not detected,
- no CV detection in the last cycle (used only for the shockwave-based detection algorithm).

The results from testing the gap-based queue spillback detection method shown in Figure 4.6 indicate that it performs well for penetration rates of 20% and higher. For example, for a 20% penetration rate the method presents correct results for about 82% of the cycles. For lower penetration rates, the method overpredicts the number of potential queue spillbacks. This is consistent with the gap-based estimation method design, since in order to limit the actual occurrence of spillbacks, we seek limiting the type II error (false negative) to less than 5%, which is the negative false rate, α, used. However, the results show that for low penetration rates our approach could be excessively conservative, even after dynamically changing the gap length based on the number of cycles without a *CV*-equipped vehicle detection.

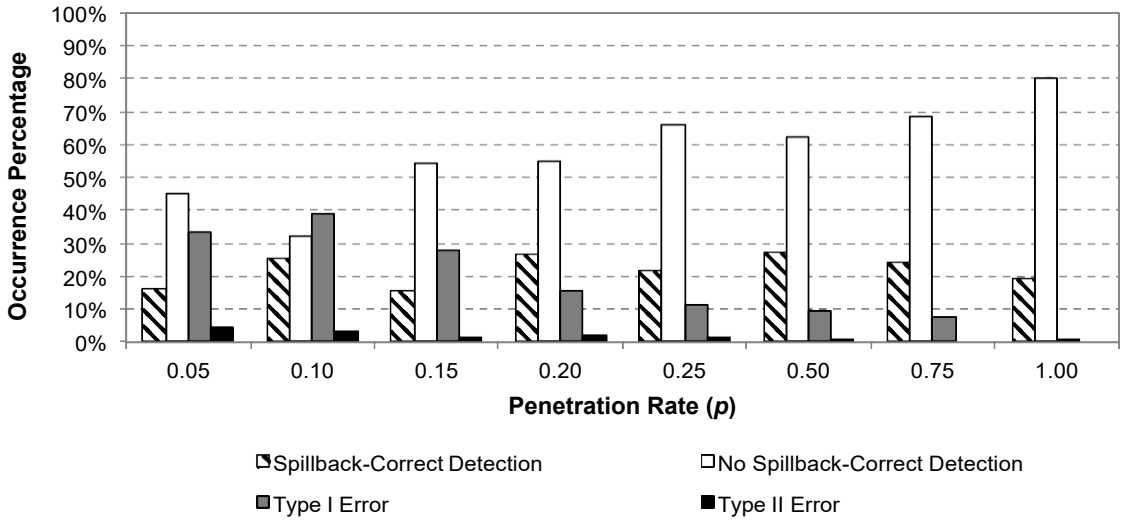


Figure 4.6 Evaluation of Gap Based Spillback Detection Method

The shockwave-based spillback detection results are shown in Figure 4.7. A free flow speed, v_f, of 60 km/hr (37 mph) was used. The shockwave-based method outperforms the gap-based for penetration rates in the range of 10 to 20%. In those cases the utilization of information on the signal settings of the upstream intersection and the assumed vehicle dynamics in addition to the position of the last *CV*-equipped vehicle in queue proves to be useful. However, for higher penetration rates, the simpler gap-based method is clearly better. This occurs because of the assumption of constant back of the queue wave speeds. The higher the penetration rate, the more likely it is that the last vehicle detected is the actual last vehicle in the queue. However, if we assume that the queue formation process lasts until the end of the cycle, errors of type I could be common. This results in a biased estimation even with 100% penetration rate. Nonetheless, with penetration rates as low as 15% this method is capable of correct detection in approximately 80% of the cycles.

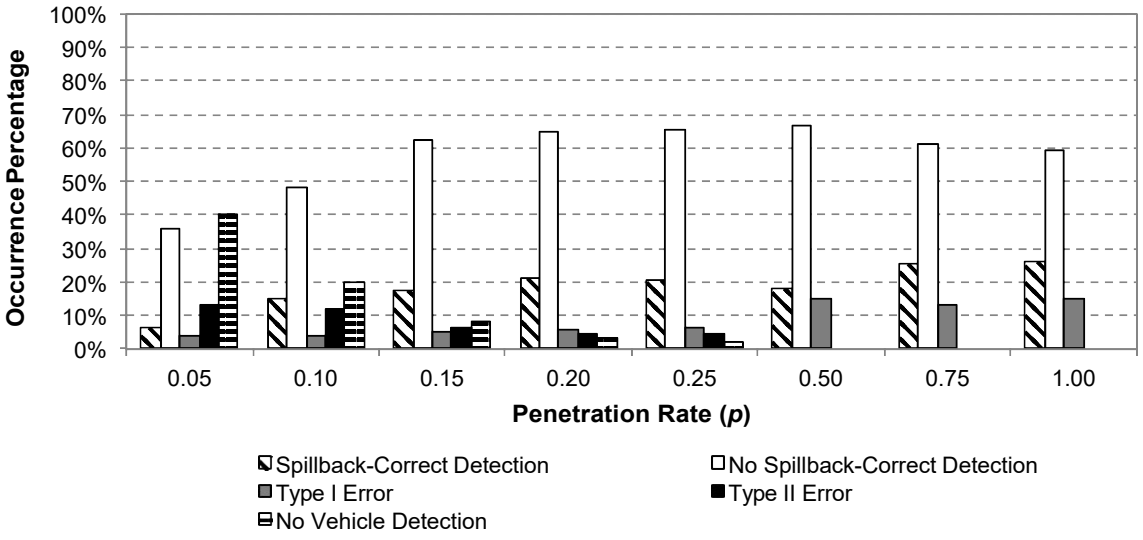


Figure 4.7 Evaluation of Shockwave-Based Spillback Detection Method

4.4.2 Alternative Signal Control combined with Gap-based Potential Spillback Detection

The gap-based spillback detection method is used to detect potential queue spillbacks and trigger the alternative signal control strategy because it resulted in the highest success percentage for a larger range of *CV* market penetration. Moreover, since the turning traffic (only left) arriving at the critical link is very small compared to the through traffic (on the order of 4%), the algorithm reduces the green time only for the through movement.

The results of the simulation tests indicate that for high penetration rates the proposed signal control strategy can effectively avoid the occurrence of spillbacks on the critical link. This is shown by the reductions achieved in the average delay and number of stops at the critical link as well as in the delays for the cross-street vehicles when compared with the performance of fixed signal settings (Tables 4.1 and 4.2). Note that the maximum queue length for the critical link does not present significant changes when the proposed signal control strategy is in place compared to the fixed signal settings scenario. This occurs because the proposed signal control strategy aims at maintaining the queue length at a constant value, rather than minimizing it. The signal control strategy's effectiveness in achieving this goal is shown by the reduced maximum queue length's coefficient of variation, which ranges between 0.06 and 0.09 with the proposed signal control strategy and has a value of 0.16 when fixed signal settings are in place.

It is also shown that for very low penetration rates (e.g., 10%) the alternative signal control strategy may be triggered in an inconsistent manner resulting in worse performance for the upstream of the critical link and the critical link itself. In addition, Table 4.2 shows that the link right upstream of the critical one suffers from higher average delays, maximum queue length, and number of stops, as it accumulates all the extra vehicles that are now prohibited to enter the critical link. The improvements observed at the cross-street link can be attributed to both the allocation of the extra green time to that approach and to the fact that spillbacks do not block the intersection as often.

Table 4.1 Queue Spillback Signal Control Evaluation--Critical Link

	Average Delay (sec/veh)	Maximum Queue Length (veh)	Number of Stops (#/ veh)
Fixed Signal Settings	37.11	13.36	0.93
p = 10%	43.98	15.77	1.01
p = 20%	39.59	14.63	0.95
p = 50%	36.86	14.27	0.88
p = 75%	36.70	13.67	0.89
p = 100%	36.19	13.90	0.89

Table 4.2 Queue Spillback Signal Control Evaluation—Upstream Links

	Average Delay (sec/veh)	Maximum Queue Length (veh)	Number of Stops (#/ veh)
Cross-street link (Addison Street)			
Fixed Signal Settings	18.79	5.50	0.72
p = 10%	12.45	4.90	0.54
p = 20%	14.30	4.80	0.60
p = 50%	15.96	5.40	0.64
p = 75%	16.30	5.20	0.64
p = 100%	16.41	5.20	0.65
Through link upstream of critical			
Fixed Signal Settings	11.86	7.00	0.44
p = 10%	42.30	16.80	1.09
p = 20%	27.59	14.17	0.77
p = 50%	21.06	13.23	0.63
p = 75%	19.18	12.13	0.60
p = 100%	17.55	11.93	0.56

Table 4.3 shows the impact of the alternative signal control strategy on the entire arterial. It appears that for high penetration rates, the alternative signal control strategy does not substantially affect the travel time over the whole arterial segment of consideration. For low penetration rates the alternative signal control is activated frequently due to the tendency of the gap-based method to overpredict the potential of queue spillbacks. Therefore, due to the conservative spillback detection method used the alternative signal control strategy is not always effective and leads to high increases in the total travel time upstream of the critical intersection. This in turn results to higher overall travel times for the northbound direction and the southbound direction. The southbound direction is also affected negatively because changing the signal timings for the northbound direction on Addison Street also affects the green time allocated to the southbound direction, since those two directions are served by the same phase.

Table 4.3 Queue Spillback Signal Control Evaluation --Four-Signal Arterial Segment

	Travel Time (sec/veh)	Number of Stops (per veh)
Northbound direction		
Fixed Signal Settings	116.54	1.52
p = 10%	154.23	2.27
p = 20%	134.25	1.85
p = 50%	124.84	1.64
p = 75%	123.03	1.63
p = 100%	120.39	1.58
Southbound direction		
Fixed Signal Settings	86.02	0.74
p = 10%	101.03	1.05
p = 20%	95.41	0.92
p = 50%	92.87	0.84
p = 75%	92.12	0.85
p = 100%	92.96	0.85

4.5 Discussion

We developed two queue spillback detection methods that utilize data from *CV*-equipped vehicles to avoid the occurrence of spillbacks within the next cycle. The first method, gap-based detection, is dependent on information obtained from the trajectories of *CV*-equipped vehicles and only uses the stopping position of the last *CV*-equipped vehicle that joins the queue in the link of interest. The second method, shockwave-based detection, utilizes information on the signal settings of the upstream intersection in addition to the *CV* information in order to predict the formation of the back of the queue based on flows measured using *CV* data.

The results indicate that both the gap-based and the shockwave-based methods can correctly detect the occurrence of spillbacks in more than 80% of the cycles for a range of penetration rates. The shockwave-based method was found to be more effective for penetration rates in the range of 10 to 20% while the simpler gap-based detection presented the best results for *CV* penetration rates that are higher than 20%. The shockwave-based spillback detection presented a tendency to overpredict the number of potential spillbacks for high penetration rates, because of the underlying assumption that the queue formation will progress until the end of the cycle. Future research should relax this assumption and analyze the detection results when the shockwave speed corresponding to vehicles joining the queue can change within a given phase. This is expected to improve the detection results for higher penetration rates.

In addition, the simulation results have shown the capability of the proposed signal control strategy to prevent the occurrence and mitigate the impact of spillbacks by maintaining shorter than the critical link length queues and by redistributing the delays at further upstream links that have more storage space for residual queues. More importantly, the proposed signal control strategy reduces the variation of the maximum queue length, as indicated by the coefficient of variation that ranges between 0.06 and 0.09 for different penetration rates when the proposed signal control strategy is in place compared to a value of 0.16 when fixed signal settings are implemented. In addition, the results at the upstream cross street show the additional benefits of the proposed signal control strategy in the operations of those vehicles.

Ongoing work includes improving the shockwave-based method for high penetration rates by accounting for the variation in the backward-wave speed during a single phase and testing the two methods as well as the signal control strategy for a variety of design and traffic demand conditions. In addition, we plan to expand the metering signal control strategy to all intersections upstream of the critical one so that spillbacks are avoided in the whole network and queues are distributed homogeneously among all affected links for more efficient traffic operations.

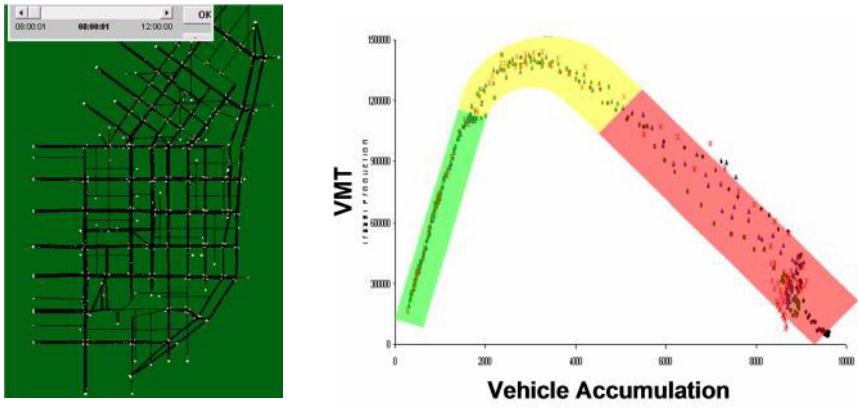
CHAPTER 5

CONTROL STRATEGIES FOR MOBILITY: CONGESTION AVOIDANCE IN GRID NETWORKS

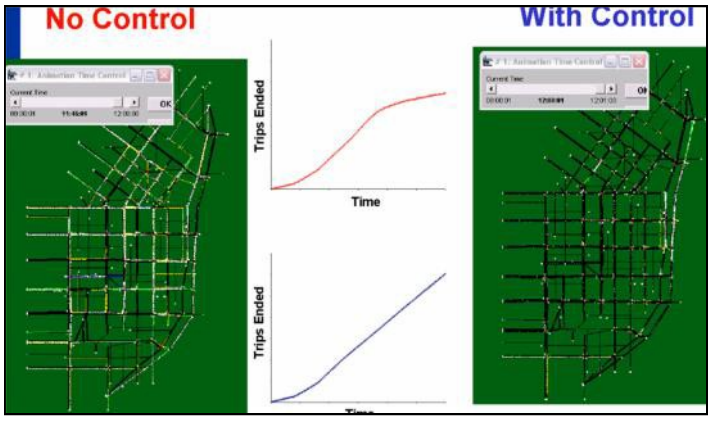
5.1 Introduction

Recent research findings show that traffic dynamics on signalized networks can be described by a macroscopic fundamental diagram (MFD) (flow-density relationship) [30]. This relationship can be obtained from CV data and/or from loop detectors. An example MFD is shown in Figure 5.1(a) for the Downtown San Francisco network with 100 signalized intersections. The existence of MFD in a network allows deployment of control strategies to maximize outflow and prevent gridlock. This is illustrated in Figure 5.1 (b) for the San Francisco network. The left side of the figure shows that the network gets congested (gridlocked) as traffic demands increase. The right side shows that the network performance improved substantially when perimeter control (or gating) is implemented.

This Chapter describes the development and testing of perimeter control and other control algorithms for avoiding congestion in grid networks based on CV data. It builds upon the methods for determining queue spillback described in Chapter 4.



a) Macroscopic Fundamental Diagram



b) Impacts of Network Gating

Figure 5.1 Queue Management Control: Downtown San Francisco

5.2 Test Network

The test site selected corresponds to the Post-Oak Area also known as the Galleria Mall area in Houston, TX (Figure 5.2). This network was originally used to develop control strategies for coordinated actuated signals in grid networks [31] and later in an NCHRP project to develop control strategies for oversaturated conditions [3]. The network was coded in the VISSIM simulation model.

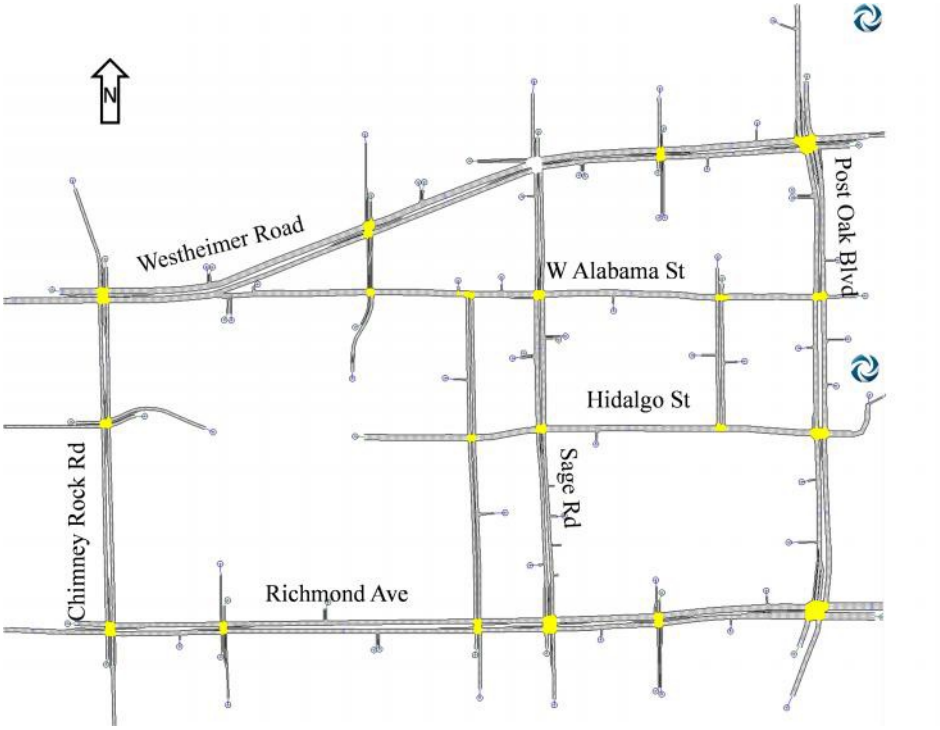


Figure 5.2 Grid Test Network: Galleria Mall, Houston, TX

The network consists of 19 signalized intersections; It includes two main arterials in the northern and southern border, Westheimer Rd and Richmond Avenue, respectively; and two arterials in the eastern and western borders, Post Oak Blvd and Chimney Rock Rd, respectively. The network also includes several major internal traffic generators because of the Galleria Mall. The network is well connected with the rest of the metropolitan area; the South of Richmond Ave there is access to the Southwest Freeway that crosses the city from East to West can be connected through the Richmond Avenue and Post Oak Blvd provides access the West Loop S, a ring across the city's downtown area.

Traffic signals operate as fixed time on a common cycle length of 150 sec. Traffic conditions are heavy especially in the pm peak period with long delays, excessive queues and average speeds less than 10 mph.

The analysis was performed with the AIMSUN microscopic simulation model. Several customized routines were developed for *CV* applications through its Application Programming Interface (API) using the Python programming language.

Fist, the input data were imported into AIMSUN from the VISSIM model. To transfer the information from one model to the other, an Origin-Destination matrix was extracted from the entry volumes and turning ratios in VISSIM and coded into AIMSUN. After transferring the information, several validation tests were made to ensure that the AIMSUN predictions are close to the VISSIM model. Comparison of the link flows from both models using the GEH statistic [32] indicated close agreement.

Following the model verification and validation, model runs were performed to obtain the baseline conditions. The baseline scenario corresponds to the PM peak hour on an average work day from 17:45 to 18:45 pm, with an initial 30 minute "warm up" period. Five model replications were performed and averaged in order to obtain representative results. Figure 5.3 shows the link densities (veh/Km) for the existing conditions. It can be seen that most of the intersections are oversaturated and several of them experience queue spillbacks as indicated by the red arrows.

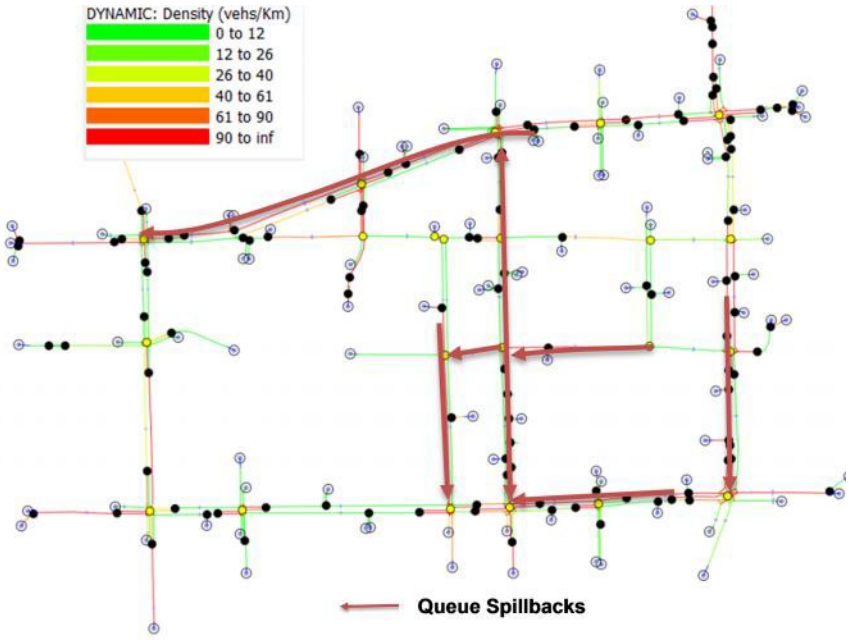


Figure 5.3 Test Network Link Density – Baseline Scenario

AIMSUN calculates several measures of effectiveness (MOEs) at the end of each model run. The following MOEs are used in the evaluation of alternative strategies:

- Total delay time (veh-hrs): The aggregated delay time that all vehicles in the network experience during the analysis period.
- Average speed (mph): This value corresponds to the harmonic mean speed for all vehicles that have left the network.
- Total distance travelled (veh-mi) : This indicator presents the total distance travelled by all vehicles.
- Unserved demand (veh): This is the number of vehicles that were not served during the period of analysis.

5.3 Queue Spillback Detection

The queue spillback detection is based on the threshold approach. The strategy defines the possibility of a spillback whenever a simulated CV vehicle is located at a distance larger than 80% of the link length from the downstream intersection and with a speed below 3 mph. When a road segment in the network presents

a vehicle with these characteristics, a spillover was considered to be imminent and the alternative control strategy was activated.

The effectiveness of this strategy depends on the penetration rate of CV vehicles in the network. A sensitivity analysis on the CV penetration rate was made and the results are shown in Figure 5.4. It is shown that with a penetration rate of 20%, over 90% of the network spillbacks are detected. These results are similar to results presented in Chapter 4. A 20% penetration rate was used to obtain the results presented in this Chapter.

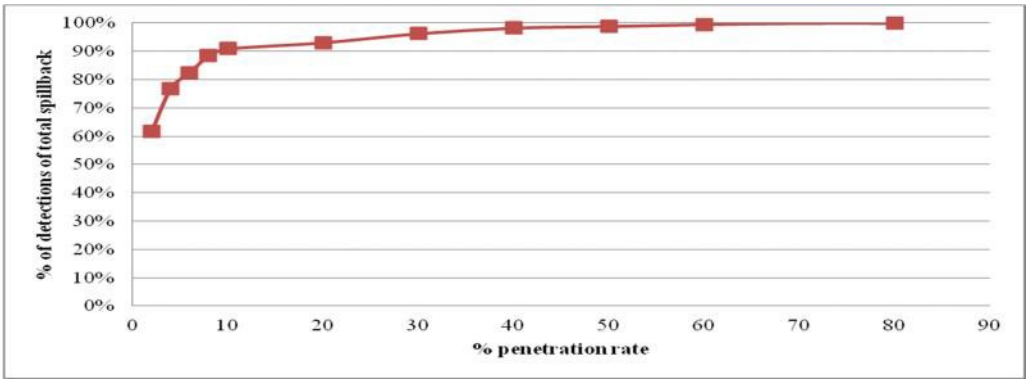


Figure 5.4 Spillback Detection vs. CV Penetration rate

5.4 Signal Control Strategies

Network Perimeter Control

The first strategy consists of reducing the vehicle entrance into the network when the occurrence of the queue spillback was detected. In this case, the strategy consisted of reducing the green time for the movements that contributed vehicle flows from external inputs into the network by 10%. Higher reductions in green times at he entrance lists was also tested but resulted in unacceptable long queues at the network boundaries. The system cycle length was kept constant at 150 seconds, and the extra green time was distributed proportionally to the rest of the signal phases.

Table 5.1 shows the performance of the existing and the proposed strategies. Most of the MOEs improved but the improvements were not statistically significant except the total distance traveled (VMT). Unserved demand (number of vehicles waiting outside the network) increased because of the reduction of green time at network entry links.

The performance of this strategy is affected by the network characteristics and traffic patterns. The study network includes several internal entrances and exits with high flows that cannot be controlled by the perimeter control strategy. This reduces the effectiveness of the strategy. Also, note that the AIMSUN reported average speed includes the speed of all vehicles in the network, including vehicles at the entry links with high delays and minimal distance traveled. This typically results in very low speed values particularly for congested networks with short links.

Table 5.1 Impacts of Perimeter Control – Total Network

Performance Measure	Existing Conditions		Perimeter Control		Benefit (%)
	Mean	St Dev.	Mean	St Dev.	
Total Delay (veh-hrs)	2079.4	56.3	2088.9	60.2	0.5%
Average Speed (mph)	4.7	0.1	4.9	0.1	4.2%
Total Distance Traveled (veh-mi)	12545.0	130.5	13083.3	88	**4.3%***
Unserved Demand (veh)	4759	109	5077	119	-6.7%

** statistically significant*

To better understand the impacts of the perimeter control we analyzed separately the performance of one major arterial: Westheimer Rd at the periphery of the network. Figure 6.5 shows the intersection delays under existing and perimeter control.

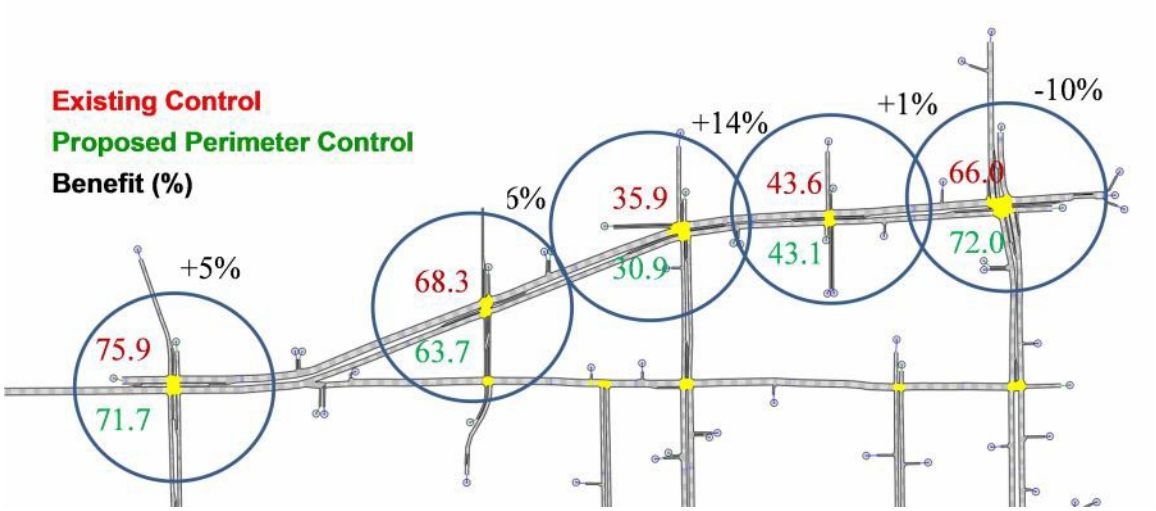


Figure 5.5 Perimeter Control – Intersection Delays (sec/veh)--Westheimer Rd Arterial

We observe from Figure 5.5 that the delay at certain intersections improved above 10% compared to the existing conditions. The highest improvements were obtained at intersections where queue spillback was present. This strategy provided no benefit to intersections where queue spillbacks are not present, and the delays increased at the intersections where the reduced green time has an important effect on the entrances to the network (e.g., intersection of Post Oak Blvd and Westheimer Rd).

Spillback-Based Phase Changing Control Strategy

The second strategy tested was a dynamic strategy where the phases reacted to queue spillbacks. Whenever a spillback was detected, the downstream intersection changed its phase to green for the through movement to clear the additional accumulation. Meanwhile, the through phase at the upstream intersection turned red to prevent additional vehicles to enter the link. Figure 5.6 illustrates the proposed control strategy, and Table 5.2 shows the network performance under each strategy.

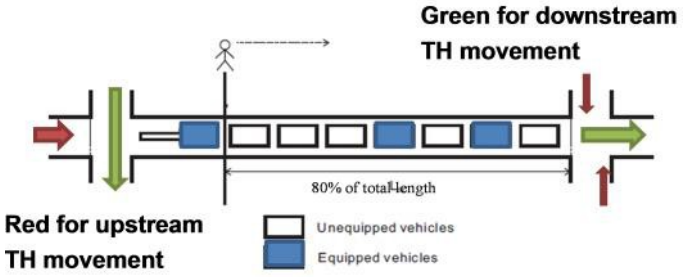


Figure 5.6 Schematic Representation of the Phase Changing Strategy

Table 5.2 Impacts of Spillback-based Phase Changing Control – Total Network

Performance Measure	Existing Conditions		Proposed Control		Benefit (%)
	Mean	St Dev.	Mean	St Dev.	
Total Delay (veh-hrs)	2079.4	56.3	2072.1	37.5	0.3%
Average Speed (mph)	4.7	0.1	5.0	0.1	**6.4%***
Total Distance Traveled (veh-mi)	12545.0	130.5	13439.4	292	**7.1%***
Unserved Demand (veh)	4759	109	4669	210	1.9%

** statistically significant*

The proposed strategy improved the average speed and the total distance travelled in the network. The percent improvements were somewhat higher than the perimeter control (Strategy 1). Again, there was no difference in the total network delay since the strategy cannot influence a significant portion of the network links.

Proposed Control Strategy With Reduced Cycle Time

This control strategy consists of identification of network links with the highest accumulation (density) through the MFD, provide priority through additional green time and at the same time reducing overall accumulation by reducing the network cycle length. The strategy tested consisted of the following steps: a) the cycle length was reduced from 150 seconds to 120 seconds maintaining the same proportional green time for every intersection, and b) provide additional 10% green time to the road segments where queue spillback first occurs. The results from the implementation of the strategy are shown in Figure 5.7 and Table 5.3.

Figure 5.7 shows the network's MFD for each strategy. The MFD was obtained from the CV vehicles in the network supplemented by loop detector (volume and occupancy) surveillance data. It is clearly shown that under the proposed strategy results in higher discharge flows strategy, hence allowing the network to complete more trips and improve traffic conditions.

Table 5.3 shows that the proposed strategy significantly improved all the MOEs. Note that the highest benefit was obtained on unserved demand; an additional 1195 vehicles completed their trip in the analysis period. This is in agreement with the results shown in the MFD (Figure 6.7) where the proposed strategy results in higher discharge flow.

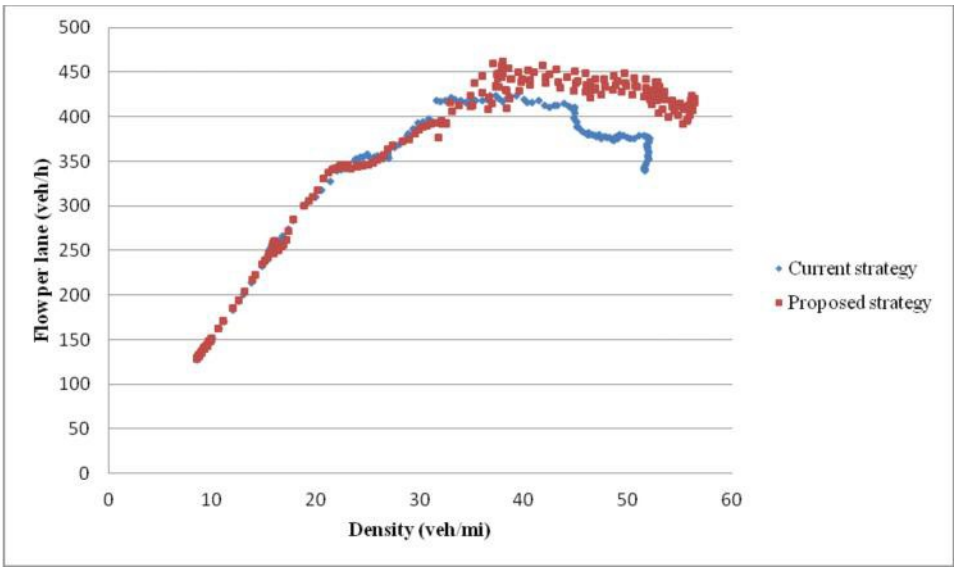


Figure 5.7 Macroscopic Fundamental Diagram—Baseline vs. Proposed Control

Table 5.3 Impacts of the Control Strategy with Reduced Cycle Time – Total Network

Performance Measure	Existing Conditions		Proposed Control		Benefit (%)
	Mean	St Dev.	Mean	St Dev.	
Total Delay (veh-hrs)	2079.4	56.3	1896.2	71.5	**8.8%***
Average Speed (mph)	4.7	0.1	5.5	0.2	**17.0%***
Total Distance Traveled (veh-mi)	12545.0	130.5	13844.0	423.8	**10.4%***
Unserved Demand (veh)	4759	109	3563.	244	**25.1%***

** statistically significant*

5.5 Discussion

Control strategies for congestion avoidance in grid networks based on CV data were developed and tested in a real-life network through simulation. The first strategy consisted of a perimeter control approach, were vehicular entrances on the perimeter of the model were limited by reducing the green times on the network entry links. The results showed small network improvements, mainly because the network includes several internal exits that cannot be controlled. The strategy was effective on the network segments that experience queue spillbacks. The second strategy consisted of signal phase changes to the upstream and downstream intersections in a network link that spillbacks are detected. The strategy resulted in small benefits because could not prevent queue spillback in the whole network. The final strategy consisted of reducing the system cycle length and providing additional green time to those network links that spillback first detected. This strategy produced the highest benefits.

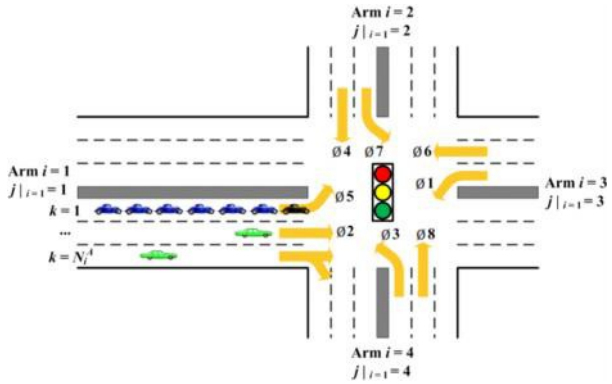
The magnitude of benefits that can be obtained from the proposed strategies depend on the characteristics of the particular network under study. Additional evaluations of the proposed strategies should be performed on networks covering a range of geometric, traffic and control conditions.

CHAPTER 6

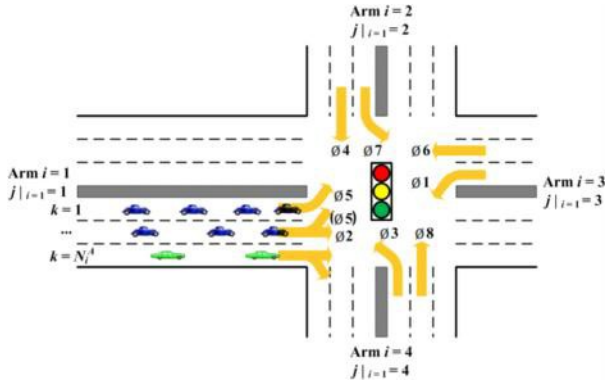
TRAFFIC CONTROL FOR MOBILITY: DYNAMIC LANE ALLOCATION

6.1 Introduction

Several control strategies have focused on signal timing optimization at traffic signals, including cycle length, phase sequence and green splits. However, these optimization strategies assume that the lane use and lane configuration have been given as exogenous inputs, which may limit the intersection's ability to handle significant variations in traffic demand. This can be addressed through dynamic lane allocation or grouping (DLG). Figure 6.1(a) illustrates an example of poor lane utilization along an approach when the traffic demand significantly deviates from the design, while Figure 6.1(b) presents an improved lane utilization by dynamically adjusting the lane allocation based on real-time traffic demands.



(a) Unbalanced lane utilization due to predefined lane grouping under demand variation



(b) Improved lane utilization due to dynamic lane grouping under demand variation

Figure 6.1 Illustration of Dynamic Lane Grouping Strategy

This Chapter presents the development and evaluation of a DLG strategy based on CV data. Detailed description of the work performed and the results are included in Appendix C.[67]

---

[6] Zhang L., and G. Wu, "Dynamic Lane Grouping at Individual Intersections: Problem Formulation and Performance Analysis," Transportation Research Record #2311, Journal of the Transportation Research Board, 2012.

[7] Wu G., K. Boriboonsomsin, L. Zhang, and M. Barth, "Simulation-Based Benefit Evaluation of Dynamic Lane Grouping Strategies at Isolated Intersections," Proceedings 15th IEEE ITSC Conference, Anchorage, AK, Sep 2012.

Previous studies on DLG include a dynamic lane assignment system tested in Houston, Texas [33] and DLG control strategies developed in the Netherlands [34]. Similar concepts to DLG have been also studied within the framework of intersection operation with fully automated autonomous vehicles [35].

The DLG strategy requires accurate information of turning movement demands, which has been difficult to achieve with the existing fixed point detection infrastructure at traffic signals. CV technology when reaches a mature level of deployment will provide accurate estimates of movement traffic demands.

6.2 Proposed Dynamic Lane Grouping Model

Consider an isolated signalized intersection with N_T approaches (e.g. $N_T = 4$), which are numbered clockwise as shown in Figure 6.1. The number of lanes for approach i is denoted as N_i^A. It is further assumed that the number of exit lanes is no less than that of approaching lanes. For tractability, a lane-level movement can be defined as

$$(i, j, k): 1 \le i = N_T \le j = N_T \text{ and } 1 \le k \le N_{iA} \quad (1)$$

representing the traffic turning from approach i to approach j (numbered locally with respect to approach i as shown in Figure 6.1) via a lane k (numbered from of the curbside lanes) in the i-th approach (without specifying which lane to use to enter the j-th approach). So j is locally defined such that different j's can be given names using movements, as follows for a four-legged intersection:

j=1 for U-turn (UT), which is not considered in this work;
j=2 for left turn (LT);
j=3 for Straight through (TH); and
j=4 for right turn (RT).

A binary function is defined to identify the permitted movements:

$$\delta(i, j, k) = \begin{cases} 0, & \text{movement from arm } i \text{ to arm } j \text{ via lane } k \text{ not allowed,} \\ 1, & \text{otherwise.} \end{cases} \quad (2)$$

For safety reasons, if the movement to approach j is allowed on lane k+1, then all movements to j+1, … N_T are prohibited on lane k.

$$\forall m = j+1, \dots, N^T; j = 1, \dots, N^T - 1; k = 1, \dots, N^i - 1; i = 1, \dots, N^T \quad (3)$$

Where m represents the index of an approach whose index is higher than j.

The traffic demand matrix is noted as

$$Q(n) = \left(Q_{i,j}(n)\right), i, j = 1, \dots, N_T \quad (4)$$

where $Q_{i,j}(n)$ is the demand from approach i to approach j (pcu / hour) at the n-th time step.

A generalized version of the estimation model for the lane saturation flow rate is used [36]:

$$s^{i,k} \triangleq \frac{}{1 + 1.5 \sum_{j=1}^{N_T} \frac{f_{i,j,k}}{r_{i,j,k}}}, \quad \text{for } k = 1, \dots, N^i, i = 1, \dots, N^T \quad (5)$$

where

$s_{i,k}$ is the saturation flow rate of lane k in approach i;

$r_{i,j,k}$ is the turning radius from approach i to approach j, (∞ for straight ahead);

$\bar{s}_{i,k}$ is the saturation flow rate for the through movement.

$f_{i,j,k}$ is the flow factor, defined as the proportion of traffic from approach i to approach j via lane k ($q_{i,j,k}$ is the flow rate from approach i to approach j via lane k)

$$f_{i,j,k} \triangleq \frac{q_{i,j,k}}{\sum_{j=1}^{N_T} q_{i,j,k}} \tag{6}$$

The flow ratio for a given approach i and lane k is then defined as

$$y_{i,k} \triangleq \frac{\sum_{j=1}^{N_T} q_{i,j,k}}{\bar{s}_{i,k}}. \tag{7}$$

Substituting equation 5 into equation 7 leads to

$$y_{i,k} = \frac{1}{\bar{s}_{i,k}} \sum_{j=1} \left(1 - \frac{1.5}{\bar{r}_{i,j,k}} \right) q_{i,j,k}. \tag{8}$$

The demand matrix $\boldsymbol{Q}$ varies over time. The change of demand matrix at time n_1 and n_2 is defined:

$$\Delta_{Q(n_1,n_2)} = \frac{|Q(n_1) - Q(n_2)|}{|Q(n_1)| + |Q(n_2)|}. \tag{9}$$

where $|Q(n_1)|$ is defined as

$$|Q(n)| \triangleq \sum_{i=1}^{N_I} \sum_{j=1}^{N_T} |Q_{i,j}(n)|. \tag{10}$$

$\Delta Q(n_1, n_2)$ is normalized to be the relative difference.

In the absence of a shared lane, the flow factor for each $f_{i,j,k}$ as defined in equation 6 is either 0 or 1. However, if a shared lane is considered, the flow factor should be estimated based on a lane selection principle, such as "equal degree of saturation", which states that available lane capacities should be used equally for adjacent lanes that share turning movements [36, 37]. Therefore, the traffic flows on adjacent lanes with common movement should be equalized, except for those exclusive lanes on which the traffic is independent of that on adjacent lanes. When the signal settings on adjacent lanes with common movements are identical, this equalization of the degree of saturation suffices to the equalization of flow ratio.

Figure 6.2 shows an example plot of the steady state estimation for a two-lane approach with one lane (k=1) as a shared left-turn and through lane and the other as a through only lane.

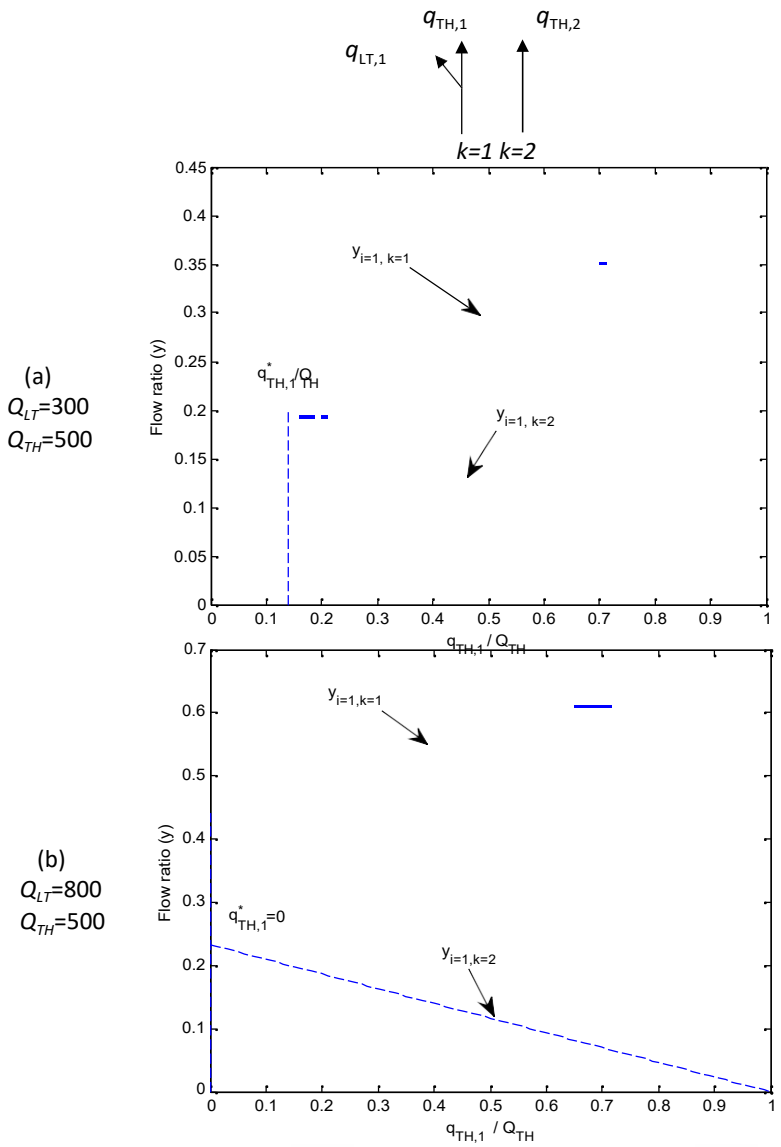


Figure 6.2 Steady State Flow Estimation on a Two-Lane Approach with One Shared Lane

In this example, the flow $q_{1,2,1}$ (which can also be noted as $q_{1,LT,1}$) always equals to $Q_{1,2}$ or $Q_{1,LT}$. The only variable is $q_{1,TH,1}/(q_{1,TH,1} + q_{1,TH,2}) \triangleq q_{1,TH,1}/Q_{1,TH}$, the ratio of through traffic in the shared lane. In Figure 6.2(a), an equilibrium solution can be found at $q^*_{1,TH,1}/Q_{1,TH}$, which equalizes the flow ratio for the two lanes ($y_{1,1} = y_{1,2}$). Figure 6.2(b) illustrates that in the case of high left turn demand, there is no equalized traffic assignment and the shared lane is used as a de facto exclusive left turn lane.

Given a lane grouping assignment, traffic is then equalized based on their flow ratios:

$$\text{equalize } y_{i,k} \triangleq \frac{1}{\bar{s}_{i,k}} \sum_{j=1} \left(1 + \frac{1.5}{\bar{r}_{i,j,k}}\right) \cdot q_{i,j,k} \qquad (11)$$

The equalization of flow ratio given a lane assignment may not always have a viable steady state equalized solution, as shown in Figure 6.2. As a result, one or more shared lanes may turn out to be a de facto exclusive lane. Therefore, the equalization problem is formulated as follows (without loss of generality, equation of approach i instead of all approaches of an intersection is considered as sub-problem to estimate steady state traffic flow):

$$\min(\max_k y_{i.k}),$$

given $\boldsymbol{\delta} = (\delta_{i,j,k})$, subject to certain constraints and conditions for equalization of traffic flow ratio. Additional details can be found in the paper by L. Zhang and G. Wu (Appendix C).

The performance index used for DLG optimization is maximum flow ratio, which is independent of the signal settings:

$$\max [illegible] = \max_k [illegible], \forall i, n \quad (12)$$

The optimization problem is formulated as follows:

Given demand matrix $\boldsymbol{Q}(n)$, find the optimum lane grouping $\boldsymbol{\delta} = (\delta_{i,j,k})$ by the objective function

$$\min \max [illegible], \dots, 4 \quad (13)$$

subject to conditions in equation 1, equation 2 and equation 3.

The optimization of the lane grouping is calculated for each intersection approach as shown in equations 1, 2, 3. In the real-world, the number of approach lanes is a small integer, which makes it possible to solve the optimization problem by enumerating all possible lane groupings.

6.3 Evaluation

The flow ratio of a lane group is used as a primary performance measure for evaluating the benefits of DLG. Additional MOEs include the average vehicle delay and the degree of saturation estimated based on the Highway Capacity Manual 2000 (HCM2000) [15].

A typical four legged isolated signalized intersection with eight-phase, dual ring control scheme is analyzed. Each approach is about 1,000 ft (300 meters) long with a speed limit of 40 mph (64 kph). The turning movement assignment for each lane is fixed for different traffic O-D patterns in non-DLG scenarios. In the DLG scenarios, the lane allocation may vary with traffic demands and the optimal solution can be obtained from the proposed mathematical model presented above. Some parameters common to all cases are listed here.

1. The saturation flows for through movement are selected as 1965 pcu/h on the curbside and 2105 for the offside lanes [34].
2. Traffic composition consists of 95% passenger cars and 5% trucks
3. The turning radii are set to 12m for left turn and 10m for right turn.
4. A fixed 120 sec cycle length is used based on Signal Timing Manual (STM) recommendations for intersections with protected left turn phases [38]
5. Minimum green duration is set to 12 seconds;
6. The lost time is assumed as 5 sec per phase (Yellow and all-red clearance intervals are 3 seconds and 1 second per phase)
7. Effective green splits are calculated based on the HCM2000 Quick Estimation Method (QEM); there is also a constraint to guarantee that green time for lane groups that share movements are grouped to the same phase

The baseline model, the demand matrix is:

$$\begin{pmatrix} arm & LT & TH & RT \\ \boldsymbol{i=1} & 375 & 600 & 225 \\ 2 & 600 & 600 & 150 \\ 3 & 300 & 750 & 150 \\ 4 & 300 & 750 & 150 \end{pmatrix}.$$

In each approach, there is one exclusive left-turn lane, one left-turn and through shared lane, one exclusive through lane, and one through and right-turn shared lane, i.e.,

$$\delta^i = \begin{pmatrix} \text{Movement/Lane} & 1 & 2 & 3 & 4 \\ \text{LT} & 1 & 1 & 0 & 0 \\ \text{TH} & 0 & 1 & 1 & 1 \\ \text{RT} & 0 & 0 & 0 & 1 \end{pmatrix}, i = 1, \ldots, N^T.$$

6.3.1 Numerical Analysis

The numerical analysis procedure shown below is applied to both fixed lane grouping and DLG under different objective functions for various demand matrices. The demand variation for one approach is enumerated by changing the left-turning demand $Q_{1,2}(n) = Q_{1,LT}(n)$ and demand of straight-through movement $Q_{1,3}(n)$ (or $Q_{1,TH}(n)$) independently, while keeping the overall demand unchanged. For each lane group the demand quota varies from 10% to 80% of the overall demand.

Numerical analysis procedure:

Given overall demand $Q_1 = Q_{1,LT} + Q_{1,TH} + Q_{1,RT}$

$n = 1$;

For each $Q_{1,LT}(n)$=0.1Q_1 to 0.8 Q_1;

For each $Q_{1,TH}(n)$=0.1Q_1 to 0.9$Q_1 - Q_{1,LT}(n)$;

$Q_{1,RT}(n) = Q_1 - Q_{1,LT}(n) - Q_{1,TH}(n)$;

Calculate $\Delta Q(n)$ using equation 10 ;

Calculate maximum flow ratio (approach *i*=1) and average intersection delay for fixed grouping; calculate green splits using QEM method based on demand *Q*(*n*) for case 1,2 and 4 (therefore is adaptive) and *Q*(0) for case 3 (fixed timing);

Find the optimum lane grouping in terms of minmax flow ratio by enumerating all possible lane groupings;

Calculate maximum flow ratio (approach *i*=1) and average intersection delay for the optimized lane grouping; calculate green splits using QEM method based on demand *Q*(n) for case 1 to 4 (adaptive timing);

$n \leftarrow n + 1$;

End for;

End for;

For all the results, obtain the cumulative statistics *F* using the following formulae:

$$F(p,t) \triangleq \frac{1}{\sum_{n=1}^{n} 1(n,t)} \sum_{n=1}^{n} 1(n,t) \cdot p(n), \text{ for any } 0 \leq t < 1. \quad (14)$$

In equation 14, *t* is a given threshold of the maximum variation of the demand matrix and the binary function 1(n,t) is defined as

$$1(\quad) \triangleq \begin{cases} 1, & \text{if } (\quad) < \\ 0, & \text{elsewise} \end{cases}. \tag{15}$$

where Δ is a resolution interval selected to get the statistics. p can be any data obtained from the numerical analysis, such as the maximum flow ratio or the delay.
In this analysis, each n corresponds to a different Q matrix at each time step.

For simplicity of illustration, numerical analysis below is conducted at an intersection with demand variation only for one approach (i=1). Demands from all other approaches are assumed to remain unchanged.

Case 1: DLG vs. fixed lane grouping with real-time adaptive timing at a four-lane approach--under-saturated conditions

When calculating delay for both the fixed grouping and DLG, real-time adaptive timing is used. It is assumed that demand matrix is known in real-time to make green time fully responsive to the demand. The cycle length remains fixed.

Case 2: DLG vs. fixed lane grouping with real-time adaptive timing at a three lane approach--undersaturated conditions

The fixed lane allocation for all approaches is given below:

$$\delta^i = \begin{pmatrix} \text{Movement\textbackslash Lane} & 1 & 2 & 3 \\ \text{LT} & 1 & 0 & 0 \\ \text{TH} & 0 & 1 & 1 \\ \text{RT} & 0 & 0 & 1 \end{pmatrix}, i = 1, \quad , N^T,$$

with one dedicated left turn lane, one dedicated straight through lane and one lane shared for straight through and right turn. The baseline demand is

$$\begin{pmatrix} arm & LT & TH & RT \\ \mathbf{i=1} & 250 & 400 & 150 \\ 2 & 400 & 400 & 100 \\ 3 & 200 & 500 & 100 \\ 4 & 200 & 500 & 100 \end{pmatrix}.$$

Case 3: DLG vs. fixed lane grouping with fixed timing at a three legged intersection

All other parameters remain the same as case 2 except that for fixed lane grouping the signal timing is also fixed. For DLG the signal timing cannot be fixed since the timing must be adjusted to avoid potential conflict when there is a change in lane allocation.

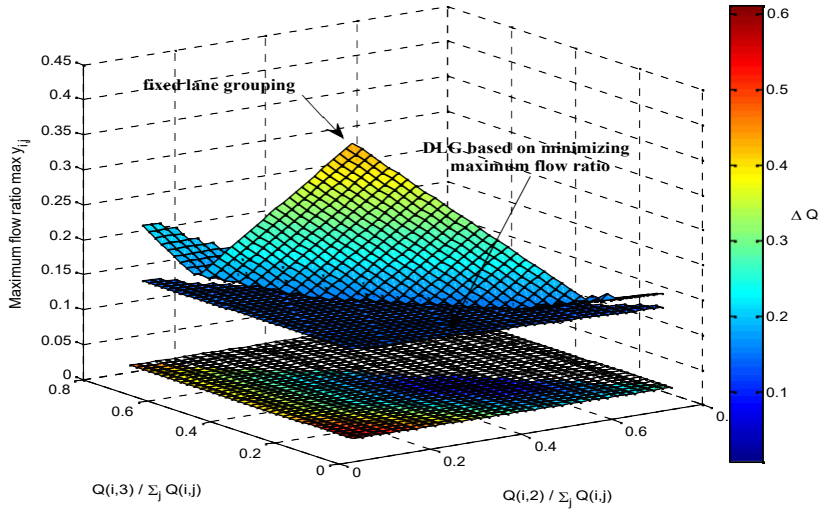
In this example, since the fixed lane grouping has fixed timing plan and the demand won't change for other approaches , we will show specifically delays for phase 2 and phase 5.

Case 4: DLG vs. fixed lane grouping three legged intersection with real-time adaptive timing and oversaturated demand.

All other parameters are the same as in case 2 except that the demand is 20% higher for all approaches.

Figure 6.3 shows the maximum flow ratio for one approach for case 1. In this Figure, the contour plots on the floor show the magnitude of the deviation of demand matrix $\Delta Q(n)$. The deviation of the demand matrix $Q(n)$ to the baseline demand matrix Q is calculated using equation 9. There are three movements per approach left turn, through and right turn. The number of independent variables for the spatial variation is only two because the sum of movement demands is fixed. The two variables chosen are left-turning traffic and straight through traffic and shown in Figure 6.3 in the x-axes and y-axes, respectively.

It can be clearly seen that with the increase of $\Delta Q(n)$, the demand deviation from the baseline assumption, the performance measured by maximum flow ratio for fixed lane grouping increases, as expected. The corresponding results of DLG remain almost constant within the full range of the variation of demand matrix.

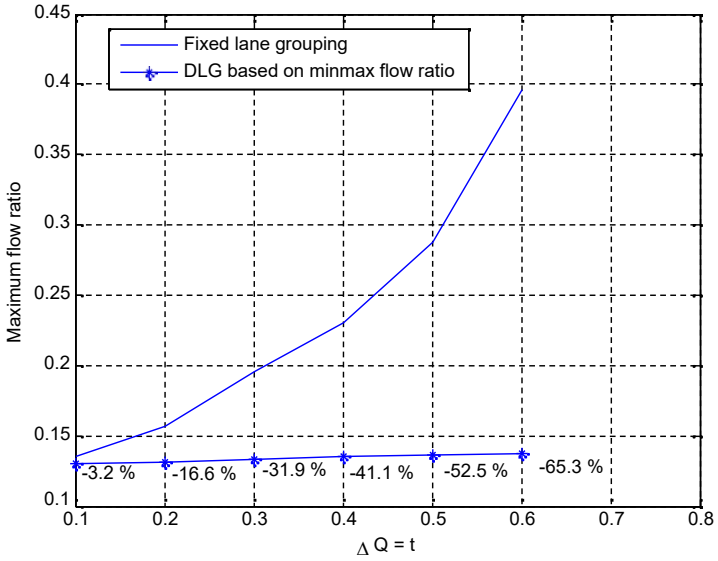


Maximum flow ratio at approach *i*=1

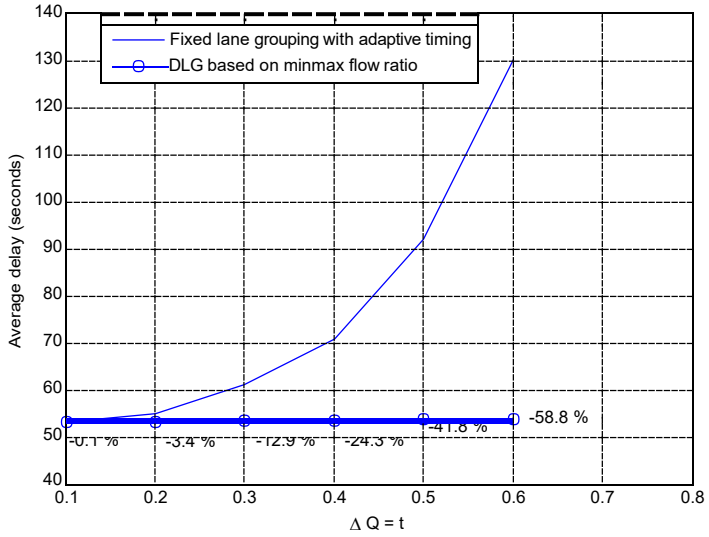
Figure 6.3 Case 1: Performance of DLG and Fixed Lane Grouping vs. Demand Variation

Figures 6.4 and 6.5 show the improvement of the DLG strategy at each threshold of demand variation for Cases 1 and 2. For example, Figure 6.4(a) shows that for 30% demand variation the maximum flow ratio is 0.19 for fixed lane grouping and 0.13 under DLG (a reduction of 31.9%). The improvement of DLG over fixed lane grouping increases when the deviation of demand matrix from the baseline model becomes higher.

Sample numerical results are also listed in Table 6.1. The improvement is quantified as a percentage of reduction from the fixed lane grouping based model of each performance measure. This reduction is shown for the DLG method based on min max $y_{i,k}(n)$. The numerical result in case 3 reveals that even more improvements of DLG can be expected when comparing to fixed lane grouping with fixed timing.

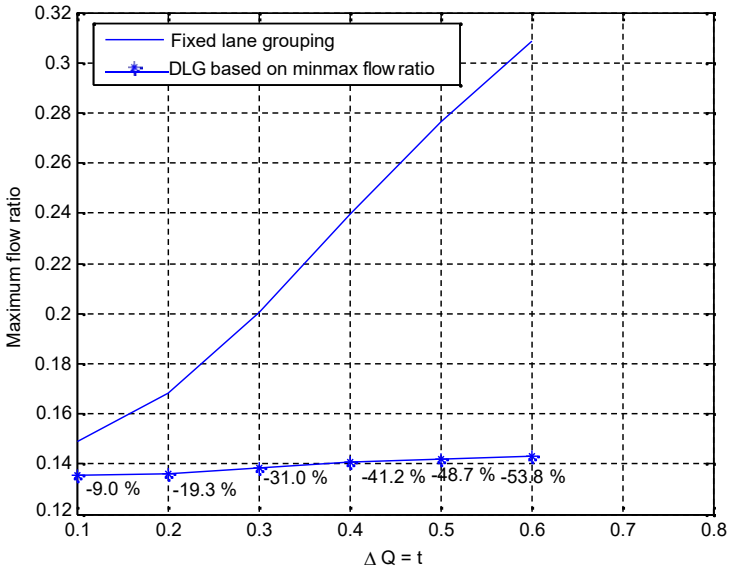


(a) Maximum flow ratio at approach *i*=1

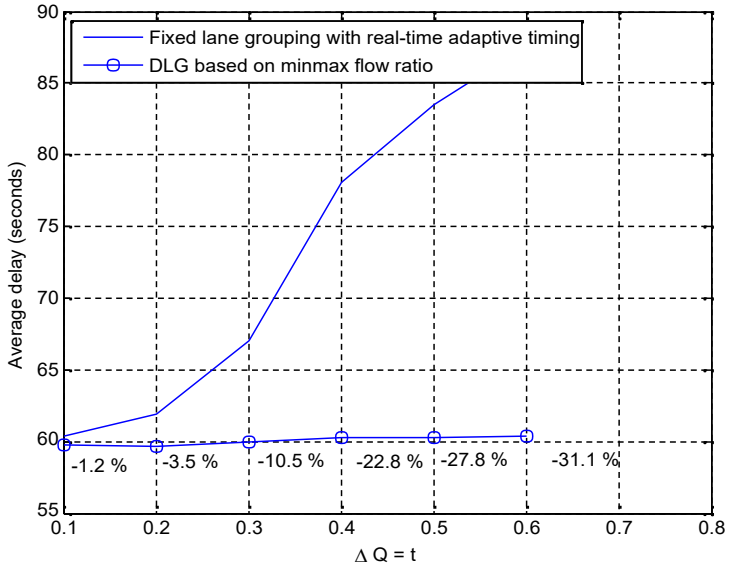


(b) Average delay

Figure 6.4 Case 1: Benefits of DLG vs. Demand Variation (max ΔQ



(a) Maximum flow ratio at approach *i*=1



(b) Average delay

Figure 6.5 Case 2: Benefits of DLG vs. Demand Variation (max ΔQ

Table 6.1 Sample Numerical Results

	Fixed-La neGrouping				DLG for Ami I Based on min 1113X, l".au{")			
aQ,{11)	Eslima1:ed Steady-S1a1.e Traffic Assignment (11,.,k,	mox,y.,	Effec, ti ·e Green Split (sJ	Delay (s)	**Saturnl ion** Degree	Es1ima1ed S te ady-S tate Traffic Assignment (q, ,},=J	0)3 XI) 'u-	Effccrive Green T ime (s)	Deloy [s {%)]	Saturation Degree
Case I. Base Lane Arm. Undersaturated. Adaptive Signal Tuning										
0.38·	[k = I 2 3 4 LT 390 426 0 0 TH O O 168 0 RT O 0 0 2 1 6	0.22	,t,5/q,2: 22	69	ume I : 1.14 Lane 2: I.IS Lane 3: *OA2* Lane 4: 0.61	k = I 2 3 275 30 1 1-10 4 0 0 71 97 RT 0 0 0 2 1 6	0.15 (- 30)	,t,5/q,2: 24.7	53 (- 23)	Lane I: 0.70 1..nne 2: 0.71 Lane 3: 0.72 Lane -I: 0.79
0.25	k= I 2 3 4 LT 217 138 0 0 TH O O 192 0 RT O O O 552	0.29	,j,5/,j,2: 25.7	69	Lane I: 0.54 Lane 2: *0.55* Lane 3: 0.41 Lane 4: 1.34	k = I 2 3 4 LT 27(, 1 80 0 0 TH 1 38 54 0 RT 0 0 *250* 301	0 . 15 (-4 5.5)	q,51,j,2: 24.5	53 (-2'.l)	1..nne I: 0.74 Lane 2: 0.76 Lnne 3: 0. 74 Lane 4 : 0. 73
Case 2. Three- Lane Arm. Unde.rsalurnl ed, Adaptive Sign, Tuning										
0.33'	k = I 2 3 LT 384 0 0 TH 0 144 0 RT O O 272	0 2 1	¢,5: 27.8 ¢,2: 15	79	Lllne I: 0.94 Lane 2: 0.53 Lone 3: LI-I	k = I 2 247 1 37 0 1-14 3 RT 0 0 271	0. 1 4 (-33)	4)5/)l2: 22	60(-24)	Lone I: 0.72 Lone 2: 0.76 Lane 3: 0.77
0.40"	k = I 2 LT *5-M* 0 0 TH 0 80 3 0 RT 0 0 176	0.31	,)l5: 35 ¢,2: 1 2	91	Lane I: 1.05 L. ne 2: 0.37 LAne 3: 0.93	k = I 2 LT 248 272 24 TH 0 0 3 SO RT 0 0 176	0.14 (- 55)	,i,s1,i,2: 22	601-30)	Lane I : 0 . 73 L.1ne 2: 0.73 Lone 3, 0.79
Case 3. 1l ircc --um e Ann. Undersntunucd , Fixed-Lane Groupi ng Using Fixed Ti mi ng										
0.33'	k = I 2 J LT 384 0 0 TH O 144 0 RT O O 272	0.21	,)15: 19 ¢,2: 23	87	Lane I: 1.34 Lru1e 2: 0.34 Lime 3: 0.74	k =cl 2 LT 247 137 0 TH 0 1-M 0	0.14 (-33)	,)15/¢12: 22	60 (- 31)	La ne I: 0.64 Lane 2: 0.66 Lune 3: 0.67
Case 4. Three-Lane Arm. Oversa1urnted. Adapuve Signal liming										
0 .4	k= I 2 LT 653 0 TH O 96 RT O 0	0.37	,t,5:35 ,)12 : 12	147	L.sne I: 2.27 Lnnc 2: 0.4-1 Lane 3: 1.12	k = I 2 299 32 28 RT 0 0 272 RT 0 0 211	0 . 17 (-55)	q>51,jl2: 22	78 (-47)	Lane I : 0.87 Lane 2: 0.88 Lane 3: 0.97

-Pcn::entngc. given in parcn -

'Q,.LT = 816; Q,.,,, = 168; Q,.ar = 216.
'Q,cr = 56: Q,.n, = 192: Q,.n = 552.
'Q,.u = 384: Q,.t. = 144: Q,.aT = 272.
'Q,LT= : Q,.n1 = 80-, Q. OT = 176.
'Qu T = 384: Q.TH = 1.14: Qu = 272.
'Q,.LT = 65): Q,n, = 96; Q"T = 21 1.

0\
0\

6.3.2 Simulation

We also evaluated the impacts of DLG strategy using the PARAMICS microscopic traffic simulation model [39] and the CMEM plug-in [40] to obtain estimates of mobility (average delay, number of stops) and environmental (energy consumption and emissions) performance measures. The duration for each simulation run is 60 minutes, but the first 30 minutes is used as a "warm-up" period.

In addition to the baseline model, the following strategies were simulated:

1) Scenario I: There is no change in lane grouping and signal timings.
2) Scenario II: There is no change in lane grouping but signal timings are adjusted based on QEM.
3) Scenario III: The lane grouping is optimized but signal timings kept intact.
4) Scenario IV: Both lane grouping and signal timings vary with the perturbed traffic demand.

The simulation results are summarized in Table 6.2. Under variable traffic demand, the DLG strategy with QEM-based adaptive signal timings (Scenario IV) provides the highest benefits. The changes in different MOEs are trivial in comparison with the baseline scenario whose lane grouping and signal timings have been optimized accordingly. On the contrary, the non-DLG strategy with fixed signal timing (Scenario I) performs the worst, where the average vehicle delay and number of stops increase as much as 11.54% and 12.23%, respectively. In addition, although improvements in MOEs can be obtained by changing signal timings, it is more effective to adjust the turning movement assignment (Scenario III vs. Scenario II), especially when the traffic demand varies significantly among different movements (e.g., left-turn vs. through). Table 6.2 also shows that the average vehicle delays obtained from the simulation are comparable with those from the numerical analysis previously described.

Table 6.2 Summary of Numerical and Simulation Results

		Num Analysis	Simulation						
Demand	**Scenario**	**Delay (sec)**	**Delay (sec)**	**# Stops**	**CO2 (g)**	**CO (g)**	**HC (g)**	**NOx (g)**	**Fuel (g)**
Baseline	Baseline	53.02	53.48	1.02	343.16	17.4	0.314	1.124	116.54
Pertubed	Scenario I	61.73	59.65	1.15	358.16	17.76	0.324	1.156	121.43
	% vs. Baseline	16.43	11.54	12.23	4.37	2.08	3.13	2.81	4.2
	Scenario II	58.48	58.09	1.12	354.64	17.51	0.319	1.154	120.2
	% vs. Baseline	8.84	8.63	9.11	3.35	0.66	1.6	2.62	3.14
	Scenario III	53.46	54.31	1.03	345.07	17.46	0.316	1.129	117.17
	% vs. Baseline	0.75	1.56	0.21	0.56	0.36	0.5	0.45	0.54
	Scenario IV	53.45	53.67	1.01	343.3	17.38	0.314	1.123	116.57
	% vs. Baseline	0.8	0.36	-1.47	0.04	-0.12	0.07	-0.12	0.03

Sensitivity Analysis

A sensitivity analysis on the proportion of left-turn traffic volume was carried out under both non-DLG and DLG strategies with adaptive timings. In the baseline model, the traffic demand matrix is

$$\begin{pmatrix} arm & LT & TH & RT \\ \boldsymbol{i=1} & 240 & 810 & 150 \\ 2 & 300 & 750 & 150 \\ 3 & 300 & 750 & 150 \\ 4 & 300 & 750 & 150 \end{pmatrix},$$

and the turning movement assignment matrix for each approach is

$$\delta_i = \begin{pmatrix} \text{lane index} & k=1 & 2 & 3 & 4 \\ \text{LT} & 1 & 0 & 0 & 0 \\ \text{TH} & 1 & 1 & 1 & 1 \\ \text{RT} & 0 & 0 & 0 & 1 \end{pmatrix}, i = 1, \ldots, 4$$

Simulation runs have been performed with four levels of left-turn traffic volume proportions,

$p_{1,LT} \triangleq \sum_k q_{1,2,k} / \sum_j \sum_k q_{1,j,k}$,

on arm 1 only ($p_{1,LT} = 0.2$, 0.4, 0.6 and 0.8) while keeping the total traffic volume and right-turn traffic volume to be 1,200 veh/hr and 150 veh/hr, respectively. More specifically, the optimal turning movement assignment for each left-turn volume proportion is:

1) For $p_{1,LT} = 0.2$, the same as the baseline model.
2) For $p_{1,LT} = 0.4$,

$$\delta_1 = \begin{pmatrix} \text{lane index} & k=1 & 2 & 3 & 4 \\ \text{LT} & 1 & 1 & 0 & 0 \\ \text{TH} & 0 & 1 & 1 & 1 \\ \text{RT} & 0 & 0 & 0 & 1 \end{pmatrix}$$

3) $p_{1,LT} = 0.6$;

$$\delta_1 = \begin{pmatrix} \text{lane index} & k=1 & 2 & 3 & 4 \\ \text{LT} & 1 & 1 & 1 & 0 \\ \text{TH} & 0 & 0 & 1 & 1 \\ \text{RT} & 0 & 0 & 0 & 1 \end{pmatrix}$$

4) $p_{1,LT} = 0.8$;

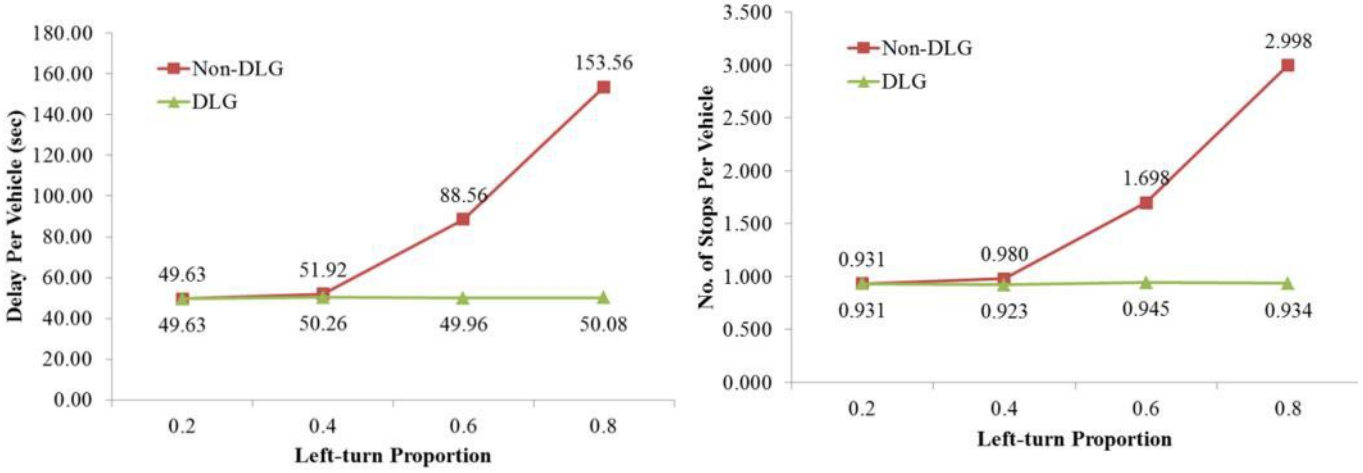
$$\delta_1 = \begin{pmatrix} \text{lane index} & k=1 & 2 & 3 & 4 \\ \text{LT} & 1 & 1 & 1 & 1 \\ \text{TH} & 0 & 0 & 0 & 1 \\ \text{RT} & 0 & 0 & 0 & 1 \end{pmatrix}$$

In both non-DLG and DLG scenarios, the QEM-based green splits for different phases may vary with different demand profiles (Table 6.3). Without changing the lane grouping of approach 1, the green time of ∅2 or ∅5 in the non-DLG strategy has to be reserved for longer as the left-turn volume proportion increases in order to clear the queuing vehicles as soon as possible. If the cycle length is fixed, then the green times for other phases have to be shortened. Thus, the average delay and number of stops may increase on the other approaches, so do the fuel consumption and emissions. On the other hand, the green splits for different phases, however, are very close to each other in the scenarios with the DLG strategy because the traffic volumes on different approaches are equal (i.e., 1,200 veh/hr) and the flow ratios in the lanes may be balanced by the DLG strategy. It should be pointed out that all signal timings in this study are obtained based on QEM, which focuses on handling the critical movement or the movement with the largest flow rate. Other tuning methods with different considerations may result in different signal timings although the results are expected to be similar.

Table 6.3 Actual Green Splits for Different Left-turn Volumes

$\bar{p}_{LT}$	Scenario	Green Time for Different Phases (sec)			
		ø2 or ø5	ø4 or ø7	ø1 or ø6	ø3 or ø8
0.2	Non-DLG	24	25	26	25
	DLG	24	25	26	25
0.4	Non-DLG	31	22	25	22
	DLG	25	25	25	25
0.6	Non-DLG	38	19	25	18
	DLG	26	25	25	24
0.8	Non-DLG	44	16	24	16
	DLG	26	25	25	24

Simulation results on average delay and number of stops are shown in Figure 6.6. As the left-turn volume deviates from the baseline model (i.e., 240 veh/hr or 20% of total volume on approach 1), the MOEs increase considerably for the non-DLG scenarios while the remain practically unchanged under the DLG scenarios. In particular, when the left-turn volume reaches 960 veh/hr (i.e. $p_{1,LT} = 0.8$), the average delay and number of stops are three folds for the non-DLG scenarios, compared with the DLG scenarios.



(a) Average vehicle delay **(b) Average number of stops**

Figure 6.6 Intersection Performance: Existing vs. DLG Strategies with QEM Adaptive Timings

Table 6.4 summarizes the results on fuel consumption and vehicle emissions. As can be seen from the table, the fuel consumption is reduced by up to 65%, and the vehicle emissions of CO, HC, and NO_x are reduced by as much as 4.75%, 18.45%, and 32.94%, respectively, if the DLG strategy is applied.

Table 6.4 Simulation Results: Fuel and Air Pollutant Emissions

Pollutant	Strategy Scenario	Left-Turn Traffic Volume Proportion			
		0.2	0.4	0.6	0.8
Fuel (g)	Non-DLG	112.28	115.00	144.07	190.85
	DLG	112.28	112.34	113.66	115.13
	% change	0	-2.31	-21.10	-65.76
CO2 (g)	Non-DLG	330.11	338.38	430.56	578.91
	DLG	330.11	329.98	333.97	338.71
	% change	0	-2.48	-22.43	-41.49
CO (g)	Non-DLG	17.07	17.34	17.68	18.28
	DLG	17.07	17.24	17.39	17.41
	% change	0	-0.57	-1.60	-4.75
HC (g)	Non-DLG	0.306	0.312	0.339	0.382
	DLG	0.306	0.309	0.311	0.312
	% change	0	-0.731	-8.145	-18.451
NOx (g)	Non-DLG	1.082	1.106	1.327	1.676
	DLG	1.082	1.075	1.098	1.124
	% change	0	-2.802	-17.231	-32.943

6.4 Discussion

The results from the numerical analysis show that the proposed DLG strategy can potentially achieve significantly better performance in terms of maximum flow ratio and average delay. For example, for the three-lane cases, the reduction of maximum flow ratio is 41.2% when the spatial variation of demand is 40%, and the reduction in average delay with fixed timing and adaptive timing is 35.1% and 22.8% respectively. DLG produces higher benefits when the intersection becomes oversaturated.

The evaluation of the DLG strategy using microscopic traffic simulation confirms the findings from the numerical analysis and shows that the DLG strategy can provide significant energy/environmental benefits. The mobility and environmental benefits increase as turning movement volumes increase from the baseline demand pattern.

The implementation of DLG may introduce adverse safety and capacity related impacts. For example intensive lane changes upstream of the intersection may reduce capacity and increase conflicts. These impacts may be mitigated with advance information to drivers regarding the lane configuration at the intersection approach through in –vehicle devices. CV technology with origin-destination information for each vehicle can support such functions. The DLG concept can be further combined with dynamic lane assignment, in which case approaching vehicles are assigned to their specific lanes based on their destination in real-time.

CHAPTER 7

TRAFFIC SIGNAL CONTROL FOR SAFETY:DYNAMIC SALL RED EXTENSION (DARE)

7.1 Introduction

Red-light-running (RLR) at signalized intersections is a major safety issue nationwide. RLR accounted for 10 percent of all intersection-related fatalities in 2009 [41]. Efforts to address the RLR problem include engineering countermeasures [42], automated enforcement [43], and most recently in-vehicle warning systems that utilize infrastructure-to-vehicle (I2V) communication [44].

Adding an all-red clearance interval is an intuitive mitigation measure for RLR related crashes, especially for those caused by willful or inattentive drivers who enter the intersection during the first few seconds of the red interval. Studies show that implementing a static longer all-red interval did reduce the crash rate at due to RLR initially, but gradually crashes returned to previous levels over time, as drivers became aware that a longer all-red phase is provided, so they enter the intersection late into yellow and/or run a red light by knowing that there is a time interval with no conflicting movements [45].

Dynamic all-red extension (DARE), on the other hand, is designed to be only triggered when the likelihood of collision to happen is high, to protect other vehicles from potential RLR related collisions (Figure 7.1). DARE improves intersection safety while avoiding drivers to game with the system as the on-demand all-red extension is not likely to be observed by drivers. The benefits of such a dynamic method therefore rely in large part on reliable detection of potentially hazardous situation caused by RLR. The information provided by CV data may provide the required data for DARE.

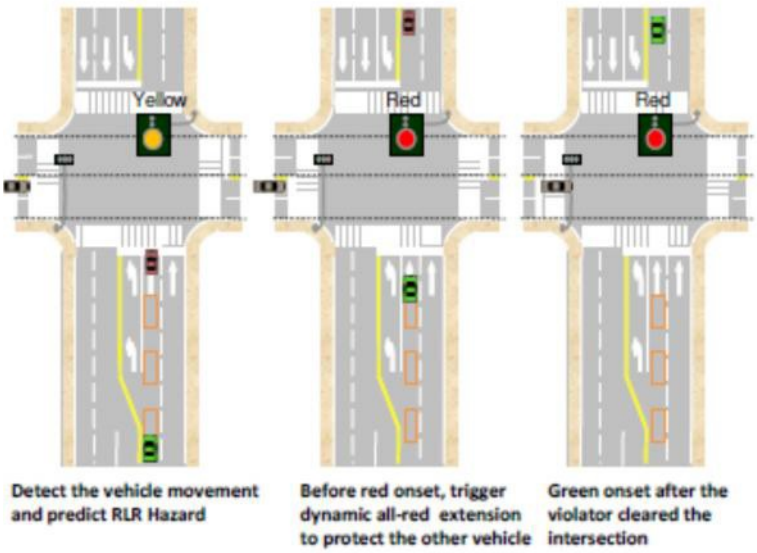


Figure 7.1 Illustration of Dynamic All-Red Extension (DARE)

This Chapter presents the development and evaluation of a DARE algorithm based on CV data. Detailed description of the algorithm is included in Appendix C.[8]

[8]Zhang L., L. Wang, K. Zhou and WB. Zhang "Dynamic All-Red Extension at Signalized Intersection: A Framework of Probabilistic Modeling and Performance Evaluation," IEEE Transactions on ITS, 2011.

7.2 Framework for Probabilistic Modeling and Performance Evaluation for DARE

Figure 7.2 illustrates the trajectory of a RLR vehicle. In the time-space diagram, t_{inred} is the violation time into the red interval, t_c is the time duration for the RLR vehicle to travel from the stop bar to clear the conflict zone, and t_e is the time a conflicting vehicle entering the conflict zone, referenced to the red onset. The Post-Encroachment Time, PET $= t_e - t_c$, provides a surrogate measure for the conflict zone. The smaller this value is, the higher the probability that a collision will result. In the extreme case, a negative value of PET indicates that the vehicle stream from the conflict movement will enter the conflict zone before the RLR vehicle could leave the zone.

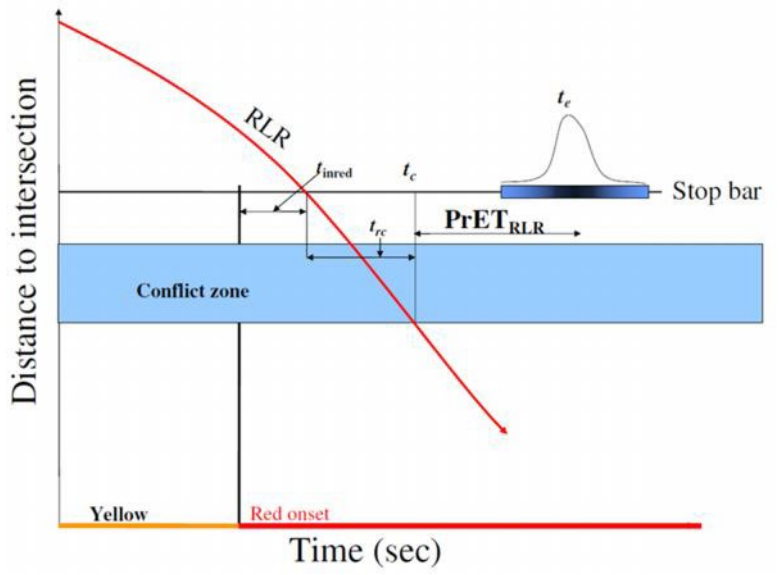


Figure 7.2 Time-Space Diagram of a Conflict Zone

PET is an objective indicator that is extensively used to quantify the severity of near-miss situations and, when below a certain level, indicates a high likelihood of crashing [46]. The traditional PET defines scenarios when a subject vehicle (SV) and a principal other vehicle (POV) are both moving at known speeds. The dynamic all-red extension (DARE) system needs to decide whether or not to extend the all-red interval before a green onset for the conflicting traffic when a POV is mostly likely still waiting to enter the intersection.

We define "Predictive Encroachment Time for RLR", or $PrET_{RLR}$ as the predicted time lapse between when the SV (RLR vehicle) clears the conflict zone and the *a priori* entry time of the POV enters at it. $PrET_{RLR}$ itself is a random variable which is a function of the speed and location of the vehicle, and the *a priori* probability distribution of the entry time of the POV. The system extends the all-red interval to avoid such when:

$$P\{\mathrm{PrET}_{\mathrm{RLR}} < \delta_0\} > P_{th}$$

where δ_0 is a pre-set constant below which the conflict will be considered as a hazardous situation likely to turn into a collision.

The empirical probability distribution of t_e can be estimated from field data, in cumulative format, as

$$F_{t_e}(x) = P\{t_e < x\}$$

Example empirical distributions of t_e is shown in Figure 7.3 based on field observations at the signalized intersection of El Camino Real and Page Mill Rd in Palo Alto, CA.

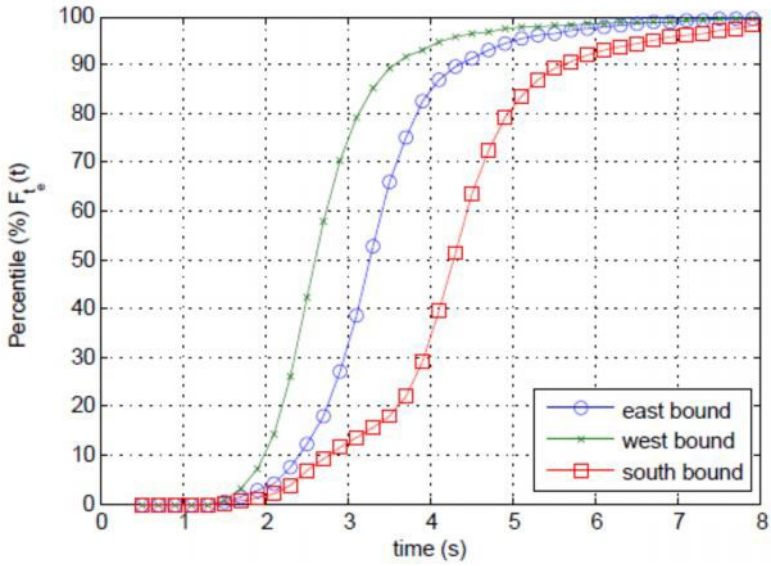


Figure 7.3 Entry Time Distribution –El Camino Real/Page Mill Rd Intersection

Given that t_c can be estimated from RLR vehicle trajectory, conditional on t_c, we have

$$P\{\text{PrET}_{\text{RLR}} < \delta_0 | t_c\} = P\{t_e - t_c < \delta_0 | t_c\} = F_{t_e}(t_c + \delta_0)$$

Therefore, the dynamic all-red extension condition can then be expressed as the followirg equation based on the time-space diagram and empirical distribution of t_e:

$$F_{t_e}(t_c + \delta_0) > P_{th} \Leftrightarrow \tau + t_{rc} + \delta_0 > F_{t_e}^{-1}(P_{th}) \Leftrightarrow \tau > F_{t_e}^{-1}(P_{th}) - \frac{d_c}{v_{\text{RLR}}} - \delta_0$$

where d_c is the distance between the stop bar and the outside boundary of the conflict zone for the RLR vehicle, v_{RLR} is the speed the RLR vehicle traveling through the intersection, and τ is the optimal threshold for violation time into red to serve as the trigger point beyond which the all-red interval needs to be extended. In other words, when $t_{\text{inred}} > \tau$, the all-red interval needs to be extended for $t_{\text{inred}} - \tau$ seconds. In operation mode, t_{inred} needs to be predicted based on vehicle trajectory at the time before the end of the current all-red interval.

The algorithm for the DARE system is illustrated in Figure 7.4. It includes two parts a) an offline parameter-setting procedure which collects empirical data to get the parameters of the probabilistic models as well as in stop-and-go prediction, and b) an online algorithm to predict RLR hazard using either advanced point detector outputs or CV data, and the parameters extracted from the offline procedure.

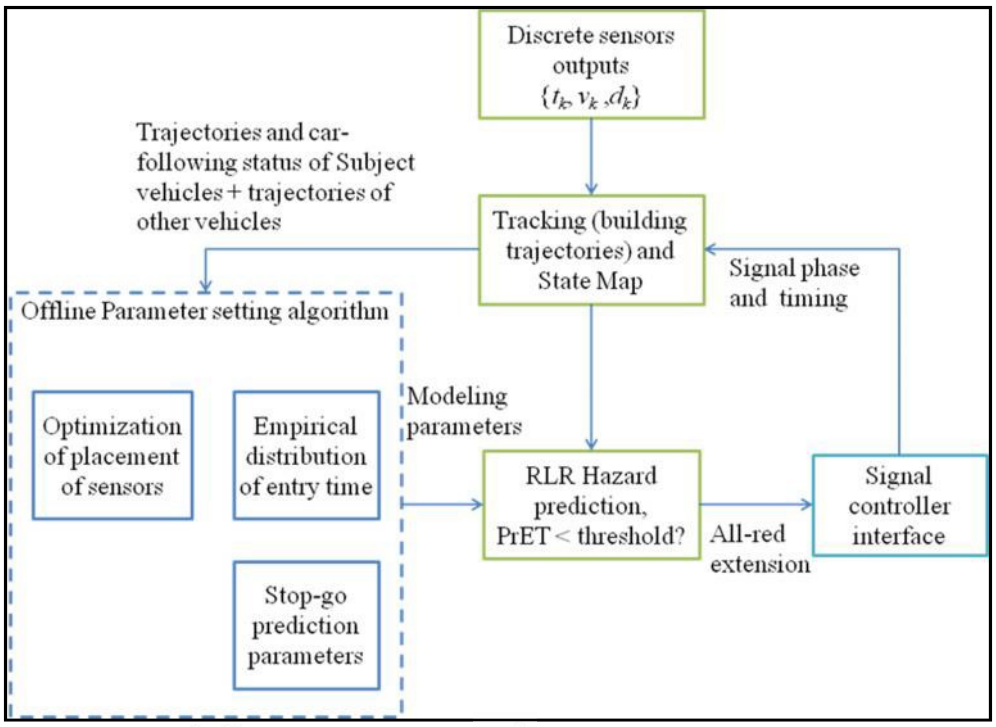


Figure 7.4 Control Algorithm for Dynamic All-Red Extension (DARE)

For each approaching vehicle, at time t_0, the state $\boldsymbol{X}$ used for RLR hazard prediction is extracted from its trajectory data $Z_{t \le t_0}$. The form of $\boldsymbol{X}$ depends on the vehicle detection means. For example, with two point detectors which are capable of detecting vehicle speed, the state $\boldsymbol{X}$ can consist of the average speed and the average acceleration, estimated from the speed measurement and the timestamps when the vehicle was detected by the point detector, and a binary indicator of car-following status, which is determined based on headway measurement obtained from the point detector.

Because many random factors contribute, $\boldsymbol{X}$ can be reasonably hypothesized to be a bivariate normal random vector. Thus to formulate the hypotheses, we have

$$
\begin{aligned}
&\mathbf{H}_s : \{\mathbf{H}_{s,\mathrm{F}}, \mathbf{H}_{s,\mathrm{NF}}\} \\
&\quad \text{where } \mathbf{H}_{s,\mathrm{F}} : \mathbf{X} \sim \mathcal{N}(\mathcal{U}_{s,\mathrm{F}}, \Sigma_{s,\mathrm{F}}) \\
&\qquad \mathbf{H}_{s,\mathrm{NF}} : \mathbf{X} \sim \mathcal{N}(\mathcal{U}_{s,\mathrm{NF}}, \Sigma_{s,\mathrm{NF}}) \\
&\mathbf{H}_g : \{\mathbf{H}_{g,\mathrm{F}}, \mathbf{H}_{g,\mathrm{NF}}\} \\
&\quad \text{where } \mathbf{H}_{g,\mathrm{F}} : \mathbf{X} \sim \mathcal{N}(\mathcal{U}_{g,\mathrm{F}}, \Sigma_{g,\mathrm{F}}) \\
&\qquad \mathbf{H}_{g,\mathrm{NF}} : \mathbf{X} \sim \mathcal{N}(\mathcal{U}_{g,\mathrm{NF}}, \Sigma_{g,\mathrm{NF}})
\end{aligned}
$$

where **Hs** and **Hg** are hypotheses for "stopping" and "going". $N_{s,F/NF}$, $\Sigma_{s,F/NF}$, $N_{g,F/NF}$, and $\Sigma_{g,F/NF}$ are the mean and covariance matrices of $\boldsymbol{X}$ under the two hypotheses, which can be obtained from empirical data. The additional subscripts $_F$ and $_{NF}$ denote the "following" and "non-following" group, respectively.

The prediction of RLR hazard is then defined as

$$
D_H(X) = \begin{cases} 1, & \text{if } D(X) = H_g \text{ and } \widehat{t_{\text{inred}}} > \tau \\ 0, & \text{otherwise} \end{cases}
$$

where $D(X) = \{H_s \text{ or } H_g\}$. There are two types of prediction errors, namely, the missing report and false alarm. The probabilities of these two types of errors can be formulated as (missing P_{M-H}, false alarm P_{FA-H})

$$P_{M-H} \triangleq P\{D(X) = H_S | H_g,\ t_{\text{inred}} > \tau\} + P\{D(X) = H_g,\ \widehat{t_{\text{inred}}} \leq \tau | H_g,\ t_{\text{inred}}\}$$

$$P_{FA-H} \triangleq P\{D(X) = H_g,\ \widehat{t_{\text{inred}}} > \tau | H_g,\ t_{\text{inred}} < \tau\} \cdot P\{H_g,\ t_{\text{inred}} < \tau\} + P\{D(X) = H_g,\ \widehat{t_{\text{inred}}} > \tau | H_s\} \cdot P\{H_s\}$$

The Neyman-Pearson criterion is used for the optimization of RLR hazard prediction. Under the NP criterion, a decision rule is found to have the maximum probability of correct detection while not allowing the probability of a false alarm to exceed a certain value, i.e.,

$$\max(1 - P_{M-H}) \quad \text{such that } P_{FA-H} < \theta$$

where θ is a preset threshold of an allowable false-alarm rate.

7.3 Field Data Collection and Processing

7.3.1 Test Site

A field data collection system was deployed at the intersection of San Pablo Avenue and Brighton Avenue in the city of Albany, CA. The collection system includes three SmartMicro radar sensors (SMS) which can collect all vehicle trajectories of approaching, inside the intersection box, and leaving the intersection. Figure 7.5 shows the placement of radar sensors at the study intersection. The maximum detection range of radar sensor is 660 ft. (200 meters) and vehicle locations are updated every 50 milliseconds.

Figure 7.5 Placement of Radar Sensors: San Pablo Avenue and Brighton Avenue, Albany, CA

At the test intersection, there are two through lanes and one left-turn lane on both southbound and northbound approaches on San Pablo Ave, and one traffic lane on Brighton Ave (Figure 7.6).

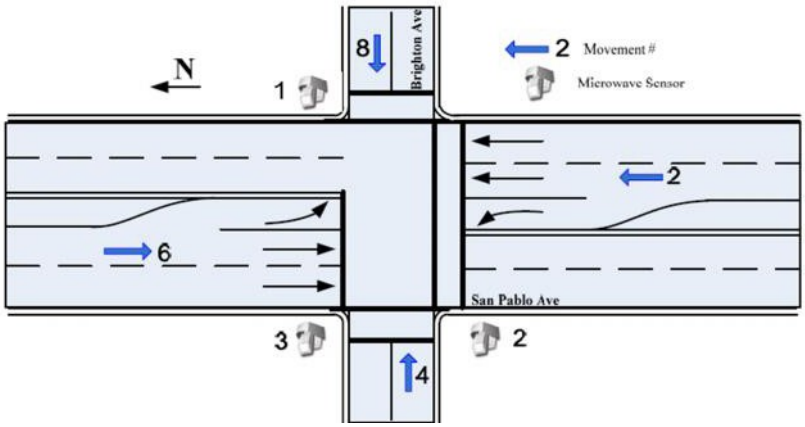


Figure 7.6 Test Intersection Geometrics and Signal Phasing

The intersection is under actuated traffic signal control with three time-of-day coordinated signal timing plans running between 6:30 AM and 8:00 PM. The signal is running as isolated fully actuated ("free") from 8:00 PM to 6:30 AM. Table 7.1 shows the signal settings and time periods that each plan is in operation. The yellow interval is 4 seconds on phase 2, 6 and 8, and 3 seconds on phase 4, which is longer than the value recommended by ITE [47], with a speed limit of 35 mph on San Pablo Ave (phase 2 and 6) and 25 mph on Brighton Ave. There are no all-red intervals. For the actuated phases 4 and 8, the minimum green interval is 6 seconds and the maximum green extension interval is 20 seconds.

Table 7.1 Signal Timing Plans at Test Site

Timing Plan #	Signal Settings			Time Period
	Cycle Length (sec)	Offset	Split (phase 4, 8)	
1	108	63	(25, 25)	6:30 – 9:00 AM 11:00 AM – 2:00 PM
2	98	65	(25, 25)	9:00 – 11:00 AM 7:00 – 8:00 PM
3	114	3	(25, 25)	2:30 – 7:00 PM

7.3.2 Data Acquisition and Processing

In addition to high-precision trajectory data, the conflict monitor card inside the signal controller cabinet was upgraded to allow output of the traffic signal status (i.e., Green, Yellow, and Red) for each permitted signal phase every 60 milliseconds. A traffic and signal data acquisition system was developed to collect synchronized vehicle trajectory data and signal status data in real-time, and archive the data in a data server located in PATH headquarters (Figure 7.7).

The data acquisition system records a snapshot of the information of all detected vehicles and signal status every 50 milliseconds. Each record includes the following data elements:

- Timestamp (milliseconds since the Epoch time)
- Signal color (Green, Yellow, or Red) for each permitted phases
- Vehicle ID, location (x, y), and speed (v_x, v_y) for every detected vehicles

The coordinate system has its origin located at the center of the intersection, with *x*-axis pointing to northbound on San Pablo Ave (Main-street) and *y*-axis pointing to westbound on Brighton Ave (Side-street).

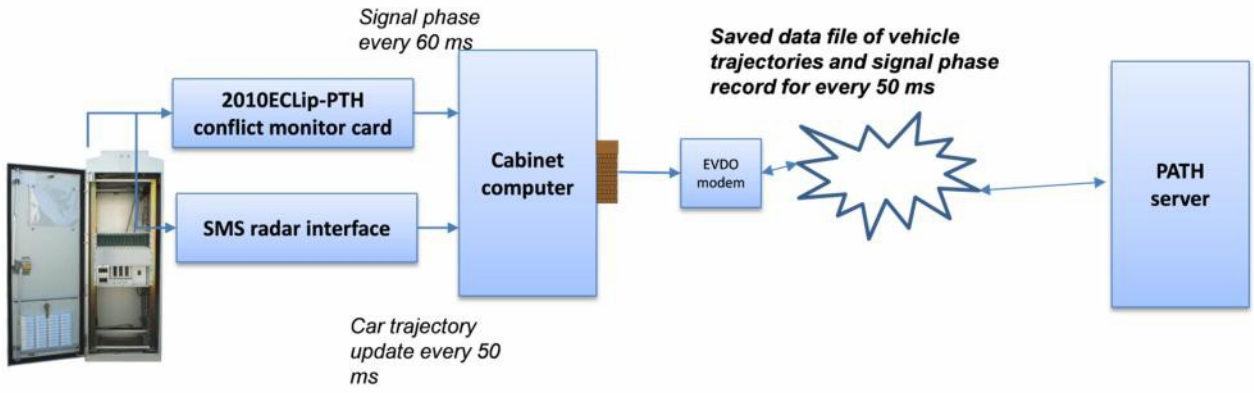


Figure 7.7 Traffic and Signal Status Data Acquisition System—Albany Intersection

Figure 7.8 plots the vehicle location (one dot represents one location measurement) in the *x*-*y* coordinate system for a 30-minute interval, and Figure 7.9 plots sample trajectories for vehicles traveling southbound on San Pablo Ave in a 10-minute interval. Phase-4 movement on Brighton Ave (the top-center part in Figure 7.8) is an entrance/exit to an off-street parking lot and had very few traffic, when compared with traffic on other movements (i.e., phase 2, 6 and 8). These two plots show that the SmartMicro radar system has good range accuracy (usually within 0.5 meters) and good tracking.

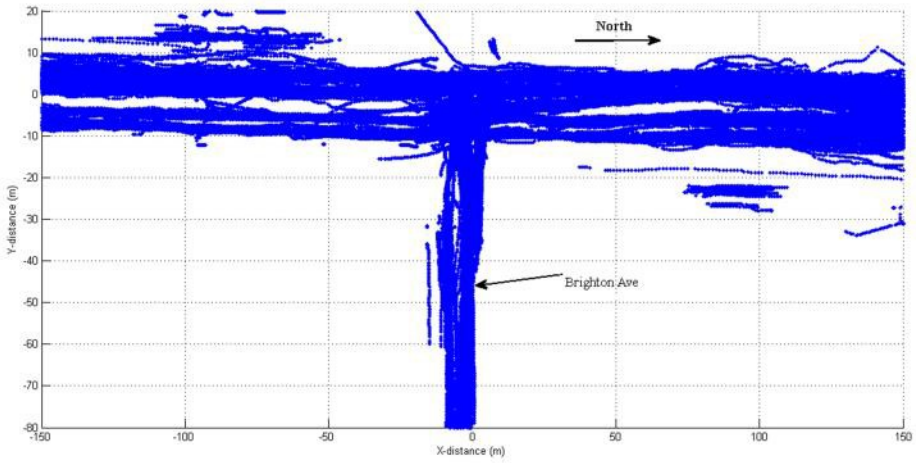


Figure 7.8 Radar Sensing Range at Test Site: San Pablo Ave/Brighton Ave

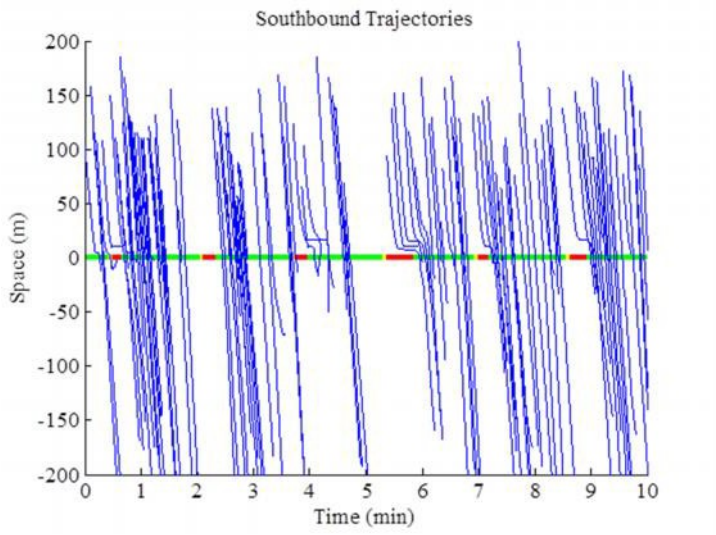


Figure 7.9 Sample Vehicle Trajectories: San Pablo Ave Southbound

Data post-processing tools were developed to filter out sensing noise, remove the trajectories that did not travel through the intersection, and to associate broken trajectories happened when switching from the detection range of one radar sensor to another. Four months of field data from March to June, 2012, were processed to extract relevant vehicle trajectories. Red-light-running (RLR) vehicle trajectories were also extracted from the filed data. A RLR event is defined as a vehicle entered the intersection 100 milliseconds after the red. The threshold of 100 milliseconds was applied in order to reduce the false RLR detection due to the different updating rates on radar sensing and signal status sensing. Table 7.2 shows the mean and coefficient of variation (CV) of daily traffic and RLR occurrences during the four-month data collection period. During the four months, a total of 349 RLR events occurred on northbound and 514 on southbound. On average, there were 38 northbound RLRs and 54 southbound RLRs per 1,000 approaching vehicles.

Table 7.2 Statistics of Daily Traffic Count and RLR Occurrences

		Daily Traffic (# of Vehicles)	Daily RLR Occurrences (# of events)
San Pablo Northbound	Mean	9,062	4
	CV	14.6%	59.6%
San Pablo Southbound	Mean	9,516	5
	CV	15.8%	57.8%

The extracted vehicle trajectory and signal status data were used as basis of CV ground truth for the development of proposed control concepts for safety.

7.4 Evaluation of the DARE Algorithm

7.4.1 El Camino Real/Page Mill Rd Test Site

Field data collected from the intersection of Page Mill Rd and El Camino Real in Palo Alto, California were applied to calibrate the probabilistic models, and to evaluate the performance of the DARE algorithm. Field data were collected using Autoscope QR cameras. Nine cameras were installed at three approaches to the intersection. We note that the signal phase and timing setting of this intersection fully

conforms to the ITE standard. We summarize the configurations into Table 7.3, where EB, WB and SB stand for east bound, west bound and south bound, respectively. As can be seen in Table 7.3, multiple virtual speed loops are placed at advanced areas of difference approaches to monitor the violation and also at stop area to get the conflict entry time behaviors.

Table 7.3 Configuration of the El Camino Real/Page Mill Rd Intersection

Page Mill rd. at El Camino Real			
	EB	WB	SB
Lanes (recorded)	2	1	2
Speed Limit (mph)	30	35	35
Average Speed (mph)	30	35	35
90% speed less than (mph)	45	45	40
Yellow Duration (s)	4	4	4
All-red Interval (s)		0.5	
# Speed Loops (Exit)/ Lane	1	1	1
# Speed Loops (advance) / Lane	3	3	3
# Speed Loops (stop / conflict entry) / Lane	3	3	2
95% RLR time-into-red less than (s)	2.5s	2.5s	2.5s
Description of the approach	0.15 Mile from upstream intersection	No upstream intersection for More than 0.7 mile upstream	0.25 mile from upstream intersection, main street

Offline performance evaluation is conducted to verify the RLR hazard prediction algorithm. Data collected from the field are used to generate the parameters for each different approach of the site. Additional data are collected to verify the algorithm. Data were collected from May to October 2008 and from April to May, 2009. Data were then divided into two groups for parameter setting and validation purposes. Figure 7.10 shows the online performance of hazard prediction for the southbound direction at the test intersection.

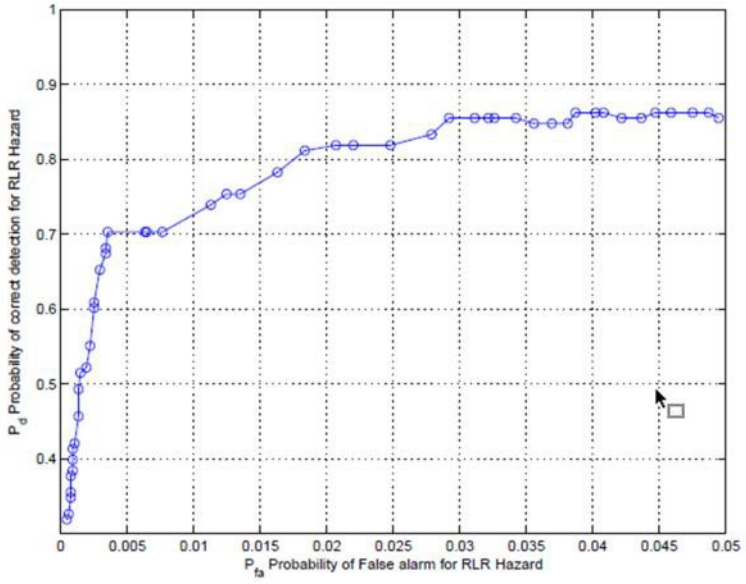


Figure 7.10 SOC Curve for RLR Prediction: El Camino Real/Page Mill Rd

In summary, the evaluation results show that for RLR hazards that meet the following condition:

$$P\{\mathrm{PrET}_{\mathrm{RLR}} < 1 \text{ sec}\} > 30\%$$

the correct prediction rate is 70 to 80% with a false alarm rate less than 5%.

The false alarm rate is calculated from the samples that could potentially contribute to RLR related safety hazards at the intersection. It is only considered when a candidate vehicle approaches the intersection during the late yellow interval, and if it does not stop, it could enter the intersection well after the red onset (e.g., 1 second after the red onset). Only this small number of candidate vehicles would possibly become a false report; the majority of the vehicles won't contribute to this false report rate.

There are mainly two types of false reports, one is that a vehicle that arrive late into yellow with high speed was predicted to be a hazard but it later stopped, or it accelerated hard enough to go through the intersection without causing hazardous situation for other vehicles.

Regarding the physical interpretation of the 5% false alarm rate as frequency, it is approximately about one false trigger of the all-red extension in about eight hours, with about 18,000 vehicles entered the intersection during that time or more than 240 signal cycles. The main negative effect of a "false trigger" for extending the all-red interval is two additional seconds of delayed green on the side street.

7.4.2 San Pablo Avenue/Brighton Avenue Test Site

CV data make it possible to extract vehicle trajectory samples at different locations (sampled before the red onset), rather than at fixed locations as with the discrete sensors case. The advantage of having continuous measurement for DARE would be the capability of always making the prediction at the “last second” before red onset, resulting in higher accuracy of stop-and-go prediction and RLR hazard prediction consequently.

Figure 7.11 illustrates the differences of applying the DARE framework with discrete sensors at fixed locations versus the case when continuous measurements are available, such as the measurements from connected vehicles.

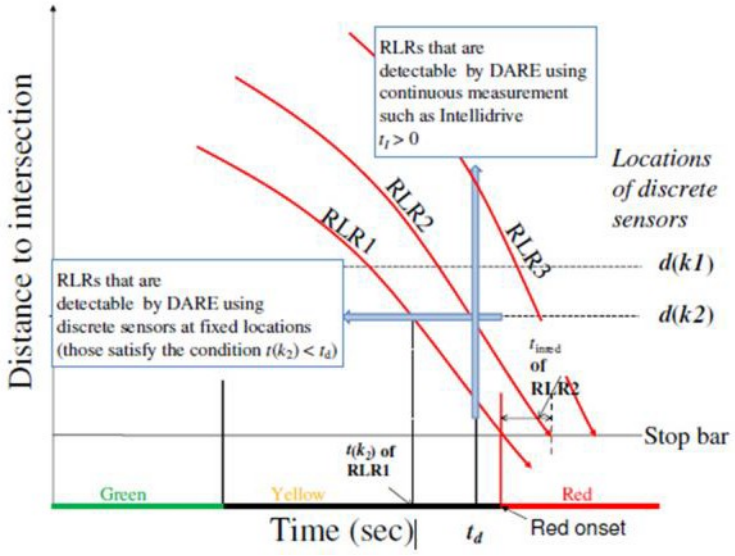


Figure 7.11 Time-Space Diagram for DARE: Discrete Sensors vs. CV Data

When discrete sensors at fixed locations are used, the DARE technology has the restriction of only being able to detect RLR vehicles that satisfy the following condition (marked as the horizontal arrow in Figure 7.11):

$$t(k_2) < t_d$$

where $t(k_2)$ is the time when the subject vehicle hits the sensor that is closest to the intersection and t_d is a preset "deadline" time after which it is too late to extend the all-red interval. Practically t_d is a time point slightly before the end of the current all-red interval on SV approach with enough buffer for the signal controller to enact the extension of an all-red period. From Figure 7.11, one can see the following benefits of Connected Vehicle based DARE relations:

- More detectable trajectories. RLR1 and RLR2 are detectable by the DARE technology using discrete sensors at $d(k_1)$ and (k_2) as predictors. For a Connected Vehicle based system, all three trajectories (RLR1, RLR2 and RLR3) can be detected
- Possibly better accuracy. For a CV-based system, it is possible for the system to detect RLR hazards based on vehicle samples that are closer to the intersection, which are the cases for both RLR1 and RLR2. The difference of the mean values of statistics of stopping trajectories versus going-through trajectories become smaller when it is closer to the intersection. Moreover, the variance of both groups of stopping and going through vehicles also decrease, and all together help to increase the accuracy of stop-versus-go prediction.

High precision microwave data collected from the Albany intersection were applied to evaluate the performance of proposed DARE algorithm, and to compare with that obtained from discrete sensors. Relevant vehicle trajectories, including the first stopped vehicle per cycle per approach, going-through vehicles during the yellow phase, and RLR vehicles, were extracted from the four-month data set. The number of extracted relevant trajectories is shown below:

Table 7.4 Vehicle Trajectories Used in the DARE Evaluation

Direction	First Stopped (#)	Yellow Through (#)	RLR (#)
Southbound	15,793	9,769	514
Northbound	12,476	7,181	349

In order to emulate discrete detection using the high precision trajectories data, two locations for discrete detectors were determined as $d_1 = 30m$ and $d_2 = 20m$ from the stop bar. The relevant vehicle trajectory data were further extracted as the same format as they were detected by these two fixed-location point detectors. When a vehicle crossed the line d_1 or d_2, we recorded the vehicle ID, timestamp and vehicle's current speed. Vehicle ID was used to link detection information from the two locations. Parameters used for RLR hazard prediction are the same as that was used for El Camino Real site, i.e., the average acceleration and average speed when crossing the two detection locations.

$$X \triangleq \left\{\bar{a} = \frac{v_2 - v_1}{t_2 - t_1}, \bar{v} = \frac{v_2 + v_1}{2}\right\} \quad (1)$$

The DARE algorithm was applied to the emulated discrete detection data. We found that about half of the RLR cases were not detectable by the discrete point detection, which happened when a RLR vehicle traveling across line d_2 after the signal turned green on the side street. Unlike the El Camino Real intersection, the Albany intersection had light traffic on the side street, and consequently, drivers could have lower threat of involving a right-angle crash when running a red light.

When applying the DARE algorithm with the CV data, the parameters as shown in Equation (1) cannot be used, as it does not contain the information of vehicle location. With CVdata, the operations of the DARE

system is assumed to make the RLR prediction for all vehicles within the communication range, 500 milliseconds before the end of the current all-red interval. The 0.5 seconds reserved as a buffer to consider the communication delay and other factors such that an all-red extension request can be responded and served by the traffic control. No existing all-red interval was implemented at the Albany intersection, thus the DARE system makes prediction of RLR 0.5 seconds before the end of yellow in every signal cycle. At time t when the DARE algorithm is performing the prediction of RLR, it extracts the latest reported vehicle location and speed data, and the earliest location and speed data within 1 second. The average speed and the average acceleration between these two location reports can be calculated as shown in Equation (1) and are used to form the parameters for RLR prediction: the difference between required deceleration to stop and the average vehicle deceleration, and the estimated vehicle's time-to-arrival at the stop bar with respective to the end of yellow (or end of the current all-red interval if it is implemented), i.e.,

$$X \triangleq \{x_1 = A_{req} + \bar{a},\ x_2 = Y_r - \frac{d}{\bar{v}}\} \qquad (2)$$

where $A_{req} = \frac{d^2}{2\bar{a}}$, and d is the latest reported vehicle's distance to the stop bar.

Figure 7.12 plots parameter $\boldsymbol{X}$ at the instant of RLR prediction on San Pablo Southbound. Each dot represents one vehicle sample, and the color of the dot corresponds to the type of the trajectory: red for RLR vehicle, yellow for yellow though vehicle, and black for the first stopped vehicle. The bottom plot is enlarge of the top plot. Figure 7.13 clearly shows different patterns for the three types of trajectories, with RLR vehicles separating the first stopped vehicles and the yellow through vehicles. At 500 milliseconds before the end of yellow, drivers already made the stop-and-go decision and are not likely to change mind (abrupt vehicle maneuver) in the "last second". This is one of the great advantages of using CA data for DARE operations that the RLR prediction accuracy could be significant improved. Figure 7.13 plots parameter $\boldsymbol{X}$ for the northbound direction, which is very similar to the southbound direction.

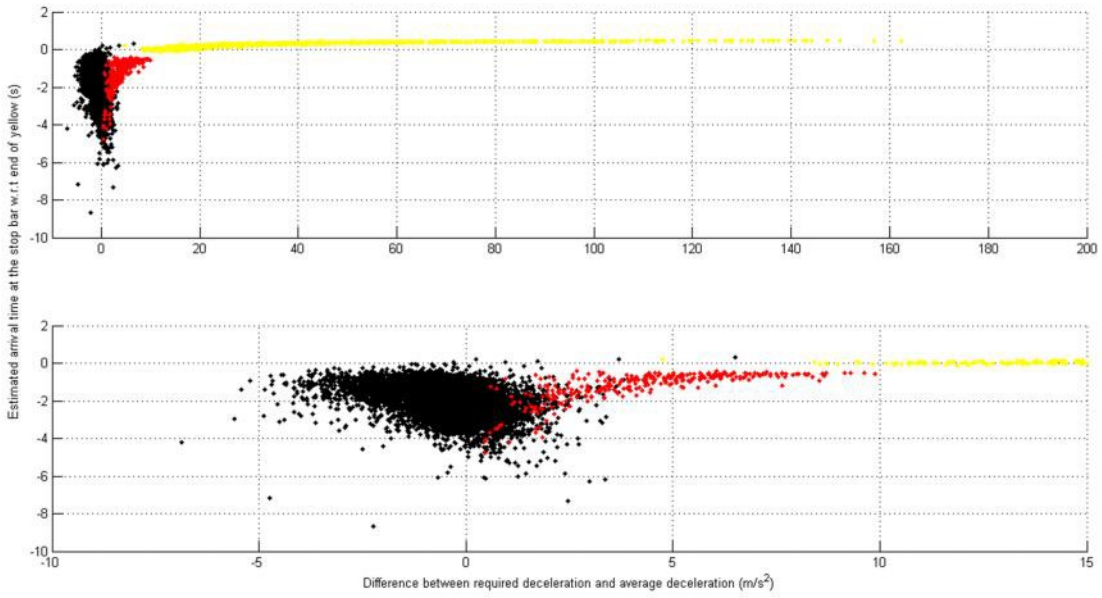


Figure 7.12 Deceleration Difference vs. Travel Time Difference: San Pablo Southbound

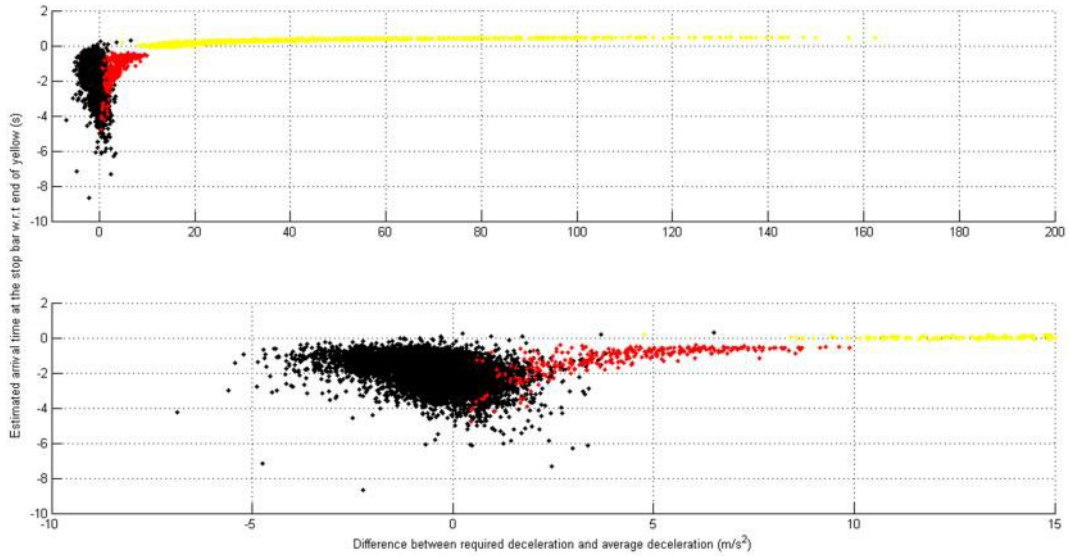


Figure 7.13 Deceleration Difference vs. Travel Time Difference: San Pablo Northbound

The performance of DARE algorithm for RLR prediction on both directions of San Pablo Avenue is shown in Figure 7.14. It shows that, with CV data, the correct prediction rate is higher than 95% with a false alarm rate less than 5%.

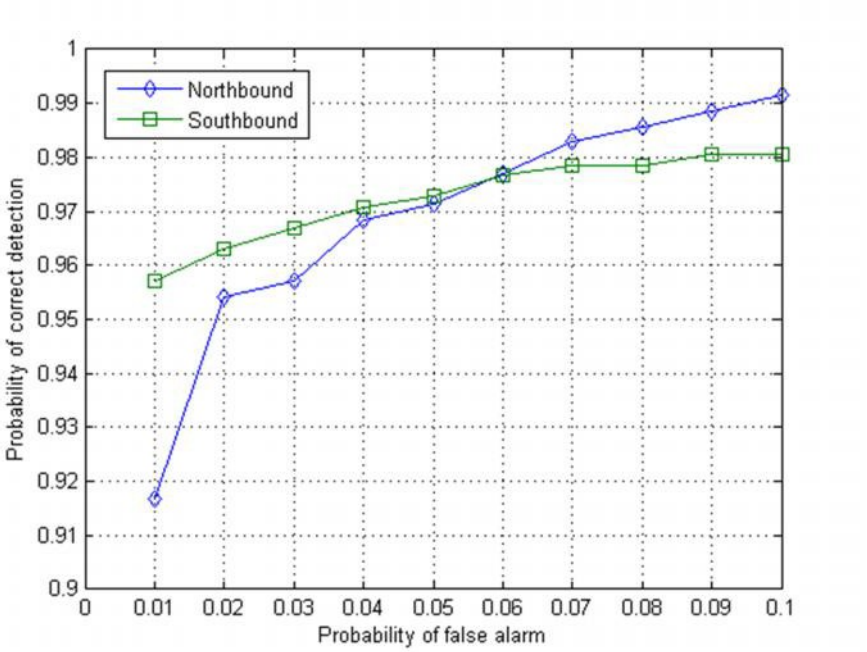


Figure 7. 14 SOC Curve for RLR Prediction: San Pablo Avenue/Brighton Avenue

7.4 Discussion

A probabilistic framework and prediction algorithm have been developed for Dynamic All-Red Extension (DARE), to counteract the intersection safety hazard caused by RLR. We defined the $PrET_{RLR}$, which is the time lapse between the time for the RLR vehicle to clear the conflict zone and the a priori entry time of the first-to-enter vehicle from conflict approach, as the primary index in the probabilistic model to identify RLR hazards. The hazard definition, which is determined by the intersection layout, empirical entry time of conflict traffic and trajectory characteristics of RLR vehicles, is linked to an important characteristic of RLR: the time into red of the violator. We then developed a model using a minimum set of two point sensors to identify the RLR hazard with speed measurements and car-following status. The prediction of the hazard is governed by the Neyman-Pearson criterion which gives the algorithm the capability of distinguishing two types of errors, namely the missing report and false alarm, and the balance of their trade- offs.

We collected detailed field data to study driver behavior related to the RLR hazard and to set the parameters of the developed model/ algorithm. The performance of the proposed framework is evaluated using field collected data from two intersections in the San Francisco Bay Area: El Camino Real/Page Mill Rd in Palo Alto, and San Pablo Ave/Brighton Ave in Albany. The evaluation results showed that the proposed algorithm can effectively detect RLR hazards. At the El Camino Real site with two fixed location point detector, at a false alarm rate of less than 5% (or equivalently once per about eight hours), the algorithm reached a correct detection rate from over 70% to over 80% for various legs, while at the San Pablo Ave site with continuous detection (which is similar to CV data), the algorithm reached a correct detection rate over 95% at a false alarm rate of less than 5%.

CHAPTER 8

TRAFFIC SIGNAL CONTROL FOR SAFETY: MINIMIZATION OF VEHICLE ARRIVALS IN THE YELLOW INTERVAL

8.1 Introduction

Empirical studies at a signalized intersection on El Camino Real in San Mateo, CA indicate that the arrival flow during the yellow change interval approximately 200 ft. (60 meters) from the stopline is the most significant factor for red light running (RLR) frequency [48]. Intuitively, higher yellow arrival flow indicates more drivers need to make stop-and-go decision and consequently correlated with higher RLR frequency.

This Chapter describes additional data analyses from other test sites to determine possible correlation between vehicle arrivals at the yellow interval and RLR frequency, and the development and evaluation of signal control strategies to minimize the arrival of vehicles during the yellow interval along signalized arterials based on CV data. Detailed description of the work performed and the results are included in Appendix C.[9]

We investigated the relationship of the arrival flow in the yellow interval on RLR frequency using the NGSIM data set from Peachtree Blvd, described in Section 3.3.1. Figure 8.1 shows vehicle trajectories traveling on the southbound direction of Peachtree Blvd during the 4:00 – 4:15 pm peak period. Three types of vehicle trajectories are plotted: red for RLR vehicles, yellow for vehicles entered the intersection during the yellow interval, and cyan for the first stopped vehicles. As it is shown in Figure 8.1, the last (4^{th}) intersection has a high RLR frequency, where the head of the approaching platoon was truncated with the onset of red. In this scenario, drivers are likely to follow the head of the platoon but ended up with running a red signal. Drivers are also likely to run into the red light when the tail of the platoon is truncated by the onset of red.

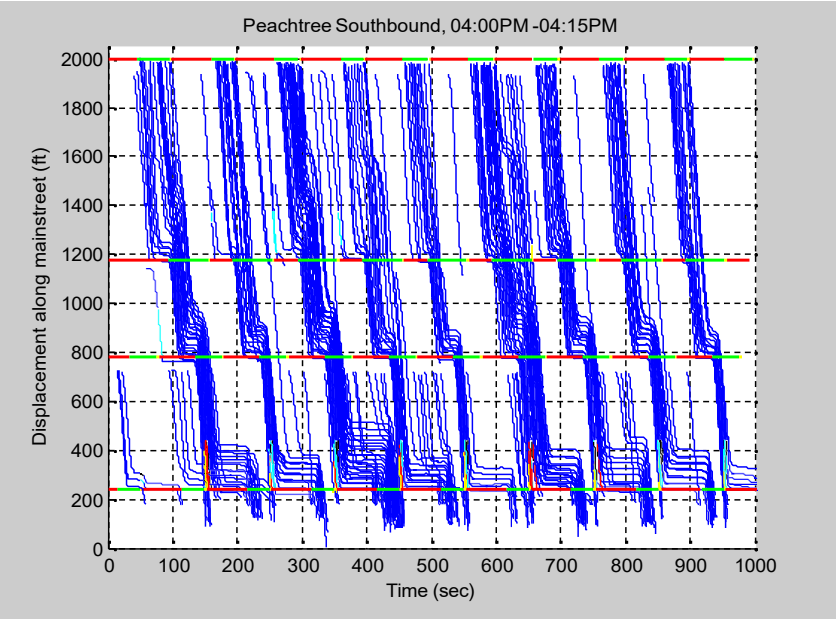


Figure 8.1 Vehicle Trajectories: Peachtree Blvd SB (4:00 – 4:15PM)

[9] Zhang, J., J. Hu, K. Zhou and WB. Zhang, "Signal Timing Optimization Based on Platoon Segmentation by Using High Precision Microwave Data," 92nd TRB Annual Meeting, Washington DC, January 2013.

In order to quantify the effect of yellow arrival flow on RLR frequency, we calculated the yellow arrival flow on each approach at the four Peachtree Blvd intersections, for both 12:45 – 1:00 pm and 4:00 - 4:15 pm time periods. The yellow arrival flow was defined as the number of vehicles arriving at 150 ft. (45 meters) upstream from the intersection stopline during the yellow interval (3 seconds). The relationship between RLR frequency and yellow arrival flow is illustrated in Figure 8.2. It is clearly shown that higher yellow arrival flow indicates higher frequency of RLR, which is consistent with the finding from the El Camino intersection.

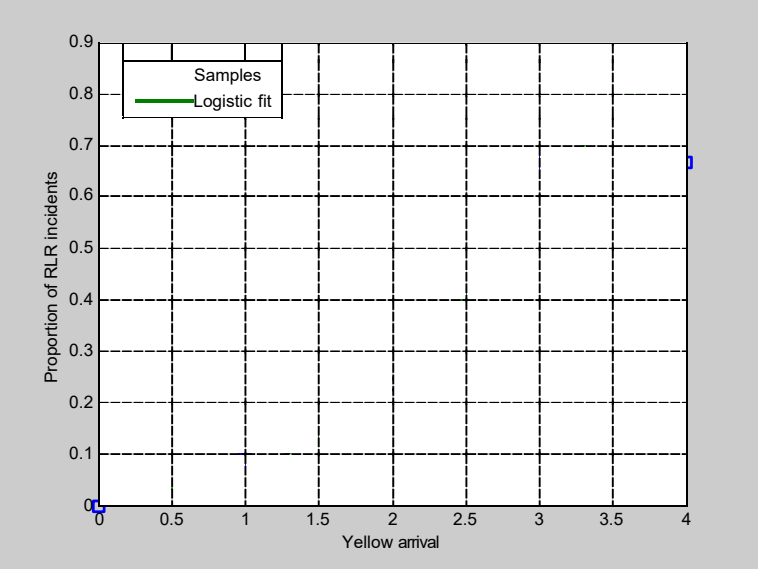


Figure 8.2 RLR Frequency vs. Arrival Flow in the Yellow Interval -- Peachtree Blvd

We also utilized the data collected at the San Pablo and Brighton Avenue intersection to test the effect of yellow arrival flow on RLR frequency. The yellow arrival flow was calculated at 200 ft (60 m) upstream from the intersection stopline (approximately 4 seconds travel time traveling at the speed limit of 35 mph). Figure 8.3 shows the comparison between the distribution of yellow arrival flow for cycles with RLR occurrences and that for all signal cycles. As it is shown in Figure 8.3, the yellow arrival flow for cycles with RLR occurrences is much higher than that on non-RLR cycles.

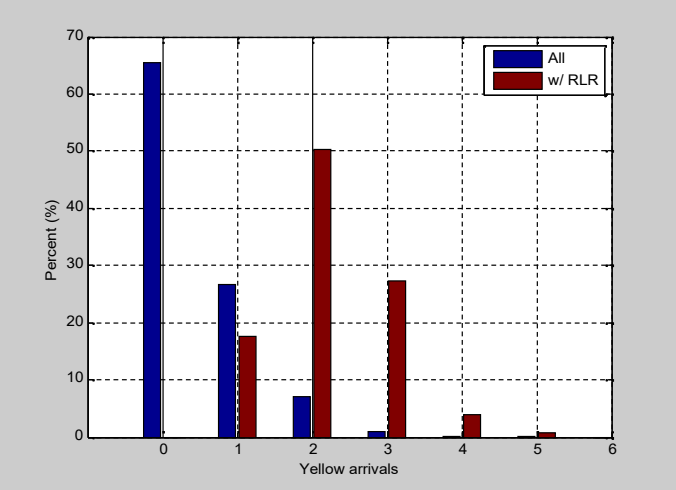


Figure 8.3 Distribution of Yellow Arrival Flow: San Pablo Ave/Brighton Ave

The San Pablo/Brighton intersection trajectory data were also used to investigate the effect of yellow arrival flow on intersection efficiency. The selected efficiency performance measures are:

- Progression ratio – proportion of vehicles arriving on green per signal cycle;
- Stop rate – proportion of vehicles stopped per signal cycle; and
- Stopped delay – Average vehicle stopped delay per vehicle per signal cycle

Figure 8.4 illustrates the effect of yellow arrival flow on the stop rate. Each dot in the Figure represents an average over a 15-minute interval. The Figure shows a positive correlation between stop rate and yellow arrival flow. In addition, there is a positive correlation between stopped delay and yellow arrival flow (Figure 8.5). These results show that higher yellow arrival flow not only results in higher RLR frequency but also into less efficient signal control.

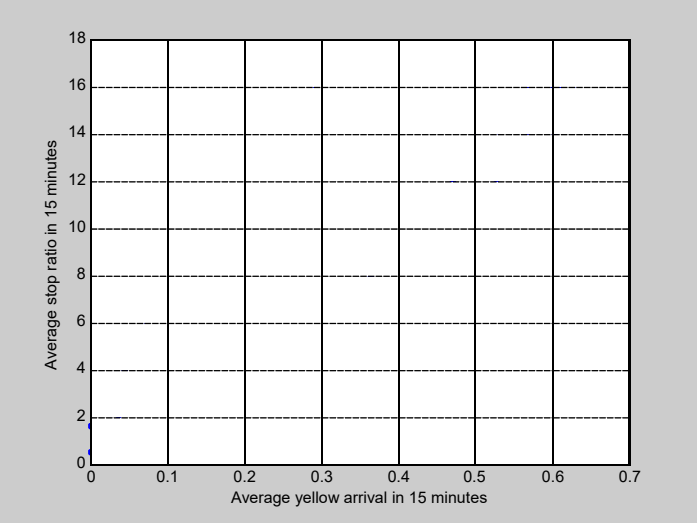


Figure 8.4 Stop Rate vs. Yellow Arrival Flow: San Pablo Ave San Pablo Ave/Brighton Ave

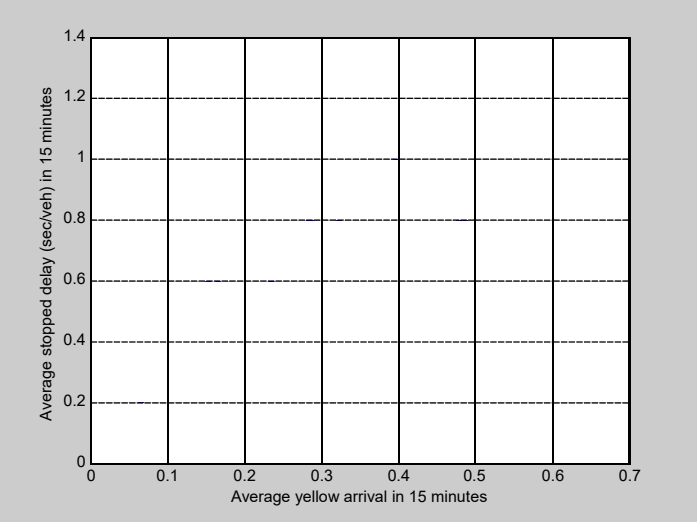


Figure 8.5 Stopped Delay vs. Yellow Arrival Flow: San Pablo Ave/Brighton Ave

The analysis of field data from three different sites, El Camino Real in San Mateo, CA, Peachtree Street in Atlanta, GA, and San Pablo Ave in Albany, CA, show that higher yellow arrival flow leads to higher RLR frequency at signalized intersections.

8.2 Proposed Control Concepts to Minimize Yellow Arrival Flow

Minimization of yellow arrival flow could be achieved with offline signal timing optimization and online signal timing refinement based on real-time traffic detection and performance measure. In a previous study an offline optimization scheme was proposed to optimize the signal offsets to minimize the weighted total delay and yellow arrival flow along an arterial [48]. Simulation results showed that the reduction of yellow arrival flow can be achieved without compromising the efficiency of intersection operation.

The proposed signal timing optimization method is based on platoon segmentation to dynamically determine the most appropriate control strategies, including green extension, phase termination, and double cycle, based on gaps between platoons and the predicted impacts of candidate control strategies on total intersection delay. The proposed methodology aims to increase the probability of platoon passing through and to reduce chances of truncating an approaching platoon with red onset, thereby reducing the yellow arrival flow.

8.2.1 Methodology and Control Strategies

Figure 8.6 shows sample vehicle trajectories approaching the intersection. At the position 1, vehicle #1, 2 and 3 are very close, and vehicle #5 to #8 get close as well. These vehicles can be segmented into three platoons, with vehicle #4 forming a single platoon by itself. Due to the influence of the traffic signal, vehicle # 4 gets close to the first platoon after the stopline and can eventually join the first platoon. All vehicles on the same movements will be partitioned into different platoons, and a platoon can include vehicles traveling on the same direction but in different lanes, as long as the time-headway between vehicles is small.

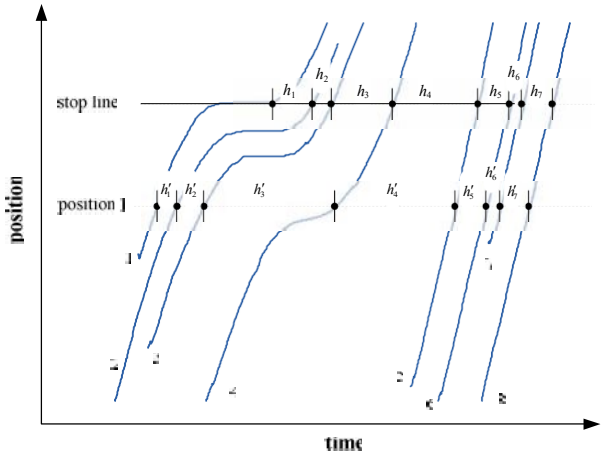


Figure 8.6 Platoon Segmentation Using Vehicle Trajectories

Each platoon is represented by its size (# of vehicles in the platoon), location (of the head and tail vehicle), and speed (average among individual vehicles of the platoon), from which the arrival times of the head and the tail of the platoon at the stopline can be predicted. Based on the current signal status and estimated arrival time gap between platoons, the following three control strategies together with continuing the existing signal timing plan are assessed in term of impact on total intersection delay. The control strategy that is predicted with the least total intersection delay will be selected to improve the existing signal timing plans. The following control strategies are proposed:

(1) Green Extension: If the green time on the arterial is ending shortly, when either the platoon that is currently crossing the stopline or a small platoon that is detected approaching the intersection along the arterial is likely to be truncated with red onset, then the green is extended to allow the platoons to clear.

Figure 8.7(a) illustrates the situation, where vehicles #1, #2 and #3, vehicle # 4, #5 and #6, and vehicle #7 and #8 formed three platoons. In the original signal timing plan, vehicle #4 will go through the intersection on yellow while vehicle #5 and #6 will have to stop and wait for the next green. Vehicle #5 could follow vehicle #4 resulting in running the red light. With green extension, both vehicle #4 and #5 will proceed without stopping thereby reducing the total delay. At the same time, both vehicles will enter the intersection either by the end of the green phase or at the beginning of the yellow interval, so the likelihood of RLR is reduced.

(2) Phase Termination: If traffic on the arterial is not busy in a short time period while vehicles are queuing in the minor road, terminating the current green signal earlier can reduce the waiting time of the vehicles in the minor road thereby reducing the total delay. Phase termination strategy can also applied for the case that a platoon is detected approaching the intersection with a high likelihood of truncating the platoon near its head with red onset, such that the whole platoon will have to stop for red leading to reduced probability of RLR. In the example illustrated Figure 8.7(b), the waiting time of vehicle #7, #8 and #9 are reduced dramatically, without any delay increment on vehicle #4, #5 and #6. In fact, the arterial could have earlier return to green as the green started earlier on the minor road, therefore delay for vehicle #4, #5 and #6 will be reduced at the same time.

(3) Double cycle: As the similar traffic situation in (2), double cycle can be applied to reduce the waiting time and delay because the vehicles in the minor road can have one more chance to pass the intersection without interrupting the arterial coordination.

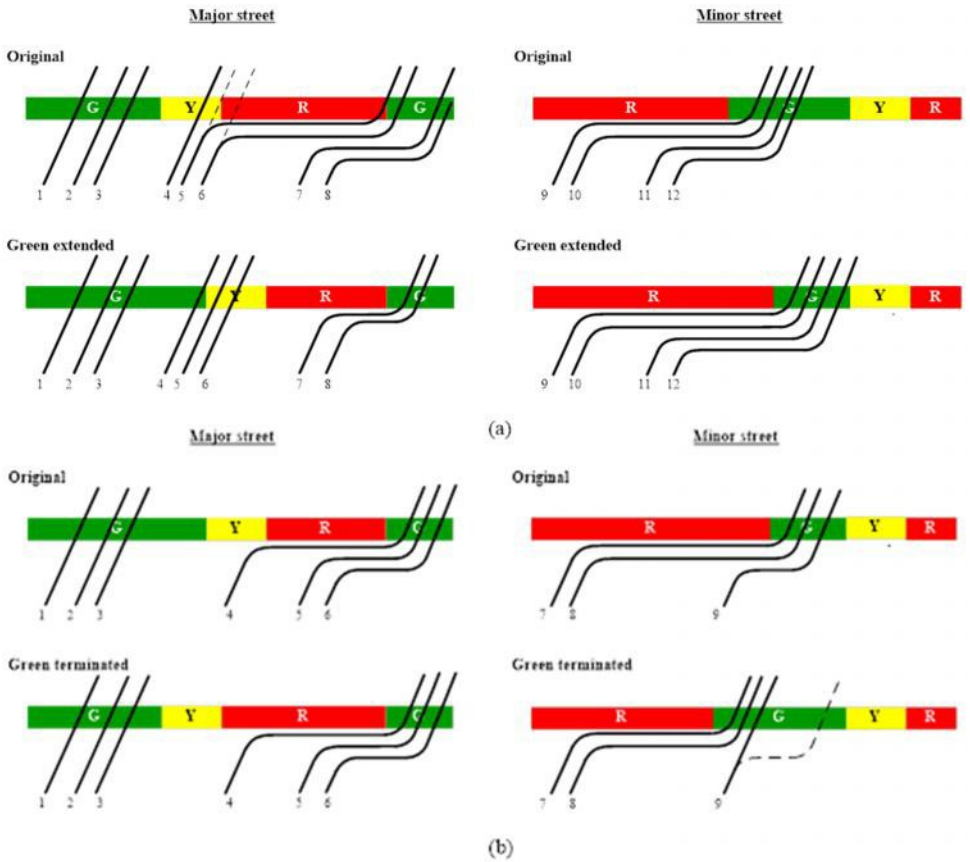


Figure 8.7 Control Strategies based on Platoon Segmentation

Regarding the platoon segmentation, time-headway was used to measure the closeness between the two adjacent vehicles and to segment platoons. According to the statistics obtained from trajectory data collected at the test site, most data of headway are distributed in the interval of 0 to 5 seconds, whereas, the headway rarely is larger than 10s. Since all the lanes in the same movements are considered together, the headway between the vehicles is quite small, even near zero. It can be inferred that most vehicles tend to keep in step with their leading vehicles.

Different threshold of headway will lead to different platoon segmentation results. The smaller is the threshold the shorter the platoon will be. The extreme case is that each platoon is only composed of one vehicle. Figure 8.8 shows the platoon segmentation results according to different headway threshold. The headway threshold for the test site is determined as 2 to 3 seconds.

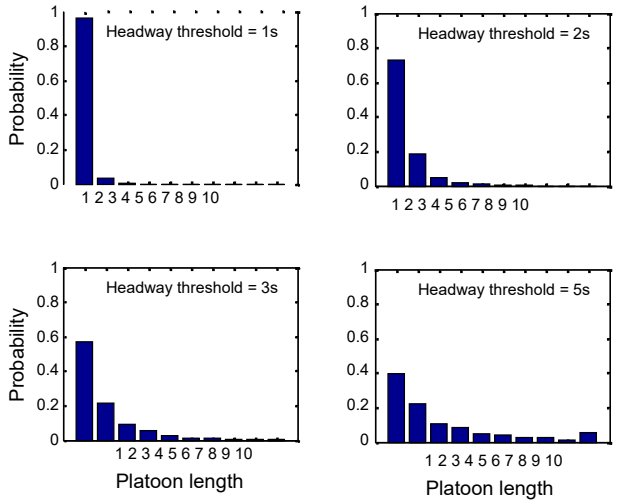


Figure 8.8 Impacts of Headway Threshold on Platoon Segmentation

8.2.2 Delay Estimation

Figure 8.9 illustrates the estimation of delay at a traffic signal from vehicle trajectories. The control delay D consists of three components: the deceleration delay D_1 (the time loss due to deceleration from the approach speed to a stop), the stopped delay D_2, and acceleration delay D_3 (the time loss due to reacceleration back to the desired speed). Each delay component can be estimated based on time points $t_1 \sim t_4$ on the time-space diagram, with t_1 as the expected time when the vehicle crosses the stop line without any interruption, t_2 as the time just when the vehicle fully stopped at the stop line, t_3 as the time when the vehicle begins to accelerate, and t_4 as the time from when the vehicle runs away at the desired speed. In general, a queuing vehicle tends to start up later than its leader, which is denoted as D_4.

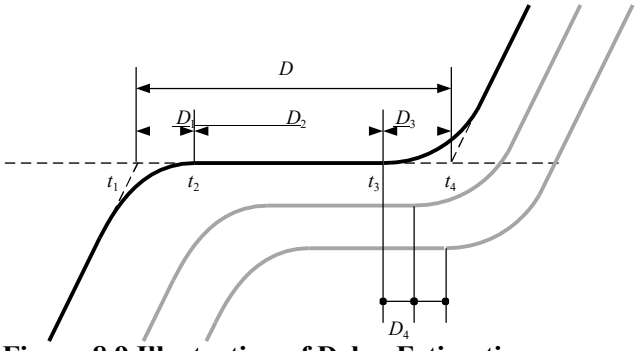


Figure 8.9 Illustration of Delay Estimation

For the most leading vehicle, its delay is $D^{(1)} = t_3^{(1)} - t_1^{(1)} + D_3^{(1)}$, where superscript represents the queuing position of the vehicle, and $t_3^{(1)}$ is the time when the first vehicle start up. For the first platoon queuing at the stop line, $t_3^{(1)}$ is assigned to the time when the light turns green; while for following platoons $t_3^{(1)}$ can be determined as $t_3^{(n)} + D_4$, where $t_3^{(n)}$ represents the time when the last vehicle in its leading platoon starts up.

For vehicles queuing in a same platoon, we have

$$t_1^{(i+1)} = t_1^{(i)} + h^{(i+1)} - \frac{d^{(i+1)}}{v^{(i+1)}} \tag{1}$$

$$t_3^{(i+1)} = t_3^{(i)} + h^{(i+1)} - \frac{d^{(i+1)}}{v^{(i+1)}} \tag{2}$$

$$D^{(i+1)} = t_3^{(i+1)} - t_1^{(i+1)} + D_3^{(i+1)} = t_3^{(i)} - t_1^{(i)} - h^{(i+1)} + \frac{d^{(i+1)}}{v^{(i+1)}} + D_3^{(i+1)} + D_4^{(i+1)} \tag{3}$$

where $h^{(i)}$ is the time-headway, $d^{(i)}$ the distance-headway, and $v^{(i)}$ the desired speed of vehicle i. The control delay for each vehicle can be calculated iteratively with Equations (1) to (3). The delay for a whole platoon (k) is then given by

$$DP^{(k)} = \sum_{i \in \text{platoon } k} D^{(i)} \tag{4}$$

Parameters $h^{(i)}, d^{(i)}, v^{(i)}$ can be easily determined from individual vehicle trajectory data, while $D_3^{(i)}$ and $D_4^{(i)}$ are determined by using aggregated trajectory data.

8.2.3 Selection of Control Strategies for Actual Scenarios

Scenario #1: The green time on the arterial is ending, and there are queuing vehicles in the minor road. In this situation both the green extension and phase termination strategies could be carried out. In order to determine the optimal signal timing strategy, the difference on control delay must be considered when taking different timing strategies.

For green extension cases, we consider three platoon types – the queuing platoon on the minor road (Type I), the just-not-passing-through platoon on the arterial (Type II) and the following platoon on the arterial (Type III). For platoon of Type I, the increment on total delay between the two timing strategies is

$$\Delta_{DP^{(I)}} = E \times N^{(I)} \tag{5}$$

where E represents for the length of green extension, $N^{(I)}$ stands for the number of vehicles in the platoon, and superscript indicates the platoon type. Similarly, the increment for platoons of Type II and Type III will be calculated according to Equation (6) and (7), respectively.

$$\Delta_{DP^{(II)}} = -DP^{(II)} \tag{6}$$

$$\Delta_{DP^{(III)}} = \left(\frac{d^{(III)}}{v^{(III)}} + (N^{(II)} + 1)D_4\right) N^{(III)} \tag{7}$$

The negative sign indicates the total delay will be reduced after carrying out the green extension strategy. If there are more than one platoons of Type III, then the term $N^{(II)}$ in Equation (7) will be replaced with the number of vehicles in the platoon right ahead of the current platoon, and every $\Delta DP^{(III)}$ should be

added up to composite the delay increment for them. Then we define the total delay increment below:

$$\Delta_{DP} = \Delta_{DP^{(I)}} + \Delta_{DP^{(II)}} + \Delta_{DP^{(III)}} \tag{8}$$

In this case, $\Delta DP < 0$ indicates that the total delay will be reduced after carrying out the green extension strategy, otherwise it would be better carrying on the original timing schedule.

For phase termination cases, we also consider three kind of platoons – Type I, the same as mentioned above; Type II, the platoons that would pass through with original strategy while would be blocked with phase termination; Type III, those would be blocked with the original strategy. The delay increment for Type I is

$$\Delta_{DP^{(I)}} = -T \times N^{(I)} \tag{9}$$

where T represents for the length of truncated green time. For Type II, it is

$$\Delta_{DP^{(II)}} = \Delta_{DP^{(II)}} \tag{10}$$

If there are more than one platoons of Type II, then just add all their delays up. In particular, if we choose an appropriate time to terminate current phase, there might be no platoons in major road be affected, thus we obtain $\Delta DP^{(II)} = 0$. And for Type III the delay increment is

$$\Delta_{DP^{(III)}} = \left(\frac{d^{(III)}}{v^{(III)}} + (N^{(II)} + 1)D4\right) N^{(III)} \tag{11}$$

where the meaning of $N^{(II)}$ is the same as defined above. Similarly if we carefully choose the time to terminate current phase, there would be no difference after terminating the current phase and $\Delta DP^{(III)} = 0$. Apparently phase termination is almost the other side of green extension, as reflected in their equations of delay increment. As expected, phase termination will be carried out when ΔDP is less than 0.

Scenario #2: The mid-point of a cycle is coming, and there are vehicles queuing in the minor road. This is a simpler case and may be occurred frequently during off-peak times. In this case, we should determine whether to carry out double cycle strategy. Two types of platoons should be taken into consideration – the queuing platoon on the minor road (Type I), the platoons that might be blocked by new timing schedule (Type II). The delays are:

$$\Delta_{DP^{(I)}} = -\frac{C}{2} \times N^{(I)} \tag{12}$$

$$\Delta_{DP^{(II)}} = -\frac{C}{2} \times N^{(II)} \tag{13}$$

where C is the cycle length of the original timing schedule. Similarly double cycle will be carried out only when the total incremental delay ΔDP is less than 0.

8.2.4 Application and Results

The proposed control strategies were applied using the data from the San Pablo Avenue/Brighton Avenue intersection from March 1, 2012. Table 8.1 summarizes the results. The proposed algorithm works well for both peak and off-peak time periods. In fact we find it more effective in peak-hours. For instance during 8:00 to 10:00 there are 34 cycles, in which green extension can be applied on 19 cycles, phase termination can be applied on 10 cycles, and double cycle can be applied on 20 cycles.

Table 8.1 Summary Results: Control Strategies for Minimum Yellow Flow

Time	Flow (veh/hr/lane)	Cycles	Green Extension # cycles	Phase Termination # cycles	Double Cycle # cycles
5:00~7:00	126	19	2	2	0
8:00~10:00	372	34	19	10	20
16:00~18:00	469	32	10	3	17
19:00~21:00	256	45	19	9	10

Detailed results are shown in Table 8.2. It can be seen that during 8:00 to 10:00, if we extend the original green time for only 1.3 second for some cycles on average, we could reduce delay for more than 200 seconds totally, and reduce delay for about 1 second for every delayed vehicle on average. The statistics for phase termination strategy indicate that it can reduce can reduce delay for the intersection, it is not as effective as green extension strategy. The double cycle strategy brings significant reduction on delays so that it can reduce at least half of the cycle length for vehicles queuing in the minor street while reduce a little time for vehicles in the major street.

Table 8.2 Detailed Results: Control Strategies for Minimum Yellow Flow

Green Extension Strategy			
Time Period	**Average extension time (s)**	**Total Δ_{DEL} (s)**	**Average Δ_{DEL} (s/veh)**
5:00~7:00	0.4	-8.1	-0.8
8:00~10:00	1.3	-201.8	-1.0
16:00~18:00	1.9	-130.8	-0.6
19:00~21:00	1.4	-206.2	-1.1
Phase Termination Strategy			
Time Period	**Average truncation time (s)**	**Total Δ_{DEL} (s)**	**Average Δ_{DEL} (s/veh)**
5:00~7:00	7.7	-6.4	-0.5
8:00~10:00	3.4	-11.4	-0.1
16:00~18:00	2.2	-17.2	-0.2
19:00~21:00	4.2	-31.6	-0.5
Double Cycle Strategy			
Time Period	**Average red time (s)**	**Total Δ_{DEL} (s)**	**Average Δ_{DEL} (s/veh)**
5:00~7:00	0	0	0
8:00~10:00	7.2	-1007.2	-10.4
16:00~18:00	6.9	-1204.5	-12.1
19:00~21:00	10.6	-524.5	-10.3

8.3 Discussion

We investigated the impacts of yellow arrival flow on the frequency of red-light-running (RLR) and on intersection operation efficiency using vehicle trajectory data from the NGSIM data set (Peachtree Blvd, Atlanta) and the San Pablo Avenue/Brighton Avenue intersection in Albany, California. The results from both sites showed higher yellow arrival flow increases the RLR frequency and reduces intersection efficiency resulting in higher stop rate and delay. These results confirm the findings from earlier studies along El Camino Real in Palo Alto.

In view of these findings, new control strategies were developed and tested to increase the probability of platoons passing through the intersection. These strategies can be deployed with CV data, which provide

trajectory information to implement platoon based signal control. Three control strategies including green extension, phase termination, and double cycle, were proposed for platoon-based control. The determination of whether to carry out an optimal timing schedule, chosen from the three proposed strategies, or to keep on the original timing schedule is based on the prediction of their impacts on total delay. A vehicle trajectory-based method was developed for the estimation of delay for individual platoons. The proposed control strategies were tested using the trajectory data collected at the San Pablo Avenue/Brighton Avenue intersection, and the experiments showed that in most cases the proposed strategy can significantly improve the intersection performance.

CHAPTER 9

DRIVER ADVISORIES FOR MINIMUM FUEL AND EMISSIONS

9.1 Introduction

A significant body of research exists on developing control strategies to reduce fuel consumption and air pollutant emissions in signal controlled networks. However, with the recent advances in ITS technology, it is now possible to develop driving strategies where vehicle velocity trajectories can be dynamically adjusted. Connected vehicle data provide detailed knowledge of vehicle trajectories as they travel in signal controlled networks. In addition, the traffic controller's signal phase and timing (SPaT) information is known and can be communicated directly to individual vehicles. This information can be used to develop algorithms for providing real-time speed advisories to drivers to minimize fuel consumption and emissions.

This Chapter presents the work performed by the project team on the development and testing of an in-vehicle prototype system to provide speed advisories to drivers for minimum fuel consumption. In addition to the speed recommendation a control strategy for signal priority to individual vehicles (APIV) was investigated alone as well as in combination with the speed advisory system. Detailed description of the work performed and the results are included in the final report prepared by BMW (Appendix A) and in journal publications included in Appendix C.[10][11]

9.2 Isolated Intersections with Fixed-Time Signals

This section presents the development and field testing of a speed advisory prototype for fixed-time signals. It also describes an operational strategy called adaptive priority for individual Vehicle (APIV) which was used in the field tests.

9.2.1 Vehicle Trajectory Planning—Single Intersection

Consider an isolated signalized intersection and trajectories of approaching vehicles (Figure 9.1). All vehicles have the same velocity $v_i(t)$ at point $d(t)$. At time t, the traffic signal's SPaT information is received by each vehicle. We consider the following cases:

Case 1: the vehicle accelerates and clears the intersection without stopping

Case 2: the vehicle continues to drive at speed $v_i(t)$, and when the traffic signal turns yellow then red, the vehicle decelerates quickly and stops rather suddenly

Case 3: the driver of the vehicle takes their foot off of the gas pedal and coasts to a stop at the intersection

Case 4: the vehicle brakes to a slower speed and then travels at a lower speed without stopping until the traffic signal turns green, and then accelerates to its desired speed

[10] Xia H., K. Boriboonsomsin, F. Schweizer, A. Winkler, K. Zhou, WB. Zhang, M. Barth, "Field Operational Testing of Eco-Approach Technology at Fixed-Time Signalized Intersection," Proceedings 15th IEEE ITSC Conference, Anchorage, AK, Sept 2012.

[11] Winkler A., "Smart Cars Meet Smart Lights," Proceedings of the ITS America Conference, Washington D.C., May, 2012.

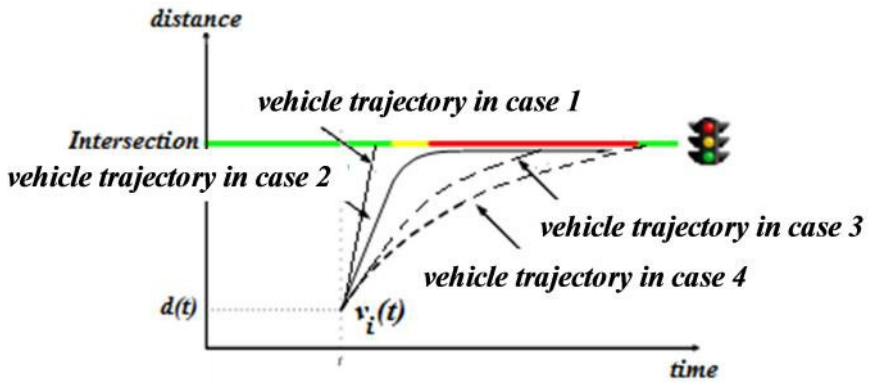


Figure 9.1 Trajectories of Vehicles Approaching a Signalized Intersection

There is a significant variation in fuel consumption and emissions for each case. In case 1, even though the vehicle did not stop at the red light, the fuel consumption and emissions are high since the vehicle had to accelerate to make the green light. In case 2, fuel/emissions are also high because the vehicle traveled at the original velocity, decelerated suddenly and then had a long idle period. In case 3, the fuel/emissions are lower, because the vehicle coasts up to the intersection. Simulation results have shown that the fuel consumption in this case is 15% lower than in case 2 [49]. Finally, the lowest fuel consumption and emissions occur in case 4, because the vehicle's acceleration from the red light isn't from a dead stop, but from a moving velocity, which requires significantly less energy to accelerate back up to speed.

Therefore, as a vehicle approaches a traffic signal, it is possible to dynamically adjust its velocity to minimize fuel consumption and emissions. A velocity planning algorithm was developed and tested [49] to provide recommended speeds subject to traveling below or at the speed limit, maintain safe headway to vehicles in front, avoid sharp accelerations and minimize the idle time at the traffic signals.

9.2.2 Implementation of the Velocity Planning Algorithm

Figure 9.2 shows a block diagram for implementing the velocity planning algorithm to calculate the recommended speed.

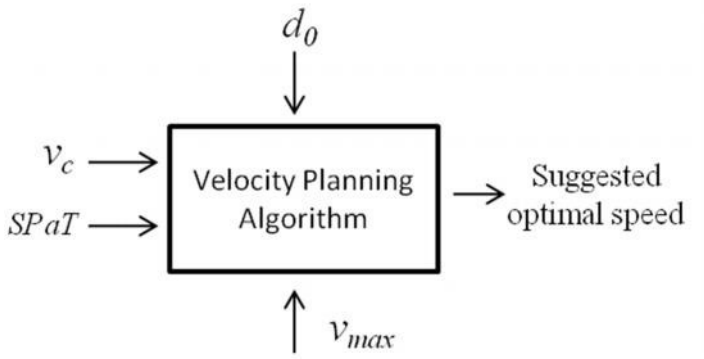


Figure 9.2 Implementation of the Velocity Planning Algorithm

The process uses several input parameters:

d_0	the distance from the vehicle to the intersection
v_{max}	the maximum speed
v_c	the current vehicle velocity
SPaT	information about signal phase and timing

There are a number of options for determining the maximum speed v_{max}. Examples include use of the local street speed limit, or taking into account the velocity of the proceeding vehicle and the safe headway distance or time. For simplicity, the algorithm uses the local speed limit as the maximum speed. Based on the input parameters provided the Velocity Planning Algorithm calculates a speed trajectory and from this trajectory a recommended optimal speed is determined.

9.2.3 Adaptive Priority for Individual Vehicles (APIV)

In the speed advisory field experiments we also tested the APIV signal control strategy. APIV adjusts the signal settings to provide additional green time to favor individual vehicles approaching a signalized intersection. The APIV strategy either 1) extends the green phase for the individual vehicle until it clears the intersection (GE = Green Extension), or 2) advances the start-of-green on the vehicle approach (EG = Early Green).

APIV is based on an adaptive transit signal priority (ATSP) system developed by California PATH [50]. The ATSP system predicts the transit vehicle's arrival times at the traffic signal, and grants priority to minimize a weighted number of stops and delay. The proposed APIV strategy is taking advantage of the CV data which will enable accurate prediction of the of individual vehicles' arrival times at intersections thereby improving the performance of signal priority. The CV equipped vehicles "here I am" messages with the information about the vehicle's location, speed and predicted arrival times at intersections allow the signals to grant priority if needed to facilitate the vehicle's movement at the signalized intersection. Figure 9.3 illustrates the APIV concept under the CV environment.

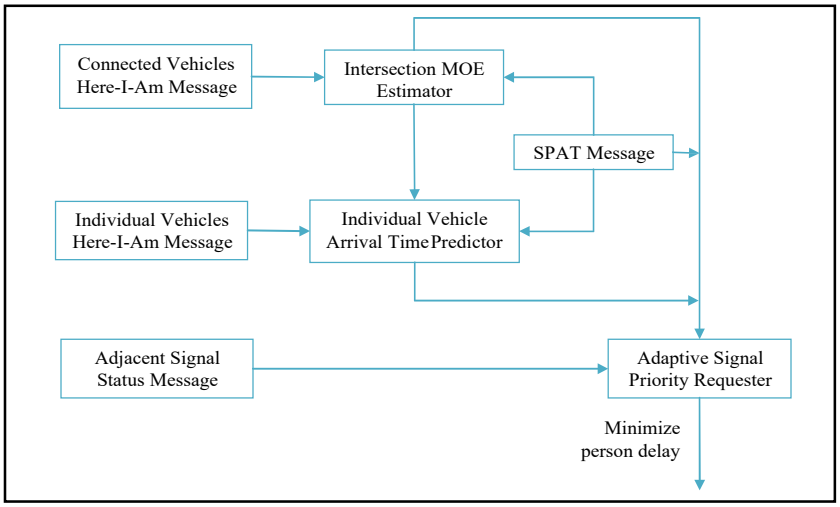


Figure 9.3 Adaptive Signal Priority for Individual Vehicles (APIV)

9.2.4 Field Test

The BMW equipped vehicle with the proposed speed advisory system was tested under controlled conditions at the Richmond Field Station campus of University of California Berkeley. The test site includes a signalized intersection and approach roadways (Figure 9.4). The blue line shows the test segment used in the test runs. The ideal selection of the test segment is the straight yellow line. However, the driveway marked by the yellow line is not mapped in the digital map of the vehicle's embedded navigation system. As the test setup uses map matched locations only, the yellow driveway was unusable for testing. Therefore, the blue line was chosen as the test segment. The total distance of the entire test "loop" is 1,000 ft. (307 m). In each test run, we started from the green arrow, passed through the intersection, and then exited the test segment at the red bar downstream, as shown in Figure 9.4. The test vehicle would travel around the test loop, at a nominal speed of around 25 mph (41kph), corresponding to the speed limit in the area.

The traffic signal controller was operating as fixed-time. The cycle length was set to be 60 seconds, with 30 seconds of green, 3 seconds of yellow and 27 seconds of red time on the test segment approach. Under APIV the maximum allowable priority interval is set to 9 seconds, i.e., the green phase can be extended for up to 9 seconds or the start of the green phase can be advanced for up to 9 seconds. The actual priority interval ranged from 0 to 9 seconds depending on the needs of the vehicle when approaching the intersection.

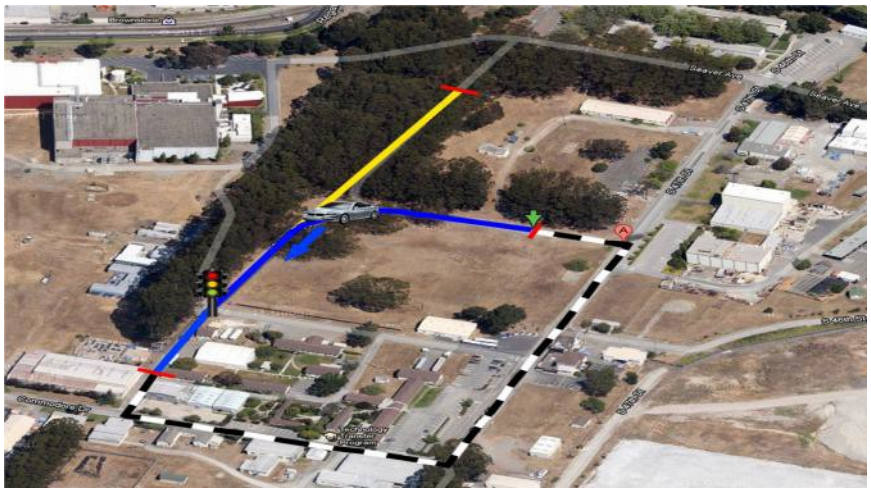


Figure 9.4 Test Track at Richmond Field Station

Scenarios

In order to compare the proposed eco-driving technology with regular driving in terms of fuel consumption, and other performance measures the following four driving scenarios were carried out:

- **"Uninformed Driver":** this is the baseline scenario. In this scenario the driver receives no speed recommendations and has no knowledge of the green time left in the current signal phase. To have a realistic comparison, the driver drives in a reasonably efficient way as much as possible in the uninformed scenario. A total of 270 test runs were made for this driving scenario.
- **"Informed Driver":** the driver is provided with the recommend speed in the programmable instrument cluster at 1 Hz as he drives through the intersection. A total of 292 test runs were made for this driving scenario.

- **"Uninformed Driver & APIV**": the driver drives uninformed but this time the APIV control strategy is used to adapt the signal timing for priority to the test vehicle. A total of 108 test runs were made for this driving scenario.
- **"Informed Driver & APIV":** signal timing is adapted by APIV and the driver is also provided with a speed recommendation. A total of 108 runs were made for this driving scenario.

For each test run, the test vehicle entered the test segment randomly in time without knowing the current signal information. Second-by-second fuel consumption were measured for each un from the point the test vehicle entered the test segment till it exited at the other end, as the blue line shows in Figure 9.4. The accuracy of field consumption measurements was 10^{-6} liter. Note that by driving through the blue curve while maintaining a speed, an additional shear force is generated at the wheels that alter the balance of forces and has to be compensated for by the engine. Thus, the engine would consume more fuel. However, this was the case for all four scenarios and thus, the effect on the relative differences between the two scenarios are negligible.

Results

Table 9.1 summarizes the results from the field test. After averaging the fuel consumption results over all the runs for all scenarios, we found that the informed driving resulted in 13.6 % less fuel compared to the uninformed driving. A second driver was also recorded to cross check the results, showing similar fuel savings compared with the first driver. The uniformed driving with APIV gained 19.06% of fuel savings compared to the base scenario. The combination of informed driving and APIV produced the best results for all performance measures.

Table 9.1 Field Test Results: Travel Time, Stops, Fuel Consumption MOEs

	Uninformed	Informed	APIV Uninformed	APIV Informed
Number of Test Runs	210	232	108	108
Stop Frequency (%)	48.57	30.60	14.81	0.93
% Change	-	-36.99%	-69.50%	-98.09%
Mean Stopped Time (sec)	15.77	10.49	5.56	2.00
% Change	-	-33.48%	-64.74%	-87.32%
Travel Time (sec/trip)	40.69	40.30	31.65	31.00
% Change	-	-0.96%	-22.22%	-23.81%
Fuel (l/100km)	10.2	8.8	8.3	7.3
% Change	-	-13.59%	-19.06%	-28.35%

Further analysis of the fuel consumption data provides more insight on where the fuel was saved. Figure 9.5 shows a frequency distribution of fuel economy for each test loop. In this figure, an initial peak can be seen in the range of 5 to 9 l/100km, corresponding to the scenario where the driver didn't need to slow down since the vehicle can pass the intersection while the signal is green. This is approximately the same for both the informed and uninformed driving cases. For the APIV cases this peak is considerably higher. This corresponds to a higher proportion of green drives and results from the fact that there was no cross-traffic and so APIV was able to grant priority green (up to 9 seconds) in most of the cases.

A second set of peaks is observed in the range of 10 to 15 l/100km, corresponding to the case when slowing down is inevitable due to a red light. In this case, the informed driver is able to coast down and consume less fuel than the uninformed driver. So the second peak is flattened (less full stops) and to the left (less waiting time). For the APIV Uniformed Scenario the shape of the second peak is almost the

same as in the base scenario, but it moved to the left which means that the behavior of the driver remained the same, but the waiting time after a full stop is reduced. In the last scenario that combines APIV and Informed Driving the second peak is very small since the driver doesn't have to stop in almost all cases.

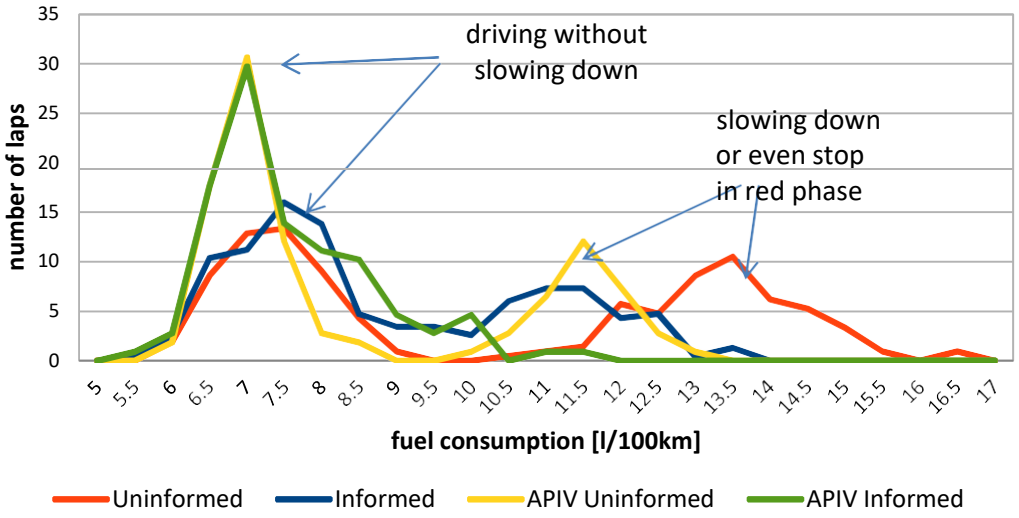


Figure 9.5 Distribution of Fuel Consumption over Test Runs

Figure 9.6 gives more insights about how "Informed Driving" influences the driving behavior. It shows the average speed (in m/s) during the test lap. It is shown that the informed driver tends to slow down earlier when approaching the intersection, and has a higher average speed to pass the intersection, compared to the uninformed driver. Additionally, the curve is much smoother than in the uniformed scenario. The APIV Uniformed scenario is similar to the base scenario again. The shape of the curve stays almost the same but moved upward. That's because with APIV, the driver sees a green signal more often, but in case of a red signal he has the same inefficient driving behavior. Combining both APIV and Informed Driving results in a smooth energy saving speed curve at a higher speeds.

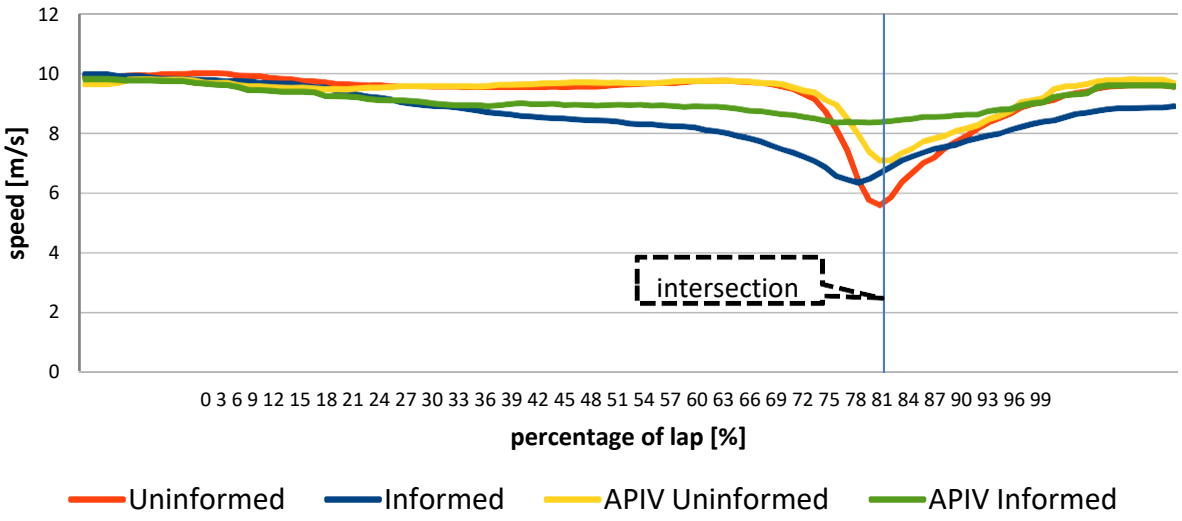


Figure 9.6 Average Speed During Test Runs vs. Driving Scenarios

Additional analyses indicate that when the vehicle is far from the intersection, an informed driver tends to decelerate earlier and uses less fuel than an uninformed driver. As the vehicle is getting close to the intersection, if the current signal is red, the informed driver has more chance to avoid a complete stop and clears the intersection at a constant speed. In contrast, the uninformed driver has to stop in most cases which results in higher fuel consumption. After the vehicle clears the intersection, the informed driver generally had higher speed than uninformed driver since informed driver tried to avoid full-stop at the intersection and kept a constant speed. Thus, the informed driver doesn't need as much fuel to accelerate back to speed limit compared to the uninformed driver.

9.3 Arterials with Coordinated Actuated Traffic Signals

The next field experiment involved testing of speed advisories along a real world arterial, El Camino Real in Palo Alto. The test section includes three signalized intersections with signals operating as coordinated actuated with fixed cycle length but variable green times. Detailed description of the test site, the SPaT messages and the instrumented vehicle field test setup is given in the BMW final report (Appendix A).

A velocity planning algorithm for fixed-time signals along arterials was developed by BMW staff and successfully tested on a section of Great Highway in San Francisco [51]. Also, researchers at the University of California Riverside developed and tested through simulation a speed advisory algorithm for a single vehicle along arterials [52]. However, providing speed-advisories for arterials with coordinated actuated signals required the development of a new speed advisory algorithm to take into account the presence of other vehicles and the operation of actuated signals in coordination.

Proposed Speed Advisory Algorithm

The proposed algorithm divides the problem into three separate problems. These three problems result in different steps of the algorithm. In each step the results of the previous step are used to solve current problem. An overview of the algorithm is shown in Figure 9.7. Detailed description of the algorithm and its components are given in the BMW Final Report (Appendix A.)

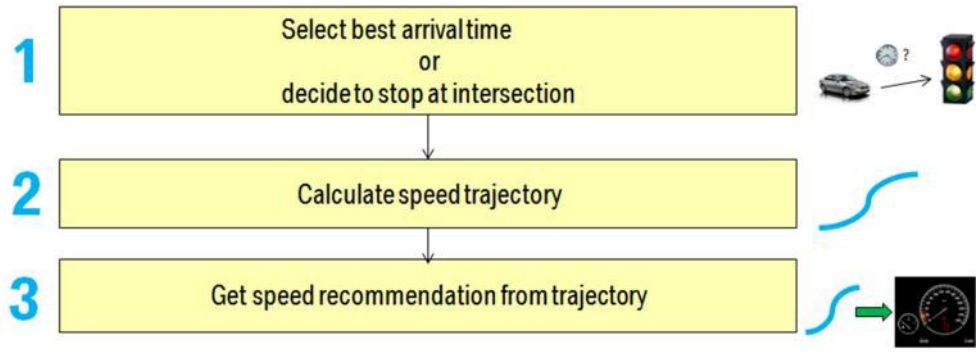


Figure 9.7 Overview of Speed Advisory Algorithm for Coordinated Actuated Traffic Signals

The first problem is to determine if it is possible to arrive at the next intersection during the green time under certain constraints: he/she must not travel above the speed limit, and the acceleration must not exceed a feasible threshold. Furthermore, the driver has to maintain an appropriate minimum speed. If is indeed possible to arrive at the next intersection during the green time, then the time when it best to arrive at the intersection is calculated in this step. The second problem is to determine a speed trajectory that allows the driver to arrive at exactly that point of time that was calculated in the first step. The last step consists of presenting a speed recommendation to the driver in a way that he can follow the desired speed profile. We implemented a simple constant forward looking mechanism to retrieve the displayed recommendation. Currently, we use one second as the forward looking interval.

During our initial test runs along the El Camino Real test site we experienced a number of shortcomings of the new algorithm and its displayed speed recommendations, described below:

Changing the speed recommendation when the vehicle is crossing the intersection is not desired. This is illustrated in Figure 9.8. The blue line shows the speed recommendation of the algorithm. As long as the driver is before the first intersection, the speed recommendation is to maintain speed as the driver arrives in guaranteed green before green band. When the driver passes the first intersection, the algorithm recognizes that with current speed the driver would arrive in the green band and would recommend to the driver to accelerate to get to the start of the green band. This recommendation to higher speed appears confusing to the driver, especially if he/she is able to see the second green light before he/she arrives at the intersection. A better trajectory would possibly be the dashed orange one. With a slight speed up, the driver arrives exactly at the beginning of the green band of the second intersection - without confusing speed changes at the first intersection. This approach requires that the algorithm takes into account the information about not only the next but the next two or even more intersections into account.

It is very difficult to follow a speed recommendation that is changing continuously. With some training it is possible to follow it when there is no other traffic, but it's impossible to follow a speed recommendation that is changing continuously when there is other traffic around. Possible resolution would be to not display the speed that the driver should have in the next immediate seconds but instead the end speed that the driver should have when passing the intersection. This speed recommendation would only change, if the driver doesn't follow the trajectory.

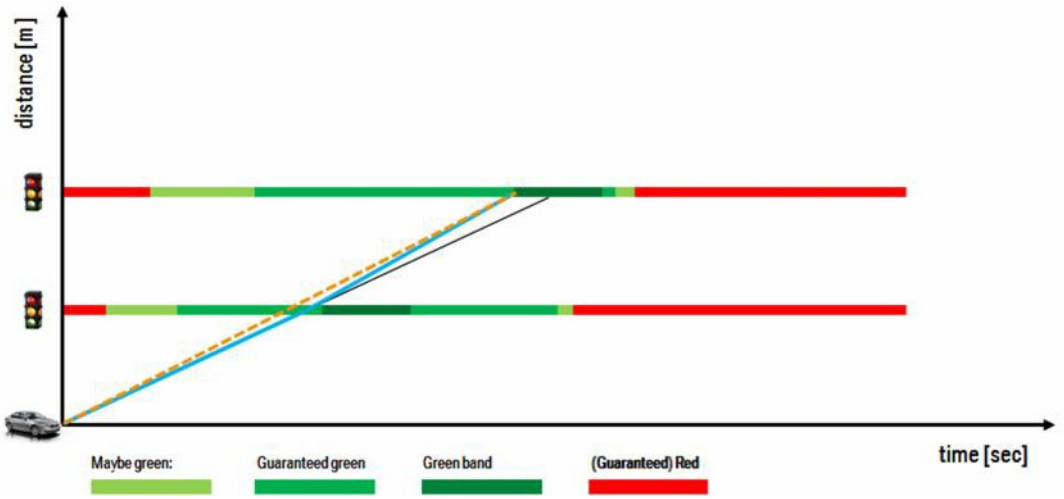


Figure 9.8 Speed Recommendation Changes

9.4 Discussion

We developed an in-vehicle prototype system to provide speed advisories to drivers for minimum fuel consumption. In addition to the speed recommendation a control strategy for signal priority to individual vehicles (APIV) was investigated alone as well as in combination with the speed advisory system.

The prototype system consists of a research vehicle, a cloud server and several traffic signals equipped with wireless communication devices. The vehicle was equipped with PCs for the computation of the speed recommendation as well as the capability to display the speed recommendation directly in a

programmable instrument cluster. The connection to the cloud server is maintained via a cellular 4G connection. The traffic signals send information about their future timings via a 3G connection to a cloud server, which forwards the appropriate information to the research vehicle.

The system was first tested in a controlled environment: a fixed time traffic signal with no cross traffic or other traffic present at the Richmond Field Station campus of UC Berkeley. Compared to the uninformed driving scenario the speed advisory system achieved significant fuel savings: using the speed advisory system, fuel savings of 13.6% were observed. For the APIV approach without the speed advisory fuel savings of 19.1% were achieved. The combination of both approaches showed fuel savings of 28.4%. Further analysis indicates that fuel savings for the informed driving scenario are mainly from the drivers' early slowing down and cruising through the intersection without having to come to a complete stop. The fuel savings due to APIV are mainly contributed by an increased proportion of driving through green.

The system was also tested on a real-world arterial El Camino Real with three coordinated actuated traffic signals. A new speed advisory algorithm was developed for actuated coordinated traffic signals and the presence of other vehicles. Due to limited time and resources, it was not possible to do any meaningful measurements on the El Camino Real field tests. Nevertheless important observations were made during the test runs performed. It was observed that is extremely difficult to follow recommended speeds that are continuously changing over time, while driving with other vehicles. This indicates that besides the speed advisory algorithm itself, the method of displaying the speed recommendation is crucial for the success of similar systems in the future.

CHAPTER 10
CONCLUSIONS

10.1 Summary of the Study Findings

The goal of this project is to develop advanced signal control strategies based on connected vehicle (CV) data, i.e., real-time information on vehicles' location, speed and characteristics as well as communication to the signal control infrastructure. A comprehensive literature review was performed with emphasis on existing adaptive traffic signal control systems and connected vehicle data and traffic control. We developed algorithms for the estimation of performance measures based on CV data. Several control concepts were developed and tested to improve mobility and safety. A prototype system was also developed and field tested to provide real-time speed advisories to drivers to minimize fuel consumption and emissions.

10.1.1 Performance Estimation with Connected Vehicle Data

We identified several performance measures (MOEs) for the development and evaluation of traffic signal control algorithms. These MOEs cover a wide range of operating environments (isolated intersections, arterials, grid networks), traffic patterns, and objectives/constraints (mobility, safety, environment). We developed and tested procedures for the estimation of MOEs from CV data, using the real-world vehicle trajectories from the NGSIM arterial Peachtree Blvd data set (for undersaturated conditions) and the El Camino Real simulation testbed (for oversaturated conditions).

Testing of the proposed estimation methods show that for undersaturated conditions penetration rates of 50% are needed for accurate estimation of average travel time, delay, and stops. However, the required penetration rate was about 20% for oversaturated conditions. This result is logical since in oversaturated conditions we have a greater amount of vehicles, which increases the number of CV sampled. Next, we developed and tested three queue length estimation methods: maximum likelihood, method of moments, and one based on Kinematic Wave Theory. The findings show that 80% of the vehicles need to be equipped to obtain accurate estimates for undersaturated conditions but only 10% for oversaturated conditions. This is promising in the sense that accurate estimates of queue length can be obtained for low penetration rates and as a result, improved control strategies can be designed for oversaturated conditions even at the first stages of the CV implementation.

We also used the simulated vehicle trajectories from the El Camino Real test bed as the baseline truth model, and are then sampled according to a variety of sampling strategies (i.e., uniform sampling, speed based sampling, acceleration based sampling, J2735 sampling strategy) that could be implemented on probe vehicles. The comparison of estimates of speeds, queue lengths and stops MOEs using the sampled data with the baseline truth values show that there is not a clear "break point" with respect to market penetration where the probe sampling results become significantly better. None of the probe sampling strategies that were analyzed showed a clear superiority to the others, and none was able to provide high-fidelity estimates of the traffic conditions at the intersection at low to moderate market penetrations. Rather, there is a close to linear relationship between the market penetration and decreasing variability in the estimates of the MOEs. The variability in the sampled data can be reduced by aggregating data from multiple signal cycles, but this introduces additional latency in the response of the traffic control system. This represents an additional trade-off in considering how to use CV data to implement real-time adaptive signal control.

We also developed a data fusion method to estimate queue lengths from fixed location loop and CV data. The queue length estimate is the weighted combination of the queue length from the two data sources

estimated separately using Kalman filtering techniques. The results from the application of the algorithm indicate that improves the queue length estimation accuracy.

10.1.2 Control Strategies for Mobility

Queue Spillback Avoidance

Queue spillbacks are a major problem in urban signalized arterials because they can lead to gridlock and excessive delays. We developed two queue spillback detection methods that utilize CV data and a signal control strategy to avoid the occurrence of spillbacks in the next signal cycle. The first method, gap-based detection, is dependent on the information obtained from the CV trajectories and only uses the stopping position of the last CV equipped vehicle that joins the queue in the link of interest. The second method, shockwave-based detection, utilizes information on the signal settings of the upstream intersection in addition to the CV information in order to predict the formation of the back of the queue based on flows measured using CV data. The proposed queue spillback detection methods and alternative signal control strategy were tested through simulation on a four-signal segment of a real-world arterial, San Pablo Avenue in Berkeley, California.

The results indicate that both the gap-based and the shockwave-based methods can correctly detect the occurrence of spillbacks in more than 80% of the cycles for a range of penetration rates. The shockwave-based method was found to be more effective for penetration rates in the range of 10 to 20% while the simpler gap-based detection presented the best results for CV penetration rates that are higher than 20%. The proposed signal control strategy prevented the occurrence and mitigated the impact of spillbacks by maintaining shorter than the critical link length queues and by redistributing the delays at further upstream links that have more storage space for residual queues. More importantly, the proposed signal control strategy reduces the variation of the maximum queue length, as indicated by the coefficient of variation that ranges between 0.06 and 0.09 for different penetration rates when the proposed signal control strategy is in place compared to a value of 0.16 when fixed signal settings are implemented.

Queue Management for Grid Networks

Control strategies for congestion avoidance in grid networks based on CV data were developed and tested in a real-life network Post-Oak in Houston, Texas. The first strategy consisted of a perimeter control approach, were vehicular entrances on the perimeter of the model were limited by reducing the green times on the network entry links. The simulation results showed small network improvements, mainly because the network includes several internal exits that cannot be controlled. The strategy was effective on the network segments that experience queue spillbacks. The second strategy consisted of signal phase changes to the upstream and downstream intersections in a network link that spillbacks are detected. The strategy resulted in small benefits because could not prevent queue spillback in the whole network. The final strategy consisted of reducing the system cycle length and providing additional green time to those network links that spillback first detected. This strategy produced the highest benefits.

The level of improvements in MOEs that can be obtained from the proposed strategies depend on the characteristics of the particular network under study. Additional evaluations of the proposed strategies should be performed on networks covering a range of geometric, traffic and control conditions.

Dynamic Lane Grouping

Dynamic lane grouping (DLG) is a control strategy that changes the lane utilization at the signalized intersection approach to accommodate spatial variations in traffic demands. The strategy is practically feasible through CV data that provide origin-destination information. The problem is formulated as a mathematical programming model to determine the optimal lane allocation of an isolated intersection in terms of minimizing the maximum lane flow ratio (defined as the assigned flow divided by the saturation

rate). The strategy was evaluated through numerical analysis and microscopic simulation for several scenarios consisting of number of lanes, traffic levels, and signal settings.

The results from the numerical analysis show that the proposed strategy can achieve significantly better performance in terms of maximum flow ratio and average delay. For example, for the three-lane approaches and spatial variation of demand of 40%, DLG reduces the maximum flow ratio by 41% and the average delay by 35% compared to fixed lane allocation. DLG produces higher benefits when the intersection becomes oversaturated. The simulation results confirm the findings from the numerical analysis and shows that the DLG strategy can provide significant energy/environmental benefits.

10.1.3 Control Strategies for Safety

Dynamic All-Red Extension (DARE)

Red-light-running (RLR) is a major safety issue at signalized intersections. Dynamic all-red extension (DARE) inserts an all-red interval only when the likelihood of collision to happen is high, to protect other vehicles from potential RLR related collisions. The effectiveness of DARE relies in large part on reliable detection of the hazardous situation caused by RLR. CV data may provide the required data for DARE.

We developed a probabilistic framework and prediction algorithm for DARE. We use $PrET_{RLR}$, which is the time lapse between the time for the RLR vehicle to clear the conflict zone and the a priori entry time of the first-to-enter vehicle from conflict approach, as the primary index in the probabilistic model to identify RLR hazards. The hazard definition, which is determined by the intersection layout, empirical entry time of conflict traffic and trajectory characteristics of RLR vehicles, is linked to the time into red of the violator. We then developed a model using a minimum set of two point sensors to identify the RLR hazard with speed measurements and car-following status.

A data acquisition and processing system was deployed at the intersection of San Pablo Avenue and Brighton Avenue in the city of Albany, CA to obtain detailed field data to study driver behavior related to the RLR hazard and to set the parameters of the proposed DARE algorithm. The system includes three SmartMicro radar sensors (SMS) which can collect all vehicle trajectories of approaching, inside the intersection box, and leaving the intersection. In addition an upgraded conflict monitor card inside the signal controller cabinet outputs the traffic signal status for each permitted signal phase. In the four months of data collection, a total of 863 RLR events occurred at the intersection.

The evaluation of the proposed framework using field collected data from two intersections in the San Francisco Bay Area showed that the proposed algorithm can effectively detect RLR hazards. At the El Camino Real site with two fixed location point detector, at a false alarm rate of less than 5% (or equivalently once per about eight hours), the algorithm reached a correct detection rate from over 70% to over 80% for various legs, while at the San Pablo Ave site with continuous detection (which is similar to CV data), the algorithm reached a correct detection rate over 95% at a false alarm rate of less than 5%.

Minimization of Vehicle Arrivals in the Yellow Interval

We investigated the impacts of vehicle arrivals during the yellow change interval on the frequency of RLR and on intersection operation efficiency using vehicle trajectory data from the NGSIM data set (Peachtree Blvd, Atlanta) and the San Pablo Avenue/Brighton Avenue intersection. The results from both sites showed that the higher arrival flow in the yellow increases the RLR frequency and reduces intersection efficiency resulting in higher stop rate and delay. These results confirm the findings from earlier studies along El Camino Real in Palo Alto.

New control strategies based on segmentation of traffic platoons were developed and tested to reduce the proportion of vehicles arriving in the yellow interval. These strategies can be deployed with CV data, which provide trajectory information to implement platoon based signal control. Three control strategies including green extension, phase termination, and double cycle, were proposed for platoon-based control. The determination of whether to carry out an optimal timing schedule, chosen from the three proposed strategies, or to keep on the original timing schedule is based on the prediction of their impacts on total delay. The proposed control strategies were tested using the trajectory data collected at the San Pablo Avenue/Brighton Avenue intersection, and the results indicate that in most cases the proposed strategies can significantly improve the intersection performance.

10.1.4 Driver Advisory for Minimum Fuel and Emission

We developed a prototype system to provide speed advisories to drivers for minimum fuel consumption. The prototype system consists of a research vehicle, a cloud server and wireless communication devices to traffic signals. The vehicle is equipped with PCs for the computation of the speed recommendation as well as the capability to display the speed recommendation directly in a programmable instrument cluster. The connection to the cloud server is maintained via a cellular 4G connection. The traffic signals send information on their signal settings via a 3G connection to a cloud server, which forwards the appropriate information to the research vehicle.

The speed advisory algorithm provides drivers the optimal speed profile based on inputs from traffic signal's SPaT information, vehicle's current state, and the maximum traffic speed constraint. In addition to the speed recommendation, a control strategy for signal priority to individual vehicles (APIV) was investigated alone as well as in combination with the speed advisory system. APIV adjusts the signal settings to provide additional green time to favor individual vehicles approaching a signalized intersection. The proposed APIV strategy is taking advantage of the CV data which enable accurate prediction of the of individual vehicles' arrival times at intersections and allow the signal to grant priority if needed to facilitate the vehicle's movement.

The system was first tested in a controlled environment: a fixed time traffic signal with no cross traffic or other traffic present at the Richmond Field Station campus of UC Berkeley. Compared to the uninformed driving scenario the speed advisory system achieved significant fuel savings: using the speed advisory system, fuel savings of 13.6% were observed. For the APIV approach without the speed advisory fuel savings of 19.1% were achieved. The combination of both approaches showed fuel savings of 28.4%. Further analysis indicates that fuel savings for the informed driving scenario are mainly from the drivers' early slowing down and cruising through the intersection without having to come to a complete stop. The fuel savings due to APIV are mainly contributed by an increased proportion of driving through green.

The system was also tested on a real-world arterial El Camino Real with three coordinated actuated traffic signals. A new speed advisory algorithm was developed for actuated coordinated traffic signals and the presence of other vehicles. Due to limited time and resources, it was not possible to do any meaningful measurements on the El Camino Real field tests. Nevertheless important observations were made during the test runs performed. It was observed that is extremely difficult to follow recommended speeds that are continuously changing over time, while driving with other vehicles. This indicates that besides the speed advisory algorithm itself, the method of displaying the speed recommendation is crucial for the success of similar systems in the future.

10.2 Future Research

The deployment of vehicle-infrastructure cooperation through connected vehicles offers the potential for significant improvements is in the efficiency of traffic signal control. This research project produced analysis procedures for estimating performance measures and several promising control concepts based on connected vehicle data for improving mobility, safety and reducing adverse environmental impacts. Ongoing and future research will cover but not limited to the following topics:

Further research is needed on the queue spillback avoidance control strategy to consider multiple intersections upstream of the critical one so that spillbacks are avoided in the whole network and queues are distributed homogeneously among all affected links for more efficient traffic operations. Also, the proposed control strategies for congested grid networks need to be further evaluated on networks a range of geometric, traffic and control conditions.

Further research is needed on the dynamic lane group (DLG) concept includes further testing of DLG on different intersection configurations and on arterial streets. Also research is needed on mitigations for adverse DLG impacts such the intensive lane changing upstream of the intersection including deployment of mid-block pre-signals, dynamic lane signs and in-vehicle advance information to drivers regarding the lane configuration at the intersection approach.

Additional data collection should be performed on vehicle trajectories at signalized intersections covering a range of conditions to further test and validate the proposed concepts for red light running avoidance, i.e.., the DARE algorithm and the control strategies for minimization of arrival flow in the yellow interval.

The findings from the field tests in a real-world arterial indicate that the speed advisory algorithm for coordinated actuated signals need to be extended and improved to explicitly consider i) the interactions with adjacent vehicles such that driver recommendations are consistent with real-world driving conditions; ii) the uncertainties in estimating the duration green phase for actuated signals; and iii) the method of providing speed recommendations, which can be easily understood and followed by drivers.

REFERENCES

1. Skabardonis A., and G. Gomes, "Measure and Field Test the Effectiveness of Adaptive Control for Traffic Signal Management," PATH Research Report UCB-ITS-PRR-2010-36, Institute of Transportation Studies, University of California, Berkeley, August 2010.
2. Stevanovic, A., "Adaptive Control Systems: Domestic and Foreign State of Practice," NCHRP Synthesis Report 403, Transportation Research Board, Washington DC, 2010.
3. Getttman, D., H. Liu, M. Abbas, and A. Skabardonis, "Traffic Control for Oversaturated Conditions," Draft Final Report NCHRP 3-90, Transportation Research Board, Washington DC, March 2011.
4. Dowling R.G., and A. Skabardonis, "Traffic Signal Control Strategies for Varying Demands and Capacities," Interim Report NCHRP 3-97, Transportation Research Board, Washington DC, 2009.
5. Urbanic T., D. Bullock, L. Head, and D. Gettman, "Traffic Signal State Transition Logic Using Enhanced Detector Information," Final Report NCHRP 3-66, Transportation Research Board, Washington DC, 2006.
6. Connected Vehicle Traffic Signal Information Exchange Meeting Notes – DRAFT, February 2011.
7. Skabardonis, A., "Fixed-Time vs. Actuated Control in Coordinated Signal Systems," paper 05-2529, 84th TRB Annual Meeting, Washington DC, January 2005.
8. Head, L. "Traffic Control in a VII Environment," Traffic Signal Systems Committee Workshop, 88th TRB Annual Meeting, Washington DC, January 2008.
9. TRAVOLUTION prototype, Ingolstadt, Germany, http://www.haeuwatchits.info/press/press_detail.asp?pid=20&aid=842.
10. Venkatanarayana, R., H. Park, B. Smith, C. Skerrit and N. Ruhter, "Application of IntelliDrive to Address Oversaturated Conditions on Arterials," 90th TRB Annual Meeting, Washington DC, January 2011.
11. Wang, L., L. Zhang, WB Zhang, and K. Zu, "Red Light Running Prediction for Dynamic All-Red Extension at Signalized Intersection," 12th IEEE ITSC Conference, St Louis, MO, Oct 2009.
12. Vasudevan, M., "VII Data Characteristics for Arterial Applications," VII/CICAS Workshop, TRB Signal Systems Committee Mid-Year Meeting, San Jose, CA, July 2007.
13. Shladover S.M., and T.M. Kuhn, "Traffic Probe Data Processing for Full-Scale Deployment of Vehicle-Infrastructure Integration," Transportation Research Record No. 2086, pp. 115- 123, 2008.
14. FHWA, "Definition, Interpretation and Calculation of Traffic Analysis Tools Measures of Effectiveness," Traffic Analysis ToolBox, Vol. VI, Report FHWA-HOP_08-054, January 2007.
15. Transportation Research Board, "Highway Capacity Manual 2000," Special Report 209, Washington DC, 2000 (Updated HCM 2010, 2010).
16. Comert, G. and M. Cetin, "Queue Length Estimation from Probe Vehicle Location and the Impacts of Sample Size," European Journal of Operational Research, Vol. 197 (1), pp. 196–202, 2009.
17. Hao P. and X. Ban, "Vehicle Queue Location Estimation for Signalized Intersections Using Sample Travel Times from Mobile Sensors," 90th TRB Annual Meeting, Washington DC, 2011.
18. Y. Cheng, Y., X. Qin, J. Jin, B. Ran, and J. Anderson, "Cycle By Cycle Queue Length Estimation for Signalized Intersections Using Sampled Trajectory Data," 90th TRB Annual Meeting, 2011.

19. Li, M., Z. Zou, F. Bu, and W. Zhang, "Application of vehicle infrastructure integration (VII) data on real-time arterial performance measurements," 87th TRB Annual Meeting, Washington DC, January 2008.

20. Cetin, M., G. List, and Y. Zhou, "Factors Affecting the Minimum Number of Probes Required for Reliable Travel Time Estimation," Transportation Research Record, # 1917, pp. 37–44, 2005

21. Alexiadis, V., J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The Next Generation Simulation Program," ITE Journal, Vol. 74, 2004, pp. 22–27.

22. Edie, L., "Traffic Stream Measurements and Definitions," Proceedings of the 2nd International Symposium on the Theory of Traffic Flow, 1963, pp. 139–154.

23. Jones T., and R. Potts, "The Measurement of Acceleration Noise-A Traffic Parameter," Operations Research Vol. 10, No. 6, 1962, pp. 745–763.

24. El Faouzi, N.-D., H. Leung, and A. Kurian, "Data Fusion in Intelligent Transportation Systems: Progress and Challenges - A Survey," Information Fusion. Vol. 12, 2011, pp. 4-10.

25. Abu-Lebdeh, G., and R.F. Benekohal. Development of traffic control and queue management procedures for oversaturated arterials," Transportation Research Record #1603, Journal of the Transportation Research Board, 1997, pp. 119–127.

26. Skabardonis, A. and N. Geroliminis, "Real-time monitoring and control on signalized arterials," Journal of Intelligent Transportation Systems, Vol. 12 (2), 2008, pp. 64–74.

27. Smith, B., R. Venkatanarayana, H. Park, N. Goodall, J. Datesh, and C. Skerrit, "IntelliDrive[SM] Traffic Signal Control Algorithms," Final Report, IntelliDrive[SM] Pooled Fund Study, Univ of Virginia, 2011.

28. He, Q., K. Head, and J. Ding, "PAMSCOD: Platoon-based arterial multi-modal signal control with online data," Transportation Research Part C, Vol. 20 (1), 2012, pp. 164–184.

29. Transport Simulation Systems (TSS), "Aimsun Users Manual," v6.1, 2010.

30. C. Daganzo C.F and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," Transportation Research Part B, Vol. 42 (9) pp. 771-781, 2007.

31. Skabardonis, A., "Progression Through a Series of Intersections with Traffic Actuated Controllers," Final Report FHWA RD-89-132, U.S. Department of Transportation, December 1988.

32. Traffic Modeling Guidelines, "TfL Traffic Manager and Network Performance Best Practice, Version 3.0, Transport for London, September 2010. Appendix III, pp. 179-180.

33. Harvey, A. and D. Bullock, "Implementation of a Distributed Control System for Dynamic Lane Assignment.," Proceedings of the Twenty-Eighth Southeastern Symposium on System Theory, 1996, pp. 524–528.

34. Middelham, F., "State of Practice in Dynamic Traffic Management in The Netherlands," Netherlands, 2006.

35. Dresner, K. and P. Stone, "A Multi-agent Approach to Autonomous Intersection Management," Journal of Artificial Intelligence Research, Vol. 31(1), 2008, pp.591–656.

36. Wong, C.K. and S.C. Wong, "Lane-based Optimization of Signal Timings for Isolated Junctions," Transportation Research Part B, Vol. 37(1), 2003, pp.63–84.

37. Akcelik, R., "On the Estimation of Lane Flows for Intersection Analysis," Australian Road Research, Vol.19 (1), 1989, pp.51–57.

38. Koonce, P. et al., "Traffic Signal Timing Manual," Report FHWA-HOP-08-024, June 2008.

39. Pitney Bowes Software, "Quadstone Paramics" http://www.paramics-online.com/, accessed on February 13th, 2012

40. Barth, M.J et al, "The Development of a Comprehensive Modal Emissions Model," Final Report for NCHRP Project 25-11, April 2000.

41. FHWA, "Red-Light Running Fatalities," http://safety.fhwa.dot.gov/intersection/redlight/data/rlr_fatal

42. Retting RA , A.F. Williams, and M.A. Greene, "Red-Light Running and Sensible Countermeasures: Summary of Research Findings," Transportation Research Record, No. 1640, Journal of the Transportation Research Board, 1998, pp. 23-26.

43. Hu, W., A.T. McCart, and E.R. Teoh, "Effects of Red Light Camera Enforcement on Fatal Crashes in Large US Cities," Journal of Safety Research, Vol. 42, 2011, pp. 272-282.

44. Maile, M. and L. Delgrossi, "Cooperative Intersection Collision Avoidance System for Violations (CICAS-V) for Voidance of Violation-Based Intersection Crashes," Enhanced Safety of Vehicles, 2009.

45. Souleyrette, R.R., M.M. O'Brien, T. McDonald, H. Preston, and R. Storm, "Effectiveness of All-Red Clearance Interval on Intersection Crashes," Final Report, Minnesota Department of Transportation, 2004.

46. Gettman, D. and L.K. Head, "Surrogate Safety Measures from Traffic Simulation Models," Transportation Research Record No. 1840, Journal of The Transportation Research Board, 2003, pp. 104-115.

47. ITE "Traffic Engineering Handbook," 6th Edition, Institute of Transportation Engineers, 2009, Washington, DC.

48. Grembek, O., I. Li, WB Zhang, and K. Zhou, "Analysis of Cycle-Based Data and Development of Enhanced Signal Timing Models to Reduce Red Light Running," Paper # 07-0774, 86th TRB Annual Meeting, Washington DC, January 2007.

49. Li., M., K. Boriboonsomsin, G. Wu, WB Zhang, and M. Barth, "Traffic Energy and Emissions Reductions at Signalized intersections: A study of the Benefits of Advanced Information," International Journal of ITS, Vol. 7(1), pp.49-58, 2009.

50. Li., M., et al, "Toward Deployment of Adaptive Transit Signal Priority (ATSP)," PATH Research Report UCB-ITS-2008-24, University of California, Berkeley, October 2008.

51. Winkler A., "Advanced Signal Control," Interim Report to the EAR Project, BMW of North America, March 2011.

52. Barth, M., S. Mandava, K. Boriboonsomsin, and H. Xia, "Dynamic ECO Driving for Arterial Corridors," Integrated and Sustainable Transportation Systems (FISTS), 2011.