Title
Improving Performance of Coordinated Signal Control Systems Using Signal and Loop Data

Authors
Li, Meng
Zhang, Liping
Song, Myoung Kyun
et al.

Publication Date
2010-03-01

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

Improving Performance of Coordinated Signal Control Systems Using Signal and Loop Data

Meng Li, Liping Zhang, Myoung Kyun Song, Guoyuan Wu, Wei-Bin Zhang, Lihui Zhang, Yafeng Yin

California PATH Research Report
UCB-ITS-PRR-2010-7

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 Task Order 6332

March 2010

ISSN 1055-1425

Improving Performance of Coordinated Signal Control Systems Using Signal and Loop Data

University of California, California PATH

Meng Li

Liping Zhang

Myoung Kyun Song

Guoyuan Wu

Wei-Bin Zhang

University of Florida

Lihui Zhang

Yafeng Yin

Draft Final Report for TO 6332

September 9, 2009

ACKNOWLEDGEMENTS

This work was performed by the California PATH Program at the University of California at Berkeley and University of Florida in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation (Caltrans). 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 contents do not necessarily reflect the official views or policies of the State of California.

The authors wish to thank Kai Leung, Jerry Kwong, Paul Chiu, James Lau, Harris Zaw, Jorge S. Fuentes, Koon Tse, Mo Ketabchi, and Sonja Sun of Caltrans for their continuing cooperation and support during the study. We also thank our colleagues, Prof. Alex Skabardonis, Dr. Kun Zhou, Scott Johnston, Lian Thang, and Bart Duncil of California PATH, for their help and discussion.

Author List

University of California, Berkeley
(Primary author for Chapter 2, 3, and 7)
Meng Li
Liping Zhang
Myoung Kyun Song
Wei-Bin Zhang

University of Florida
(Primary author for Chapter 1, 4, 5, and 6)
Lihui Zhang
Yafeng Yin

Abstract

In this project, a traffic data collection system based-on the iDEN wireless network has been developed, lab tested and preliminarily tested in the field. The objective of the system is to provide a cost-effective and easy-to-maintain system that could still reliably provide traffic data over the wireless link.

The project team developed an arterial performance measurement method that is based on the signal infrastructure data collected at PT^2 Lab, U.C. Berkeley. The performance of the proposed model is illustrated by using a simulation study. The six-signal simulation network covers both heavily congested and light traffic intersections. The proposed model works well at both the intersection level and the arterial level. Estimation errors of travel time, number of stops and travel time reliability are insignificant.

A general approach for robust signal optimization under demand uncertainty or flow fluctuations has been developed. The approach has been demonstrated in two different settings. The first one deals with the problem of synchronization of actuated signals along arterials along arterials. The other demonstration is to optimize the signal settings including the cycle length, green splits, offset points and phase sequences in an integrated manner, taking into account the day-to-day demand variations or uncertain further demand growth. The robust timing plans resulted from both models have been demonstrated in numerical tests to perform better against high-consequence scenarios without losing optimality in the average sense. Although the robust signal timing approach is applicable more widely, this report has been focused on timing models for pre-timed arterials.

Executive Summary

A traffic data collection system based-on the iDEN wireless network has been developed, lab tested and preliminarily tested in the field. The objective of the system is to provide a cost-effective and easy-to-maintain system that could still reliably provide traffic data over the wireless link. The mobile wireless network has its inherent characteristics of less reliable than the wired network. We have built an adaptive flow control layer over the wireless TCP communication to address the occasional outage problem. The validity of the method is in the fact that the collection second by second traffic signal status data does not necessarily achieve 0% loss rate. So with allowing some minimized data loss, we could control the data flow to avoid exceeding the allowed data rate and thus leading to major outage. The system is able to continuously provide over 2.68kbps upload data rate per remote handset for over 95% of the time, i.e., one remote handset could deliver data fetched from signal controllers at a period of 200ms continuously with about 5% of data loss rate. On the cost side, one set of client hardware costs ~$70 and $10 monthly, while covering up to 8 local signal controllers. These combined features, low cost and high performance, make the system a unique solution for traffic data collection.

This report has described the development of an arterial performance measurement method that is based on the signal infrastructure data collected at PT^2 Lab, U.C. Berkeley. The performance of the proposed model is illustrated by using a simulation study. The six-signal simulation network covers both heavily congested and light traffic intersections. The proposed model works well at both the intersection level and the arterial level. Estimation errors of travel time, number of stops and travel time reliability are insignificant.

The findings of the study together with the data collection means developed by PT^2 Lab provide a cost-effective way to achieve an arterial performance measurement system. The data and analysis results will support transportation researchers on various research topics such as traffic control and operations; help planners and local agencies on daily management and system monitoring; and provide travelers real-time information when scheduling their trips.

For the next step, we will further calibrate, validate and demonstrate our model by using field data and conducting field experiments. Some given parameters in this study, such as the demand factor α and time window T_W when queue spillback happens, average deceleration and acceleration rates, free flow speed, and turning ratios, should be calibrated or measured based on field data and/or observations. Effective adaptive models should be developed to dynamically estimate parameters such as saturation flow and start-up lost time. Sensitivity analysis on some of the key parameters will be studied. Moreover, the model to address the over-saturated scenarios will be developed and validated by the simulation network and then the field data.

This report has presented a general approach for robust signal optimization under demand uncertainty or flow fluctuations. The approach has been demonstrated in two different settings. The first one deals with the problem of synchronization of actuated signals along arterials along arterials. The formulation is a mixed integer linear program easily solvable using the state-of-the-art solvers. The computational time only increases polynomially as the number of scenarios increases. The approach can be used to either design a new coordination plan for implementation or fine-tune the plan offline after implementation. In the latter case, the specification of scenarios is an easy task with the archived signal status data. One may randomly select 50 to 200 red time realizations from the data and assume equal probability of occurrence. To design a new coordination plan where the

distributions of red times are normally unknown, we suggest specifying 50 to 200 scenarios as the points that equally divide the red-time intervals into K+1 segments and assume equal probability of occurrence of $1/K$. The suggestion is based on our observation from the numerical experiments that the robust formulation is not overly sensitive to the specification of scenarios. Even with biased scenarios, the formulation may still produce meaningful robust plans.

The other demonstration is to optimize the signal settings including the cycle length, green splits, offset points and phase sequences in an integrated manner, taking into account the day-to-day demand variations or uncertain further demand growth. Considering a large number of binary variables in the formulation, we have developed a simulation-based GA to solve the problem. It should be mentioned that the setting of the GA-based algorithm, such as the fitness function, may influence the quality of the final plan and the convergence speed. Numerical experiments are needed to fine-tune the setting. We also note that the simulation-based model is broadly applicable, particularly when the objective function is difficult or time-consuming to evaluate.

The robust timing plans resulted from both models have been demonstrated in numerical tests to perform better against high-consequence scenarios without losing optimality in the average sense. Although the robust signal timing approach is applicable more widely, this report has been focused on timing models for pre-timed arterials. Future study can be conducted to expand the proposed models for more sophisticated corridors and grid networks.

1 Introduction

This report constitutes the final deliverables for California PATH Task Order 6332—"Improving Performance of Coordinated Signal Control Systems Using Signal and Loop Data." The project has investigated the following:

- developing a cost-effective data collection system for arterials;
- developing an arterial performance measurement system;
- and proposing a stochastic programming approach for robust signal timing optimization using signal and loop data.

1.1 Motivation

A large segment of traffic signal control systems in California are closed-loop and actuated, further improvements in the efficiency (e.g., delay per vehicle) and robustness (e.g., variance of delay per vehicle) of these systems can yield significant improvements in the management of traffic flows and mitigation of congestion. The advancement and deployment of telecommunication and ITS technologies make real-time traffic and signal status data more readily available. These data provide tremendous opportunities to allow signal control systems to operate more efficiently and robustly.

Under TO5325, preliminary investigation has been made on the use of traffic and signal status data along two research directions [1-1]. The first is to gradually adjust or refine the signal settings *in operations* to make the signals more responsive to the traffic. An offline offset refiner has been developed using archived signal status data to fine-tune the signal offsets to provide smoother progression in either one-way or two-way coordination, addressing the so-called problem of "early-return-to-green". The application of the refiner to a stretch of El Camino Real shows that it reduces the red-meeting probability from

21.8% to 2.3% without affecting adversely the bandwidth. The other research direction is to explicitly consider traffic flow fluctuations in signal timing optimization *prior to operations* and develop robust timing plans that are able to tolerate those fluctuations and perform stably. It has been demonstrated at an isolated intersection that, when compared with traditional approaches, robust timing plans may reduce standard deviations of delay per vehicle by 11.3% to 16.7%, and 28.2% to 44.6% under over-saturated and under-saturated conditions respectively, without worsening off the average performance. Those robust timing plans also allow slower deterioration of performance. We note that the signal timing process is normally time-consuming and expensive. Thus most local transportation departments or state DOTs cannot afford to frequently fine-tune their traffic signals unless changes in traffic conditions are so significant that the system begins performing poorly. It has been estimated that traffic experiences an additional 3%-5% delay per year as a consequence of not retiming signals as conditions evolve over time [1-2]. Therefore, it is desirable to have timing plans that accommodate or tolerate these changes in traffic to a greater extent.

1.2 Proposed Research

Motivated by the reduced costs of wireless communication services and increasing processing capabilities of cell phone handsets, the Parsons Traffic and Transit Laboratory (PT^2L) at University of California, Berkeley has been developing a Motorola iDEN phone based communication device for the coordinated signal control system. Without changing the existing infrastructure in control cabinets, the communication device with customized Java programs is capable of pulling traffic signal status and traffic detection data from the local controllers through a RS232 connection using AB3418 protocol and forwarding the data to TMC through wireless communication. The entire process can be automatic and in real time. The costs for installation, maintenance, and monthly service

are very low comparing with state-of-the-art application. The initial cost per signalized intersection for such communication setup is as low as $5 only $1 for monthly communication service.

Built upon this ongoing work and previous results obtained from Task Order 5325, the purpose of this task order is to materialize the Motorola iDEN phone based communication devices and develop a cost-effective data collection system. The signal and loop data made available from the system are further utilized to develop an arterial performance measurement system. Moreover, the notion of robust signal timing proposed in Task Order 5325 is extended to be a general approach to determining signal timing plans that improve efficiency and robustness of currently deployed closed-loop signal control systems

1.3 Report Outline

This is the first of seven chapters of the final report. Chapter 2 describes the development of a cost-effective data collection system for arterials and presents some results from field experiments. Given the loop and signal status data, Chapter 3 develops an arterial performance measurement system and verifies the system using data from probe vehicles. Utilizing the same set of loop and signal data, Chapters 4, 5 and 6 start another front of development. Chapter 4 introduces the basic concept of a stochastic programming approach to signal timing optimization that explicitly and proactively considers the demand uncertainty. Chapter 5 applies the approach to synchronize actuated signals along arterials, addressing the so-called “early-return-to-green” issue. Chapter 6 further formulates a scenario-based stochastic programming model to optimize the timing of actuated signals along arterials under day-to-day demand variations or uncertain traffic

future growth. Conclusions and a brief description of future work are provided in chapter 7.

1.4 Reference

1-1 Yin, Y., Liu, H.X., Laval, J.A., Lu, X.-Y., Li, M. Pilachowski, J. and Zhang, W.-B. *Development of an Integrated Microscopic Traffic Simulation and Signal Timing Optimization Tool*, Final Report for Task Order 5325, 2005.

1-2 Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D. and Mirchandani, P. ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems. Design Guidelines for Deploying Closed-Loop Systems. In *Transportation Research Record* No. 1856, TRB, National Academies Council, Washington, D.C., 2003, pp. 175-184.

2 Development of Cost-effective Data Collection System for Arterials

2.1 System Overview

The data collection system for arterials is developed based on the existing mobile communication networks and Motorola® iDEN[1] mobile handsets. It is designed to provide a cost-effective and reliable means to remotely collect traffic data in real time from the roadway network. The choice of using wireless mobile network ensures an easy-to-maintain and cost-effective way for the field to center data communication.

2.1.1 Components and Features

The data collection system includes remote handsets for real time traffic data collection from both the field master and local signal controllers, a reliable wireless link based on the iDEN mobile network and highly scalable data centers with web based system management support.

The moderately priced Motorola® iDEN series phones are used as remote data modems. The phones feature a standard RS232 serial port, the iDEN wireless data connection support and Global Positioning System (GPS) support.

The iDEN wireless network serves as a cost-effective and reliable communication link for the system. The channel capacity limit is 9.6kbps. This data rate is adequate for the

[1] **Integrated Digital Enhanced Network (iDEN)** is a mobile telecommunications technology, developed by Motorola, which provides its users the benefits of a trunked radio and a cellular telephone. iDEN places more users in a given spectral space.
(http://en.wikipedia.org/wiki/Integrated_Digital_Enhanced_Network)

traffic application and the service contract pricing for this network is superior to other available rate plans.

Highly scalable data centers and a web-based management system are also parts of the system. MySQL, an open–source, high performance database is used to store the collected data for further processing. The system is designed with a flexible architecture so that multiple data centers can be incorporated directly into the system as need or requirements dictate.

In summary, the system has the following features:

- A continuous real-time traffic data collection system structured for thousands of master controllers;
- Conformity to the AB3418 standards and the ability to work with various kinds of signal controllers;
- Special support for Caltrans 170E signal controller and CTNet server program;
- Synchronization of signal controllers' local clocks to the GPS reference;
- Low system deployment and operational cost;
- Reliable communication based on an adaptive wireless link;
- Web-based system management that simplifies maintenance efforts;

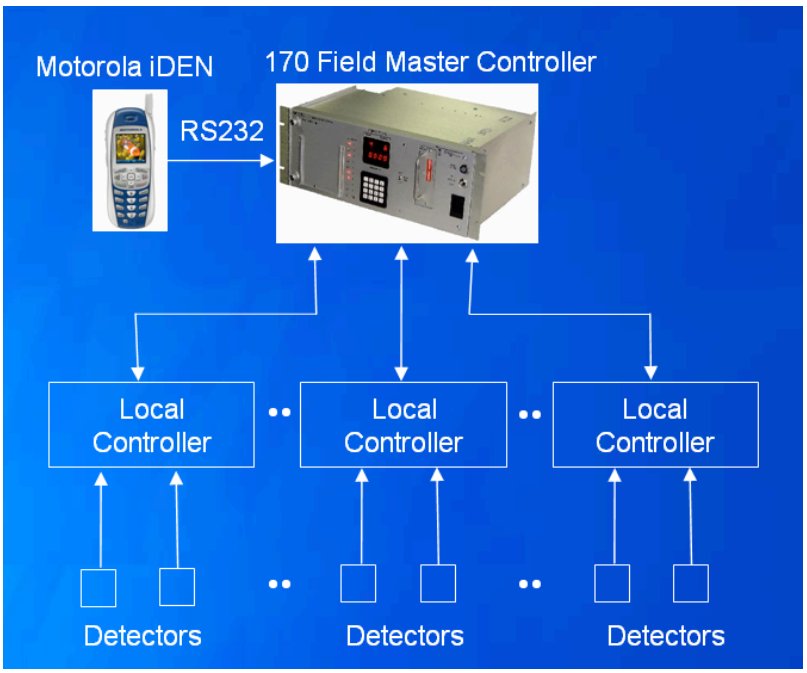


Figure 2-1 Field Set-up between Motorola iDEN Phones and 170 Controller

As shown in Figure 2-1, a typical field set-up is as simple as connecting the handset to a master signal controller. All of the traffic data from the connected local signal controllers are automatically forwarded to a data center and stored in the database. Meanwhile, using the same configuration and equipment, the traffic data are also automatically forwarded to any designated CTNet server by the data center.

2.1.2 System architecture

The data collection system consists of three layers: field layer, data server layer, and data application layer. The architecture of the system is illustrated in **Figure 2-2**.

At the field layer, the Motorola iDEN series handsets with customized Java programs are capable of pulling traffic signal status and traffic detection data from the local controllers through an RS232 connection using AB3418 protocol and forwarding the data to the

remote traffic data server via wireless communication. The developed field set-up does not require any hardware changes in the existing infrastructure in control cabinets. The entire data collection process can be automatic and in real time.

Data obtained by the remote handsets are transmitted to the data center(s) directly using the mobile communications network. Currently the Nextel® data service is selected to forward the data. Using the mobile telecommunications network enables a very cost effective way to interconnect the distributed traffic controllers. The major disadvantage of the wireless communication is the highly variable nature of the link quality, especially for the data service. Adaptive flow control technique is employed to cope with the link variation problem. The achieved system performance measurements, including the throughput, data loss percentage, etc. are presented in Section 2.3.

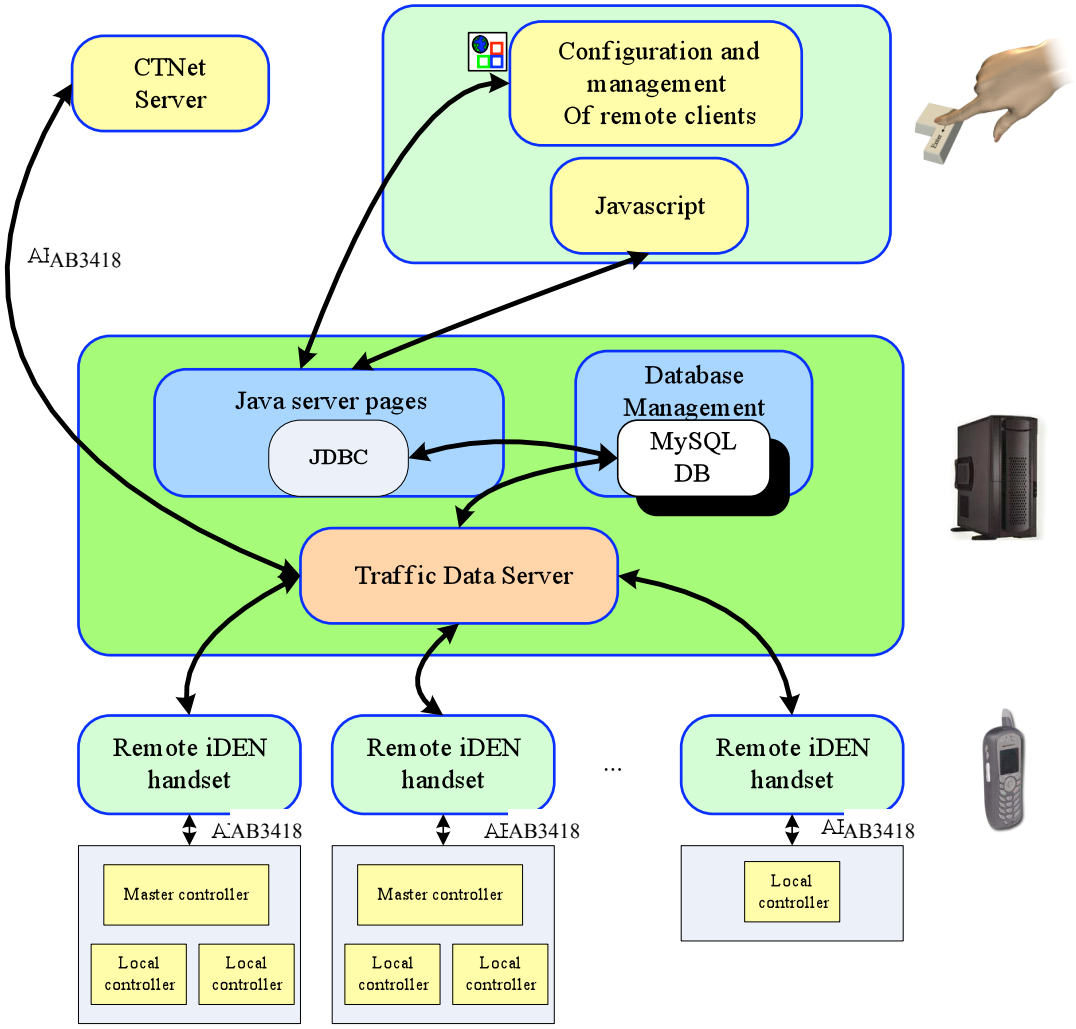


Figure 2-2 System architecture

Data from various intersections are sent to one or more data centers and stored in MySQL® database. Data are organized and processed upon standardized database application interfaces and connectivity technologies, thus dramatically reducing efforts in maintaining, interpreting and analyzing the data.

Application of the data collection system can be, but definitely is not limited to, the proposed signal control optimization for coordinated intersections. Those applications are built upon the DataBase Connectivity (DBC) technology and the web server technologies

such as Java Server Pages (JSP). Also an easy-to-use, web-based management tool is provided for the maintenance and management of possibly large numbers of handsets in the future.

The traffic data server program developed for this system also features a secure TCP/IP link with the CTNet server program. With this support, the CTNet server program can directly obtain data from this system and send command packages to the remote signal controllers using the reverse data link.

Testing with Caltrans 170E signal controller at RFS has been carried out to evaluate the data throughput and reliability. One data center server with MySQL database server and Sun Java System Application Server has been installed and multiple iDEN cell phones have been employed for testing. One cell phone works with a Caltrans 170E master signal controller with two local controllers, additional eight cell phones generate test data on their own (loop test) due to a lack of signal controllers. The obtained or simulated data are sent to the data center and stored into the database in real time. Two-way communication has been successful tested with CTNet server and one master signal controller.

2.2 Features

The data collection system features real time traffic data collection from the field master controller and local signal controller, a reliable wireless link based on iDEN mobile network, high scalability data center and easy-to-use system management tools.

2.2.1 Data collection using Motorola iDEN phones

2.2.1.1 Components

Motorola iDEN cell phones are employed to collect data from signal controllers. The cell phones collect the field data using a standard RS232 serial port, which is an ideal interface for type 170 signal controllers. These type 170 controllers are the most widely implemented controller type in California. **Figure 2-1**, illustrates the Motorola iDEN phone connected with a type 170 field master controller through its RS232 port. For the coordinated actuated signal control system, one 170 field master can poll up to 31 local signal controllers in real time for their signal status and system detector data. It can then push the data to the iDEN phones through the RS232 connection. For the isolated control system, the Motorola iDEN phone can directly connect to the signal controller through the RS232 port. In either approach, the wireless channel can forward all the field data to the remote database server. Finally, the build-in GPS receiver can provide accurate coordinated universal time, which can then be used to synchronize local controllers and time-stamp the field data.

All the data collection set-up is straightforward for small and large-scale implementations. First of all, it does not require any major hardware changes inside the controller cabinet. Only one RS232 serial cable needs to be plugged into the 170 controller. Secondly, all the equipment, namely an iDEN265 cell-phone, a SONY 4.2Volt 2Amp AC/DC power supply, and a serial cable, are low-cost and commercial-off-the-shelf (COTS) products.

2.2.1.2 Features of data collection

There is Java 2 Micro Edition (J2ME) embedded firmware running in the cell phone to enable the data collection. All of the embedded programs in the cell phones are identical. This unified software approach may ease the maintenance, upgrade and repair.
The features of the data communication between iDEN265 and master signal controller are summarized in **Table 2-1**.

Table 2-1 Features of iDEN265 data collection

	Feature	Specification	Description
1	**Maximum Serial Data rate**	9600bps	This data rate conforms that of the signal controller and is also reconfigurable with the iDEN265 J2ME (Java 2 Micro-Edition) API. Maximum value of 115200bps was tested to be applicable.
2	**Power supply**	AC	The cell phone uses an AC/DC converter to provide power.
3	**Power consumption**	Peak <15W Avg: < 5W	Note that the peak current at the DC end (4.2 volt) is about 2Amp and average less than 0.5 Amp.
4	**Working temperature**	0°C-60°C	As specified by the manufacturer, not tested

The master signal controller polls the local controllers either in focus mode or normal mode. The data collection system can handle data from master controller at an interval of as short as 200ms.

The length of messages varies from several bytes to as long as 67 bytes for the test carried out at PATH. So, for example, when all sentences are 67 bytes long and the master controller is working in focus mode, the maximum requirement for air data rate will be 335byte/s = 2680bps. According to the PATH lab tests, the system meets this requirement over 95% of the time. Details of the test results can be found in Section 2.3.

2.2.1.3 Synchronization of signal controller's local clock

The GPS feature of iDEN265 enables it to update the universal time (UTC time) at a resolution of one second. The received UTC time is then sent to the signal controller for synchronization of the local clock. The new firmware for the Caltrans 170E signal controller that supports this functionality is now available.

By synchronizing the signal controllers using accurate UTC time from GPS, it is possible to coordinate signal controllers for a large area without frequent manual adjustments or a secondary UTC clock.

The UTC time is sent to the signal controller every 10 seconds. It is retrieved from the standard NMEA sentence and repacked into a standard AB3418 message, the format of which is shown in **Figure 2-3**. Note that *dw* is the day in week, *mo* is the month, *d* the day, *y* the year, *h* the hour, *m* the minute, *s* the second and *ms* the millisecond, all from GPS reception. Also note that although it is available, the millisecond data here is not usable.

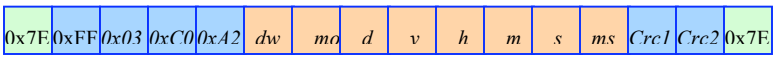


Figure 2-3 Format of the synchronization sentence

2.2.2 Wireless link based on mobile network

The Nextel® iDEN wireless network provides a low-cost wireless full-duplex data service with a minimum data rate of 9.6kbps. The cell phone is utilized as a wireless data modem to access Internet using iDEN technology. Due to the inherent natural variation of the wireless data link, the wireless service is usually not so reliable. In order to achieve high-data-rate and reliable communications, we developed high level control protocols. An adaptive flow control protocol was employed in order to maintain a data link channel with variable but highly reliable capacity on top of the TCP/IP over the iDEN network.

2.2.2.1 TCP/IP over iDEN network

Basically there are two options to address this issue: TCP/IP and UDP/IP solutions. For wireless communication, TCP/IP supplies a reliable solution but requires better channel link quality while UDP/IP relies less on the channel quality. Here is a brief comparison of these solutions for the data collection purpose.

Table 2-2 comparison of TCP and UDP over iDEN wireless network

	TCP/IP	UDP/IP
Connection	Connection	Connectionless
Throughput	Lower	Higher
Security	Better (with connection)	Poor
Flow control	Better	Worse
Conclusion	√	

Using UDP/IP over iDEN network can achieve a higher throughput at the cost of less reliable communication. Since the setup of TCP/IP over iDEN already provides enough bandwidth while outperforming in reliability, it is chosen as the preferred approach in this system.

2.2.2.2 Automatic configured wireless link

The traffic data are sent to the data center using a TCP/IP protocol set. The server program automatically configures the settings of each client. No manual operations are required to configure the remote cell phone in order to transmit data to the data center. The embedded program in the cell phone starts automatically after a power cycle, and begins to listen to the TCP/IP port that is dedicated to the application. Thus the deployment of the cell phone at field is as simple as connecting the serial and power cables and then turning on the phone.

2.2.2.3 Adaptive real time communication

There are two factors related to the variations of the communication data rate. One is that the input data rate from the signal controller may vary from time to time and from site to site. The other factor is the unstable nature of the wireless channel. At times, it is impossible to keep transmitting at the demanded data rate. There is also an outage probability, such that during some short period, no data can be transmitted. Even when there is no outage, the wireless data communication is still vulnerable to multi-path, rain degradation and other factors. These are factors common to all wireless networks, and while the probabilities are low, they are considerations for deployment.

Great amount of effort has been made to ensure a continuous, high throughput data communication over the wireless iDEN network using TCP/IP. The resolution employed in the system is a set of adaptive flow control protocols. On the one hand, the cell phone transmits more when the signal controller is sending more data to the cell phone. On the other hand, the cell phone delivers data in a "best effort" way, which means it sends as much as the channel currently allows during a given period and discards the data, which fails to be delivered after a few seconds.

It is possible to temporarily send more data than the channel allows, but constantly doing so will also cause problems. It is believed that the iDEN network will depreciate the resources of the aggressive client until it is restarted. This is a low-level management control protocol built in to the wireless network to deal with network issues. With adaptive flow control, the remote handsets always achieves nearly best data rate they can get automatically.

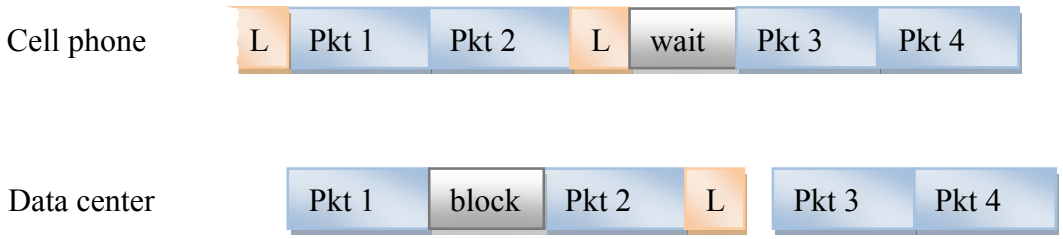


Figure 2-4 Adaptive flow control

Figure 2-4 briefly shows the procedure of the flow control. At the cell phone end, every several data packets (pkt), it will send to the data center a label packet (L) which serves as a virtual timestamp and indicates the relative sequence and absolute numbering of the data packets sent. At the data center end, the server program maintains a slide window buffer and monitors label sequence. The server sends acknowledgement label packets back to the cell phone, so that it could tell the link quality from the delay of the labels and discarding data from its internal slide-window buffer whenever necessary. The flow control mechanism guarantees that the wireless channel itself won't be saturated by greedy client programs without awareness of the instantaneous link quality measurements. Details of the flow control mechanism can be found in Section 0.
It is possible to lose some data with this adaptive flow control approach. During lab testing, in general, less than 1% of data are lost due to application of flow control. More details of the test results are shown in Section 2.3.

2.2.3 Data center and system management

The system works with either single or multiple data centers. Evolving from one data center to multiple data centers requires only a simple on-line reconfiguration process. The data centers can also work seamlessly with the CTNet traffic data server. This is implemented by the traffic server program (c.f. **Figure 2-2**, the system architecture) automatically forwarding data to and receiving data from a CTNet server.

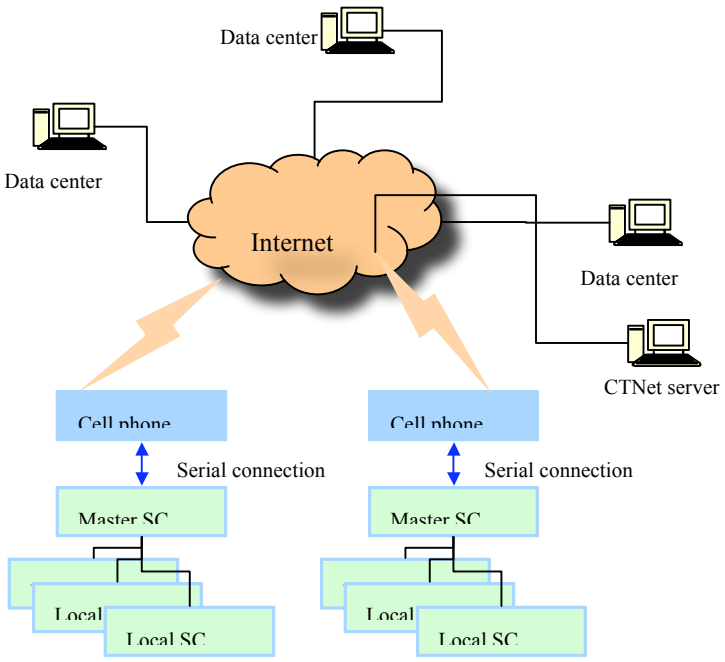


Figure 2-5 Architecture of the data collection system with multiple centers

Currently, the cell phone works as a socket server. The database center PC works as the socket client. The benefits of this setting allow the distribution of the workload of data collection and processing over multiple data centers when the system is massively deployed.

2.2.3.1 Web-based management tool

The system features a web-based management tool that can remotely manage the data centers and remote data collection handsets. **Figure 2-6** shows the user interface of the web-based data collection management tool. For better security, the Secured Shell (SSH) protocol is used for management purposes and only password-authorized users can login.

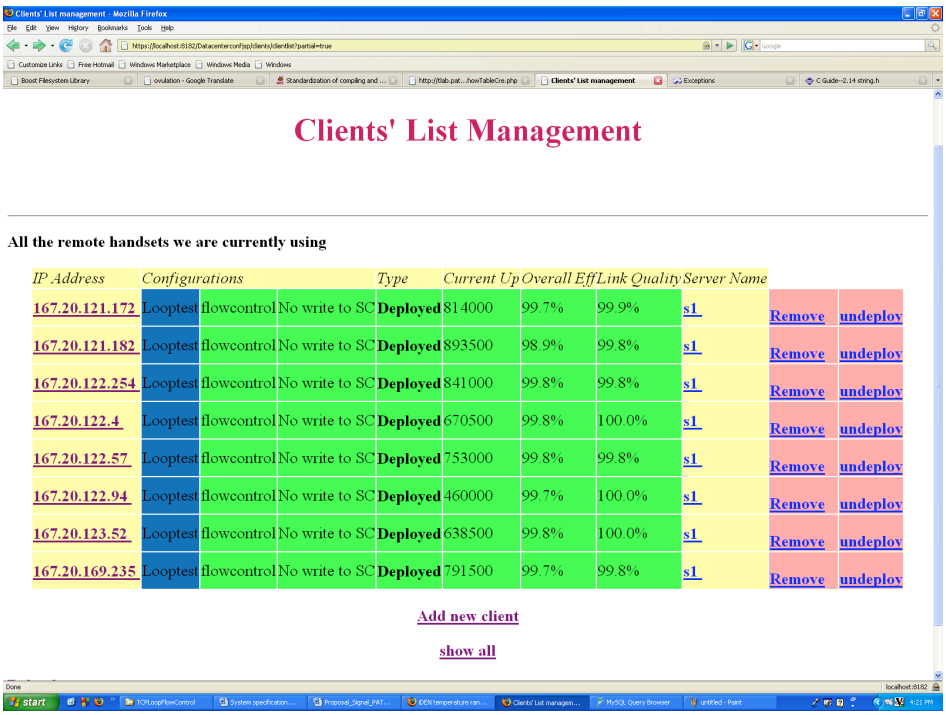


Figure 2-6 The web-based data collection management

The data center management tool includes the following features:

- Adding new cell phone with specified working mode;
- Assigning a cell phone to a given server;
- Monitoring the real-time wireless link status at a 10 seconds updating rate;
- Viewing the statistics of the data received and lost, for the past hour and for cumulated statistics;
- Remotely diagnosing the communication status between phones and signal controllers;

Figure 2-7 Reconfiguring a selected data collection handset

2.2.3.2 Security of remote management

System communication and management are designed to be highly secured. Firstly, the data link between the cell phone and the server is two-way firewalled. By "two-way" firewalling, both the data collection server AND the cell phone are protected from impersonated attacks. The remote client, instead of the server, is usually more vulnerable since for most systems there do not exist a protection for the client side. Our Java ME program in the cell phone client filters out any data connection request from other entities than the designated server. Authentication is further employed on the client side to make sure the connection is not impersonated. Secondly, the system management java server uses secured socket layer (SSL) protocol that ensures all the data are encrypted before running between the web server and the web user. Also, username and password authentications are required to conduct any modifications to the system configuration (for example, to allow a remote handset to send data packets to a signal controller).

Furthermore, there is an optional secured link option that can encrypt the link between the cell phones and the traffic data server. Also the link between the traffic data server, the database and the CTNet server are all based on SSL links. Based on these efforts, the system design fully demonstrates the security of its data communications and remote management.

2.2.4 System costs

The system is designed and developed with a major emphasis on the cost. The current device and service costs of the remote handsets are shown in **Table 2-3**.

Table 2-3 Device cost and monthly service cost

Item	Cost	Description
iDEN265	~$30 each without contract	One cell phone for one master controller with one power supply
RS232 Cable	~$15 each	
Power Supply	~$15 each	
Installation	Simply put into the cabinet No extra device required	
Resetting circuit	Optional, <$10	A simple circuit for resetting the cell phone when there is a failure in embedded software
Nextel® Service:	$10/month with static IP	2007 cost figure

Assuming a master signal controller is connected to 5 local signal controllers, the total estimated cost to deploy such a system to 1,000 intersections is illustrated in **Table 2-4**.

Table 2-4 Total cost for 1000 intersections per year

Item	Cost	Description
Device cost	**$12000 -14000**	200 set of field devices
Service cost	**$24000/year**	
Labor	Maintenance for a field set requires less than half an hour (estimated), plus travel time.	

2.3 Performance Characteristics

The system performance characteristics described in this report are the throughput and the instantaneous link quality (which shows the service availability). Preliminary results on the reliability of the system are also reported based on tests conducted at University of California at Berkeley Richmond Field Station.

2.3.1 System performance indexes

The statistics of eight cell phones and one data center over 10 days were averaged to form the following performance indexes as shown in **Table 2-5**.

Table 2-5 Average system performance indexes

Performance	Average	Definition
Instantaneous throughput	**619Bytes/s**	Number of bytes received per second by the data center from one cell phone, measured every 10 seconds. *Note: these statistics do not include measurements taken when there is a communication outage.*
Hourly	**533Bytes/s**	Number of bytes received per second

throughput		by the data center from one cell phone. Measured every hour.
Instantaneous System Availability	**99.8%**	The number of bytes received by the data center divided by the number of original bytes sent by the signal controller to the cell phone, measured every 10 seconds
Hourly system availability	**99.6%**	The number of bytes received by the data center divided by the number of original bytes sent by the signal controller to the cell phone, measured every hour
Latency	**2 s**	The time a packet takes to travel from the source (only the GPS message has its original time stamp, so the source originates from the GPS satellites) to the data center. Due to a lack of high-resolution timestamp, the latency is estimated to be roughly 2s in most observations.

Detailed statistics of these results are presented in the following subsections.

2.3.2 System throughput

Both the hourly average throughput and instantaneous throughput at 10 seconds period were obtained for all the clients under test. The tests were carried out at Richmond Field Station, where the communication network condition was worse than at several of the

arterial intersections we tested. The measured system throughput and service availability at the field test locations were higher than those obtained at the Richmond Field Station. **Figure 2-8** and **Figure 2-9** show the cumulative distributions of instantaneous and hourly throughput, respectively. It shows that, the instantaneous rates (regardless of the communication outage) of the cell phones are highly probably greater than 335B/s most of the times. That rate is the throughput required when the master controller works in focus mode and is polling with a 200ms period. This is accomplished with a probability of over 96%, while rates higher than 335B/s over 90% of the time can be sustained over the long term when outage and other losses are taken into account.

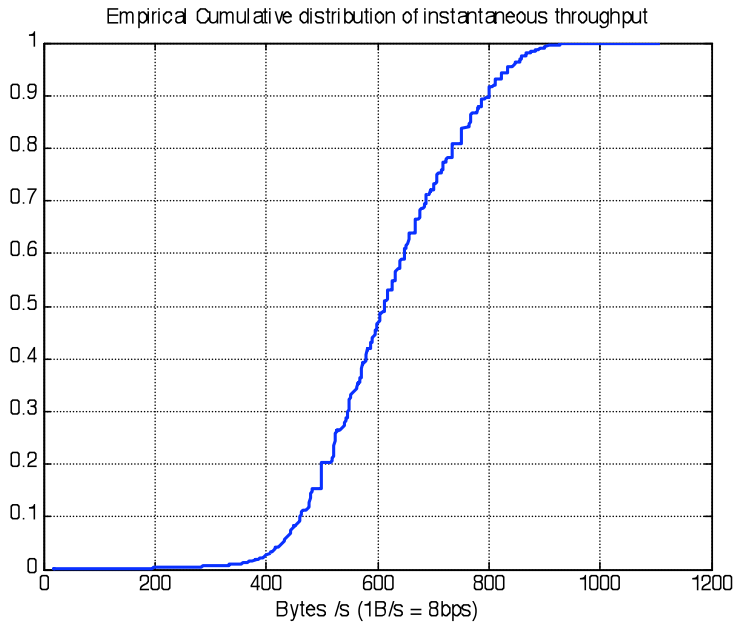


Figure 2-8 Cumulative distribution of the instantaneous throughput (Bytes/s)

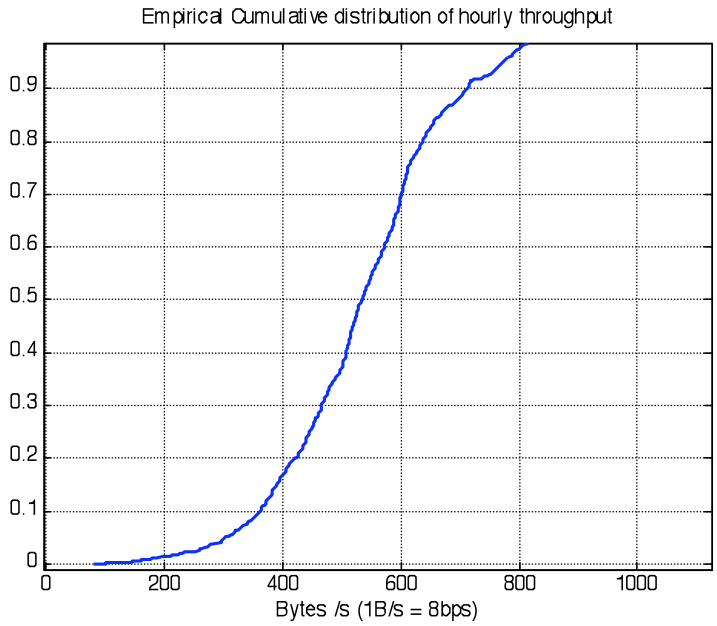


Figure 2-9 Cumulative distribution of hourly throughput (Bytes/s)

2.3.3 System Service availability

The service availability is defined as the number of bytes received by the data center divided by the total number of bytes the original signal controller sent to the client (cell phone). It is always less than 1.0, so hereafter service availability is presented in percentage.

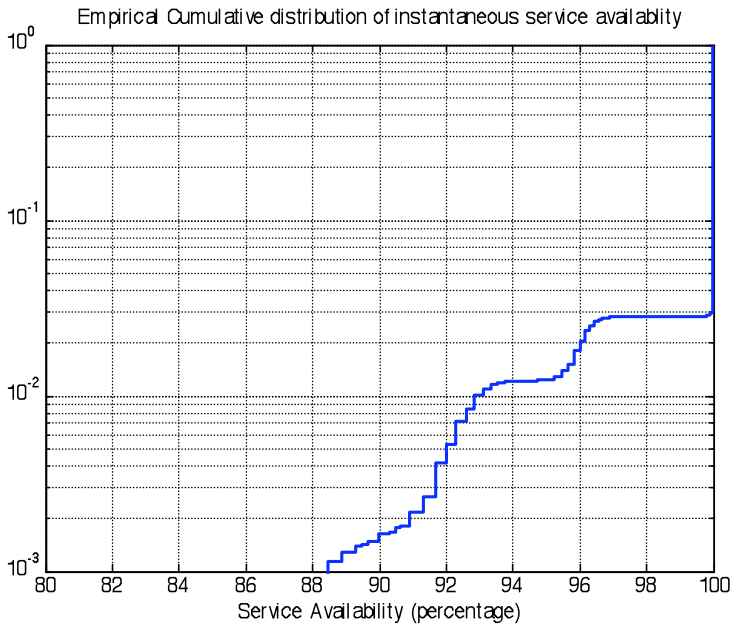


Figure 2-10 Instantaneous system service availability

From **Figure 2-10**, the probability of data lost due to flow control being greater than 0 is only 2%. **Figure 2-11**, illustrates the data loss due to flow control and outage being greater than 2% is about 2%.

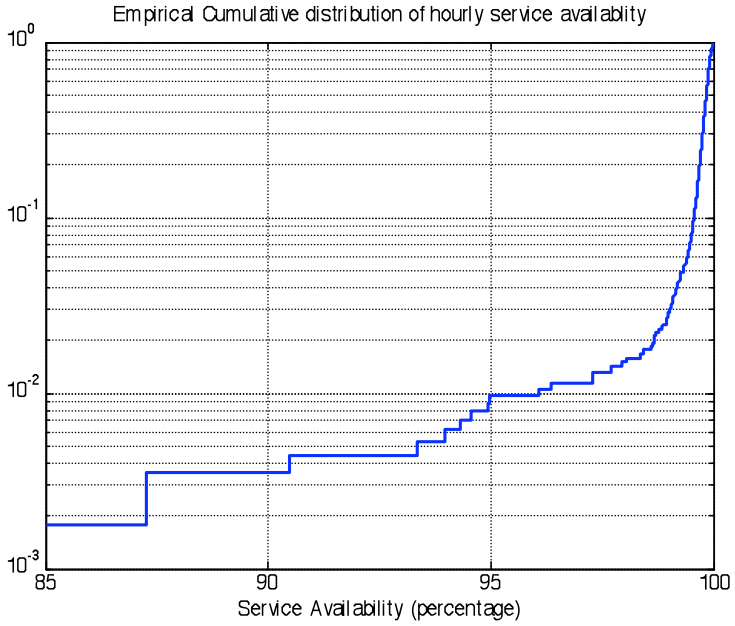


Figure 2-11 Hourly system service availability

2.3.4 Preliminary results on reliability

The reliability of data collection is measured by the probability of the duration of system running free of failure being greater than a given threshold. When the failure is caused by a runtime embedded software error, the system can recover from this failure by using an automatic reset circuit. The cost of this kind of failure is about one hour of lost of data from the specific area served by that cell phone. This occurrence is also counted in the preliminary reliability results presented here.

An exponential reliability (denoted by *r*) model is used for test, which is

$$R(t) = \Pr\{r > t\} = \exp(-t/\theta), \tag{2-1}$$

where θ is a unknown parameter. The test is carried out that for N (N=9) equivalent setups (one cell phone communicating with the data center is treated as one setup), measuring the first failure of the N setups and calculating the reliability based on this period.

Table 2-6 Reliability test result lookup table

Reliability R (*N*=9)		
Expected life t (R(t) = 90%) weeks	**Expected life t (R(t) = 80%) weeks**	**Test period (*x*)(*weeks*)**
10.0	4.8	5
20.1	9.5	10
30.1	14.2	15
60.2	28.5	30

Currently the preliminary reliability test has been run for over five weeks. The testing shows that the mean time between failures for each setup in the system is at least over 10 weeks with 80% of probability and 4.8 weeks with 90% of probability.

2.4 Technical Aspects Related to System usage

The communications links in the system include: (1) between the cell phone and the data centers (duplex); (2) the cell phone and the signal controllers; (3) the data center and the CTNet server; and (4) from data center to data center; only parts (1) through (3) are covered in this document.

2.4.1 Communication between cell phone and the data center(s)

Underlying communication protocols are standard TCP/IP based on the iDEN wireless network. The protocol stacks of the data collection (cell phone to remote data center PC) without details of the iDEN network are illustrated in **Figure 2-14**.

Note that due to use of the low level socket communication, the higher level protocols dedicated for this application are operating directly over TCP/IP.

List of these protocols are for:

- Message Link control
- Wireless Flow control;
- Data collection Management;
- Signal controller data formatting;
- Optional security control;

The message link control protocol is a data packaging protocol which groups raw data into individual data packages. The packages can easily been converted to AB3418 messages at the data centers.

The wireless flow control protocol is for adaptive flow rate control. This keeps the wireless link operating at an optimized data rate. The optimized data rate is an adaptive

rate that adapts to the available bandwidth of the wireless channel at the time of communication. Package loss may occur due to the flow rate control.

A more detailed illustration is shown below in **Figure 2-12** of this flow control procedure. Also **Figure 2-13** is a flow chart of the cell phone program flow control feature.

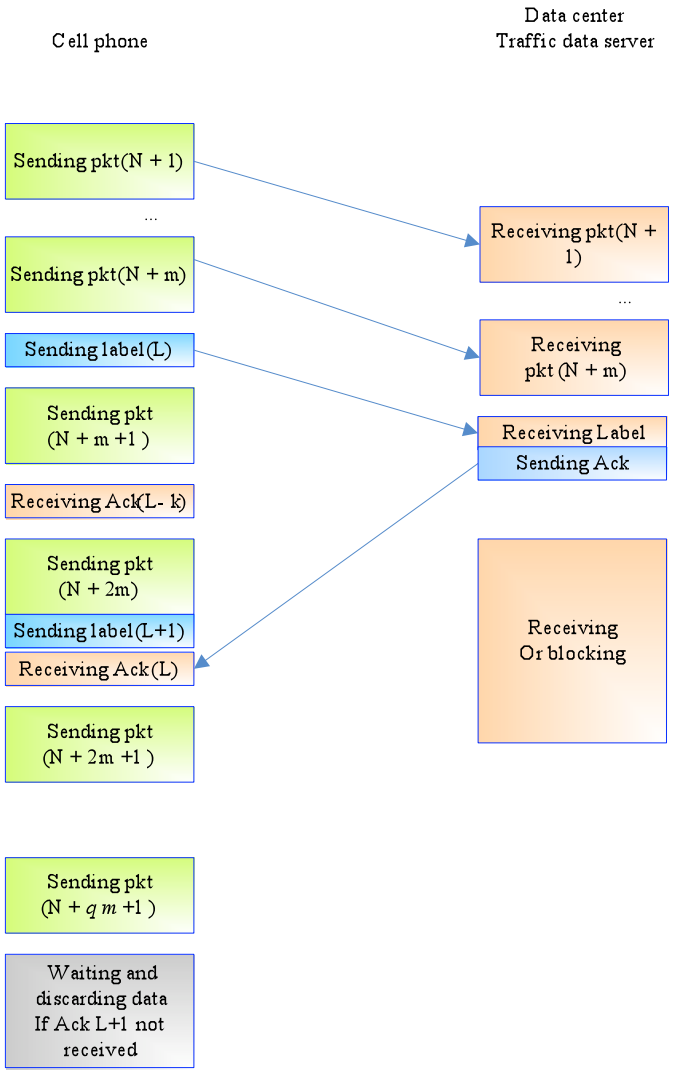


Figure 2-12 Adaptive flow control mechanism

The *m* packets are grouped to form macro packets, which are indexed and used as "hand-shake" labels between the cell phones and the traffic data server. The cell phone

side then determines the instantaneous link quality from this pseudo timestamp label by comparing the received acknowledgement label and internal outgoing label. Under good communication conditions, this difference can be kept low, but with system outages, the difference will increase quickly. The macro packet size *m* and the threshold of label difference *q* are all selected according to empirical data obtained through experiments. The loss due to flow control is kept to a minimum by carefully selecting parameters and using a windowed buffer.

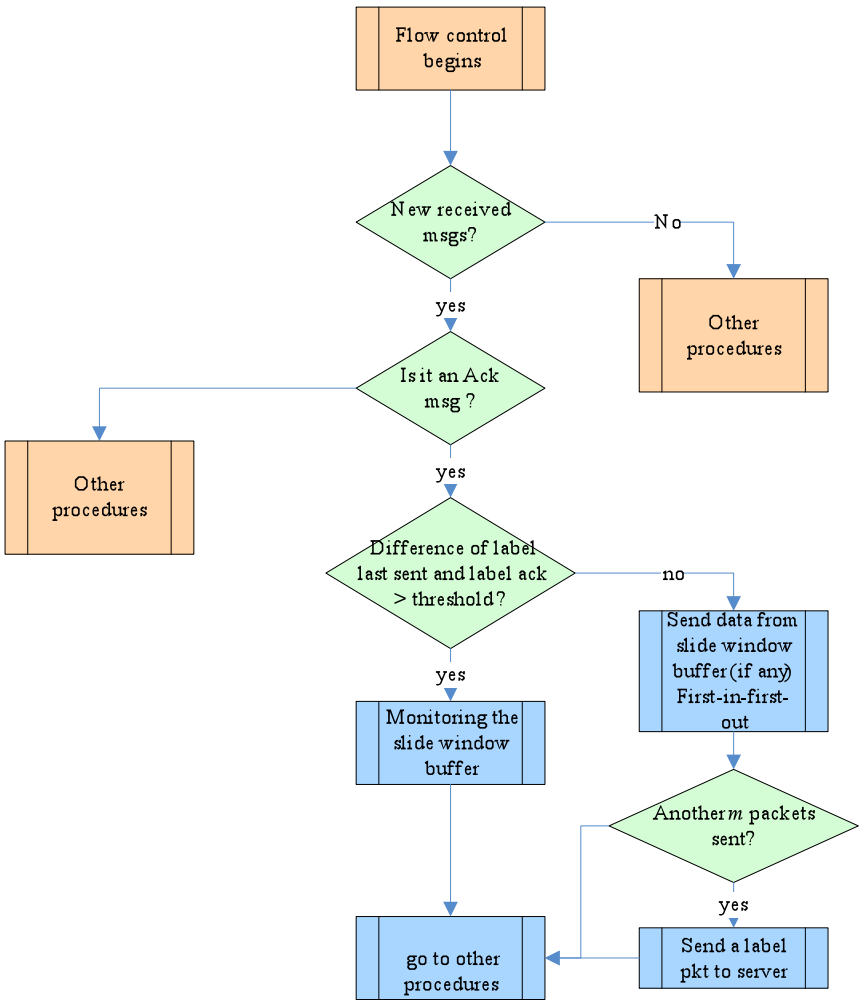


Figure 2-13 Flow control procedure in the cell phone program

The data management control protocol allows the data center to set the work mode of the data collection, obtain statistics of the data collection process from the cell phone and manage the way the cell phones talk with the signal controller.

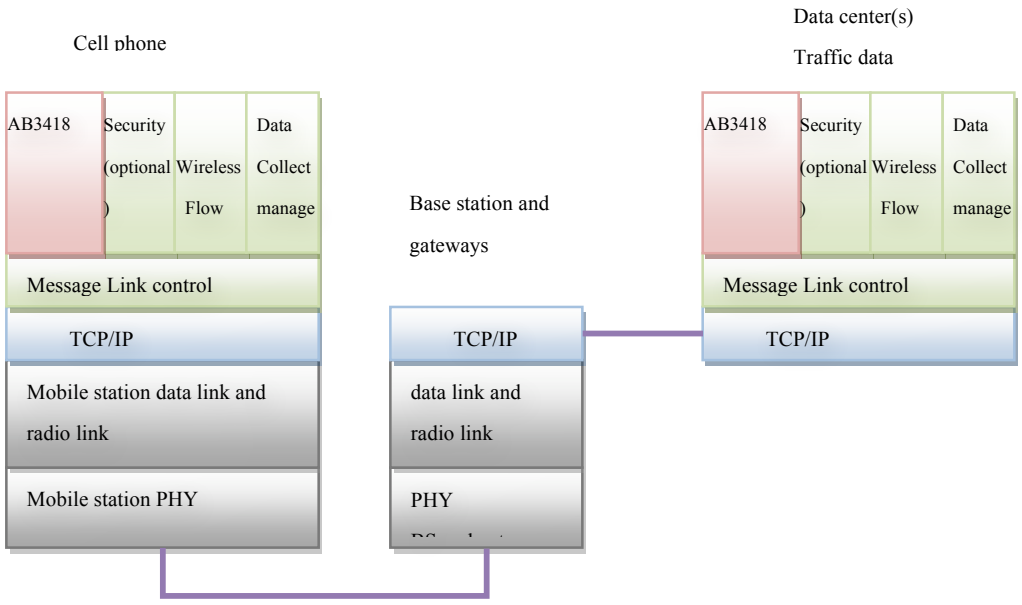


Figure 2-14 Protocol stack for the cell phone and traffic data server

2.4.2 Protocol between cell phone and the signal controller

The communication between the cell phone and the signal controller are full-duplex serial communication based on the RS232 protocol. The two-way communications operates in a complete asynchronized manner. Data messages conform to the AB3418 protocol.

2.4.3 Proposal for data center communication with CTNet server

Currently the CTNet server implements the protocol stack with AB3418 directly on top of TCP/IP. Without incorporating new protocols, presently the CTNet server won't

distinguish unique origin from the traffic data servers where AB3418 sentences from many different master controllers are forwarded by a single source address.

To address this issue, a simple solution to solve this problem is proposed. The traffic server emulates itself as multiple master signal controllers with different TCP port on its own IP. This is the basic idea of the source addressing solution. The traffic server itself works as multiple virtual clients of the CTNet server. Each virtual client of the CTNet server has different TCP port number and the same IP address which is the address of the traffic data server. The traffic data server allocates the different TCP ports automatically in order to ease the system management procedure.

Figure 2-15 Protocol stack for CTNet server and Traffic data server

2.4.4 Full protocol stack of traffic data server

The traffic data server incorporates a multiple-threaded server side protocol stack with the remote cell phones, and a single tunnel server side protocol stack with the CTNet server.

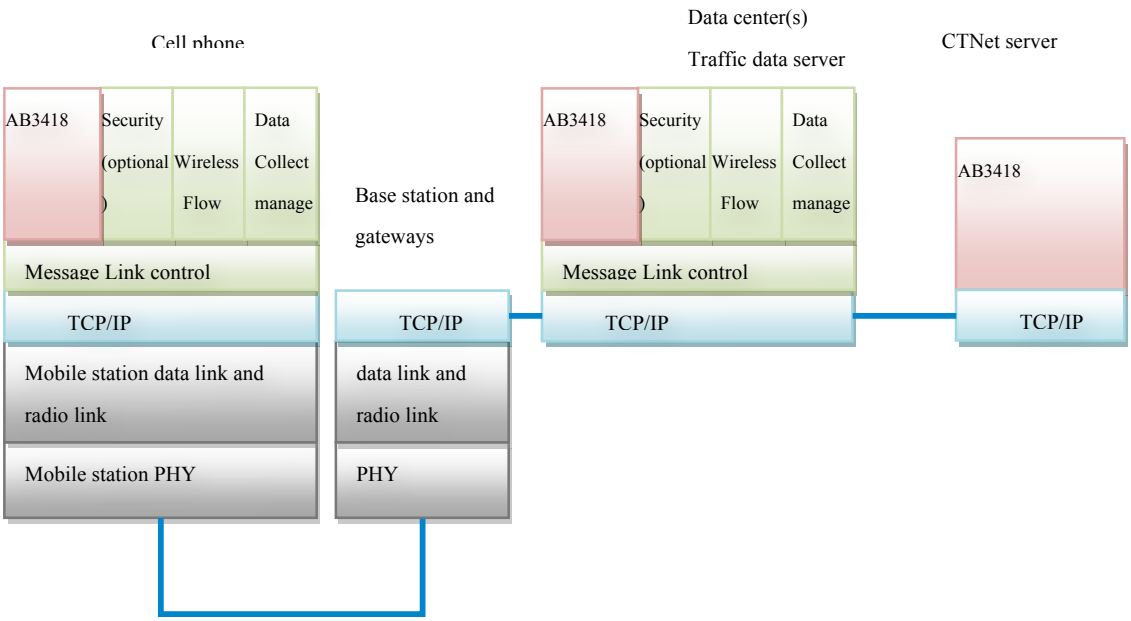


Figure 2-16 Protocol stack and traffic data server

2.5 Preliminary field testing

With all the development work completed, the project team assembled a bunch of prototype devices, which consist of an iDEN Motorola phone, a DC power adapter, and a customized serial cable to connect iDEN phone and the serial port (C2S) on 170 traffic signal controller. The connector assignment for the serial cable between iDEN phone and 170 signal controller is shown in Table 2-7.

The iDEN cell phone, as a very low cost communication device, bears the disadvantage of low profile in its power supply. Using it as a data modem for relatively high data rate imposes an excessive power need to the device. Our lab experiment showed that even the OEM charger could not replenish the battery fast enough for the data communication needs (Cell phone charger does not supply power directly to the device). Therefore, we studied the power consumption need of the device with peak data rate through several

rounds of lab testing and picked up a COTS (made by SONY) power adaptor to directly supply power for the device, to work around the limited power supply issue. By this workaround, the device was made successfully working on peak data rate continuously during our lab test.

Table 2-7 Connector Assignment between Phone and Signal Controller

PIN	Function	Wire Color
A	Audio IN	White
B	Audio IN	Black
C	Audio Out	Red
E	Audio Out	Green

With all the communication software installed, we tested all twenty-four phones in the laboratory. After the three-week lab test, we split all prototype devices into two groups. A dozen of phones were assigned to Caltrans headquarter and local districts including District 1 (Eureka, CA), District 4 (San Francisco Bay Area, CA), District 7 (Los Angeles, CA), and District 11 (San Diego, CA). The field testing last more than three months. Caltrans engineers from traffic operation department were excited about the new solution on traffic data collection. They actively involved in the testing and interacted with the project team with questions and comments on how to improve the system design and development. With the installation of the devices inside controller cabinets, Caltrans engineers don't have to go to the field to troubleshoot controller operation. Some simple tasks, e.g. timing change, time-of-day change, etc., can be done remotely through the iDEN interface. It saved many trips, time, and money for Caltrans.

Through the field testing, we proved the concept that the iDEN phone can be used for traffic data collection and remote control of controller operation. Although the wireless connection was not very reliable, our communication software and data server were proved to be running robustly.

2.6 Conclusion

A traffic data collection system based-on the iDEN wireless network has been developed, lab tested and preliminarily tested in the field. The objective of the system is to provide a cost-effective and easy-to-maintain system that could still reliably provide traffic data over the wireless link. The mobile wireless network has its inherent characteristics of less reliable than the wired network. We have built an adaptive flow control layer over the wireless TCP communication to address the occasional outage problem. The validity of the method is in the fact that the collection second by second traffic signal status data does not necessarily achieve 0% loss rate. So with allowing some minimized data loss, we could control the data flow to avoid exceeding the allowed data rate and thus leading to major outage. The system is able to continuously provide over 2.68kbps upload data rate per remote handset for over 95% of the time, i.e., one remote handset could deliver data fetched from signal controllers at a period of 200ms continuously with about 5% of data loss rate. On the cost side, one set of client hardware costs ~$70 and $10 monthly, while covering up to 8 local signal controllers. These combined features, low cost and high performance, make the system a unique solution for traffic data collection.

3 An Online Performance Measurement Method based on Arterial Infrastructure Data

3.1 INTRODUCTION

The uninterrupted growth in traffic demand has led to a continuous need not only for the development of traffic surveillance systems, but also for the development of real-time traffic management systems. The estimation of performance measurement plays an indispensable role when analyzing traffic conditions and evaluating different traffic signal control strategies.

Compared with freeway travel time estimation, estimation for arterial performance is more challenging because of the nature of traffic signal operations and stochastic disturbances caused by pedestrians, bicycles, and other factors. As a result, arterial performance measurement requires much more data collection, particularly at or near signalized intersections. However, real-time data collection along arterials is still very rare in the current practice (Balke and Herrick, 2004; Liu and Ma, 2007).

Due to the status of arterial data collection, the majority of previous research and applications can be categorized into two types. The first type is built upon existing signal control infrastructure. But they are mostly based on offline and aggregated point-based measurement (Highway Capacity Manual, Transportation Research Board, 2000; Skabardonis and Geroliminis, 2007; Zhang, 1999; Robinson and Polak, 2005; Sisiopiku and Rouphail, 1994; Lucas et al., 2004). The other type is based on advanced detection technologies, such as automatic vehicle identification (AVI) on probe vehicles and buses (Bertini and Tantiyanugulchai, 2003; Cetin et al., 2005; Wang, 2004) or inductive loop

detectors (ILDs) with advanced detection cards or video sensors which can output identifiable vehicle signatures (Coifman and Cassidy, 2002; Coifman et al., 1998). The recent advancement and deployment of telecommunication and ITS technologies make real-time traffic and signal status data more readily available. For example, Parsons Traffic and Transit Laboratory (PT^2 Lab) at University of California has been working together with California Department of Transportation (Caltrans) on developing an interface for 170E controllers. With the cost-effective interface and the general packet radio service (GPRS) communication devices, PT^2 Lab is now collecting second level traffic signal status, ILD counts and occupancies in the real time from more than 50 intersections along the El Camino Real Corridor at Santa Clara County, CA and the San Pablo Corridor at Alameda County, CA (Li et al., 2007). Such development enables an online arterial performance measurement system using the existing signal control infrastructure without significant modifications and investments. It is also the motivation of this research.

Defining measures of effectiveness (MOEs) is an essential requirement of a performance measurement system. Some previous studies have discussed MOEs for arterial performance from the perspective of traffic engineers (Kloos, 2006), traffic and transit researchers/planners/agencies and travelers (Li et al., 2008; Koonce, 2006; Kloos, 2006 and NCHRP 3-79). Of all the performance measurements, travel time is particularly critical to assessing traffic system operations, providing traveler information, making decisions on selecting routes and designing control strategies for traffic signal operation. Due to the importance of estimating the travel time, many researchers have developed various methodologies for estimating trips either on freeways or along signalized arterials [1-5]. Therefore, three essential MOEs: arterial travel time, number of stops and travel time reliability are the main focus of this paper. Some other conventional MOEs, such as

saturation degree (volume-to-capacity ratio), width of green band, arrival ratio during green, pedestrian delay, etc., can be easily measured or calculated from the study data. In this project, we propose an online arterial performance measurement method that is based on signal infrastructure data. The remainder of the chapter is organized as follows. A virtual loop detector method is introduced first to construct virtual arrival and departure curves at isolated intersections. The second section describes a model to create imaginary trajectories crossing multiple intersections based on the previously constructed N-T curves. The third section demonstrates the application of the proposed model in a microscopic simulation environment. The last section concludes this chapter with recommendation for future research.

3.2 MODEL FOR ISOLATED INTERSECTIONS

The proposed method is based on signal infrastructure data. More specifically, the method is applied to the actuated signal control system that is the most popular signal control system in U.S. (according to Gettman et al., 2007, over 90%). Figure 3-1 illustrates a typical layout for a semi-actuated signal. The main street, which in most cases is coordinated direction, has advance loops embedded at about 56 meters (184 feet) away from stop-bar. In contrast, the minor street only has presence loop installed at stop-bar. In this study, only the arterial street is chosen for performance measurements.

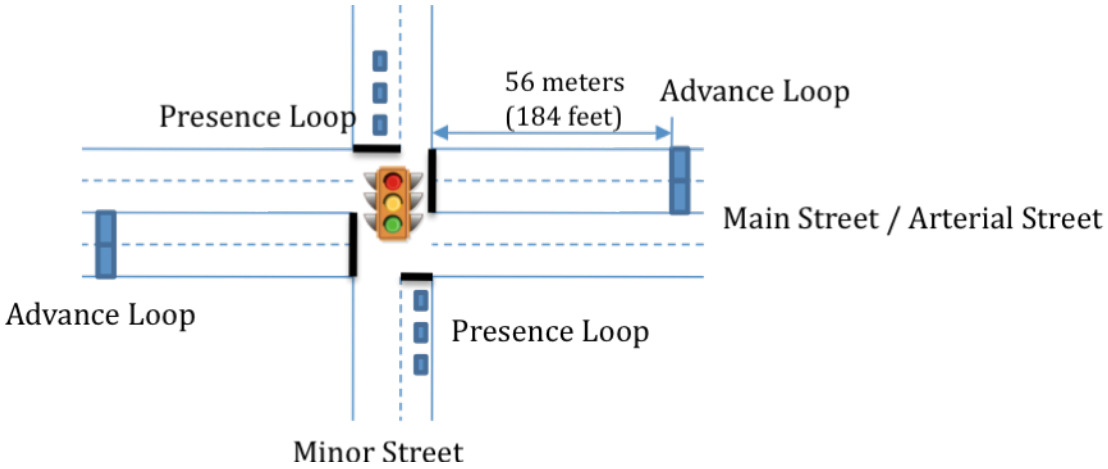


Figure 3-1 Typical layout for an intersection under semi-actuated control

The proposed model takes both online and offline inputs. The online inputs consist of two data sources from traffic signal controllers. One is the count and occupancy from ILDs. The other is the signal status and pedestrian button information. The offline inputs include geometry information, historical turning ratio, average free flow speed or speed limit, and saturation flows.

The arrival curve is constructed based on the count and occupancy data at the advance loop. The conventional method used to calculate the arrival curve at stop-bar is to shift the detected arrival curve at the advance loop by a free flow travel time. The underlying assumption is that vehicles arrive at advance loops without any impacts from either the downstream signal or the waiting queues. However, it is not the case for the layout illustrated in Figure 1, where the advance loop is not far enough from stop-bar. Considering a general case with 35MPH speed limit and 2.5 MPSS average deceleration rate, the average deceleration distance calculated by equation (1) is 61 meters that is farther than the advance loop even there is no waiting queue. Accordingly, all vehicles, which will experience delays at the signal, are decelerating when passing the advance loop. Thus the aforementioned conventional method cannot be directly applied.

$$D_{dec} = \frac{1}{2} \times \frac{{V_{FF}}^2}{a_{dec}} \qquad (1)$$

where: D_{dec} is the deceleration distance; V_{FF} is the free flow speed; and a_{dec} is the deceleration rate.

The conventional input-output method on delay estimation assumes that all queued vehicles stop at stop-bar vertically without taking any physical space, i.e. vertical queue $q(t)$. In this study, we assume that dilemma zone does not exist. More specifically,

drivers can always make the perfect stop-or-go decision. Upon the signal status and queue lengths, drivers will face three scenarios. (I) If vertical queue exists when vehicles arrive at advance loops, drivers will choose to slow down with a pre-defined gentle deceleration rate a_{dec}; (II) Otherwise if no vertical queue and signal is green when they are at advance loops, drivers will run through the intersection with the free flow speed; (III) Lastly, if no vertical queue exists and signal is amber or red, drivers will start to decelerate at the advance loop and finally stop at stop-bar with a braking deceleration rate a_{brake}.

Scenario I is depicted in Figure 3-2. Vehicle i arrived at advance loop at time t. $q(t)$ is the number of vehicles in the vertical queue. The driver estimates the queue length as $q(t) \times L_v$ where L_v is the effective space for a queuing vehicle. It should be noted that a driver's observation of queued vehicles when he/she is at the advance loop might not represent the exact number of queued vehicles when he/she eventually reaches the end of queue. The vehicle starts to decelerate at the deceleration start location and attempts to stop at the end of queue. Thus, the deceleration distance D_{dec} can be calculated by equation (1). The distance between the advance loop and stop-bar is D_{AL} and the distance between virtual loop and stop-bar is D_{VL}. The vehicle's travel time from the virtual loop to the advance loop consists of two parts: travel time from the deceleration start location to the advance loop τ_1 and travel time from the virtual loop to the deceleration start location τ_2. The two trip times can be derived from equation (2a-2c) and (3).

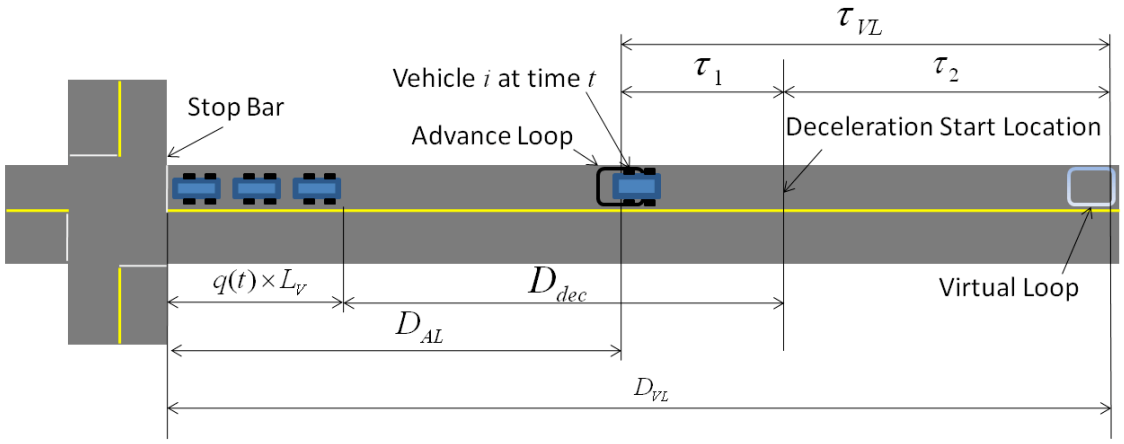


Figure 3-2 Virtual loop method for Scenario I

$$\tau_1 = \frac{V_{FF}}{a_{dec}} - \frac{v_{AL}}{a_{dec}} \qquad (2a)$$

$$v_{AL} = \sqrt{2 \times a_{dec} \times (D_{AL} - q(t) \times L_V)} \qquad (2b)$$

From Equation (1a) and (1b),

$$\tau_1 = \frac{V_{FF} - \sqrt{2 \times a_{dec} \times (D_{AL} - q(t) \times L_V)}}{a_{dec}} \qquad (2c)$$

where: v_{AL} is the vehicle speed at advance loop and D_{AL} is the distance from the advance loop to stop-bar.

$$\tau_2 = \frac{(D_{VL} - q(k) \times L_V - D_{dec})}{V_{FF}} \qquad (3)$$

where: D_{VL} is the distance from the virtual loop to stop-bar.

For Scenario II and III, the travel time from the virtual loop to the advance loop τ_{VL} is just the free flow travel time. Thus the travel time τ_{VL} for all three scenarios can be obtained by equation (4). Given τ_{VL}, the arrival curve at the virtual loop can be constructed by shifting the detected arrival curve at the advance loop backwards as much as τ_{VL}. It is noted that τ_1 is a function of time t in Scenario I, therefore, the shape of the arrival curve at the advance loop might be different from that at the virtual loop.

$$\tau_{VL} = \begin{cases} \tau_1 + \tau_2, & \text{Scenario I} \\ \dfrac{D_{VL} - D_{AL}}{V_{FF}}, & \text{Scenario II and III} \end{cases} \quad (4)$$

The instant arrival count at virtual loop can be derived by equation (5). When queue spillbacks to the advance loop, a consistent high occupancy (higher than a threshold OCC_{TH}) with a low flow (lower than F_{TH}) can be detected for time T_{TH}. In this case, the historical count data during T_W seconds (e.g. a signal cycle) ahead the queue spillback can be applied. The shifted arrival at stop-bar at time step t, $\lambda(t)$, can be estimated by formula (6).

$$count^V(t) = count(t - \tau_{VL}) \quad (5)$$

$$\lambda(t) = \begin{cases} count\left(t - \tau_{VL} + \dfrac{D_{VL}}{V_{FF}}\right), & \text{when no queue spillback} \\ \dfrac{\alpha}{T_W}\sum_{k=1}^{T_W} count\left(t - \tau_{VL} + \dfrac{D_{VL}}{V_{FF}} - k\right), & \text{otherwise} \end{cases} \quad (6)$$

where: $count^v(k)$ and $count(k)$ are instant counts at time step t and at advanced loop and virtual loop, respectively; $\lambda(t)$ is the arrival at stop-bar at time step t; α is the parameter to reflect excessive arrival during queue spillback and T_W is the time window to calculate historical arrival rate.

Given the instant arrival $\lambda(t)$ at stop-bar and time step t, the arrival curve $N_A(t)$ can be constructed by equation (7). The departure curve $N_D(t)$ can be constructed by equation (8) and (9).

$$N_A(t) = \sum_{k=T_0}^{t} \lambda(k) \quad (7)$$

where: T_0 is the start time for the cumulative arrival curve.

$$f(t) = \begin{cases} f_G(t) = \begin{cases} \mu, & \sum_{k=T_0}^{t-1} f(k) + \mu \le \sum_{k=T_0}^{t} \lambda(k) \\ \lambda(t), & \text{otherwise} \end{cases}; & \text{in effective green} \\ 0; & \text{otherwise} \end{cases} \quad (8)$$

$$N_D(t) = \sum_{k=T_0}^{t} f(k) \tag{9}$$

where: $f(t)$ is the instant departure at stop-bar and time step t; $f_G(t)$ and $f_R(t)$ is the instant departure at stop-bar during effective green and red, respectively; μ is the saturation flow rate; the effective green is the green period after the start-up lost time. The number of queuing vehicles can be calculated by equation (10). The control delay for each individual vehicle $delay^c(i)$ can be calculated by the inverse functions of the arrival and departure curves, as shown in equation (11).

$$q(t) = \begin{cases} q(t-1) + \lambda(t) - f(t), & if \quad q(t-1) + \lambda(t) > f(t) \\ 0, & otherwise \end{cases} \tag{10}$$

$$delay^c(i) = N_A^{-1}(i) - N_D^{-1}(i) \tag{11}$$

$delay^c(i)$ is the delay between the virtual loop and stop-bar. Once the vehicle departs from stop-bar, there will be an acceleration delay before the vehicle reaches V_{FF}. In the model, vehicles which experience delays at the signal are assumed to decelerate with a constant rate a_{dec} towards stop-bar. For vehicles arriving during the beginning of green, some of them might not need to decelerate to a full stop. Given a_{dec}, we can calculate the travel time and delay for a deceleration from V_{FF} to v by equation (12) and (13). Accordingly, the minimum speed can be calculated based on the estimated control delay by equation (14). Then, the acceleration delay can be calculated by equation (15) with the acceleration rate a_{acc}. Finally, total signal delay for vehicle i $delay(i)$, and number of stop $stop(i)$ can be estimated by equation (16) and (17), respectively. A stop is defined by the running speed lower than a threshold speed V_{stop}, e.g. 5MPH.

$$t = \frac{V_{FF} - v}{a_{dec}} \tag{12}$$

$$delay = t - \frac{vt + a_{dec}t^2/2}{V_{FF}} \tag{13}$$

$$V_{MIN}(i) = max\left(V_{FF} - \sqrt{2 \times a_{acc} \times V_{FF} \times delay^{c}(i)}, 0\right) \quad (14)$$

$$delay^{acc}(i) = \frac{(V_{FF} - V_{MIN}(i))^2}{2 \times a_{acc} \times V_{FF}} \quad (15)$$

$$delay(i) = N_A^{-1}(i) - N_D^{-1}(i) + delay^{acc}(i) \quad (16)$$

$$stop(i) = \begin{cases} 1, & V_{MIN}(i) \le V_{stop} \\ 0, & otherwise \end{cases} \quad (17)$$

3.3 ARTERIAL MODEL

Most of arterials are controlled by coordinated traffic signal systems. The green bandwidth and the arrival ratio during green are the basic measurements for signal progression. The efficiency of signal progression plays an important role in the performance of arterial systems, e.g. travel time and number of stops at traffic signals. Therefore, the sum of time-average link travel times cannot represent the arterial travel time.

In this study, we measure the performance of arterial travel time by creating imaginary trajectories from an origin to a destination. At the origin, imaginary trajectories start with a constant headway (e.g. 5 seconds). The link travel times for each trajectory on different links are estimated by finding the best "fit" connection, which means the closest departure time from the upstream link and arrival time at the downstream link. Finally, an imaginary trajectory is built by those "connected" link trips.

3.4 SIMULATION STUDY

To validate our proposed model, we started with microscopic traffic simulator instead of the field data and field experiments. There are a couple of reasons. First of all, it is not quite realistic to get continuous vehicle trip data for the majority of traffic along an arterial. The microscopic traffic simulator provides a cost-effective method, which

enables such detailed data collection. Secondly, field measurements are normally imperfect. Particularly for arterials, the count measurement from ILDs is not so critical in actuated signal control. Thus, many of field implemented ILDs are not well tuned. According to some field experiments (Li et al, 2007), the absolute relative error for fine tuned ILDs can be reduced by 25%. Therefore, we plan to first use a fully calibrated simulation model to fine-tune our approach.

In this study, we selected a section of El Camino Real (ECR or SR-82) which is the major arterial connecting San Francisco, CA with San Jose, CA. Based on the geometry layout, field observation and survey data, we coded and calibrated the network in PARAMICS, a microscopic simulation model. As shown in Figure 3, the simulation network consists of six signalized intersections in Palo Alto, CA: Churchill Ave @ ECR, Park Blvd @ ECR, Stanford Ave @ ECR, Cambridge Ave @ ECR, California Ave @ ECR, and Page Mill Rd. @ ECR. Among them, Page Mill Rd. @ ECR and California Ave. @ ECR have relatively high traffic volumes. All these six traffic signals are under semi-actuated control and coordinated along El Camino Real.

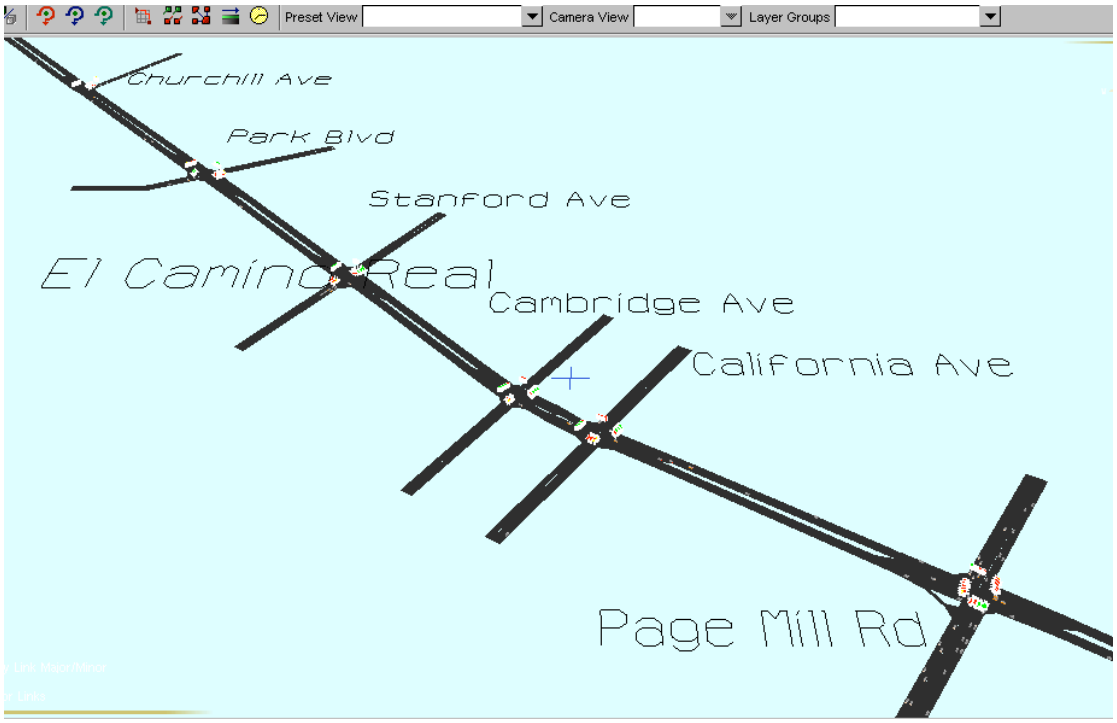


Figure 3-3 A six-signal simulation network

In the simulation, we chose the morning peak traffic period from 07:15 a.m. to 09:30 in a typical weekday as the testing interval. Besides the network coding, we also wrote programs through the application programming interface (API) provided by PARAMICS. The purpose of such APIs is two-fold: one is to implement the semi-actuated and coordinated traffic signal control logic, while the other is to collect the simulation data to support and evaluate our proposed model. In the simulation, we mimic the data collection system and communication system, so the data is collected by the API program in the exact same format and frequency as ILDs at the same locations. An API program was written to collect the trajectory for each vehicle. Such data serve as the baseline to evaluate our model performance.

To run the proposed model, the distance between the virtual loop and stop-bar (D_{VL}) is set as 250 meters, and the occupancy threshold (OCC_{TH}) is determined as 0.35. 20

seconds is used as the time window (T_W). Lastly, the α value is set as 1.2 without further calibration in this study.

With the first five minutes as the simulation warm-up time, signal delays are estimated for each vehicle travelling southbound and aggregated for every five minutes. Then the estimation results are compared with the observations from the simulation (baseline). Figure 3-4 and Figure 3-5 illustrate the estimation results for the heavy traffic situation at Page Mill Rd @ ECR and the light traffic situation at Park Blvd @ ECR, respectively. For both of the two scenarios, the model works pretty well. For most of the time intervals, the estimated results trace the trend of the signal delay quite closely.

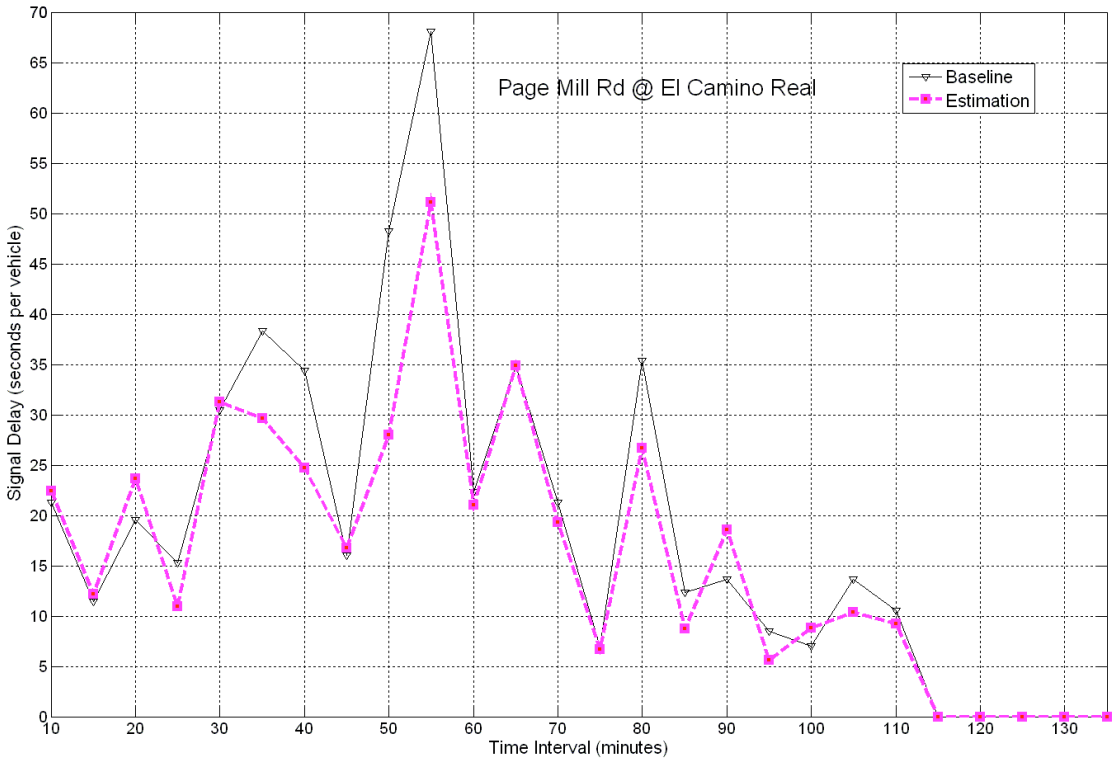


Figure 3-4 Estimation of signal delay for Page Mill Rd @ ECR (heavy traffic)

Table 3-1 shows the RMSE (Root Mean Square Error) and RMSPE (Root Mean Square Percentage Error) of the delay estimation. For the intersections that have relatively light traffic, the model shows very low RMSE and relatively high RMSPE due to the low average delay from the simulation. For the busy intersections such as California Ave and Page Mill Rd, on the other hand, the results show relatively high RMSE and low RMSPE. Although with relatively high RMSPE, the model is still effective because RMSE is within the reasonable range. The model can be further improved by calibrating some parameters such as α and start-up lost time. Moreover, because the difference between the model and the simulation may result from variant turning ratio and saturation flow rates, adding the estimations for turning ratios and saturation flows can further improve the model.

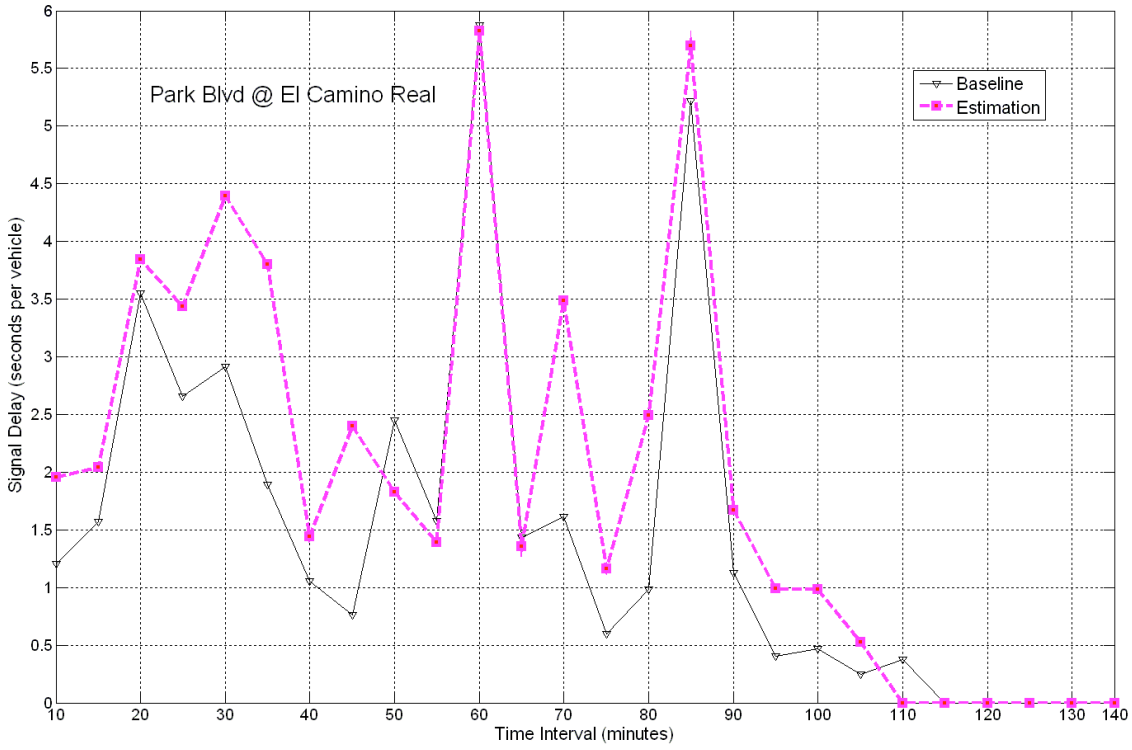


Figure 3-5 Estimation of signal delay for Park Blvd @ ECR (light traffic)

Table 3-1 Summary of Estimation Results for Signal Delays

	Mean (sec/veh)		RMSE	RMSPE
	Baseline	Estimation	(sec./veh.)	(%)
Churchill Ave	0.46	0.78	0.75	367%
Park Blvd	1.46	1.95	0.83	81%
Stanford Ave	2.13	1.98	0.89	47%
Cambridge Ave	1.82	1.73	0.76	36%
California Ave	8.81	6.25	3.51	36%
Page Mill Rd	18.83	16.17	6.35	20%

Every five seconds, a southbound imaginary trajectory is generated from upstream of Churchill Ave. The average arterial travel time is calculated for every five minutes and compared with the baseline collected by the API program in PARAMICS. As shown in , the developed model works pretty well. The estimated average arterial travel time closely traces the trend of baseline. The calculated RMSE and RMSPE for the travel time estimation are only **9.5** seconds and **5.9** percent, respectively. The largest error 24 seconds happens at time interval 10. Interestingly enough, it is also the time when the standard deviation of travel time (53seconds) reaches the maximum, as shown in Figure 3-7. When the travel time variance is large, the estimation result is sensitive to the "selection" of the "fit" connection. In other words, the model selected trajectory is more likely not representative of average trips when the trip travel times are varies a lot. Figure 3-7 also demonstrates that the proposed method can estimate well the reliability of arterial travel times, which is considered as one of the most important measures of effectiveness (MOEs) to evaluate arterial performance.

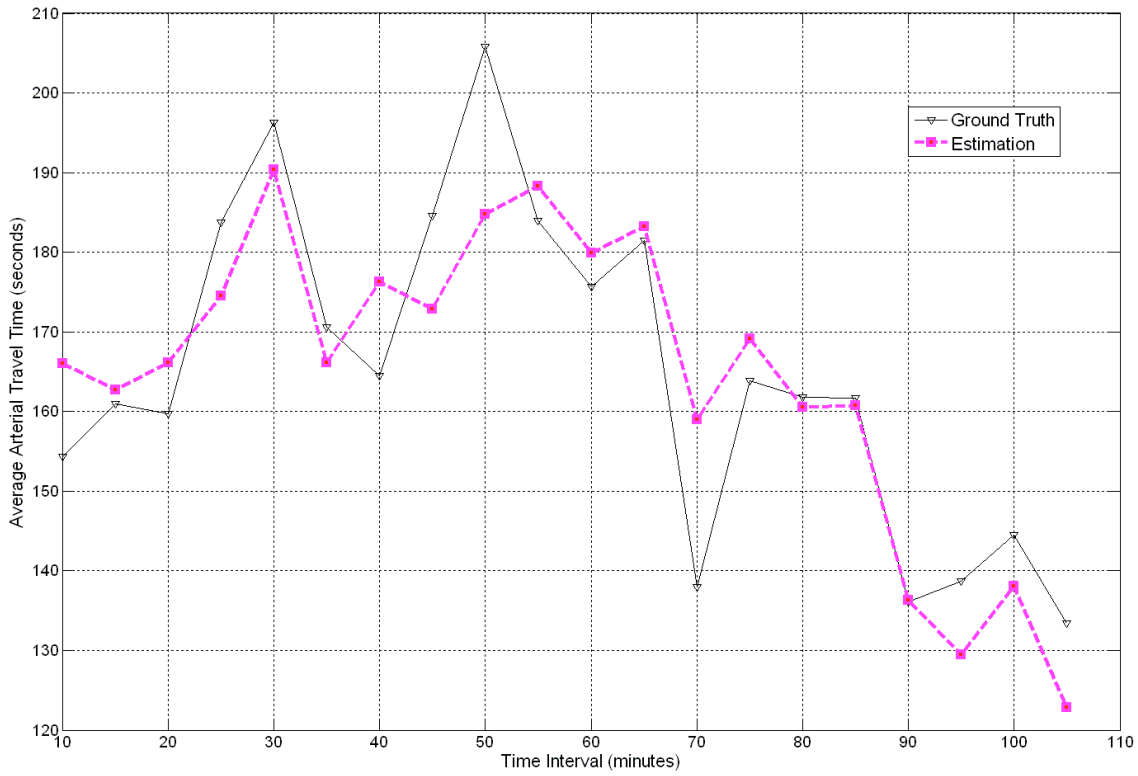


Figure 3-6 Estimation of average arterial travel time

The average number of stops at signals per arterial trip is important for evaluating arterial signal coordination. It is also a critical factor when drivers are comparing different route selections for their trips. Figure 3-8 demonstrates the capability of the model on estimating the average number of stops per arterial trip. For most time intervals, the model can accurately estimate the average number of stops. The RMSE of the stop estimation is 0.27, which is insignificant.

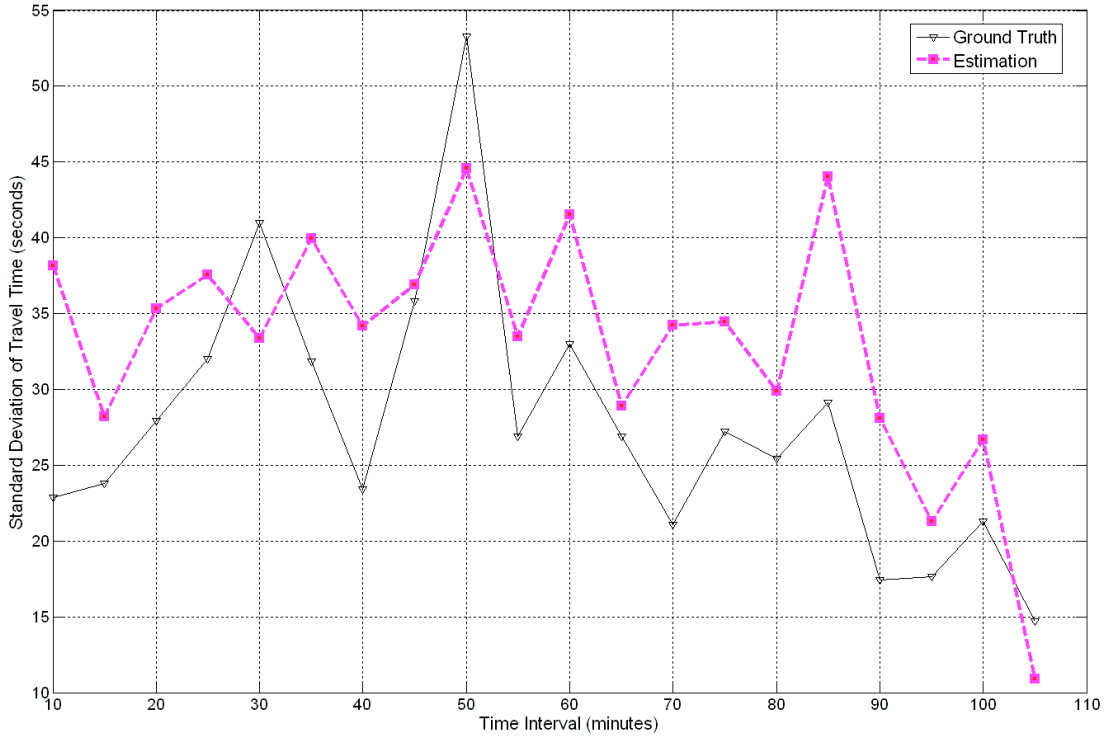


Figure 3-7 Estimated standard deviations for arterial travel times

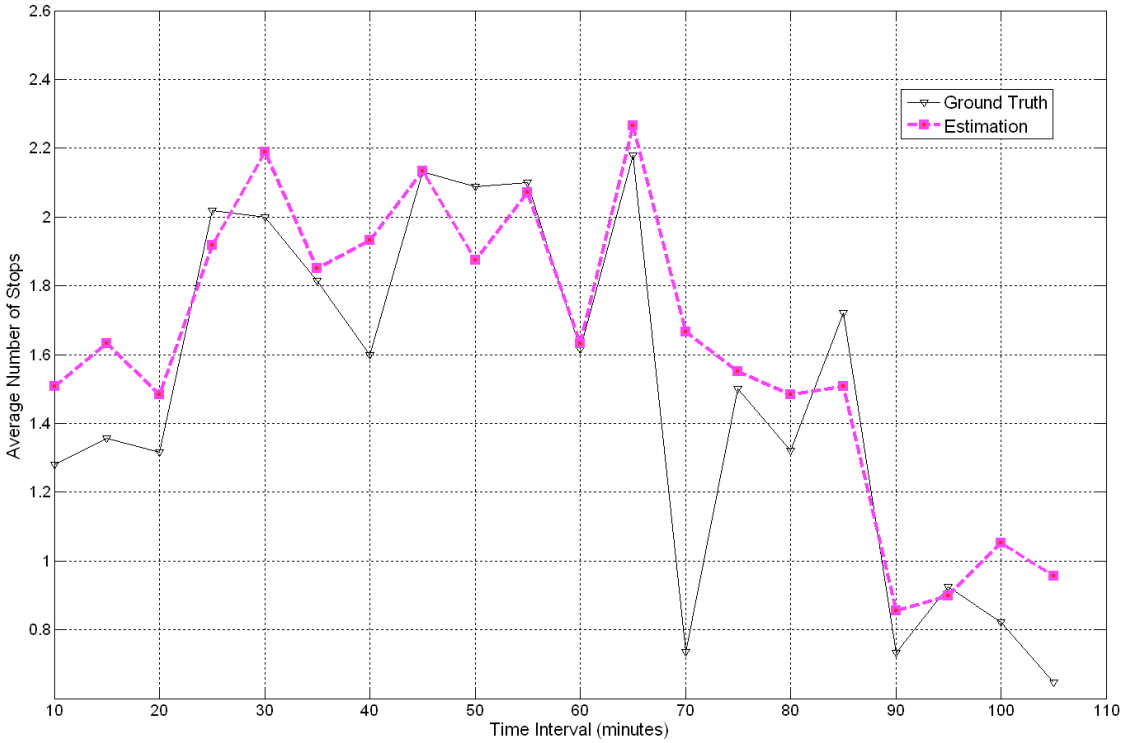


Figure 3-8 Estimated number of stops for arterial trips

3.5 CONCLUSION AND FUTURE RESEARCH

This paper has described the development of an arterial performance measurement method that is based on the signal infrastructure data collected at PT^2 Lab, U.C. Berkeley. The performance of the proposed model is illustrated by using a simulation study. The six-signal simulation network covers both heavily congested and light traffic intersections. The proposed model works well at both the intersection level and the arterial level. Estimation errors of travel time, number of stops and travel time reliability are insignificant.

The findings of the study together with the data collection means developed by PT^2 Lab provide a cost-effective way to achieve an arterial performance measurement system. The data and analysis results will support transportation researchers on various research topics such as traffic control and operations; help planners and local agencies on daily management and system monitoring; and provide travelers real-time information when scheduling their trips.

For the next step, we will further calibrate, validate and demonstrate our model by using field data and conducting field experiments. Some given parameters in this study, such as the demand factor α and time window T_W when queue spillback happens, average deceleration and acceleration rates, free flow speed, and turning ratios, should be calibrated or measured based on field data and/or observations. Effective adaptive models should be developed to dynamically estimate parameters such as saturation flow and start-up lost time. Sensitivity analysis on some of the key parameters will be studied. Moreover, the model to address the over-saturated scenarios will be developed and validated by the simulation network and then the field data.

3.6 REFERENCES

3-1 Rice, J. and E. van Zwet. A Simple and Effective Method for Predicting Travel Times on Freeways. *Proceedings of IEEE, Intelligent Transportation Systems*, 2001.

3-2 Kwon, J. and K. Petty. A Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes. *Proceedings of the 84th Transportation Research Board Annual Meeting*, Washington, D.C., January, 2005.

3-3 Mousa, R. M. Analysis and Modeling of Measured Delays at Isolated Signalized Intersections. *Journal of Transportation Engineering*, 347-354 2002

3-4 Lucas, D. E., P. B. Mirchandani, and N. Verma. Online Travel Time Estimation Without Vehicle Identification. *Transportation Research Record No. 1867, Freeway Operations and Traffic Signal Systems*, 2004.

3-5 Kothuri, S. M., K. A. Tufte, E. Fayed, and R. L. Bertini. Toward Understanding and Reducing Errors in Real-Time Estimation of Travel Times. *Proceedings of the 87th Transportation Research Board Annual Meeting*, Washington, D.C., January, 2008.

3-6 Zhang, H. M. Link-Journey-Speed Model for Arterial Traffic. *Transportation Research Record No. 1676*, 1999.

3-7 Transportation Research Board. Highway Capacity Manual 2000. *National Research Council*, Washington D. C., 2000.

3-8 Skabardonis, A. and N. Geroliminis. Real-Time Estimation of Travel Times along Signalized Arterials. *16th ISTTT conference*, Maryland, 2005.

3-9 Liu, H. X. and W. Ma. A Real-Time Performance Measurement System for Arterial Traffic Signals. *Proceedings of the 87th Transportation Research Board Annual Meeting*, Washington, D.C., January, 2008.

3-10 Wang, Z. Using Floating Cars to Measure Travel Time Delay – How Accurate Is the Method? *Transportation Research Record No. 1870*, Washington D. C. 2004. pp.84-93.

3-11 Li, M., Zhang, W.B., Zhou, K., Leung, K., and Sun, S., "*Parsons Traffic and Transit Laboratory (Parsons T^2 Lab)*", ITS World Congress, Beijing, China, October, 2007

3-12 Balke, K., Charara, H., and Parker, R., "*Development of a Traffic Signal Performance Measurement System (TSPMS)*", Texas Transportation Institute, Report 0-4422-2, 2005

3-13 Robinson S. and Polak, J. W., "*Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method*", Transportation Research Record, 1935, 47-56, 2005

3-14 Sisiopiku, V. and N. Rouphail, "*Travel Time Estimation from Loop Detector Data for Advanced Traveler Information Systems Applications*", Technical Report in Support of the ADVANCE Project, Urban Transportation Center, University of Illinois, Chicago, 1994

3-15 Lucas, D.E., Mirchandani, .P.B. and Verma, N., "*Online Travel Time Estimation Without Vehicle Identification*", Transportation Research Record, 1867, 193-201, 2004

3-16 Cetin, M., List, G.F., and Zhou, Y., "*Factors Affecting Minimum Number of Probes Required for Reliable Estimation of Travel Time*", Transportation Research Record, 1917, 37-44, 2005

3-17 Coifman, B., and M. Cassidy. *"Vehicle Reidentification and Travel Time Measurement on Congested Freeways",* Transportation Research, Vol. 36A, No. 10, 2002, pp. 899-917

3-18 Coifman, B., D. Beymer, P. McLauchlan, and J. Malik., *"A Real-Time Computer Vision System for Vehicle Re-identification",* Transportation Research, Vol. 6C, No. 4, 1998, pp. 271-288.

3-19 Gettman, D., Shelby, S.G., Head, L., Bullock, D. M. and Soyke, N. (2007) Data-driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems. *Transportation Research Record*, 2035, 1-9.

3-20 Yue Li, Peter Koonce, Meng Li, Kun Zhou, Yuwei Li, Scott Beaird, Wei-Bin Zhang, Larry Hegen, Kang Hu, Alex Skabardonis, and Z. Sonja Sun, Transit Signal Priority Research Tools, California PATH Research Report, UCB-ITS-PRR-2008-4

3-21 Peter Koonce, Performance Measures for Urban Street Applications, 2006 Workshop on Performance Measures, AHB25: Traffic Signal System Committee

3-22 Bill Kloos, Signal System Performance Measures – a traffic engineer's perspective, 2006 Workshop on Performance Measures, AHB25: Traffic Signal System Committee

4 A stochastic programming approach for robust signal timing optimization

4.1 Introduction

The performance of signal timings obtained from traditional approaches for closed-loop pre-timed control systems is sensitive with the fluctuation of traffic demand and the deviation of the actual and the designed flow rates. In current practice, closed-loop control systems typically segment a day into a number of time intervals, each of which is assigned a best suited signal timing plan as determined by applying Webster's formula [4-1] or using optimization tools such as TRANSYT-7F [4-2]. Typically, three to five signal timing plans are used in a given day. For such a system to work well, the traffic pattern within each interval should remain relatively constant.

Unfortunately, travel demands and traffic arrivals to intersections can vary significantly even for the same time of day and day of week. As an example, Figure 4.1 displays hourly arrivals at two crossing streets, 34th Street and University Avenue, in Gainesville, Florida, during an AM peak on weekdays over a period of four months. The flows present significant day to day variations.

A consequent issue that traffic engineers may be confronted with is to determine which flows to use to optimize signal timings. This issue was hardly a concern in old days since the data collection used to be resource demanding, and traffic data were only collected for a couple of days. As the advancement of portable-sensor and telecommunications technologies make high-resolution traffic data more readily available, chances for traffic engineers to raise such a question become more prevalent. For example, the data

collection system developed in Chapter 2 provides a cost-efficient way to collect field data of traffic flow and signal status in a real time manner.

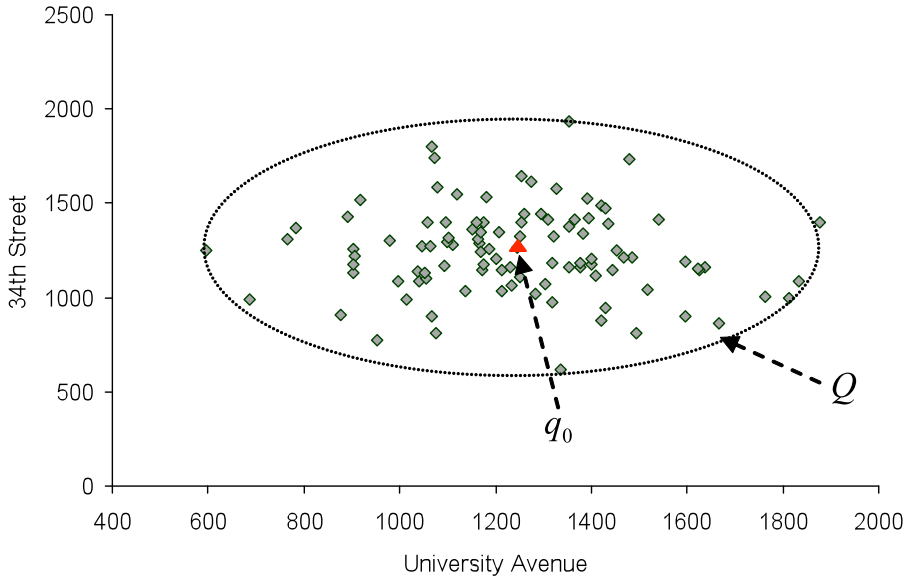


Figure 4-1 AM-Peak Hourly Flow Rates at One Intersection in Gainesville, FL

Use of the average flows (i.e., q^0 in Figure 4.1) may not be a sensible choice. Heydecker [4-3] pointed out that if the degree of variability of traffic flows is significant, optimizing signal timing with respect to the average flows may incur considerable additional delay, compared with the timing obtained by taking this variability into account. If the degree of variability is small, use of the average flows in conventional timing methods will only lead to small losses in average performance (efficiency). However, as we observed in our preliminary investigation [4-4], it may still cause considerable losses in the performance against the worst-case scenarios or the stability of performance (robustness), thereby causing motorists' travel times to be highly variable. On the other hand, if the highest observed flows are used instead, the resulting timing plans may be over-protective and unjustifiably conservative. The average performance is very likely to be inferior. Smith et

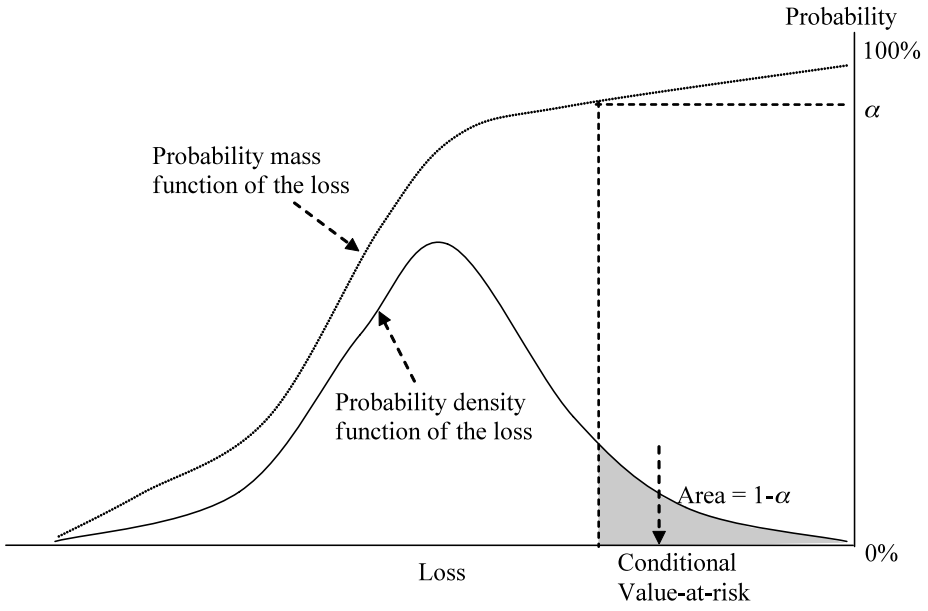
al. [4-5] suggested using 90th percentile volumes as the representative volumes to generate optimal timing plans and further noted that if time permits, other percentile volumes should be used to compare the results. However, it is well known in the statistical literature that extreme value estimates can be easily biased and highly unreliable when not computed properly [4-6].

Motivated by recent developments in robust optimization ([4-7] and [4-8]), this report proposes a methodology to design a robust optimal signal timing plan based on the collected traffic data. The performance of such timing plan is near optimal in an average sense and also stable under any realization of uncertain traffic flows in arterials or grid networks. As a preliminary investigation, Yin [4-4] developed two robust timing approaches for isolated fixed-time signalized intersections. The first approach assumes specific probabilistic distributions of traffic flows and then formulates a stochastic programming model to minimize the mean of the delays exceeding the α -percentile (e.g., 90th percentile) of the entire delay distribution. In contrast, the second approach assumes uncertain traffic flows to be unknown but bounded by a likelihood region, and then optimizes signal timing against the worst-case scenario realized within the region. It has been demonstrated that, when compared with traditional timings, robust timings may reduce the worst-case delay per vehicle by 4.9% and 11.3% respectively as well as the standard deviation of delay per vehicle by 12.0% and16.3% respectively, without adversely affecting the average performance at a real-world intersection.

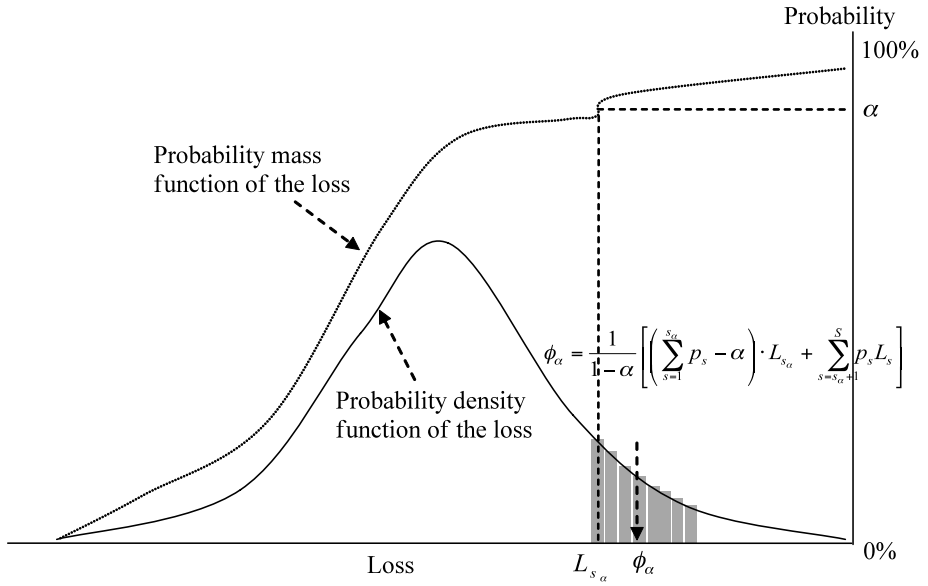
In this research, we extend the first approach, i.e., stochastic programming approach, to a general setting, optimizing the timings of actuated signals along arterials.

4.2 Scenario-Based Stochastic Programming Approach for Signal Optimization

In this approach, we fully recognize the uncertainty of traffic flows and assume that they follow certain probability distributions. To represent the uncertainty of traffic flows, a set of scenarios $\Omega = \{1,2,3,\cdots,K\}$ is introduced. For each scenario $k \in \Omega$, the probability of occurrence is p_k. With these random generated scenarios, it is feasible to formulate a stochastic program to maximize the average performance of the robust signal plan across all scenarios. However, practically travelers and system managers may be more concerned with the adverse system performance, and are less likely to complain if system performs better than expected. To address such a risk-averse attitude and avoid being too conservative, we attempt to determine a robust timing plan that performs better against high-consequence scenarios. More specifically, we minimize the expected loss (to be defined. Loss may represent different things in different settings, e.g., delay per vehicle) incurred by those high-consequence scenarios whose collective probability of occurrence is $1-\alpha$, where α is a specified confidence level (say, 80%). In financial engineering, the performance measure is known as conditional value-at-risk, or mean excess loss [4-9]. See Figure 4.2(a) for an illustration of the concept. The probability density function of a continuous regret and the probability mass function for discrete case are shown in the figure. The right tail area has an area size 1-α, which contains relatively higher losses. And the conditional value-at-risk is simply the mean of the losses in this area. By minimizing the conditional value-at-risk, it can be claimed that the losses incurred by the high-consequence scenarios are minimized.



(a) A continuous loss function



(b) A discrete loss function

Figure 4-2 Illustration of Concept of Conditional Value-at-Risk

For each scenario *k* and one particular feasible signal plan, the loss can be computed according to the problem settings, which we denote as L_k. Consider all the scenarios and order the loss as $L_1 < L_2 < ... < L_k$, let k_α be the unique index such that:

$$\sum_{k=1}^{k_\alpha} p_k \geq \alpha > \sum_{k=1}^{k_\alpha - 1} p_k$$

In words, L_{k_α} is the maximum loss that is exceeded only with probability $1-\alpha$, called as α-value-at-risk. Consequently, the expected loss exceeding the α-value-at-risk, i.e., conditional value-at-risk (CVAR) is:

$$\phi_\alpha = \frac{1}{1-\alpha}[(\sum_{k=1}^{k_\alpha} p_k - \alpha)L_{k_\alpha} + \sum_{k=k_\alpha+1}^{K} p_k L_k] \qquad (4\text{-}1)$$

The second component in the bracket is simply to compute the mean value, and the first is to split the probability 'atom' at the delay point L_{k_α} to make the collective probability of scenarios considered in the bracket exactly equal to $1-\alpha$. See Figure 4.2(b) for an illustration of the concept. It can be seen that the probability mass function has a jump at the point L_{k_α} due to the associated probability of p_{k_α}, which makes $\sum_{k=1}^{k_\alpha} p_k > \alpha$. To make the collective probability of scenarios exactly equal to $1-\alpha$, we need to split the probability of delay L_{k_α}. Note that if p_{k_α} makes $\sum_{k=1}^{k_\alpha} p_k = \alpha$, then "split" is not needed, and Equation (4-1) reduces to

$$\phi_\alpha = \frac{1}{1-\alpha}(\sum_{k=k_\alpha+1}^{K} p_k L_k).$$

For each feasible signal plan, Equation (4-1) can be used to compute the resulting conditional value-at-risk and our intention is to find a signal plan that leads to the minimum conditional value-at-risk. Rockafellar and Uryasev [4-9] showed that minimizing Equation (4-1) is equivalent to minimizing the following equation:

$$\min_{t,\xi} Z_\alpha = \xi + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k \cdot \max\left(L_k(t) - \xi, 0\right)$$

where ζ is a free decision variable. Subjective to a set of specific constraints, the optimal value of the objective function is the minimum conditional value-at-risk and the optimal solution (t^*, ζ^*) represents the robust signal timing plan and α -value-at-risk respectively.

4.3 Reference

4-1 Webster, F. V. *Traffic Signal Settings*. Road Research Technical Paper, No. 39, Her Majesty's Stationary Office, London, U.K., 1958.

4-2 Wallace, C. E., Courage, K. G., Hadi, M. A. and Gan, A. G. *TRANSYT-7F User's Guide*, University of Florida, Gainesville, FL, 1998.

4-3 Heydecker, B. Uncertainty and Variability in Traffic Signal Calculations. *Transportation Research*, Part B, Vol.21, 1987, pp. 79-85.

4-4 Yin, Y. Robust Optimal Traffic Signal Timing. *Transportation Research*, Part B, Vol.42, 2008, pp. 911-924.

4-5 Smith, B. L., Scherer, W. T., Hauser, T. A. and Park., B. B. Data-Driven Methodology for Signal Timing Plan Development: a Computational Approach. *Computer-Aided Civil and Infrastructure Engineering*, Vol.17, 2002, pp. 387-395.

4-7 Ben-Tal, A. and Nemirovski, A. Robust Optimization—Methodology and Applications. *Mathematical Programming*, Ser.B, Vol. 92, 2002, pp. 456-480.

4-8 Bertsimas D. and Sim, M. Robust Discrete Optimization and Network Flows. *Mathematical Programming*, Ser.B, Vol. 98, 2003, pp. 49-71.

4-9 Rockafellar, R.T. and Uryasev, S. Optimization of Conditional Value-at-Risk. *Journal of Risk*, Vol.2, 2000, pp. 21-41.

4-10 Rockafellar, R.T., and Uryasev, S. Conditional Value-at-Risk for General Loss Distribution. *Journal of Banking and Finance*, Vol. 26, 2002, pp. 1446-1471.

5 Robust Synchronization of Actuated Signals on Arterials

This chapter applies the scenario-based robust optimization approach introduced in Chapter 4 to synchronize actuated signals along arterials for smooth and stable progression under uncertain traffic conditions, mainly addressing the issue of uncertain (not fixed) starts/ends of green of sync phases. The model developed is based on Little's mixed-integer linear programming (MILP) formulation [5-1], which maximizes the two-way bandwidth to synchronize signals along arterials by determining offsets and progress speed adjustment etc. By specifying scenarios as realizations of uncertain red times of sync phases, we define the regret associated with a coordination plan with respect to each scenario, and then formulated a robust counterpart of Little's formulation as another MILP to minimize the average regret incurred by a set of high-consequence scenarios. The numerical example shows that the resulting robust coordination plan is able to increase the worst-case and 90^{th} percentile bandwidths by approximately 20% without affecting the average bandwidth.

5.1 Background

An increasing number of traffic signal controllers used in the United States are traffic-actuated. It has been a common practice to operate these controllers in coordinated systems to provide progression for major traffic movements along arterials and networks. Compared with fixed-time coordinated systems, these semi-actuated coordinated systems offer additional flexibility in responding to fluctuations in traffic demand. Under signal coordination, traffic actuated signals operate on a common background cycle length. Coordination is provided through a fixed reference point, which defines the start of the

controller local clock, and can be set to the start of green, end of green (yield point) or other time interval for the sync phases (e.g., beginning of the flashing don't walk interval). Note that the controller local clock definition varies among signal controller manufacturers [5-2].

To ensure operation efficiency of coordinated actuated systems, attention should be paid to determining appropriate signal settings, particularly offsets due to the fact that the start of green of the sync phases (typically Phases 2 and 6) is not fixed. Several approaches have been proposed in the literature to address such a so-called "early return to green" problem in the determination of offsets.

Jovanis and Gregor [5-3] suggested adjusting the end of green of the sync phases to the end of the through-band for non-critical signals. Skabardonis [5-4] proposed three methods for determining offsets for actuated signals from the optimal fixed-time splits and offsets. Although the three methods differ in the procedure and applicable situation, the concepts are essentially the same: making best estimates on average starting point of the sync phases and then optimizing the offset based on the estimates. Chang [5-5] offered a similar suggestion of obtaining the offsets from a second optimization run that uses the *anticipated* green times on the noncoordinated phases, as constraints on their maximum green times.

The above studies have focused on determination of appropriate offsets in the design stage of signal timing plans. Certainly after implementing the timing plans in the field, there are still opportunities for fine-tuning. Shoup and Bullock [5-6] examined a concept of using the link travel times observed for the first vehicle in a platoon to adjust offsets. The concept could lead to an online offset refiner, if vehicle identification technologies had been deployed in arterial corridors. Abbas et al. [5-7] developed an online real-time

offset transitioning algorithm that continually adjusts the offsets with the objective of providing smooth progression of a platoon through an intersection. More specifically, the objective was achieved by moving the green window so that more of the current occupancy actuation histogram is included in the new window. A greedy search approach was used to determine the optimal shift of the green window. In the ACS-Lite system [5-8], a run-time refiner can modify in an incremental way the cycle, splits and offsets of the plan based on observation of traffic conditions. Gettman et al. [5-9] elaborated the data-driven algorithm in the ACS-Lite for tuning offsets. The algorithm uses upstream detectors to construct cyclic profiles of traffic arrivals and then adjusts the offset to maximize the number of vehicles arriving during the green phase. As aforementioned, Yin et al. [5-10] proposed an offline refiner to fine-tune signal offsets, making use of a large amount of archived signal status data from real-time signal operations. Based on a more realistic estimate of distributions of starts/ends of green of the sync phases from the data, the refiner adjusts the offsets to minimize the red-meeting probability of the leading vehicle as well as maximize the average bandwidth.

This chapter attempts to use the scenario-based approach to synchronize actuated signals along corridors. More specifically, we determine a robust coordination plan that provides wider bandwidths in high-consequence scenarios. The approach can be used to either design a new timing plan for implementation or fine-tune the plan offline after implementation.

5.2 Bandwidth Maximization for Arterial Signal Coordination

Generally speaking, there are two approaches of generating coordination plans to synchronize signals along arterials and grid networks. One aims at bandwidth maximization, e.g., *MAXBAND* [5-11] and *PASSER-II* [5-12] while the other is

performance-based optimization, synchronizing signals to minimize the performance measures such as control delay and corridor travel time, e.g., *TRANSY-7F* [1-2]. To facilitate the presentation of our robust approach, this paper bases the model development on Little's MILP formulation [5-1], which maximizes the two-way bandwidth to synchronize signals along arterials by determining offsets and progress speed adjustment etc. The model has been proven to be a flexible and robust approach for signal synchronization and actually lays the foundation for *MAXBAND*. It has been later extended to consider variable bandwidth, phase sequencing and grid network synchronization [5-13], [5-14] and [5-15].

In the following we briefly reiterate Little's MILP formulation. Given a two-way arterial with an arbitrary number of signals, a common background cycle length, and the split information for each signal, the formulation attempts to synchronize the signals to produce a maximum sum of the inbound and outbound bandwidths. Let S_h and S_i be any pair of adjacent signals, and S_i follows S_h in the outbound direction. Figure 2.1 presents the geometry of the green bands between S_i and S_h. The horizontal lines indicate when the sync phases are red, and the zigzag lines represent the vehicle trajectories.

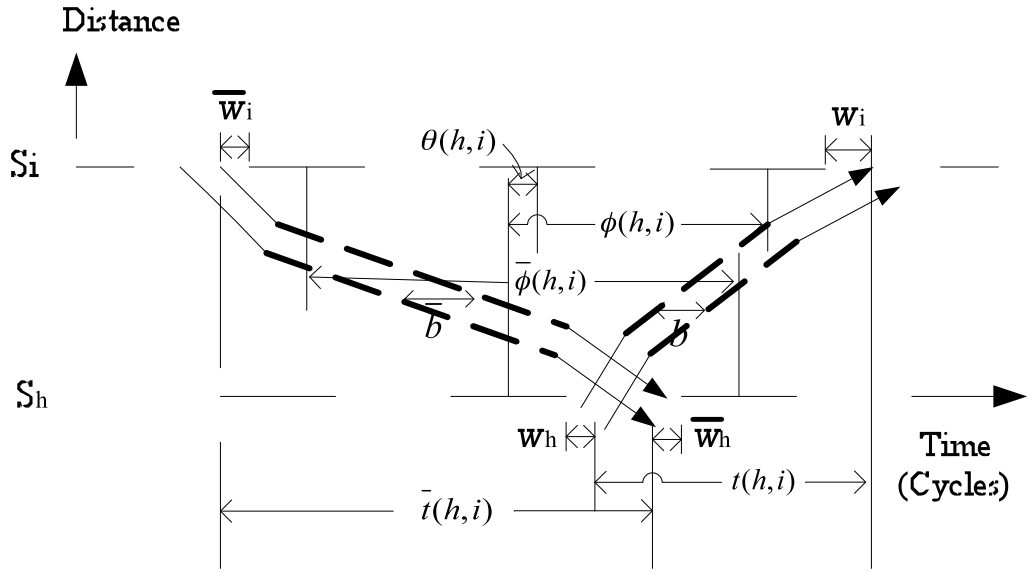


Figure 5-1 Geometry of Green Bands

Notations to be used together with those in Figure 5.1 are introduced as follows:

r_i	red time of signal i on the corridor (cycles)
$b(\bar{b})$	Outbound (inbound) bandwidth (cycles).
$t(h,i)[\bar{t}(i,h)]$	Outbound (inbound) travel time from signal $h(i)$ to $i(h)$ (cycles)
$\phi(h,i)\lfloor\bar{\phi}(i,h)\rfloor$	time from the center of red at S_h to the center of a particular red at S_i. The two reds are chosen so that each is immediately to the left (right) of the same outbound (inbound) green band (cycles)
$w_i(\overline{w_i})$	time from the right (left) side of S_i's red to the green band (cycles)
$m(h,i)$	$\phi(h,i)+\bar{\phi}(h,i)$. According to Figure 2.1, $m(h,i)$ must be integer (cycles)
T_1, T_2	lower and upper bounds on cycle length (s)
z	signal frequency (cycles/s), the inverse of cycle length
$d(h,i)$	distance from S_h to S_i (m). $d_i = d(i,i+1)$
$v_i(\overline{v_i})$	speed between S_i and S_{i+1} outbound (inbound) (m/s)
$e_i, f_i(\overline{e_i}, \overline{f_i})$	lower and upper bounds on outbound (inbound) speed (m/s)

Note that in Little's formulation, in addition to offsets, cycle length, travel speed, and change in speed between street segments are also decision variables, constrained by upper and lower limits. The bandwidth maximization problem is mathematically written as follows:

$$\max_{(b,\bar{b},z,w,\bar{w},t,\bar{t},m)} B = b + \bar{b}$$

s.t.

$$1/T_1 \le z \le 1/T_2 \tag{5-1}$$

$$w_i + b \le 1 - r_i \qquad \forall i = 1,...,n \tag{5-2}$$

$$\bar{w}_i + \bar{b} \le 1 - r_i \qquad \forall i = 1,...,n \tag{5-3}$$

$$(w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + (t_i + \bar{t}_i) = m_i - (r_i - r_{i+1}) \qquad \forall i = 1,...,n-1 \tag{5-4}$$

$$m_i = \text{integer} \qquad \forall i = 1,...,n-1 \tag{5-5}$$

$$(d_i / f_i)z \le t_i \le (d_i / e_i)z \qquad \forall i = 1,...,n-1 \tag{5-6}$$

$$(d_i / \bar{f}_i)z \le \bar{t}_i \le (d_i / \bar{e}_i)z \qquad \forall i = 1,...,n-1 \tag{5-7}$$

$$(d_i / h_i)z \le (d_i / d_{i+1})t_{i+1} - t_i \le (d_i / g_i)z \qquad \forall i = 1,...,n-2 \tag{5-8}$$

$$(d_i / \bar{h}_i)z \le (d_i / d_{i+1})\bar{t}_{i+1} - \bar{t}_i \le (d_i / \bar{g}_i)z \qquad \forall i = 1,...,n-2 \tag{5-9}$$

$$b, \bar{b}, w_i, \bar{w}_i \ge 0 \qquad \forall i = 1,...,n$$

In the objective function, $b, \bar{b}, w, \bar{w}, t, \bar{t}$ and m represent vectors whose elements are the scalar decision variables $b_i, \bar{b}_i, w_i, \bar{w}_i, t_i, \bar{t}_i$ and m_i. Equation (5-1) is the constraint on cycle length. Equations (5-2) and (5-3) are the constraints on green bandwidth. Equations (5-4) and (5-5) are the integer constraints due to the fact that $\phi(h,i) + \bar{\phi}(h,i)$ must be integer. Equations (5-6) and (5-7) are the constraints associated with travel speed. Equation (5-8) and (5-9) are the constraints on speed changes between adjacent street segments. All of these constraints and the objective function are linear, and (5-4) and

(5-5) are integer, therefore the problem is an MILP, which can be efficiently solved by using, e.g., the branch-and-bound algorithms. From the solutions, offsets can be easily obtained as follows according to the geometry in Figure 5.1:

$$\theta_{i+1} = [w_i - w_{i+1} + t_i + 1/2(r_i - r_{i+1})]_+ \qquad \forall i = 1,...,n-1$$

where θ_{i+1} is the offset of signal *i*+1 with respect to *i*; $[x]_+$ is defined as $x - \text{int}(x)$ and $\text{int}(x)$ represents the largest integer not greater than *x*. Hereafter, we denote the vector of offsets as θ.

5.3 Scenario-Based Approach for Robust Synchronization of Actuated Signals

Little's formulation assumes that the durations of minor phases (red times of sync phases) are deterministic. Such an assumption does not hold for actuated signal control, where the phase durations of minor phases vary between zero (skipped) and the corresponding maximum greens. That is the reason for a so-called "early return to green" problem. To illustrate how prevailing the problem of "early return to green" is, Figure 5.2 depicts the histograms for starts of green of Phase 2 and 6 at two selected intersections, Page Mill and Stanford, along El Camino, Palo Alto, CA. The data were collected from 11:00 am - 3:00 pm between February 6 and March 7, 2005, for a total of 14 weekdays with more than 1,600 cycles. Page Mill is a critical intersection for the corridor with almost equal amounts of mainline and cross-street traffic. Still, the probabilities of "early return to green" are 61% for Phase 2 and 39% for Phase 6. Stanford has low volume of minor-phase traffic, thus the probabilities are as pretty high as 92% for Phase 2 and 94% for Phase 6 respectively. The histograms confirm the assertion made in the previous studies that the problem of "early return to green" should be recognized and explicitly addressed in actuated signal synchronization. Note that in addition to uncertainty of start

of green, end of green is also uncertain, especially under lead-lag phase sequence, due to skip or gap-out of the left-turn phase. Figure 5.3 presents the histograms for green terminations of Phase 2 and 6 at Page Mill and Stanford. It can be seen that compared with starts of green, terminations of green have much narrower spans. Under many circumstances, the termination is the force-off point.

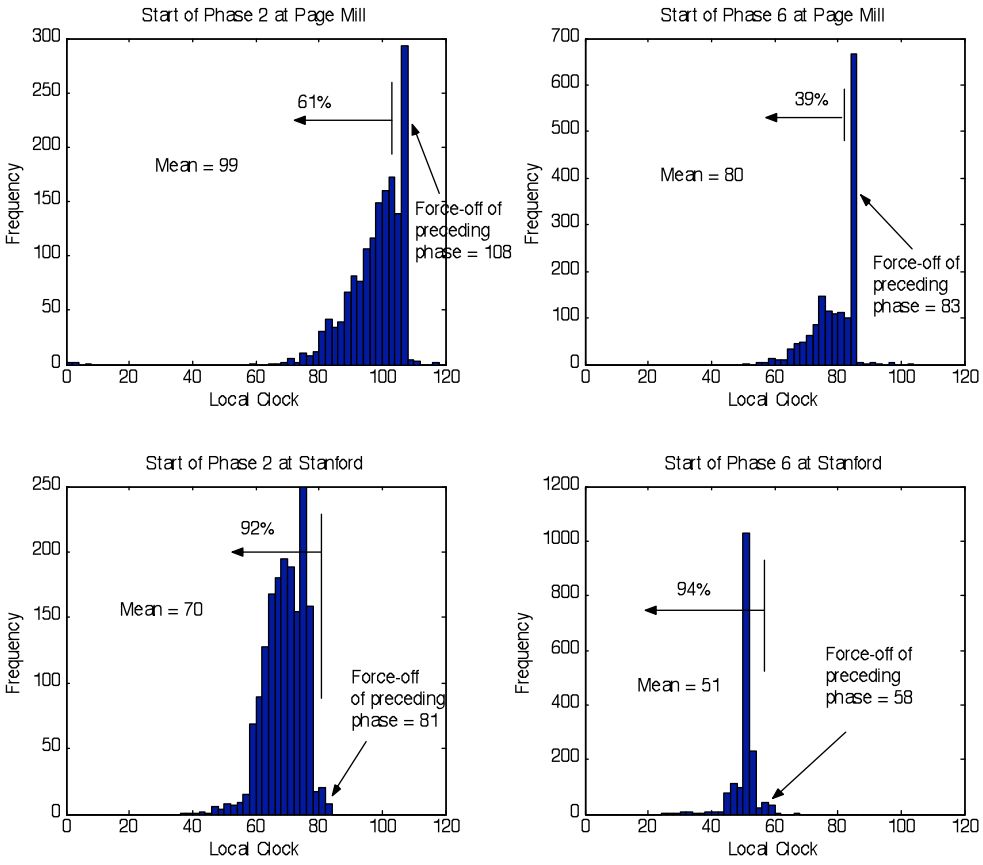


Figure 5-2 Probability of Early-Return-to-Green at Two Intersections

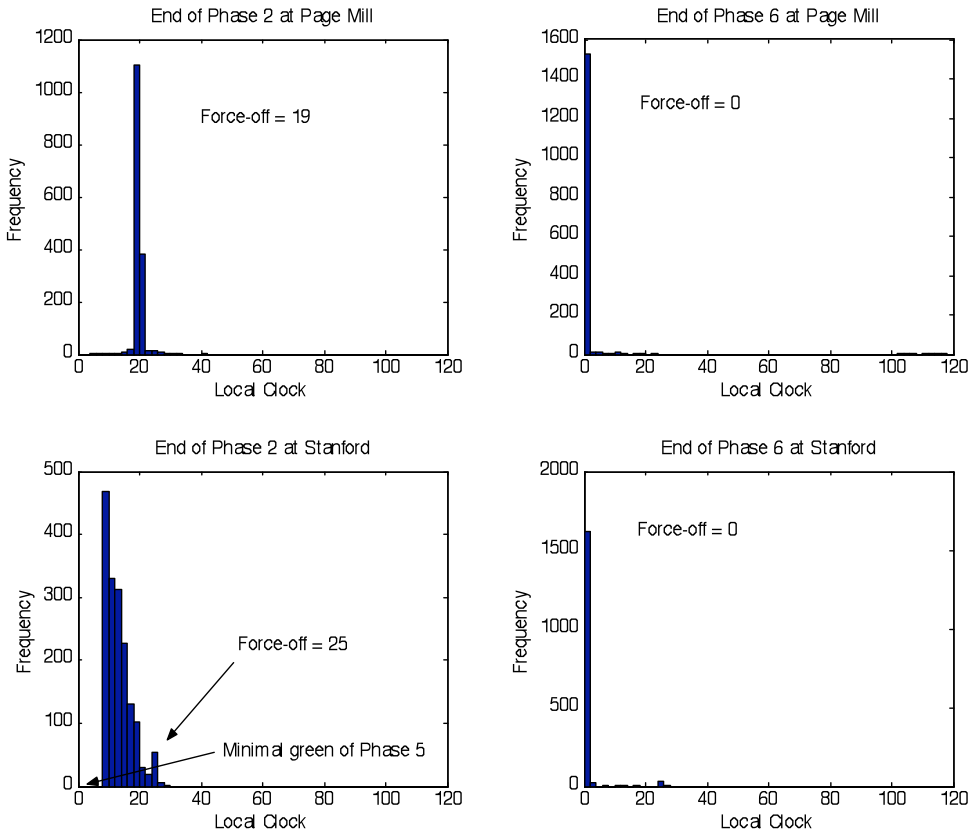


Figure 5-3 Uncertain Termination of Green at Two Intersections

The above empirical data shows that for actuated signals, the red times, r_i, in Little's formulation should follow random distributions with supports between zero and the sum of maximum greens of the conflicting minor phases. To represent the uncertainty, a set of scenarios $\Omega = \{1,2,3,\cdots,K\}$ in introduced, and each individual scenario is composed of the red time of the sync phases[2] at all the signals, which for example at intersection i, is denoted as r_i^k.

[2] Normally Phases 2 and 6. For simplicity, we assume here that both phases start and end at the same time.

We now define the regret function. For each scenario k, we solve Little's MILP formulation to obtain the maximum two-way bandwidth, denoted as B_k^*. Consequently, for any other feasible coordination plan (offsets θ and cycle frequency z) that may not be optimal for scenario k, the regret or loss can be defined as:

$$L_k = B_k^* - B_k(\theta, z) \qquad \forall k$$

where L_k is the regret or loss of the coordination plan (θ, z) with respect to red time scenario k, and B_k is the two-way bandwidth resulted by (θ, z) under scenario *k*.

Then the robust signal coordination plan can be obtained by minimizing the following equation:

$$\min_{\theta, z, \xi} Z_\alpha = \xi + \frac{1}{1-\alpha} \sum_{k=1}^{K} p_k \cdot \max\left(L_k(\theta, z) - \xi, 0\right)$$

Therefore, the scenario-based robust synchronization model for actuated signals can be written as follows:

$$\min_{(b^k, \bar{b}^k, z, w^k, \bar{w}^k, t^k, \bar{t}^k, m^k, \theta, \xi, U_k, L_k)} Z_\alpha = \xi + \frac{1}{1-\alpha} \sum_{k=1}^{K} p_k U_k$$

s.t.

$$U_k \geq L_k - \xi \qquad \forall k = 1, \ldots, K$$

$$U_k \geq 0 \qquad \forall k$$

$$L_k = B_k^* - (b^k + \bar{b}^k) \qquad \forall k$$

$$1/T_1 \leq z \leq 1/T_2$$

$$w_i^k + b^k \leq 1 - r_i^k \qquad \forall i = 1, \ldots, n, \ \forall k$$

$$\bar{w}_i^k + \bar{b}^k \leq 1 - r_i^k \qquad \forall i = 1, \ldots, n, \ \forall k$$

$$(w_i^k + \bar{w}_i^k) - (w_{i+1}^k + \bar{w}_{i+1}^k) + (t_i^k + \bar{t}_i^k) = m_i^k - (r_i^k - r_{i+1}^k) \qquad \forall i = 1, \ldots, n, \ \forall k$$

$$m_i^k = \text{integer} \qquad \forall i = 1,...,n, \ \forall k$$

$$\theta_{i+1} = w_i^k - w_{i+1}^k + t_{i+1}^k + 1/2(r_i^k - r_{i+1}^k) \qquad \forall i = 1,...,n-1, \ \forall k$$

$$(d_i / f_i)z \le t_i^k \le (d_i / e_i)z \qquad \forall i = 1,...,n-1, \forall k$$

$$(d_i / \overline{f}_i)z \le \bar{t}_i^k \le (d_i / \bar{e}_i)z \qquad \forall i = 1,...,n-1, \forall k$$

$$(d_i / h_i)z \le (d_i / d_{i+1})t_{i+1}^k - t_i^k \le (d_i / g_i)z \qquad \forall i = 1,...,n-2, \forall k$$

$$(d_i / \bar{h}_i)z \le (d_i / d_{i+1})\bar{t}_{i+1}^k - \bar{t}_i^k \le (d_i / \bar{g}_i)z \qquad \forall i = 1,...,n-2, \forall k$$

$$b^k, \bar{b}^k, w^k, \bar{w}^k \ge 0 \qquad \forall k$$

where U_k is an auxiliary decision variable, equal to $\max(L_k(\theta,z) - \xi, 0)$. Note that B_k^* is pre-determined by solving Little's formulation for each scenario. As formulated, the problem is another MILP and can be efficiently solved. As demonstrated in pervious studies [5-16], [5-17] and [5-18], such a formulation can offer computational advantages and allow handling a large number of scenarios. Our numerical example below also shows that the problem can be solved in polynomial time.

5.4 Numerical Example

5.4.1 Plan Generation

We solve the robust synchronization formulation for an arterial with six signals, which is included in CORSIM as one of the tutorial examples, named "ActCtrl Example", shown in Figure 5.4. The outbound signals, starting from signal 1, are located at 0, 314, 554, 759, 1012, 1317m respectively. The timing plan for each isolated signal is determined using the Webster's equation. The red times of the sync phases are assumed to be independently normally distributed with means of 0.27, 0.24, 0.50, 0.35, 0.35, 0.43 cycles respectively and the same standard deviation of 0.05 cycles across all the signals. The

scenarios are specified by random sampling and are assumed to have equal probability to occur. The upper and lower limits of the cycle length are 100 s and 45 s. Limits on travel speed for different street segments are set to be the same with the upper limit of 17.9 m/s (40 mph) and the lower limit of 13.4 m/s (30 mph). Changes in reciprocal speeds across all the segments are limited between -0.0121 and 0.0121$(m/s)^{-1}$, corresponding to a maximum possible change in speed of ±2.3 m/s (±4.9 mph) at the lower limit of the speed and ±3.9 m/s (±8.8 mph) at the upper limit. The confidence level α is selected to be 0.90.

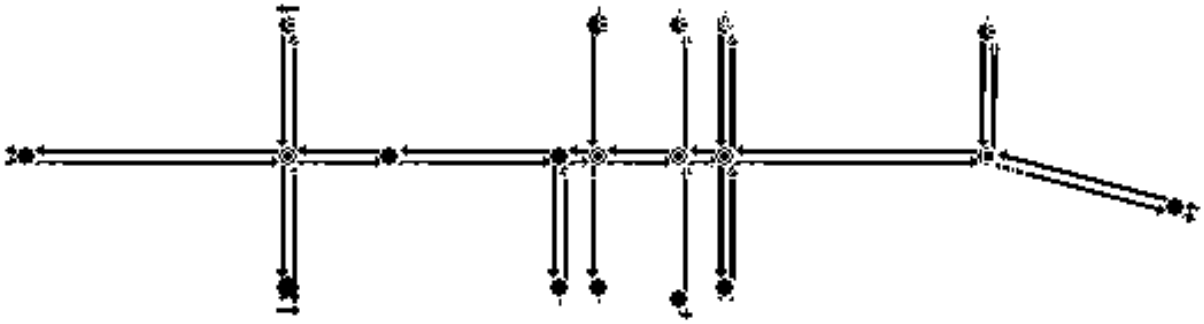


Figure 5-4 A Snapshot of the "ActCtrl Example" in CORSIM

We use an algebraic modeling system called GAMS [5-19] and CPLEX solver [5-20] to solve the robust synchronization formulation with the number of scenarios varying from 10 to 250. The computation times (in CPU seconds) and the plan differences are presented in Table 5.1, plotted in Figure 5.5.

Table 5-1 Computation Time and Coordination Plan Difference

Number of Scenario	Time*(sec)	Plan Difference
10	2.88	169.3%
20	5.11	2.9%
50	11.67	0.7%
100	28.58	1.0%
150	45.67	1.1%
200	61.97	0.4%

250	81.81	0.0%

*: including the times for solving Little's formulation for each scenario

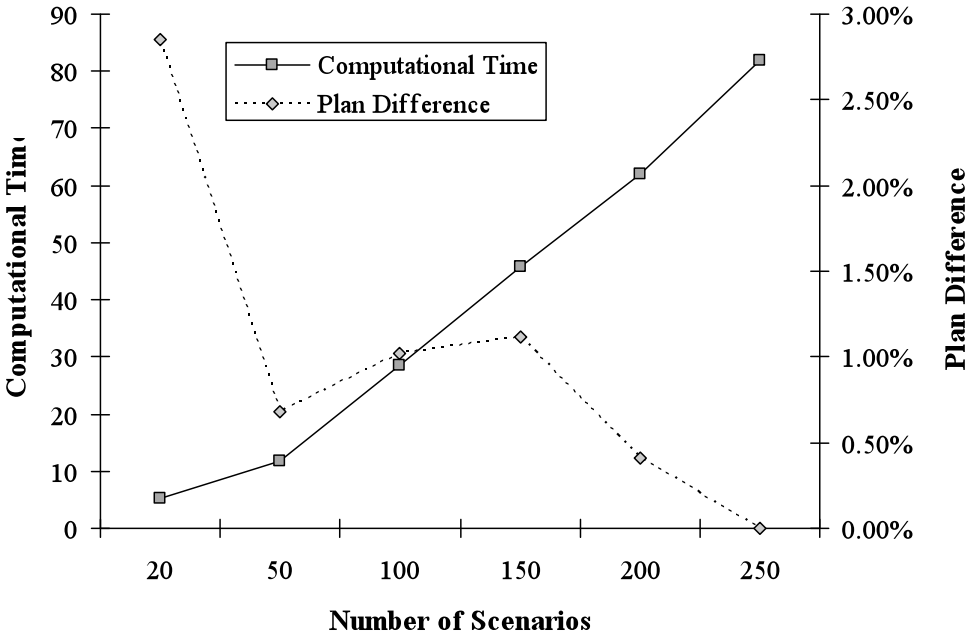


Figure 5-5 Computational Time and Plan Difference

The plan difference is defined as $\left\|\theta^K - \theta^{250}\right\|_2 / \left\|\theta^{250}\right\|_2$, where θ^K is the plan generated with *K* scenarios. Within expectation, the difference tends to decrease as the number of scenarios increases. However, it can be observed that relatively small number of scenarios is already enough to produce similar robust plans.

The reported computation times include the times needed to solve Lttile's formulation for each scenario to obtain B_k^*. Regressing the computational times against the number of scenarios, we obtain the following equation, suggesting that the problem may be solved in polynomial time:

Time = 0.3286*(number of scenario) - 2.6556 , R^2 = 0.9962.

For the comparison purpose, we also generate a nominal plan following the procedure of deterministic signal coordination, in which we use the mean red times as our estimates, and then use Little's formulation described in part two of the section to obtain the nominal plan.. Both the robust plan (with 250 scenarios) and the nominal plan are reported in Table 5.2.

Table 5-2 Robust Plan and Nominal Plan

	Offset (sec)					Cycle Length (sec)
Intersection	2	3	4	5	6	
Robust plan	33	2	2	35	4	80
Nominal plan	53	74	72	47	68	77

5.4.2 Evaluation

To evaluate the performance of both nominal and robust plans, we conduct macroscopic Monte Carlo simulation and microscopic simulation in CORSIM. In the Monte Carlo simulation, 2000 samples of red times are drawn from the same normal distributions previously used to generate scenarios for solving the robust synchronization formulation. With each sample, the bandwidths resulted by both the nominal and robust plans are computed. Consequently, several performance measures, including the average, worst-case (minimum) and 90^{th} percentile minimum bandwidths, and the 90% conditional value-at-risk, are calculated and reported in Table 5.3.

Table 5-3 Monte Carlo Simulation with Normal Distribution

Bandwidth (sec)	Mean	Worst case	90th percentile	90% CVaR*	Change			
					Mean	Worst case	90th percentile	90% CVaR
Nominal Plan	28.7	17.0	23.8	40.1	-	-	-	-
Robust Plan	35.8	21.8	30.6	34.5	24.7%	28.2%	28.6%	-14.0%

The results indicate that the robust coordination plan performs better against high-consequence scenarios, with the worst-case bandwidth increasing by 23.1%, and the 90th percentile bandwidth by 23.9%, and the conditional value-at-risk decreasing by 17.3%. At the same time, the average bandwidth also increases by 20%. However, since the robust plan is designed to guard against high-consequence scenarios, an improvement of the average performance is not what we should expect and will not be necessarily obtained.

To further validate the robust synchronization formulation, we evaluate the plans with 2000 samples randomly drawn from independent *uniform* distributions with the minimum and maximum values described in Table 5.4. The resulting performance measures are summarized in Table 5-5. The robust plan still outperforms the nominal plan, with the average, worst-case and 90th percentile bandwidths increasing by 16.7%, 22.1% and 22.5%, and the regret decreasing by 16.5%. This examination suggests that the robust formulation is not overly sensitive to the specification of scenarios and using distorted distributions to generate scenarios may still result in robust timing plans.

Table 5-4 Critical Values for Uniform Distribution

Signal	Red time	
	Minimum duration (cycles)	Maximum duration (cycles)
# 1	0*	0.42

# 2	0	0.25
# 3	0	0.56
# 4	0	0.53
# 5	0	0.56
# 6	0	0.45

*: Minor phases skipped.

Table 5-5 Monte Carlo Simulation with Uniform Distribution

Bandwidth (sec)	Mean	Worst case	90th percentile	90% CVaR	Change			
					Mean	Worst case	90th percentile	90% CVaR
Nominal Plan	36.6	17.8	24.3	54.1	-	-	-	-
Robust Plan	44.3	22.6	31.0	46.9	21.0%	27.0%	27.6%	-13.3%

We recognize the limitation of the bandwidth-based synchronization that traffic flows and intersection capacities are not considered in the optimization criterion [5-21], and thus bandwidth maximization does not necessarily optimize other delay-related performance measures. To examine how the robust plan affects those measures, we conduct a microscopic simulation using CORSIM. The robust and nominal plans are implemented respectively in the semi-actuated corridor of "ActCtrl Example" in CORSIM. We select the control delay, corridor travel time, and vehicle stop ratio as the performance measures where vehicle stop ratio is defined as the total number of stops (when speed is lower than 3 mph) divided by the total number of vehicles served by the corridor within the simulation period. The means and standard deviations calculated from ten simulation runs are reported in Table 5.6. We conduct *t* tests and *F* tests to examine whether those performance measures are statistically different. As shown in Table 5.6, the three *t* values are all greater than the critical *t* value of 2.552 at the significance level of 1%, suggesting

that the corridor has better average performance with the robust plan. The *F* values are all smaller than the critical *F* value of 2.44 at the significance level of 10%, indicating that we can not reject the hypothesis that the variances are the same. It should be pointed out that the intention of the CORSIM simulation is to examine whether the robust plan makes the delay-related performance measures worse off. Although the simulation results actually suggest otherwise, we do not expect to always obtain such improvements, since it is not what the robust plan is designed for. The only conclusion we draw from the CORSIM simulation is that the robust plan seems unlikely to worsen the delay-related performance measures.

Table 5-6 Microscopic Simulation Results and Hypothesis Test

Performance Measure		Control delay (sec)	Travel time (sec)	Stop ratio
Robust plan	Mean	96.2	318.4	0.148
	Std. Deviation	12.2	12.7	0.008
Nominal plan	Mean	122.6	346.8	0.176
	Std. Deviation	9.9	12.4	0.010
Hypothesis test	*t* value	5.31	5.06	7.08
	F value	1.54	1.05	1.32

5.5 Reference

5-1 Little, J. D. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. *Operations Research*, Vol. 14, No. 4, 1966, pp. 568-594.

5-2 *FHWA Traffic Control Systems Handbook*. Report FHWA-SA-95-032. U.S. Department of Transportation., Washington D.C., 1996.

5-3 Jovanis, P. P. and Gregor, J. A. Coordination of Actuated Arterial Traffic Signal Systems. *Journal of Transportation Engineering*, Vol.112, No. 4, 1986, pp. 416-432.

5-4 Skabardonis, A. Determination of Timings in Signal Systems with Traffic-Actuated Controllers. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1554, TRB, National Research Council, Washington, D.C., 1996, pp. 18-26.

5-5 Chang, E. C. P. Guidelines for Actuated Controllers in Coordinated Systems. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1554, TRB, National Research Council, Washington, D.C., 1996, pp. 61-73.

5-6 Shoup, G. E. and Bullock, D. Dynamic Offset Tuning Procedure Using Travel Time Data. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1683, TRB, National Research Council, Washington, D.C., 1999, pp. 84-94.

5-7 Abbas, M., Bullock, D. and Head, L. Real-Time Offset Transitioning Algorithm for Coordinating Traffic Signals. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1748, TRB, National Research Council, Washington, D.C., 2001, pp. 26-39.

5-8 Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D. and Mirchandani, P. ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems. Design guidelines for deploying closed loop systems. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1856, TRB, National Research Council, Washington, D.C., 2003, pp. 175-184.

5-9 Gettman, D., Shelby, S. G., Head, L., Bullock, D. M. and Soyke, N. Data-Driven Algorithms for Real-Time Adaptive Tuning of Offsets in Coordinated Traffic Signal Systems. In *Transportation Research Record*: Journal of the Transportation

Research Board, No. 2035, TRB, National Academies Council, Washington, D.C., 2007, pp.1-9.

5-10 Yin, Y., Li, M. and Skabardonis, A. An Offline Offset Refiner for Coordinated Actuated Signal Control System. *Journal of Transportation Engineering*, Vol. 133, No. 7, 2007, pp. 426-432.

5-11 Little, J.D., Kelson, M.D. and Gartner, N.H. MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 795, TRB, National Research Council, Washington, D.C., 1981, pp. 40-46.

5-12 Chang, E., Lei, J.C. and Messer, C.J. *Arterial Signal Timing Optimization using PASSER-II-87 Microcomputer User's Guide*, TTI Research Report 467-1, Texas A&M University, College Station, Texas, 1988.

5-13 Gartner, N.H. and Stamatiadis, C. Arterial-Based Control of Traffic Flow in Urban Grid Networks. *Mathematical and Computer Modeling*, Vol.35, 2002, pp. 657-671.

5-14 Gartner, N.H. and Stamatiadis, C. Progression Optimization Featuring Arterial- and Route-Based Priority Signal Networks. *ITS Journal*, Vol.8, 2004, pp. 77-86.

5-15 Messer, C.J., Hogg, G.L., Chaudhary, N.A. and Chang, E. *Optimization of Left Turn Phase Sequence in Signalized Networks using MAXBAND 86, Vol. 1 Summary Report.* Technical Report FHWA/RD-87/109, FHWA, Washington DC.

5-16 Krokhmal, P., Palmquist, J. and Uryasev, S. Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints. *Journal of Risk*, Vol.4, 2002, pp. 46-68.

5-17 Chen, G., Daskin, M.S., Shen, Z.J.M. and Uryasev, S. The α-Reliable Mean-Excess Regret Model for Stochastic Facility Location Modeling. *Naval Research Logistics,* Vol.53, 2006, pp. 617-626.

5-18 Yin, Y. A Scenario-Based Model for Fleet Allocation of Freeway Service Patrols. *Network and Spatial Economics*, Vol.8, No.4, 2008, pp. 407-417.

5-19 Brooke, A., Kendirck, D., and Meeraus, A. *GAMS: A User's Guide*. The Scientific Press, South San Francisco, California, 1992.

5-20 *CPLEX*, Version 9.0. CPLEX Optimization Inc., Nevada, 2004.

5-21 Gartner, N.H., Assmann, S.F., Lasaga, F. and Hou, D.L. A Multi-Band Approach to Arterial Traffic Signal Optimization. *Transportation Research*, Part B, Vol.25, 1991, pp. 55-74.

6 Simulation-based robust optimization for signal timing

This chapter applies the scenario-based approach to optimize the timings of actuated signals along arterials under day-to-day demand variations or uncertain traffic future growth. Based on a cell-transmission representation of traffic dynamics, a stochastic programming model is formulated to determine cycle length, green splits, phase sequences and offsets to minimize the expected delay incurred by high-consequence scenarios of traffic demand. The stochastic programming model is simple in structure but contains a large number of binary variables. Existing algorithms, such as branch and bound, are not able to solve it efficiently. Consequently, a simulation-based genetic algorithm is developed to solve the model. The model and algorithm are validated and verified in two networks. It is demonstrated that the resulting robust timing plans perform better against high-consequence scenarios without losing optimality in the average sense under both congested and uncongested traffic conditions.

6.1 Background

Since the seminal work of Webster [6-1], significant efforts have been devoted to improving signal timing for saturated isolated intersections, coordinated arterials and grid networks etc. For example, Robertson and Bretherton [6-2] described the evolution of an adaptive traffic control system SCOOT from the TRANSYT method. The advantage of SCOOT is that the system can measure the cycle flow profile in real time and then update the coordination plan in an online manner. Gartner [6-3] proposed another adaptive control strategy for synchronizing traffic signals using the virtual-fixed-cycle concept. The strategy can continuously optimize signal settings in response to demand fluctuations,

which is achieved via executing a distributed dynamic programming algorithm by means of a three-layer architecture.

However, only a few studies have been conducted in the literature to directly address signal timing under flow fluctuations for pre-timed control systems. Heydecker [6-4] investigated the consequences of variability in traffic flows and saturation flows for the calculation of signal settings, and then proposed an optimization formulation that minimizes the mean rate of delay over the observed arrivals and saturation flows. Sensitivity analysis was carried out to test the benefit of taking into account the variability of arrival rate when optimizing the signal settings. Following the same notion, Ribeiro [6-5] proposed a novel technique called Grouped Network for using TRANSYT to calculate timing plans that are efficient even when demand is variable. Both studies focus on optimizing the average performance.

This chapter applies the scenario-based approach to optimize signal timings along corridors, explicitly considering the demand uncertainty. More specifically, we determine a robust coordination plan that results in smaller delay in high-consequence scenarios. Based on a cell-representation of traffic dynamics, the model attempts to optimize cycle length, green splits, phase sequences and offsets simultaneously.

6.2 Cell-Transmission Model

6.2.1 Model introduction

Modeling traffic dynamics is particularly important for signal timing optimization because realistic evaluation of each feasible timing plan cannot be performed without a realistic traffic flow model. At the same time, the evaluation should be efficient such that it can be incorporated into an optimization procedure. For these reasons, we select the

macroscopic cell-transmission model (CTM) proposed by Daganzo [6-6] and [6-7] in order to fully capture traffic dynamics, such as shockwaves, and queue formation and dissipation.

CTM is a finite difference solution scheme for the hydrodynamic theory of traffic flow or the Lighthill-Whitham-Richards (LWR) models. Mathematically the theory can be stated as the following equations:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{6-1}$$

$$q = Q(k, x, t) \tag{6-2}$$

Where the q and k are two macroscopic variables: flow and density. Equation (6-1) is the flow conservation equation and Equation (6-2) defines the traffic flow (q), at location x and time t, as a function of the density (k).

For a homogeneous roadway, Daganzo [6-6] and [6-7] suggested using the time-invariant flow-density relationship:

$$q = \min\{Vk, Q, W(k_{jam} - k)\}$$

where V = the free flow speed;

Q = the inflow capacity;

k_{jam} = the jam density;

W = the backward wave speed.

Figure 6.1 shows the flow-density relationship in a piecewise linear diagram.

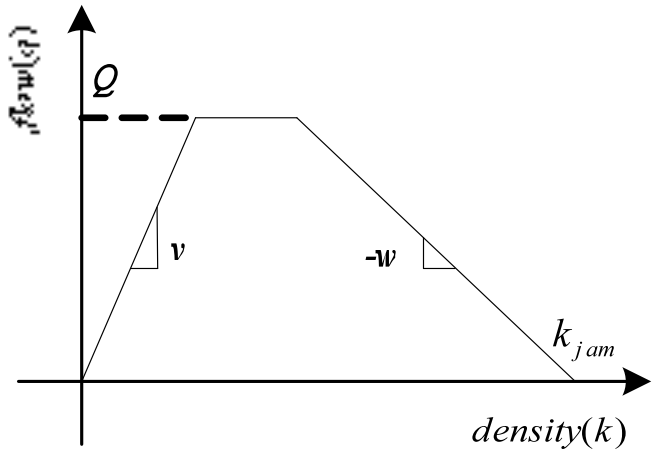


Figure 6-1 Piecewise Linear $q-k$ Relationship

By dividing the whole network into homogeneous cells with the cell length equal to the duration of time step multiplied by the free flow speed, the results of the LWR model can be approximated by a set of recursive equations:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t) \tag{6-3}$$

$$y_i(t) = \min\{n_i(t), Q_i(t), \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]\} \tag{6-4}$$

where $n_i(t)$ = the number of vehicles in cell i during time step t;

$y_i(t)$ = the number of vehicles that leave cell i during time step t;

$N_{i,\max}$ = the maximum number of vehicles that can be accommodated by cell i;

$Q_i(t)$ = the minimum of the capacity flows of cell i and $i+1$;

$\omega = W/V$.

Equation (6-3) ensures the flow conservation that the number of vehicles in cell i during time step $t+1$ equals to the number of vehicles in cell i during time step t plus the inflow and minus the outflow. Equation (6-4) determines the outflow for each cell during each time step, which is a piecewise linear function.

6.2.2 Encapsulating CTM in Signal Timing Optimization

Lin and Wang [6-8], Lo [6-9] and Lo et al. [6-10] have successfully incorporated CTM in their signal timing optimization formulations.

Lin and Wang [6-8] formulated a 0-1 mixed integer linear program, considering the number of stops and fixed or dynamic cycle length. In the model, cells in the network are categorized into four groups: ordinary, intersection, origin and destination cells. The objective is to minimize a weighted sum of total delay and total number of stops. In their model, Equation (6-4) is replaced by three linear inequalities, which do not accurately replicate flow propagation and may suffer the so-called "vehicle holding problem". To address this issue, one additional penalty term is added to the objective function. The authors demonstrated the model capable of capturing traffic dynamic using an emergency vehicle problem. However, the model is developed only for one-way streets and neither merge nor diverge of traffic is considered.

Lo ([6-9] and [6-10]) and his colleagues [6-11] developed dynamic signal control formulations based on CTM. By introducing binary variables, Equation (6-4) is equivalently converted into a linear system. The models proposed are able to generate dynamic or fixed timing plan and optimize cycle length, phase splits and offsets explicitly. Unfortunately, the models are again proposed for one-way streets.

This section expands Lo's models to a more general and realistic setting, including modeling two-way traffic, phase sequence optimization and applying new technique to equivalently transforming CTM for a general signal-controlled network to be a linear system of equalities and inequalities with integer variables.

6.3 Enhanced Deterministic Signal Optimization Model

Assuming deterministic constant or time-variant demands within the optimization horizon, this section presents a CTM-based deterministic signal timing optimization model. The model extends Lo's models in the following aspects:

- Modeling two-way traffic: this extension not only increases the size of the problem but also introduces another layer of complexity in representing signalized intersections and signal settings. For example, as the number of traffic movements increases, the number of phase combinations and sequences increase significantly;
- Optimization of phase sequence: the left-turn leading or lagging control is modeled explicitly;
- New formulation: we transform the CTM of a general signal-controlled arterial to be an equivalent linear system of equalities and inequalities with integer variables using a technique recently proposed by Pavlis and Recker [6-12] and formulate a mixed-integer linear program to optimize cycle length, green splits, offsets and phase sequences.

It is assumed that every intersection along the arterial is signalized, and all the cells comprising the network can be categorized into six groups: ordinary, origin, destination, non-signalized diverge, signalized diverge and signalized merge cells, as shown in Figure 6.2 from (a) to (e). Each group has a different configuration to be discussed below.

6.3.1 Objective Function

In the deterministic setting, we aim to optimize signal timing to minimize the total system delay of an urban arterial. The objective is to minimize the total area (as in Figure 6.3) between the cumulative arrival curves of the origin cells and the cumulative departure curves of the destination cells, expressed as the following linear function:

$$L = \min(\sum_{i \in O} \sum_{t=1}^{T} \sum_{j=1}^{t} d_i(j)) - \sum_{i \in D} \sum_{t=1}^{T} \sum_{j=1}^{t} y_i(j))$$

where, O is the set of origin cells and D is the set of destination cells; T is the duration of the optimization horizon; $d_i(j)$ is the demand at origin cell i during time step j. It is straightforward to observe that if the demands at origin cells are given, the objective function is equivalent to maximizing the second component, i.e., the area under the cumulative departure curves.

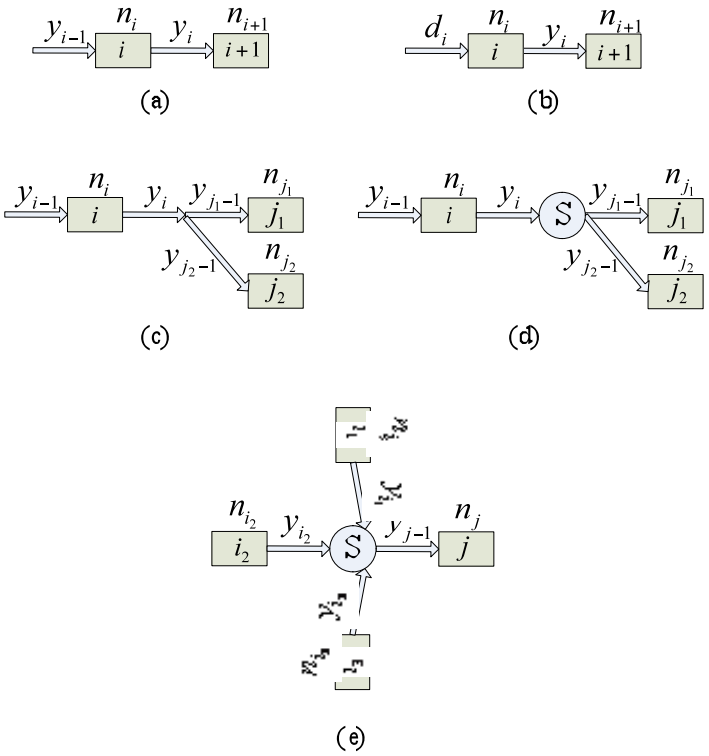


Figure 6-2 Cell Configurations

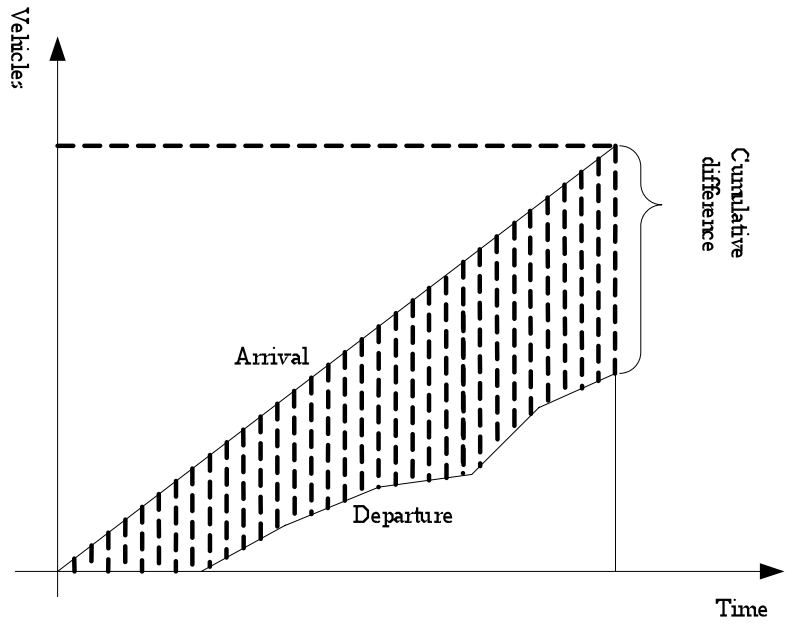


Figure 6-3 Interpretation of the Objective Function

6.3.2 Constraints

6.3.2.1 Constraints for ordinary cells

The ordinary cells are those with only one inflow and one outflow as cell i in Figure 6.2(a). According to the cell transmission model, the flow constraints are as follows:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t)$$

$$y_i(t) = \min\{n_i(t), Q_{i,\max}(t), Q_{i+1,\max}(t), \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]\}$$

$y_i(t)$ is determined by the min function, which is essentially a linear conditional piecewise function (CPF). Pavlis and Recker [6-12] provided a scheme to transform this kind of CPF into mixed integer constraints with the least number of integer variables. By introducing two binary variables, i.e., χ_1 and χ_2, and a sufficiently large negative constant, i.e., U^-, the CPF can be equivalently translated into the following constraints:

$$(\chi_1 + \chi_2)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1 + \chi_1 - \chi_2)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1 - \chi_1 + \chi_2)U^- \le y_i(t) - Q_{i+1,\max}(t) \le 0$$

$$(2 - \chi_1 - \chi_2)U^- \le y_i(t) - \omega \cdot [N_{i+1,\max} - n_{i+1}(t)] \le 0$$

To see the equivalence, Table 6.1 enumerates all possible 0-1 combinations and evaluates the value of $y_i(t)$.

Table 6-1 CPF Result under Different 0-1 Combinations

0-1Combination (χ_1, χ_2)	Constraint Representations			
(0, 0)	$y_i(t) = n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(0, 1)	$y_i(t) \le n_i(t)$	$y_i(t) = Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(1,0)	$y_i(t) \le n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) = Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(1,1)	$y_i(t) \le n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) = \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$

6.3.2.2 Constraints for origin cells

The origin cells, as shown in Figure 6.2(b), have the same structure as the ordinary cells, except that the inflow is fixed as the corresponding demand input. These cells perform as valves that control the traffic volume flowing into the network. The above constraints are slightly changed to incorporate the demand:

$$n_i(t+1) = n_i(t) + d_i(t) - y_i(t)$$

$$(\chi_3 + \chi_4)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1+\chi_3-\chi_4)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1-\chi_3+\chi_4)U^- \le y_i(t) - Q_{i+1,\max}(t) \le 0$$

$$(2-\chi_3-\chi_4)U^- \le y_i(t) - \omega \cdot [N_{i+1,\max} - n_{i+1}(t)] \le 0$$

6.3.2.3 Constraints for destination cells

The destination cells are those with outflow unlimited, implying that all the vehicles currently reside in the cells are able to flow out of the system at the next time step. The constraints are as follows:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t)$$

$$y_i(t) = n_i(t)$$

6.3.2.4 Constraints for non-signalized diverge cells

Non-signalized diverge occurs at certain roadway segments where the geometry or capacity changes and traffic diverse to different lanes for their respective destinations. Figure 6.2(c) is a typical configuration for non-signalized diverge: traffic in cell i diverges to cells j_1 and j_2 according to proportion parameters β_{j_1} and β_{j_2}. The constraints can be stated as follows:

$$y_{j_1-1} = \beta_{j_1} \cdot y_i$$

$$y_{j_2-1} = \beta_{j_2} \cdot y_i$$

$$\beta_{j_1} + \beta_{j_2} = \cdot 1$$

$$(\chi_5+\chi_6+\chi_7)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1+\chi_5+\chi_6-\chi_7)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1+\chi_5-\chi_6+\chi_7)U^- \le y_i(t) - Q_{j_1,\max}(t)/\beta_{j_1} \le 0$$

$$(1-\chi_5+\chi_6+\chi_7)U^- \le y_i(t) - Q_{j_2,\max}(t)/\beta_{j_2} \le 0$$

$$(2-\chi_5-\chi_6+\chi_7)U^- \le y_i(t)-\omega\cdot[N_{j_1,\max}-n_{j_1}(t)]/\beta_{j_1}] \le 0$$

$$(2-\chi_5+\chi_6-\chi_7)U^- \le y_i(t)-\omega\cdot[N_{j_2,\max}-n_{j_2}(t)]/\beta_{j_2}] \le 0$$

$$\chi_5+\chi_6+\chi_7 \le 2$$

$$\chi_6+\chi_7 \le 1$$

6.3.2.5 Constraints for signalized diverge cells

Signalized diverge is the diverge that happens within a signalized intersection, when traffic from one direction enters the intersection during a corresponding green phase and leaves the intersection while diverging into two or more bounds of traffic. Figure 6.2(d) sketches a configuration of the signalized diverge, where the sign *S* indicates a traffic signal. The constraints are as follows:

$$y_{j_1-1} = \beta_{j_1}\cdot y_i$$

$$y_{j_2-1} = \beta_{j_2}\cdot y_i$$

$$\beta_{j_1}+\beta_{j_2} = \cdot 1$$

$$(\chi_8+\chi_9+\chi_{10})U^- \le y_i(t)-n_i(t) \le 0$$

$$(1+\chi_8+\chi_9-\chi_{10})U^- \le y_i(t)-Q_i(t) \le 0$$

$$(1+\chi_8-\chi_9+\chi_{10})U^- \le y_i(t)-Q_{j_1,\max}(t)/\beta_{j_1} \le 0$$

$$(1-\chi_8+\chi_9+\chi_{10})U^- \le y_i(t)-Q_{j_2,\max}(t)/\beta_{j_2} \le 0$$

$$(2-\chi_8-\chi_9+\chi_{10})U^- \le y_i(t)-\omega\cdot[N_{j_1,\max}-n_{j_1}(t)]/\beta_{j_1}] \le 0$$

$$(2-\chi_8+\chi_9-\chi_{10})U^- \le y_i(t)-\omega\cdot[N_{j_2,\max}-n_{j_2}(t)]/\beta_{j_2}] \le 0$$

$$\chi_8+\chi_9+\chi_{10} \le 2$$

$$\chi_9+\chi_{10} \le 1$$

The set of constraints is identical to those for non-signalized diverge cells except that $Q_{i,\max}(t)$ is replaced by $Q_i(t)$. The value of $Q_i(t)$ depends on the status of the signal phase associated with cell i and will be discussed in Section 3.2.7.

6.3.2.6 Constraints for signalized merge cells

Figure 6.2(e) is an example of traffic merge under signal control. According to the signal settings, these three streams of traffic entering the intersection are associated with three individual signal phases that conflict with each other. Therefore, practically there is only one stream of traffic entering the intersection at one time step. The constraints are thus as follows:

$$y_{j-1}(t) = y_{i1}(t) + y_{i2}(t) + y_{i3}(t)$$

Approach 1:

$$(\chi_{11} + \chi_{12})U^- \le y_{i_1}(t) - n_{i_1}(t) \le 0$$

$$(1 + \chi_{11} - \chi_{12})U^- \le y_{i_1}(t) - Q_{i_1}(t) \le 0$$

$$(1 - \chi_{11} + \chi_{12})U^- \le y_{i_1}(t) - Q_{j,\max}(t) \le 0$$

$$(2 - \chi_{11} - \chi_{12})U^- \le y_{i_1}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

Approach 2:

$$(\chi_{13} + \chi_{14})U^- \le y_{i_2}(t) - n_{i_2}(t) \le 0$$

$$(1 + \chi_{13} - \chi_{14})U^- \le y_{i_2}(t) - Q_{i_2}(t) \le 0$$

$$(1 - \chi_{13} + \chi_{14})U^- \le y_{i_2}(t) - Q_{j,\max}(t) \le 0$$

$$(2 - \chi_{13} - \chi_{14})U^- \le y_{i_2}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

Approach 3:

$$(\chi_{15} + \chi_{16})U^- \le y_{i_3}(t) - n_{i_3}(t) \le 0$$

$$(1 + \chi_{15} - \chi_{16})U^- \le y_{i_3}(t) - Q_{i_3}(t) \le 0$$

$$(1 - \chi_{15} + \chi_{16})U^- \le y_{i_3}(t) - Q_{j,\max}(t) \le 0$$

$$(2 - \chi_{15} - \chi_{16})U^- \le y_{i_3}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

where $Q_i(t)$ is to be discussed next.

6.3.2.7 Constraints for connection between signal and flow

At signalized intersections, the capacity flow of a cell depends on the status of the corresponding signal phase, because only when this phase turns green can the traffic propagates forward or makes a turn. The capacity flow satisfies the following statement:

$$\text{If } b(p) < t \le e(p), \text{ then } Q_i(t) = s \text{ ; otherwise, } Q_i(t) = 0,$$

where s is saturation flow rate;

$b(p)$ is the beginning of green phase p;

$e(p)$ is the end of green phase p.

The above if-then relationship can be translated into a system of equalities and inequalities by introducing two binary variables $z_1(p,t)$ and $z_2(p,t)$. The system is stated as follows:

$$-U \cdot z_1(p,t) + \varepsilon \le t - e(p) \le U \cdot [1 - z_1(p,t)]$$

$$-U \cdot z_2(p,t) \le b(p) - t \le U \cdot [1 - z_2(p,t)] - \varepsilon$$

$$z_1(p,t) + z_2(p,t) - z(p,t) = 1$$

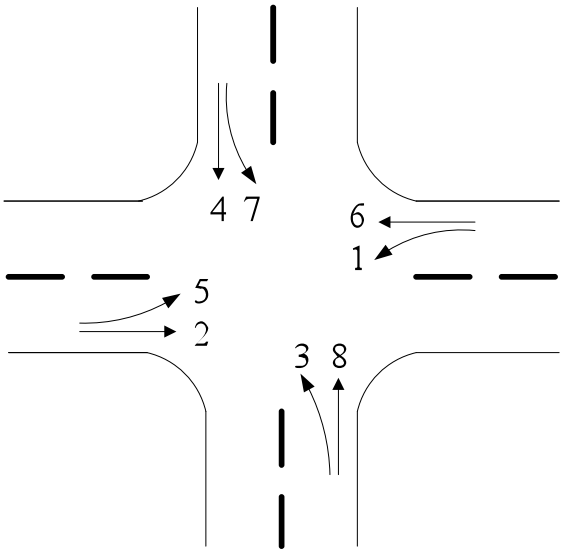
$$Q_i(t) = (z_1(p,t) + z_2(p,t) - 1) \cdot s$$

$$\sum_p (z_1(p,t) + z_2(p,t)) \le 2$$

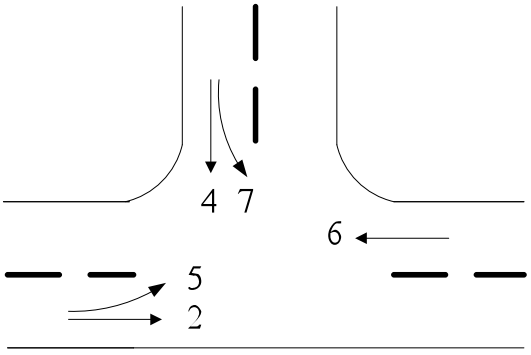
where U is a sufficiently large positive number and ε is an arbitrary small number. The last constraint ensures that there are at most two phases that can be green at the same time.

6.3.2.8 Constraints for signal phase sequence

The model intends to explicitly optimize phase sequences under the NEMA phasing structure. Two types of intersections are considered as shown in Figure 6.4: (a) Four-way intersection and (b) T-intersection.



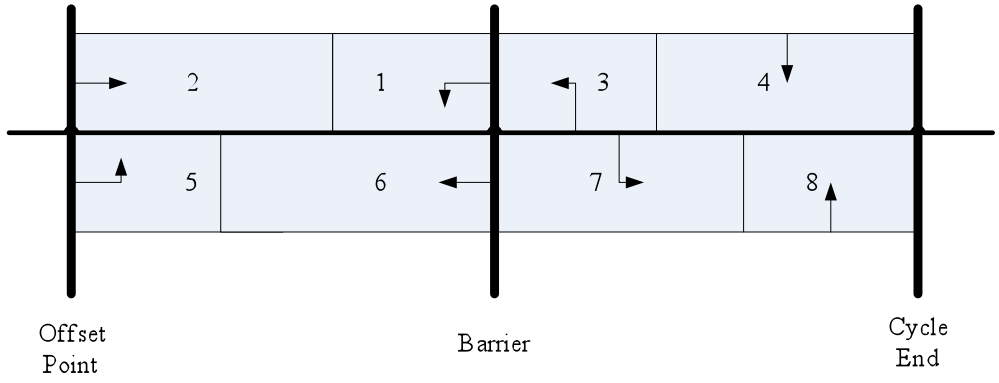
(a) Four-way intersection



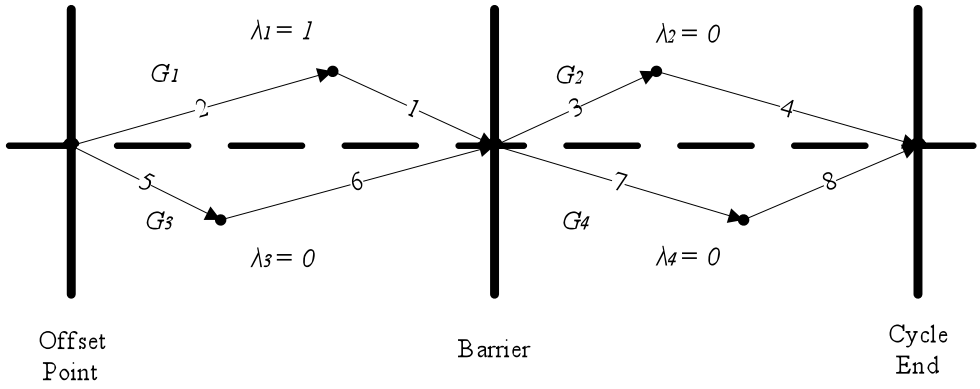
(b) T-intersection

Figure 6-4 Traffic Intersection Configurations

Four-Way Intersection: Figure 6.5(a) illustrates the standard NEMA phasing for a four-way intersection. With the barrier in the middle, the structure can be divided into four portions and phase sequence is determined within each portion.



(a) NEMA phasing



(b) Transformation of NEMA phasing

Figure 6-5 NEMA Phasing Structure for a Four-Way Intersection

A binary variable is introduced for each portion as shown in Figure 6.5(b). Consider phase 1 and 2 as an example. Let g denote the green time duration; o denote the offset point of each signal; l be the cycle length; k indicate the signal identification number; c represent the cycle identification number and h be the barrier time point. The following constraints are included for determining the phase sequence for phase 1 and 2:

$$b(k,"1",c) = \lambda_1 \cdot o(k) + \lambda_1 \cdot l \cdot (c-1) + (1-\lambda_1) \cdot e(k,"2",c)$$

$$e(k,"1",c) = b(k,"1",c) + g(k,"1")$$

$$b(k,"2",c) = (1-\lambda_1) \cdot o(k) + (1-\lambda_1) \cdot l \cdot (c-1) + \lambda_1 \cdot e(k,"1",c)$$

$$e(k,"2",c) = b(k,"2",c) + g(k,"2")$$

$$g(k,"1") + g(k,"2") = h(k)$$

It can be seen that when λ_1 equals 1, phase 1 starts at the offset point of the signal and phase 2 follows phase 1. It is also true reversely. Similarly, by introducing another three binary variables λ_2, λ_3 and λ_4 respectively, constraints can be constructed to determine phase sequence for the pairs of phase 3 and 4, 5 and 6, 7 and 8. Once the value of (λ_1, λ_2, λ_3, λ_4) is determined, the left-turn leading and lagging information can be obtained explicitly. Figure 6.5(b) presents a particular phase sequence corresponding to $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 0$ and $\lambda_4 = 0$.

T-Intersection: T-intersection can be modeled the same way as the four-way intersection but is much simpler. Figure 6.6 illustrates the NEMA phasing structure of a particular T-intersection where the only phase sequence needs to be determined is between phase 5 and 6. Therefore one binary variable λ_3 is introduced for the whole intersection. Correspondingly, only one set of constraints is needed for the entire structure, listed as follows:

$$b(k,"5",c) = \lambda_3 \cdot o(k) + \lambda_3 \cdot l \cdot (c-1) + (1-\lambda_3) \cdot e(k,"6",c)$$

$$e(k,"5",c) = b(k,"5",c) + g(k,"5")$$

$$b(k,"6",c) = (1-\lambda_3) \cdot o(k) + (1-\lambda_3) \cdot l \cdot (c-1) + \lambda_3 \cdot e(k,"5",c)$$

$$e(k,"6",c) = b(k,"6",c) + g(k,"6")$$

$$g(k,"5") + g(k,"6") = h(k)$$

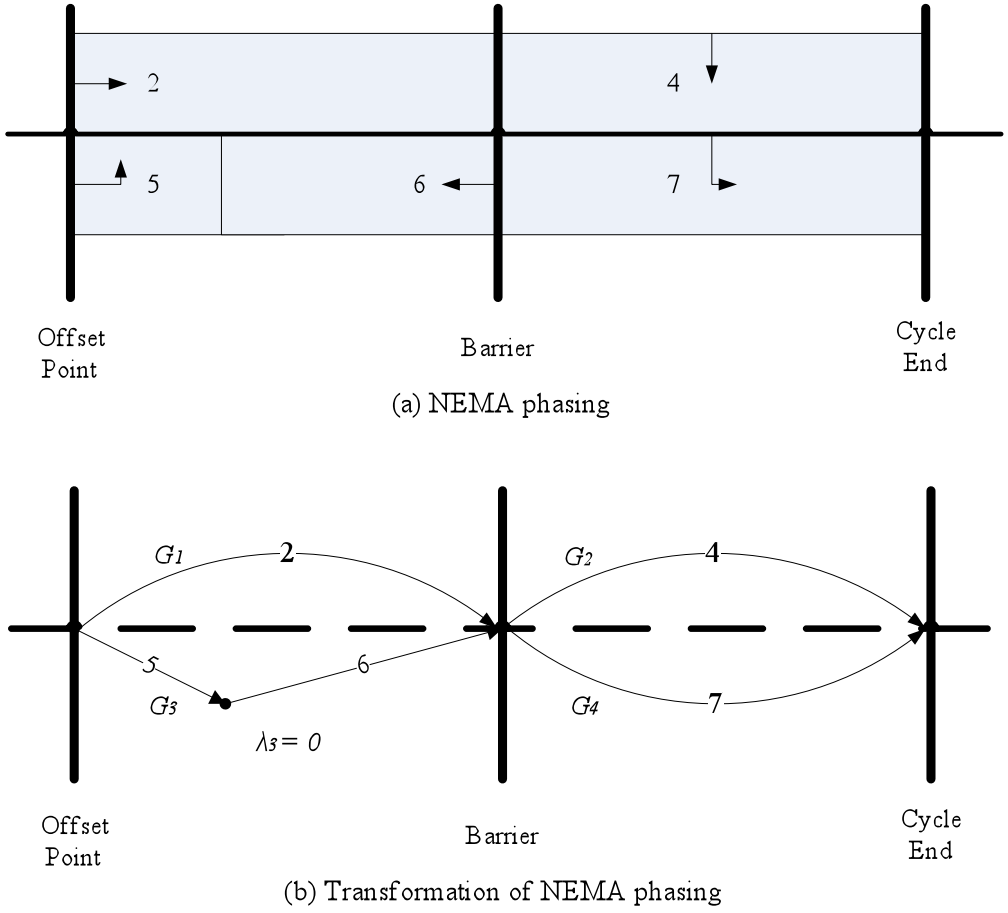


Figure 6-6 NEMA Phasing Structure for a T-Intersection

6.3.2.9 Model formulation

Given a particular network, the cell representation should be first constructed according to the geometry and signal setting. The cells are then classified into six categories and the corresponding set of constraints can be written for each cell as previously presented. The constraints comprise a linear system with integer variables. With the linear objective function to minimize total system delay, the optimization problem is a mixed-integer linear program. One portion of the optimal solution to the program specifies the signal

timing, denoted as a vector $(l^*, o^*, \lambda^*, g^*)$, where l^*, o^*, λ^* and g^* are vectors of optimal cycle length, offsets, phase sequences and green splits.

6.3.3 Stochastic Signal Optimization Model

In the above deterministic case, traffic demand is assumed to be fixed within the optimization horizon. However in reality, demand may varies significantly, like in Figure 1.1. We assume that the demand at each origin cell follows a certain stochastic distribution. To capture the joint stochastic distribution of traffic demands, also a set of scenarios $\Omega = \{1,2,3,\cdots,K\}$ is introduced. A typical scenario consists of demand realizations at all origin cells. More specifically, a scenario is a vector $d^k = (d_1^k, d_2^k, \cdots d_r^k \cdots)^T, \forall r \in O$.

For each demand scenario *k* and one particular feasible signal plan (l, o, λ, g), the total system delay can be computed, as described in the previous section 3.2. We denote the resulting delay as $L_k(l, o, \lambda, g)$. By applying the scenario-based approach, the robust signal coordination plan can be obtained by minimizing the following equation:

$$\min_{l,o,\lambda,g,\xi} Z_\alpha = \xi + \frac{1}{1-\alpha} \sum_{k=1}^{K} \pi_k \cdot \max\left(L_k(l, o, \lambda, g) - \xi, 0\right)$$

Each demand scenario requires a set of the constraints as discussed in the previous section, so the final stochastic optimization problem will include multiple sets of such constraints depending on the number of scenarios generated.

6.4 Numerical Examples

6.4.1 Simulation-Based Genetic Algorithm

The stochastic programming model formulated above is simple in structure but contains a large number of binary variables. Therefore, existing algorithms, such as branch and bound, are not able to solve it efficiently, particularly when the optimization horizon is long and the network size is large. We thus develop a simulation-based binary genetic algorithm (GA) to solve the model. Here the "simulation-based" means that the fitness function in the GA is evaluated through macroscopic simulation using CTM.

GAs have been widely used in different fields such as engineering, economics and physics to solve problems that are not analytically solvable or cannot be solved by traditional search methods. In the transportation literature, researchers have developed GA-based solution algorithms to solve problems including equilibrium network design, dynamic traffic assignment and second-best congestion pricing and traffic control problems. In Lo et al. [6-11], GA was used to solve the signal optimization problem.

GA is a global search technique. It starts from an initial group of randomly generated feasible solutions, and then employs operations like crossover and mutation to generate the new solution pools. The iteration continues until some criterion is satisfied, e.g., the maximum number of generation. The simulation-based GA proposed in this report follows the general framework of GA, and Figure 6.7 presents the flow chart of the algorithm. There are two loops: the outer loop for counting the number of generations while the inner is to track the number of individuals within each generation. Other core components of the algorithm will be discussed next.

6.4.1.1 Chromosome configuration

The chromosome is defined according to the decision variables, which include the cycle length, offsets, phase sequences and phase splits. Each chromosome defines a solution, and only the feasible solutions can be selected as the individuals in each generation. Figure 6.8 presents an example of the chromosome that represents a three-signal arterial. There are in total144 bits, of which the first six bits 1-6 represent the cycle length. A length of six binary numbers can represent a decimal value from 0 to 63. If traffic dynamic is modeled at two seconds per time step, then the cycle length can vary from 0 to 126 seconds. The rest bits are equally divided into three portions, 46 bits for each signal.

Consider the first signal. Bit 7, 8, 9, and 10 define the phase sequences for the signal as discussed in Section 3.2.8. Generally, if a bit has a value of 0, then the corresponding odd phase is activated before the even phase. Otherwise, the even phase comes first. A four-way intersection will require determining values in all four bits, while a T-intersection only needs one bit information. The next seven bits, i.e., 11-17, represent the offset for the signal. A seven-bit binary number can represent a decimal number from 0 to 127. Because an offset is expressed as a percentage of the cycle length in this report, one additional constraint on the binary number is in place to ensure the feasibility of the offset. Bits 18-24 represent the barrier point, which is also expressed as a percentage of the cycle length. Another additional constraint is required as well to ensure that the newly generated barrier point stays in the current cycle.

The next four clusters of bits represent four green times G_1, G_2, G_3 and G_4 (see Figure 6.5(b)), which are the green durations of the phases that lead in the respective portions. G_1 and G_3 are in percentage of the barrier time while G_2 and G_4 are in percentage of the difference between cycle length and barrier time.

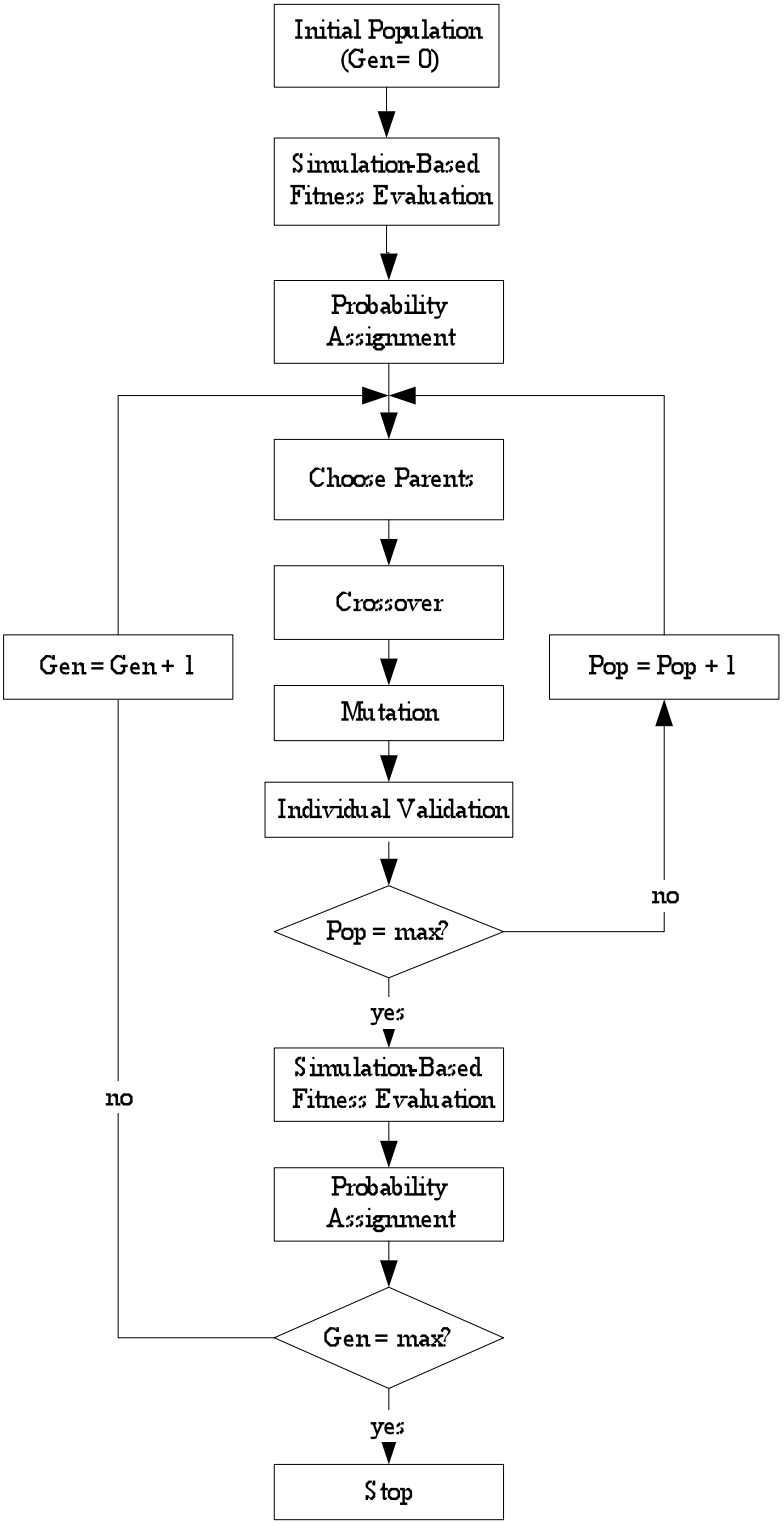


Figure 6-7 Flow Chart of the Simulation-Based GA

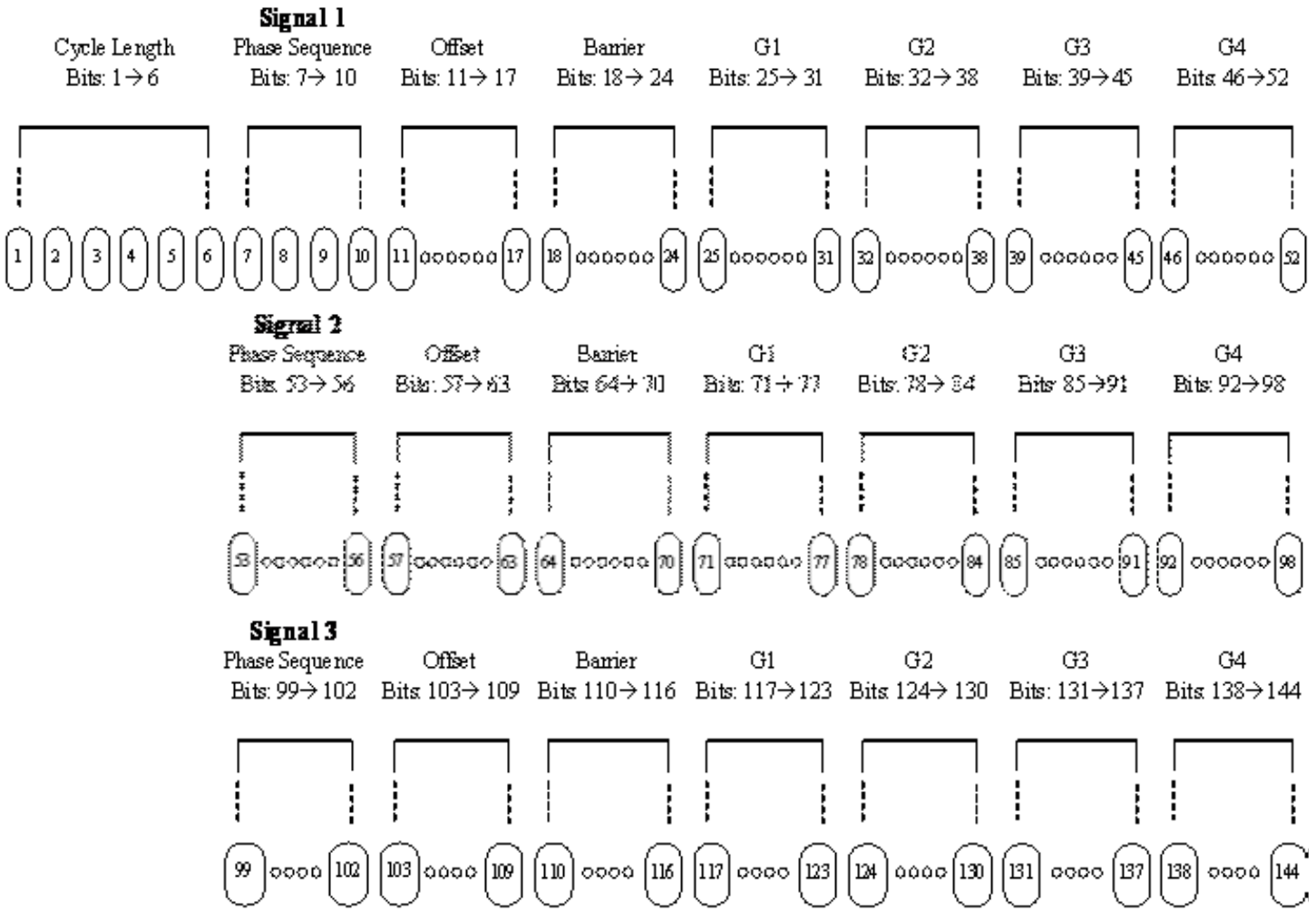


Figure 3.8 Configuration of the Chromosome

6.4.1.2 Fitness evaluation

For each generation, every individual needs to be evaluated so as to decide its priority to breed the next generation. Here each individual represents a specific signal plan and the fitness evaluation is to determine its corresponding mean excess delay. More specifically, for each individual signal plan, we run a macroscopic simulation based on CTM with all demand scenarios and calculate the corresponding system control delays. The mean excess delay can be computed and will be used to determine its priority for breading the next generation.

6.4.1.3 Probability assignment

Generally, the smaller the mean excess delay is, the larger probability the corresponding signal plan will be chosen to breed the next generation. To calculate these probabilities, we have tested a variety of fitness functions and the following two generally show good performance:

$$\begin{cases} \dfrac{1}{(CVAR)^5} \\ \dfrac{1}{\ln(CVAR)} \end{cases}$$

The crossover probabilities are calculated proportionally to the fitness function value.

Crossover

Crossover is the main procedure to generate new chromosomes. To increase diversity, we use multi-point crossover, developed according to the chromosome structure, other than using one-point crossover. After the selection of two parents according to the crossover probability, several crossover points will be randomly generated but ensure that one is among the first six bits, which influences the cycle length, and one for each signal, which may change the setting for each signal. Therefore, if there are n signals, there will be $n+1$ crossover points in total.

6.4.1.4 Mutation

Each crossover operation will generate two offspring and the mutation operation is subsequently conducted. Mutation randomly changes the value of the bit value in the chromosomes to increase the diversity in the population, so that the GA will have the chance to find a better solution rather than stop at one local optimum. The mutation rate to be used is 5‰.

6.4.1.5 Individual validation

New individual produced through crossover and mutation operations may not be appropriate, i.e., the corresponding timing plan may not be technically feasible. Therefore, additional constraints need be set to ensure the validity of each individual. Our algorithm mainly checks the followings:

- The individual newly added will not repeat any individual contained in the population to maintain the diversity of the population;
- An individual with a cycle length smaller than a certain value will not be considered to ensure the cycle length to be in a reasonable range;
- The offsets must be smaller than the cycle length;

- The barrier points must be in the corresponding cycle;
- And each signal phase maintains a certain minimum green.

6.4.2 Numerical Example I

6.4.2.1 Test network and demand data

The first numerical experiment is carried out on an artificial arterial with three intersections, whose cell representation is shown in Figure 6.9. The design speed limit is 35mph, which is approximately equivalent to 50 feet/second. Since traffic dynamics is modeled in a resolution of two seconds per time step, 100 feet is the cell length for all 117 cells. We implement the stochastic signal timing model under both uncongested and congested traffic conditions. The low demand in Table 6.2 is for the uncongested situation while the high demand is for the congested cases. The turning percentages at each signal are also given in the table.

Table 6-2 Traffic Data for Three-Node Network

Traffic Volume		Westbound	Northbound	Eastbound	Southbound
Signal 1	Low Demand	--	**162 ±27***	**1223 ±180**	**125 ±18**
	High Demand	--	**462 ±45**	**1523 ±270**	**325 ±36**
	Left	0.1445	0.2876	0.0291	0.6637
	Through	0.5772	0.1373	0.8960	0.2389
	Right	0.2783	0.5752	0.0750	0.0973
Signal 2	Low Demand	--	**117 ±18**	**--**	**169 ±27**
	High Demand	--	**317 ±36**	**--**	**269 ±45**
	Left	0.0061	0.7664	0.0778	0.1923
	Through	0.9703	0.0935	0.7243	0.0577
	Right	0.0237	0.1402	0.1979	0.7500
Signal 3	Low Demand	**75 ±18**	**--**	**--**	**400 ±45**
	High Demand	**1075 ±180**	**--**	**--**	**400 ±45**
	Left	--	--	--	0.6000

	Through	0.8000	--	1.0000	--
	Right	0.2000	--	--	0.4000

*: **a ± b** means that demand is uniformly distributed in the interval (a-b, a+b) vehicles per hour.

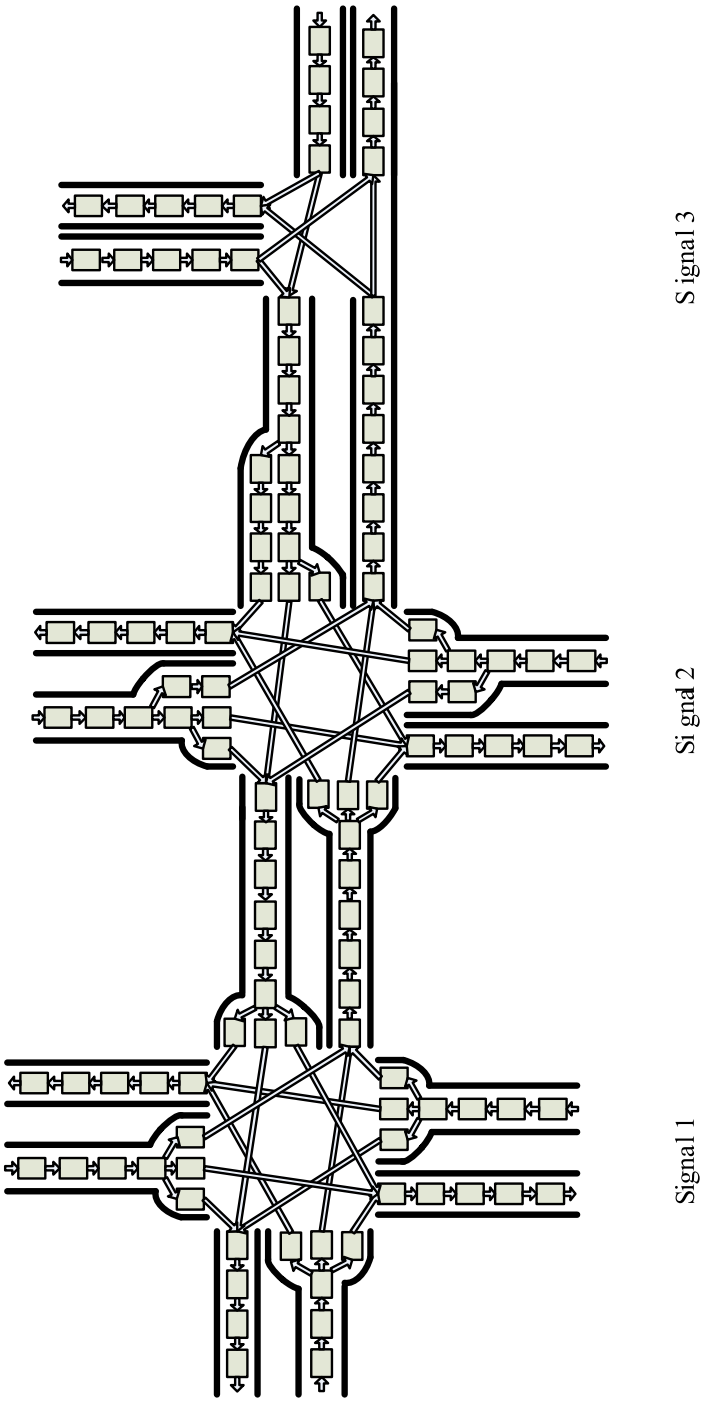
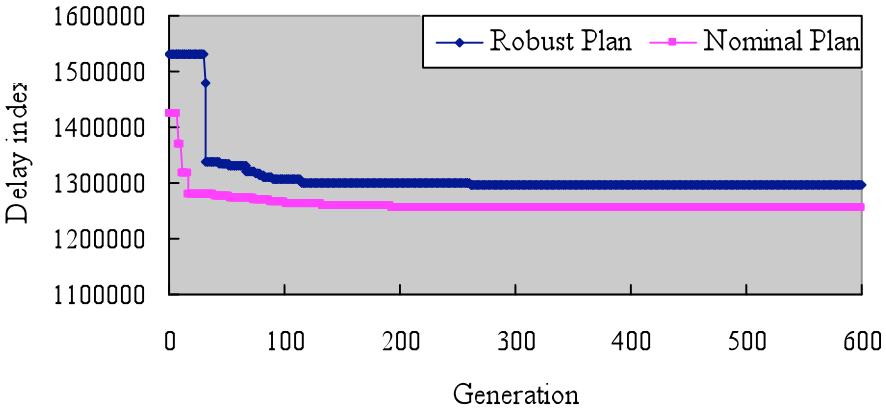


Figure 3.9 Cell Representation of the Three-Node Network

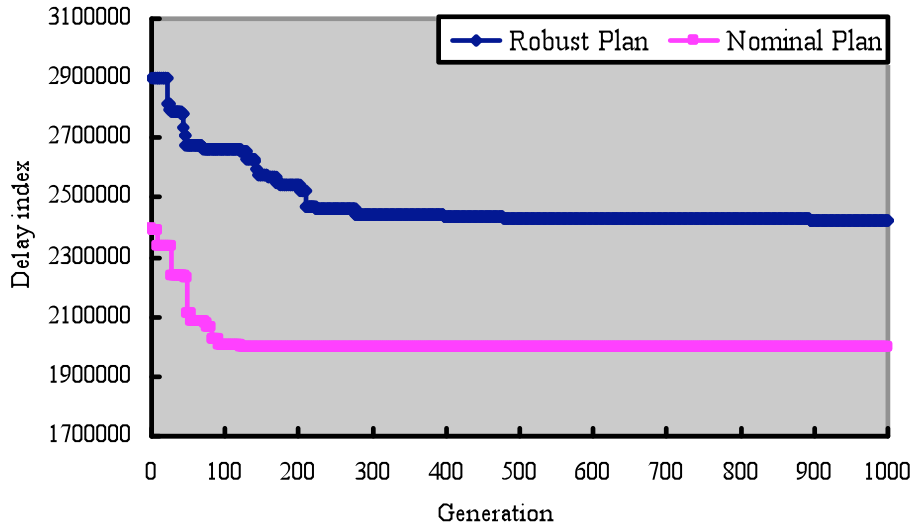
6.4.2.2 Plan generation

For the comparison purpose, two plans are generated under both uncongested and congested conditions: one is called as robust plan, which is generated by solving the stochastic signal timing model with the demand scenarios created from the uniform distributions shown in Table 6.2; the other is called nominal plan, derived by solving the deterministic signal timing model with the mean demand. Because the two fitness functions have similar convergence speed and generate timing plans with similar performance, we only report the result from using the log-form fitness function.

According to the convergence performance of the algorithm, we set the maximal number of generations to 600 and 1000 for the uncongested and congested case respectively. Figure 6.10 presents the convergence tendency of the algorithm in both cases. The algorithm converges faster in the uncongested case, particularly in the early stage of the iterations.



(b) Uncongested case



(a) Congested case

Figure 6-8 Convergence of GA under Both Traffic Conditions

Table 6.3 presents the resulting signal plans. The phase sequence is given in form of binary vector while others are decimal numbers in the unit of second. The minimum green for each phase is set as four seconds. P1 to P8 stand for the phases in NEMA phasing.

Table 6-3 Signal Plans for Three-Node Network

Uncongested Case		Cycle Length	Phase Sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust Plan	S1	80	(1, 0, 1, 1)	0	4	58	4	14	4	58	8	10
	S2	80	(0, 1, 1, 0)	76	6	58	4	12	4	60	6	10
	S3	80	(1, 0, 0, 1)	26	--*	50	--	30	32	18	30	--
Nominal Plan	S1	80	(1, 1, 0, 1)	0	4	62	4	10	22	44	4	10
	S2	80	(1, 0, 0, 1)	16	4	46	6	24	6	44	4	26
	S3	80	(0, 1, 0, 0)	28	--	50	--	30	22	28	30	--
Congested Case		Cycle Length	Phase Sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust Plan	S1	108	(1, 1, 1, 1)	0	8	78	8	14	18	68	8	14
	S2	108	(0, 0, 0, 0)	98	14	72	8	14	28	58	8	14
	S3	108	(1, 0, 1, 0)	16	--	68	--	40	32	36	40	--
Nominal Plan	S1	80	(1, 1, 0, 1)	0	4	66	4	6	20	50	4	6
	S2	80	(1, 0, 1, 0)	10	4	46	4	26	4	46	6	24
	S3	80	(0, 0, 1, 1)	4	--	76	--	4	34	42	4	--

*: phase not applicable.

6.4.2.3 Plan evaluation

We compare the robust and nominal signal plans using the microscopic CORSIM simulation, and the system delay is selected as the performance measure. Figure 6.11 is a snapshot of the CORSIM network. In the simulation, demand scenarios are obtained by sampling the uniform distributions provided in Table 6.3. Table 6.4 summarizes the simulation result. It can be seen that the robust timing plans reduce the mean excess delay by 28.68% in the uncongested case and 7.46% in the congested case. In both cases, it also improves the average delay across all demand scenarios by over 20%.

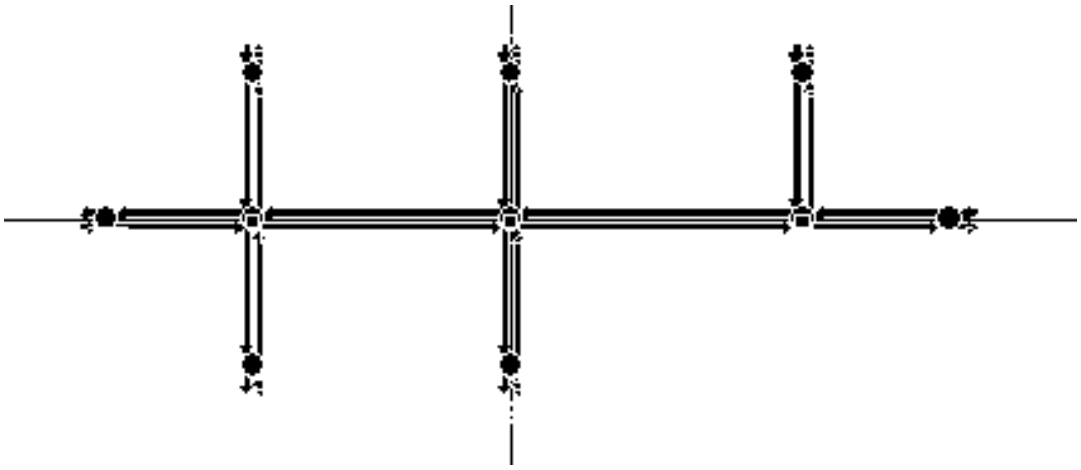


Figure 6-11 A Snapshot of the Three-Node Network in CORSIM

Table 6-4 CORSIM Result for Three-Node Network

Taffic Condition	Index Measure	Robust Plan	Nominal Plan	Change
Uncongested Case	Mean Delay	13.15*	17.70	-25.69%
	Mean Excess Delay	14.02	19.66	-28.68%
Congested Case	Mean Delay	79.16	98.97	-20.06%
	Mean Excess Delay	106.58	115.18	-7.46%

*: in vehicle hours.

6.4.3 Numerical Example II

6.4.3.1 Test network and demand data

The second numerical experiment is carried out on a stretch of El Camino Real in the San Francisco Bay Area of California, starting from Crystal Springs Rd to 5th Ave. Figure 6.12 is the cell representation of the arterial. The speed limit on the major street is 35 mph or 50 feet per second while 25 mph or 36 feet per second on the side streets. Because traffic dynamics is modeled second by second, the cell length for the major and side

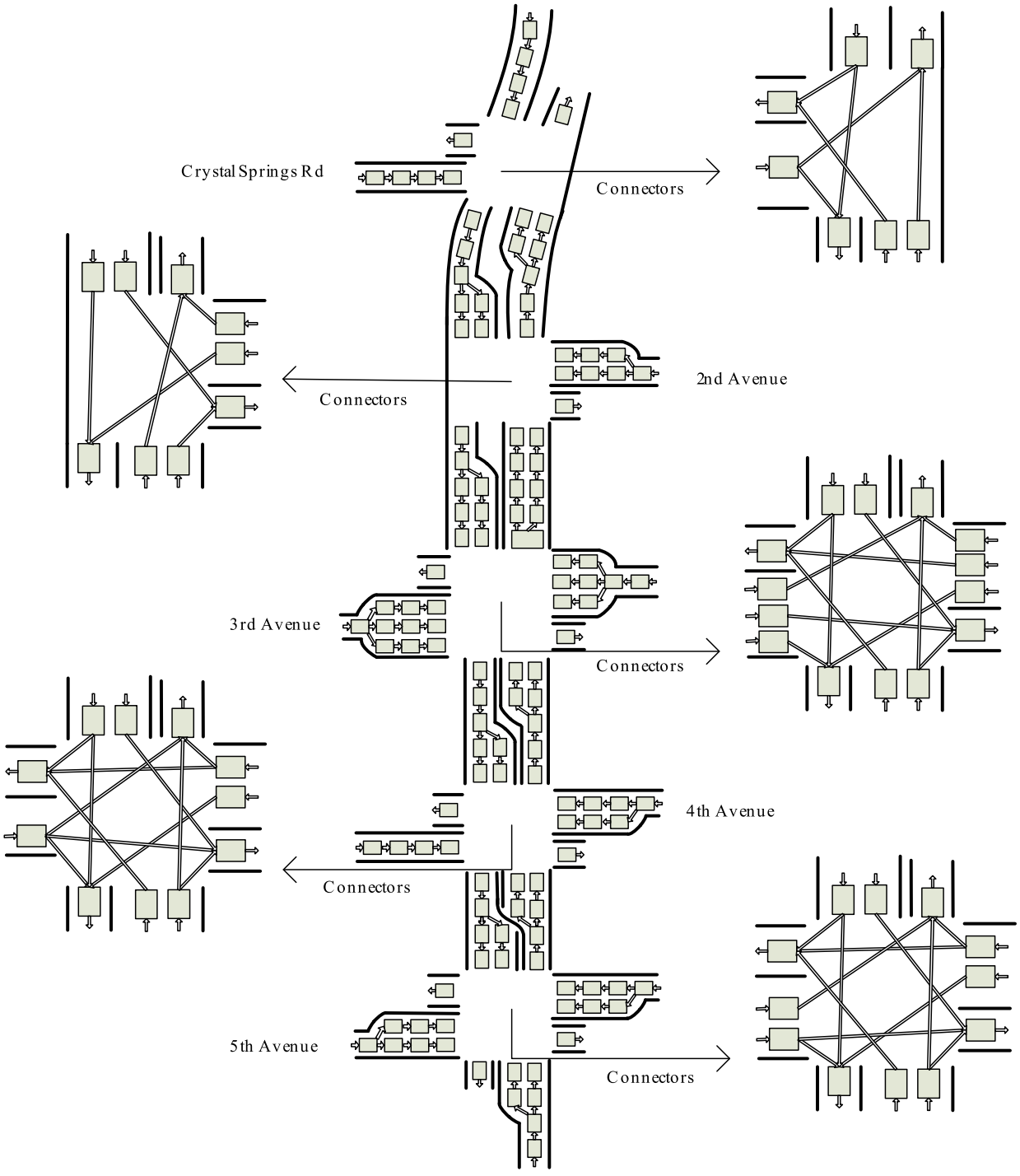


Figure 6-9 Cell Representation of El Camino Real Arterial

streets is 50 and 36 feet respectively. Traffic demand data were collected from loop detectors for peak hours in a duration of 10 working days in July 2008. Table 6.5 provides a summary of the flow data and the turning proportions at the intersections.

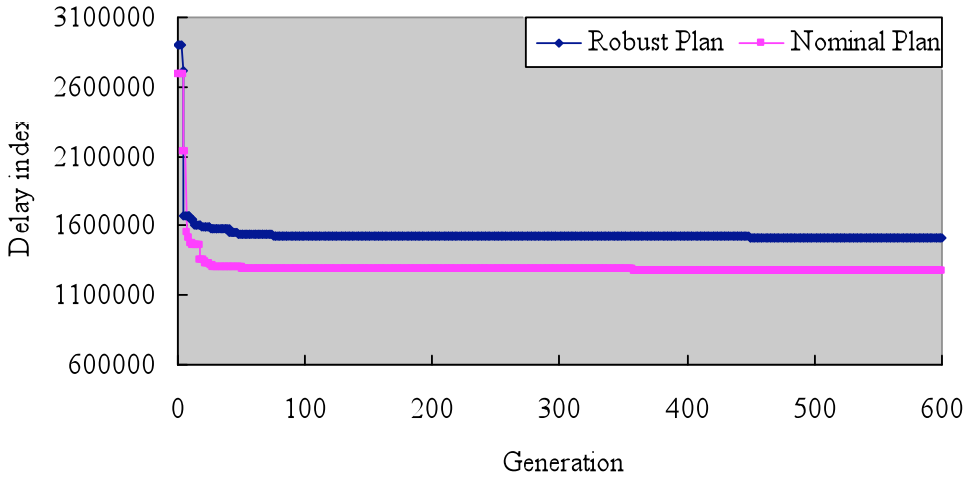
6.4.3.2 Plan generation

The observed flow rates are used directly as demand scenarios with equal probability of occurrence, to generate the robust plans by solving the stochastic programming model using the simulation-based GA approach. For comparison, a nominal plan is generated by solving the deterministic model with the mean demands presented in Table 6.5. Both robust and nominal plans are generated after 600 generations. Figure 6.13 shows the convergence of the GA with both 5^{th}-form and log-form fitness functions. It can be observed the 5^{th}-form fitness function converges faster than the log-form counterpart.

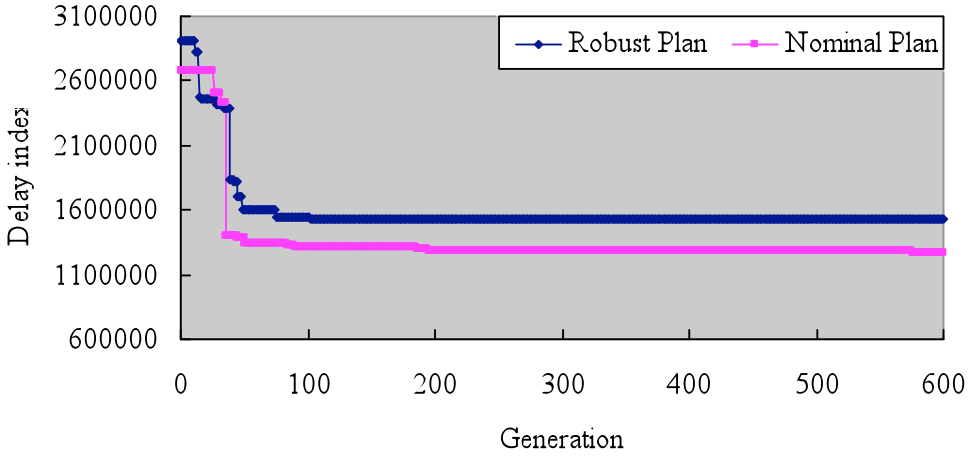
Table 6-5 Traffic Data for El Camino Real Arterial

Traffic Volume		Westbound	Northbound	Eastbound	Southbound
Crystal Spring	Demand Mean	-- *	--	**179**	**1112**
	Demand SD	--	--	**13**	**44**
	Left	--	0.0603	0.7205	--
	Through	--	0.9397	--	0.8369
	Right	--	--	0.2795	0.1631
2nd Ave	Demand Mean	**174**	--	--	--
	Demand SD	**18**	--	--	--
	Left	0.6647	--	--	0.1118
	Through	--	0.7818	--	0.8882
	Right	0.3353	0.2182	--	--
3rd Ave	Demand Mean	**270**	**--**	**238**	--
	Demand SD	**22**	**--**	**18**	--
	Left	0.4545	0.0600	0.1911	0.0247
	Through	0.2557	0.8679	0.4837	0.8983
	Right	0.2898	0.0722	0.3252	0.0770
4th Ave	Demand Mean	**528**	**--**	**101**	--
	Demand SD	**32**	**--**	**10**	
	Left	0.3351	0.0237	0.1272	0.1193
	Through	0.2990	0.8732	0.6301	0.8593
	Right	0.3660	0.1031	0.2428	0.0214
5th Ave	Demand Mean	**219**	**1443**	**184**	--
	Demand SD	**10**	**82**	**18**	--
	Left	0.3853	0.0447	0.2889	0.0312
	Through	0.4391	0.9407	0.5804	0.8982
	Right	0.1756	0.0147	0.1307	0.0706

*: -- means data not applicable or available.



(a) 5th-form fitness function



(b) Log-form fitness function

Figure 6-10 Convergence of GA with Both Fitness Functions

Table 6.6 presents the signal plans generated from both fitness functions, and their performances will be compared next. The minimum green for each phase is set as eight seconds.

Table 6-6 Signal Plans for El Camino Real Arterial

5^{th}-Form Fitness Function		Cycle Length	Phase Sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust Plan	S1	90	(0, 1, 1, 1)	0	11	58	8	13	47	22	12	9
	S2	90	(0, 0, 1, 0)	15	16	34	26	14	34	16	9	31
	S3	90	(0, 1, 1, 1)	13	24	33	23	10	38	19	17	16
	S4	90	(0, 0, 0, 0)	67	32	18	40	--	--	50	--	40
	S5	90	(0, 0, 0, 0)	0	--	35	--	55	15	20	55	--
Nominal Plan	S1	112	(0, 0, 0, 0)	0	14	76	14	8	78	12	9	13
	S2	112	(1, 0, 0, 1)	34	32	55	17	8	70	17	10	15
	S3	112	(1, 0, 0, 1)	38	26	43	25	18	57	12	24	19
	S4	112	(0, 0, 1, 1)	83	29	21	62	--	--	50	--	62
	S5	112	(1, 1, 0, 0)	43	--	68	--	44	21	47	44	--
Log-form fitness function		Cycle Length	Phase Sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust Plan	S1	94	(0, 0, 0, 0)	0	8	63	15	8	62	9	10	13
	S2	94	(0, 0, 1, 1)	24	16	35	35	8	35	16	27	16
	S3	94	(1, 0, 0, 1)	60	21	32	30	11	26	27	28	13
	S4	94	(1, 1, 1, 0)	13	29	10	55	--	--	39	--	55
	S5	94	(1, 0, 0, 0)	69	--	55	--	39	44	11	39	--
Nominal	S1	112	(1, 0, 1, 1)	0	8	80	16	8	79	9	14	10
	S2	112	(0, 1, 1, 0)	12	11	60	3	33	45	26	31	10

Plan	S3	112	(1, 1, 1, 0)	26	26	45	13	28	62	9	19	22
	S4	112	(1, 1, 1, 1)	10	32	22	58	--	--	54	--	58
	S5	112	(0, 0, 1, 1)	50	--	27	--	85	19	8	85	--

6.4.3.3 Plan evaluation

The comparison is also conducted via microscopic simulation with demand profiles randomly generated based on Table 6.5 assuming truncated normal distributions. Figure 6.14 is a snapshot of the CORSIM network for the corridor. Table 6.7 presents the CORSIM simulation result. The traffic condition is very congested through the whole simulation period. It can be seen that the robust plans outperform the corresponding nominal plan, with the mean delay reduced by 23.69% and 17.82%, and the mean excess delay reduced by 22.80% and 17.34%. It demonstrates that the robust plans perform much better against high-consequence scenarios. As a side effect, the average performance is also improved. Although the 5th-form fitness function leads to a faster convergence, it does not improve the performance as much as the log-form fitness function does.

Table 6-7 CORSIM Result for El Camino Real Arterial

Fitness Function	Index Measure	Robust Plan	Nominal Plan	Change
Log-Form	Mean Delay	234.36	307.13	-23.69%
	Mean Excess Delay	240.92	312.08	-22.80%
5^{th}-Form	Mean Delay	228.02	277.46	-17.82%
	Mean Excess Delay	232.30	281.02	-17.34%

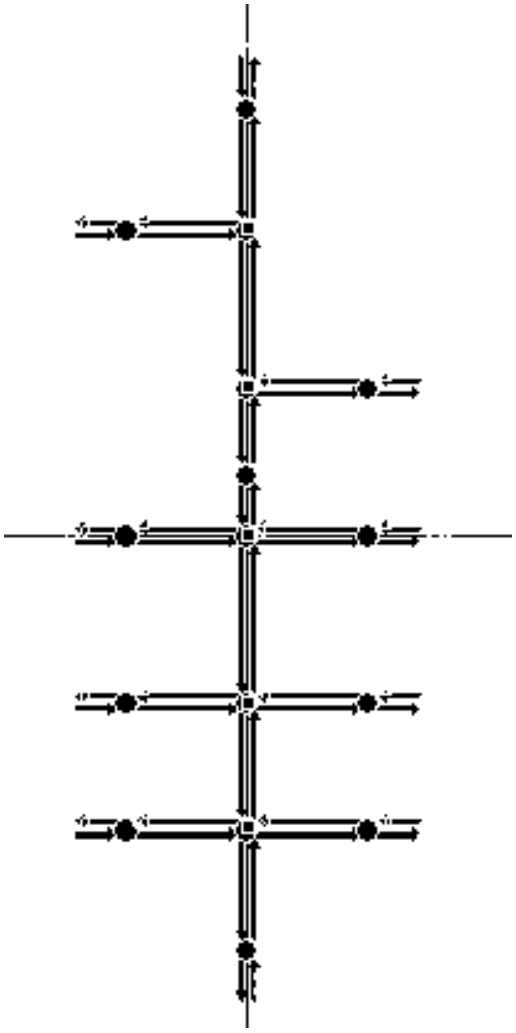


Figure 6-11 A Snapshot of the EL Camino Real Arterial in CORSIM

6.5 Reference

6-1 Webster, F. V. *Traffic Signal Settings*. Road Research Technical Paper, No. 39, Her Majesty's Stationary Office, London, U.K., 1958.

6-2 Robertson, D. I. and Bretherton, R. D. Optimizing Networks of Traffic Signals in Real-Time: the SCOOT Method. *IEEE Trans. Vehicular Tech*, Vol.40, 1991, pp.11-15.

6-3 Gartner, N. H. Development and Implementation of an Adaptive Control Strategy in a Traffic Signal Network: the Virtual-Fixed-Cycle Approach. In *Proceeding of 15th International Symposium on Transportation and Traffic Theory*, 2002, pp.137-155.

6-4 Heydecker, B. Uncertainty and Variability in Traffic Signal Calculations. *Transportation Research*, Part B, Vol.21, 1987, pp. 79-85.

6-5 Ribeiro, P. C. M. Handling Traffic Fluctuation with Fixed-Time Plans Calculated by TRANSYT. *Traffic Engineering and Control*, Vol.35, 1994, pp. 365-366.

6-6 Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research*, Part B, Vol. 28, 1994, pp. 269-287.

6-7 Daganzo, C. F. The Cell Transmission Model, Part II: Network Traffic. *Transportation Research*, Part B, Vol. 29, 1995, pp. 79-93.

6-8 Lin, W. and Wang, C. An Enhanced 0-1 Mixed Integer LP Formulation for Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, 2004, pp. 238-245.

6-9 Lo, H. A Novel Traffic Signal Control Formulation. *Transportation Research*, Part A, Vol. 44, 1999, pp. 436-448.

6-10 Lo, H. A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plan. *Transportation Science*, Vol. 35, 2001, pp.1 48-164.

6-11 Lo, H., Chang, E. and Chan. Y. C. Dynamic Network Traffic Control. *Transportation Research*, Part A, Vol. 35, 2001, pp. 721-744.

6-12 Pavlis, Y. and Recker, W. A Mathematical Logic Approach for the Transformation of the Linear Conditional Piecewise Functions of Dispersion-and-Store and Cell Transmission Traffic Flow Models into Mixed-Integer Form. *Transportation Science*, Vol. 43, 2009, pp. 98-116.

7 Concluding Remarks

A traffic data collection system based-on the iDEN wireless network has been developed, lab tested and preliminarily tested in the field. The objective of the system is to provide a cost-effective and easy-to-maintain system that could still reliably provide traffic data over the wireless link. The mobile wireless network has its inherent characteristics of less reliable than the wired network. We have built an adaptive flow control layer over the wireless TCP communication to address the occasional outage problem. The validity of the method is in the fact that the collection second by second traffic signal status data does not necessarily achieve 0% loss rate. So with allowing some minimized data loss, we could control the data flow to avoid exceeding the allowed data rate and thus leading to major outage. The system is able to continuously provide over 2.68kbps upload data rate per remote handset for over 95% of the time, i.e., one remote handset could deliver data fetched from signal controllers at a period of 200ms continuously with about 5% of data loss rate. On the cost side, one set of client hardware costs ~$70 and $10 monthly, while covering up to 8 local signal controllers. These combined features, low cost and high performance, make the system a unique solution for traffic data collection.

This report has described the development of an arterial performance measurement method that is based on the signal infrastructure data collected at PT^2 Lab, U.C. Berkeley. The performance of the proposed model is illustrated by using a simulation study. The six-signal simulation network covers both heavily congested and light traffic intersections. The proposed model works well at both the intersection level and the arterial level. Estimation errors of travel time, number of stops and travel time reliability are insignificant.

The findings of the study together with the data collection means developed by PT^2 Lab provide a cost-effective way to achieve an arterial performance measurement system. The data and analysis results will support transportation researchers on various research topics such as traffic control and operations; help planners and local agencies on daily management and system monitoring; and provide travelers real-time information when scheduling their trips.

For the next step, we will further calibrate, validate and demonstrate our model by using field data and conducting field experiments. Some given parameters in this study, such as the demand factor α and time window T_W when queue spillback happens, average deceleration and acceleration rates, free flow speed, and turning ratios, should be calibrated or measured based on field data and/or observations. Effective adaptive models should be developed to dynamically estimate parameters such as saturation flow and start-up lost time. Sensitivity analysis on some of the key parameters will be studied. Moreover, the model to address the over-saturated scenarios will be developed and validated by the simulation network and then the field data.

This report has presented a general approach for robust signal optimization under demand uncertainty or flow fluctuations. The approach has been demonstrated in two different settings. The first one deals with the problem of synchronization of actuated signals along arterials along arterials. The formulation is a mixed integer linear program easily solvable using the state-of-the-art solvers. The computational time only increases polynomially as the number of scenarios increases. The approach can be used to either design a new coordination plan for implementation or fine-tune the plan offline after implementation. In the latter case, the specification of scenarios is an easy task with the archived signal status data. One may randomly select 50 to 200 red time realizations from the data and assume equal probability of occurrence. To design a new coordination plan where the

distributions of red times are normally unknown, we suggest specifying 50 to 200 scenarios as the points that equally divide the red-time intervals into K+1 segments and assume equal probability of occurrence of $1/K$. The suggestion is based on our observation from the numerical experiments that the robust formulation is not overly sensitive to the specification of scenarios. Even with biased scenarios, the formulation may still produce meaningful robust plans.

The other demonstration is to optimize the signal settings including the cycle length, green splits, offset points and phase sequences in an integrated manner, taking into account the day-to-day demand variations or uncertain further demand growth. Considering a large number of binary variables in the formulation, we have developed a simulation-based GA to solve the problem. It should be mentioned that the setting of the GA-based algorithm, such as the fitness function, may influence the quality of the final plan and the convergence speed. Numerical experiments are needed to fine-tune the setting. We also note that the simulation-based model is broadly applicable, particularly when the objective function is difficult or time-consuming to evaluate.

The robust timing plans resulted from both models have been demonstrated in numerical tests to perform better against high-consequence scenarios without losing optimality in the average sense. Although the robust signal timing approach is applicable more widely, this report has been focused on timing models for pre-timed arterials. Future study can be conducted to expand the proposed models for more sophisticated corridors and grid networks.