STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

ADA Notice

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

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA06-0703		

4. TITLE AND SUBTITLE

Toward Deployment of Adaptive Transit Signal Priority (ATSP)

5. REPORT DATE

October 1, 2008

6. PERFORMING ORGANIZATION CODE

7. AUTHOR

Meng Li, Kun Zhou, Yafeng Yin, Chin-Woo Tan, Wei-Bin Zhang, Sonja Sun, Kai Leung, Jam

8. PERFORMING ORGANIZATION REPORT NO.

UCB-ITS-PRR-2008-24

9. PERFORMING ORGANIZATION NAME AND ADDRESS

California PATH Program
University of California, Berkeley
1357 46th St.
Richmond, CA 94804

10. WORK UNIT NUMBER

11. CONTRACT OR GRANT NUMBER

65A0161 TO 5404

12. SPONSORING AGENCY AND ADDRESS

California Department of Transportation
Division of Research and Innovation
P.O. Box 942873, MS 83

13. TYPE OF REPORT AND PERIOD COVERED

Final Report
11/21/2003 to 06/30/2005

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document reports the continuous efforts on the development and implementation of an Adaptive Transit Signal Priority (ATSP) system. The ATSP system has three distinguished features, including: 1) providing priority to transit vehicles while making a tradeoff between bus delay savings and the impacts on the rest of the traffic, 2) utilizing existing advanced vehicle location (AVL) and communications system (ACS) already instrumented on buses to continuously monitor bus locations and predict bus arrival times to intersections and to request signal priority, and 3) building upon closed-loop signal control systems with 170 E controllers. These features allow ATSP to have potential for wide-scale implementation. This report describes the development of ATSP algorithms, the field test results of the prototype ATSP system and the feasibility analysis for utilizing existing transit communication for ATSP.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Transit Signal Priority, Bus Signal Priority, Signal Optimization, Adaptive,, ATSP, AVL, advanced vehicle location, ACS, priority, monitor, bus, locations, intersections, closed-loop, signal control systems, wide-scale implementation, feasibility analysis.	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	161	

DISCLAIMER STATEMENT

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

Toward Deployment of Adaptive Transit Signal Priority Systems

Meng Li, et al.

California PATH Research Report
UCB-ITS-PRR-2008-24

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 5404

October 2008

ISSN 1055-1425

Toward Deployment of Adaptive Transit Signal Priority Systems

Final Report for **Task Order 5404**

Prepared by:
California PATH
University of California, Berkeley
And
California Department of Transportation
In collaboration with
San Mateo Transit District

Acknowledgements

This work was performed by the California PATH Program at the University of California at Berkeley 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 thank Don Dean, Ron Deleon and Greg Larson of Caltrans' Division of Research and Innovation, Paul Chiu and Lindy Cabugao of Caltrans District 4, Frank Burton of San Mateo County Transit District for their support and advice during the project.

Author List

University of California, Berkeley:
Meng Li
Kun Zhou
Yafeng Yin
Chin-Woo Tan
Wei-Bin Zhang

Division of Research and Innovation, Caltrans:
Sonja Sun

Division of Traffic Operation, Caltrans:
Kai Leung

District 4, Caltrans:
James Lau

Abstract

This document reports the continuous efforts conducted by California PATH Program on the development and implementation of an Adaptive Transit Signal Priority (ATSP) system. The ATSP system has three distinguished features, including: (1) providing priority to transit vehicles while making a tradeoff between bus delay savings and the impacts on the rest of the traffic, (2) utilizing existing AVL/communication system already instrumented on buses to continuously monitor bus locations and predict bus arrival times to intersections and to request signal priority, and (3) building upon closed-loop signal control systems with 170E controllers. These features allow ATSP to have potential for wide-scale implementation. This report describes the development of ATSP algorithms, the field testing results of the prototype ATSP system and the feasibility analysis for utilizing existing transit communication for ATSP.

Keywords: Transit Signal Priority, Bus Signal Priority, Signal Optimization

Executive Summary

Transit signal priority (TSP), a critical technology for bus rapid transit systems and for improving traditional transit services, has been implemented in many cities across the country. TSP deployments have demonstrated positive effects on reducing bus intersection delays and improving service schedule adherence. However, concerns have been frequently raised that TSP operations may interrupt the normal operation of signal control and thus increase delays to other traffic, particularly those served by non-prioritized phases (often minor phases, including cross-street and main street left turns).

On the other hand, active TSP systems normally adopt selective vehicle detection means that sense the presence of an approaching bus only at a fixed location, and thus have difficulty obtaining exactly bus arrival time to an intersection, if the detection means are placed far from the intersection. Many transit agencies have installed or procured Global Positioning System (GPS)-based Automatic Vehicle Location (AVL) system to their fleet, so the deployment of TSP upon GPS/AVL system will be cost-effective, because it allows all buses instrumented with GPS/AVL system to become TSP capable without requiring any additional equipment on buses.

In light of the above considerations, California PATH, in collaboration with Caltrans Headquarters and District 4 has been developing an adaptive TSP (ATSP) concept. The main features of the ATSP concept include:

- Providing priority to transit vehicles if warranted while trying to make a tradeoff between bus delay savings and the impacts on the rest of the traffic
- Making real time decisions adaptive to the movements of transit vehicles, traffic conditions and signal status
- Utilizing GPS/AVL system instrumented on buses to continuously monitor bus locations and predict bus arrival times to intersections
- Built-upon closed-loop signal control systems with 170E controllers, which may have potential for wide-scale implementation

The development and implementation of a prototype ATSP system were conducted under PATH Project RTA 65A0026. The project has successfully validated the GPS-based ATSP concept, and demonstrated preliminarily the effectiveness of the ATSP system. As a continuation of the above research, this project addresses the issues that still remain for eventual deployment of the ATSP system. First, a more advanced priority request generator (PRG) has been developed in order to make an *explicit* and *graceful* tradeoff between bus delay savings and traffic delays. Second, the ATSP concept takes advantage of the infrastructure of deployed GPS/AVL systems. However, the commercial GPS/AVL systems used by most of transit agencies have limited communication resources. Each bus in a fleet is sampled every two minutes. Bus position sampling at this low rate most likely would not support a timely request for priority.

In the improved ATSP system, the TTA predictor has been improved by adding an observer to smooth out measurement noises and applying an adaptive mechanism on historical model to have its parameters dynamically updated based on real-time bus movement. An arrival time flow prediction model, based on an adaptive recursive least-squares method, has been developed to provide PRG an estimation short-term traffic arrival flow, so that a more advanced PRG that is able to adaptively and optimally select either early green or green extension and determine the corresponding signal timing strategies for the TSP operation. The objective that guides the decision is to make a tradeoff between bus intersection delays and other traffic delay. The level of the tradeoff can be adjusted via a weighting factor and should be determined through negotiations among the stakeholders on how much preference the transit operation should be given. Some case studies, hardware-in-the-loop simulation tests, and a field operational field test have been conduced to demonstrate and verify the validity of the enhanced adaptive TSP system.

The field operational tests of the ATSP system at the seven intersections along the state highway 82 (El Camino Real) have shown the following results:

- The operations of ATSP saved bus trip travel time by 13% (51 seconds) at northbound and by 9% (32 seconds) at southbound. The changes are statistically significant.

- ATSP increased bus average traveling speed by 11% (1.5 m/s or 3.4 MPH) at northbound and by 7% (1.1 m/s or 2.4 MPH) at southbound. The changes are statistically significant.
- The average intersection delay for testing bus runs was reduced by 50%

It is concluded that the results of the field operational tests of the ATSP system are very positive. Bus delay savings and passenger intersection delay savings are statistically significant. The major phase traffic delays are reduced, while the incurred minor phase delays are statistically insignificant.

The next steps of the research are:

- To investigate a robust and efficient system architecture suitable for a large-scale implementation and identify the corresponding communications links and means;
- To conduct an extensive field operational test that covers the entire SamTrans bus routes along the El Camino Real corridor in San Mateo County. The field operational test could serve three purposes: 1) through the test, the ATSP system can be further improved and refined to be suitable for large-scale implementations. 2) field data analysis and interviews with personnel from SamTrans and city traffic authorities would be able to fully reveal the impacts of the system; 3) demonstration of the system would facilitate the widespread deployment of the ATSP concept and thus lead to more significant benefits.

Table of Contents

Table of Figures

List of Tables

1 BACKGROUND AND OVERVIEW

1.1 Background

Transit signal priority (TSP) has been implemented in many cities across the country as one of the critical technologies for successful bus rapid transit system deployment and as a tool for improving traditional transit services,. Prior deployments of TSP have demonstrated the positive effects of reducing bus intersection delays and improving service schedule adherence. However, concerns have been frequently raised that transit priority operations may interrupt the normal operation of signal control and thus increase delays to other traffic, particularly traffic served by the non-prioritized phases (often minor phases, including cross-street traffic and main street left-turns).

Under many circumstances, the above concerns are in fact legitimate. The state-of-practice active TSP systems often adopt ad-hoc or heuristic TSP control logics, which are not adjustable in a real-time manner. For example, early green operation may truncate all of the preceding phases to minimum green time for other traffic and pedestrian phases. Such operations could cause severe delay to the minor-phase traffic and residual queues could last for several cycles. In order to address the concerns, priorities should be granted only when they are warranted and in a way that impacts to other traffic can be minimized. Those priority strategies should be adaptive to real-time traffic, pedestrian conditions and bus arrivals.

Active TSP systems normally adopt selective vehicle detection technologies that sense the presence of an approaching bus only at a fixed location. As a result, these systems have difficulty obtaining exact bus arrival time at an intersection.if the means of detection are placed far from the intersection. To ensure efficient priority treatments, the detection location has to be in the close proximity of the intersection. Consequently, the resulting "short notice" only gives the signal control system limited lead time to change signal settings to provide priority. This short lead time may cause noticeable delay to the non-prioritized traffic. In contrast, continuous detection means, such as Global Positioning System (GPS) and Automatic Vehicle Location (AVL), are able to actively monitor a bus's movement. With the improved capability of predicting bus arrivals to intersections, TSP systems should operate more efficiently.

Many transit agencies have installed or procured GPS-based AVL systems in their fleets. For example, by 2005 there were more than 2,500 AVL equipped buses in the San Francisco Bay Area. If the TSP system can be built upon the GPS/AVL system, it would allow a continuous means of bus detection. More importantly, the deployment will be cost-effective, because it allows all buses instrumented with GPS/AVL system to become TSP capable without requiring any significant additional equipment or expense on the buses

In light of the above considerations, California PATH, in collaboration with Caltrans Headquarters and District 4 has been developing an Adaptive Transit Signal Priority (ATSP) concept. The main features of the ATSP concept include:

- Providing priority to transit vehicles if warranted by accepted criteria. This attempts to make a trade-off between bus delay savings and the impacts on the rest of the traffic

- Making real time decisions adaptive to the movements of transit vehicles, traffic conditions and signal status. Again, this reflects the desire to balance the criteria of signal priority with competing interests at the intersection.

- Utilizing the existing GPS/AVL systems installed on buses to continuously monitor bus locations and predict bus arrival times at intersections. This builds functionality on existing technology in a cost-effective manner.

- Building the system using closed-loop signal control systems with 170E controllers. This may have potential for wide-scale implementation given an existing installed technology base.

The development and implementation of a prototype ATSP system have been conducted under PATH Project RTA 65A0026. The project has successfully validated the GPS-based ATSP concept and preliminarily demonstrated the effectiveness of the ATSP system. Reference 1-1 has documented the findings from the project and provides further detail on the demonstration. As a continuation of the previously mentioned research, this project addresses the issues that still remain for eventual deployment of the ATSP system. First, a more advanced priority request generator (PRG) has been developed in order to make a definable, repeatable and smoother tradeoff between bus delay savings and traffic delays. Second, the ATSP concept takes advantage of the infrastructure of deployed GPS/AVL systems. However, a challenge exists with the commercial GPS/AVL systems used by most of transit agencies. Most systems have very low "polling" rates of location, speed and direction. Each bus in a fleet is typically sampled every two minutes. Bus position sampling at this low rate most likely would not support a timely request for priority at normal operating speeds. Section 1.2 briefly introduces the prototype system design and operation of the ATSP concept. Sections 1.3 and 1.4 briefly introduce how these two issues are addressed, and the following chapters will introduce key developments and recommendations.

1.2 Prototype ATSP System

1.2.1 System Overview

The system architecture of the developed prototype ATSP system in project RTA 65A0026 is illustrated by Figure 1-1.

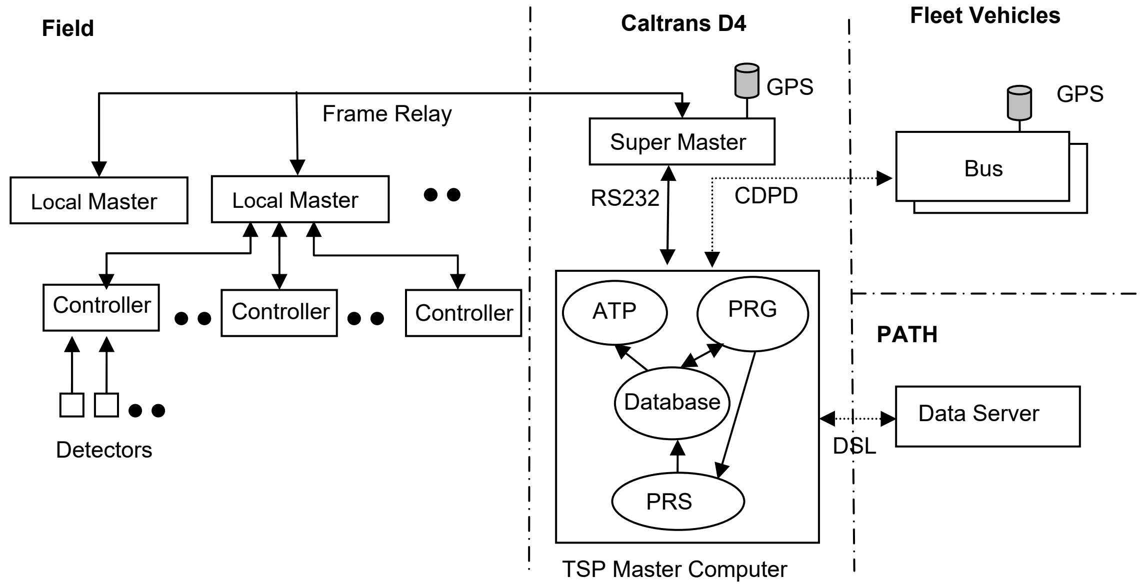

Figure 1-1 System Architecture of the Prototype ATSP System

In the prototype system, portable GPS installed on buses transmits second-by-second bus location information to a TSP master computer physically located in Caltrans District 4 via cellular digital packet data (CDPD). A bus arrival time predictor (ATP) hosted in the master computer use the historical and real-time GPS information to predict bus arrival times at intersections. The master computer also hosts a real-time database, a priority request generator (PRG) and a priority request sever (PRS). The TSP master computer is connected with the master computer of the signal control system by a direct serial port connection, allowing traffic data and signal status to be received by the real-time database. The PRG uses bus arrival prediction, signal status and pedestrian presence information to select TSP strategies. The PRG sends a priority request message to the PRS whenever a bus needs it and a check-out request after the vehicle has passed the signalized intersection. Upon receiving priority requests from multiple buses, the PRS prioritizes all the different priority requests based on the requested priority treatments, requested phase, and desired service time. It then generates a service request and sends the service request to the 170E type signal controller at the intersection for execution.

As its name suggests, the TSP master computer that initiates TSP requests, determines TSP strategies, prioritize TSP requests, and sends service requests to signal controllers for execution.

1.2.2 Bus Arrival Time Predictor (ATP)

The ATP uses real-time bus location and bus wheel speed information, together with historical AVL data to predict bus arrival times for upcoming traffic lights. The predictor consists of two models: 1) a historical model that uses linear regression to predict the arrival time solely based on historical data and 2) an adaptive model that uses recursive regression and adaptively adjusts its filter gain from the real-time AVL data. The final prediction generated by the predictor is a weighted average of the predictions from these two models. The weighting is also adaptively adjusted according to error variances obtained from the historical and adaptive models.

1.2.3 Priority Request Generator (PRG)

A PRG was developed to generate TSP strategies that minimize bus intersection delays. At the same time the generated strategy limits negative impacts on the minor-phase traffic and ensures pedestrian safety.

By knowing accurate bus arrival time a fairly long distance away from the intersection, it is possible to initiate TSP operation and adjust signal timing early. If there is enough time available, the negative impacts to minor-phase traffic can be reduced by distributing signal recovery across multiple cycles. However, in order to maintain the coordination in the actuated signal control in the prototype system, the cycle length remains fixed and the TSP treatment is only initiated at the start of the cycle when the bus is supposed to enter the intersection. The time for a bus to actually enter an intersection is the summation of the predicted bus arrival time and the queue discharging time. Using real-time traffic counts from loop detectors, the queue discharging time can be estimated using the assumption of "point queue" or "vertical queue".

There are two candidate priority treatments in the ATSP concept: green extension and early green. "Green extension" requests an extended green for the bus phase (main phase) while "early green" requests an earlier start of green for the bus phase. If the predicted time for a bus to enter the intersection is in a pre-defined time window, a green extension request will be generated. Otherwise, an early green request will be generated. The pre-defined time window starts at the end of the main phase (the yield point). Its length is the maximum green time that can be extended for buses. The maximum extension is determined with consideration of ensuring enough time for all pedestrian calls in the subsequent cycle and limiting impacts of controller clock transition or recovery. In the prototype system it is preset as 10 seconds and the resulting maximum signal recovery time will be less than 40 seconds, which is about half to one-third of a normal signal cycle length.

The PRG determines how much priority is appropriate for an approaching bus. In the case of green extension, the requested phase would be extended from its original end to the time point that either the maximum allowable green extension time is reached or the bus cleared the intersection. When providing early green treatment, in order to fairly distribute the impacts across the minor phases, the PRG re-allocates the available green time among the remaining phases (at the current time instant) before the bus phase. The available green time for reallocation is defined as:

$$t_{avail} = \max(0, t_{arrival} - t_{discharge} - t_{guarantee}) \qquad \textbf{1-1}$$

where t_{avail} is the green time available for allocation; $t_{arrival}$ is the bus arrival time given by the bus arrival time predictor; $t_{discharge}$ is the queue discharge time and $t_{gurantee}$ is the total guaranteed green time that the remaining phases should have to accommodate minimum green, yellow time, intersection clearance time, and pedestrian walking time if the pedestrian button on that approach is actuated. The available green time is calculated and updated second by second, taking into account the latest information on bus arrival, queue length, duration of elapsed phases and pedestrian presence.

If $t_{avail} = 0$, the early green treatment will truncate the phases before the bus phase into their corresponding guaranteed green times. If the phases are not truncated, they will be allocated based on the following approach: Allocation will be based on the historical signal status data and the phase duration distributions. The distribution values vary between the corresponding minimum greens and maximum greens. With these phase duration distributions, a probability level can be determined such that:

$$t_{avail} = \sum_{i=1}^{n} cdf_i^{-1}(L) \quad \textbf{1-2}$$

where n is the number of the remaining minor phases before the bus phase; cdf_i is the cumulative density function for phase i, and L is a probability level. $cdf_i^{-1}(L)$ yields the time that corresponds to the probability level L, and it is the time that is allocated to phase i in addition to the guaranteed green time. In this sense, L can be viewed as the confidence level for each phase that fully accommodates the approaching traffic with the allocated green time plus minimum green time. There are two reasons in support of using Equation 1-2. First, it is a fair allocation and evenly distributes the impacts among multiple phases. Second, this method of allocation can ensure that minimum percentages of vehicles are blocked by the priority treatments if departure flows at signalized intersections strictly decrease from the saturation flow at the end of minimum green to zero at the end of the maximum green. Although it is not generally true, the departure flows do present a decreasing pattern and therefore Equation 1-2 may implicitly lead to fewer impacts on the minor-phase traffic.

Figure 1-2 presents a simplified flow chart for the proposed TSP algorithm.

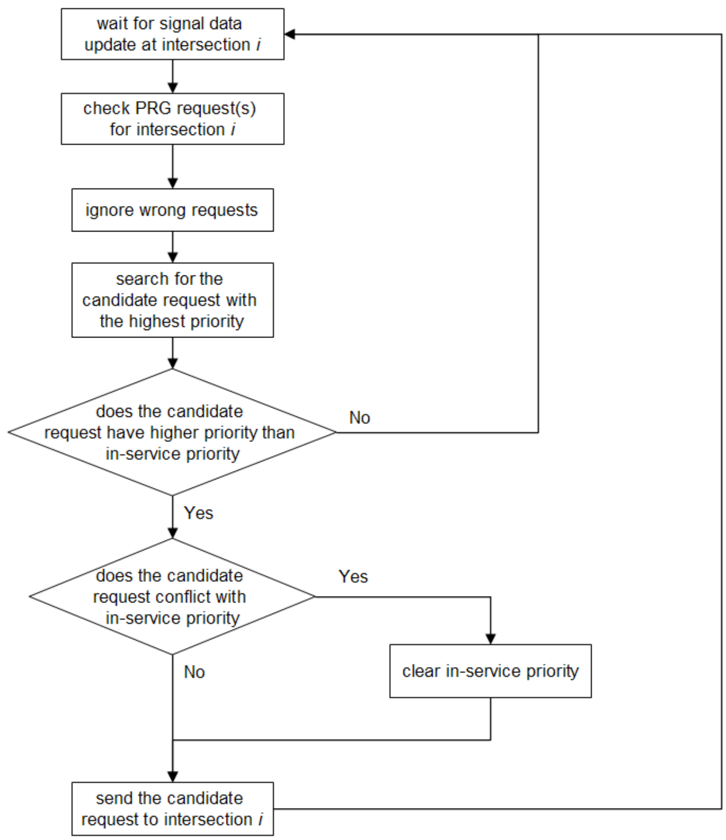

Figure 1-2 Flow Chart of TSP Algorithm

1.2.4 System Implementation and Field Operational Test

A prototype TSP system was been implemented on a 0.7-mile-long highway corridor along El Camino Real, San Mateo, CA, including two four-way intersections, the 25th Ave and the 27th Ave respectively and one three-way intersection at 28th Ave. The existing signal control system was a distributed closed-loop coordinated system with Model 170E controllers.

The original Caltrans C-8 controller software was updated to allow the options of green extension and early green on the controllers. Specifically, the controller grants green extension by “freezing” the local clock at the yield point while early green is implemented by moving default force-off points to requested locations. A master controller, installed at 25th Ave, polls local controllers every second to obtain signal status, vehicle detector and pedestrian pushbutton data. In addition, once receiving a priority service request from the super master, it forwards the request to the corresponding local controller. It also forwards the receipt acknowledgement back to the super master once it is received from the local controller.

A field operational test was conducted between May and July, 2004. “Before” and “after” data analyses show that the prototype ATSP system reduced bus signal delay by 12% to 87% per intersection. At the same time, the minor-phase traffic delay increased by 1% to 17% and the major-phase traffic delay decreased by 1% to 72%.

In summary, under the project RTA 65A0026, an ATSP concept was developed and the system implementation and field operational test demonstrated its feasibility and validity. Issues still remain for eventual large-scale deployment of the concept, including the need for a more advanced priority requestor and the problem of low “polling rate” with the commercial GPS-based AVL systems.

1.3 Overview of Advanced TSP Requestor

This section gives an overview on the development of an advanced TSP requestor or PRG. Chapters 2, 3 and 4 present respectively the technical details on bus arrival time prediction, arrival flow prediction and TSP timing optimization models.

A PRG is the key to determining which TSP strategies, either early green or green extension, to implement once a request is received. It is also fundamental to determining the extent of the TSP operation. The prototype system developed in RTA 65A0026, makes the decision in an adaptive, but still heuristic manner. More specifically, the decision is made without explicitly considering the TSP impacts to the other traffic. Moreover, the conditions under which early green or green extension should be adopted are predetermined, and not necessarily appropriate under the actual traffic and pedestrian circumstances. In view of all of these, we have developed a more advanced PRG that is able to adaptively and optimally select either early green or green extension to implement and determine the corresponding signal timing strategies for the TSP operations. The objective that guides the decision is to make a tradeoff between reducing bus intersection delay and affecting other traffic delay. The level of the tradeoff can be adjusted via a weighting factor and should be determined through negotiations among the stakeholders on how much preference the transit operation should be given.

The framework of the new PRG is depicted by Figure 1-3.

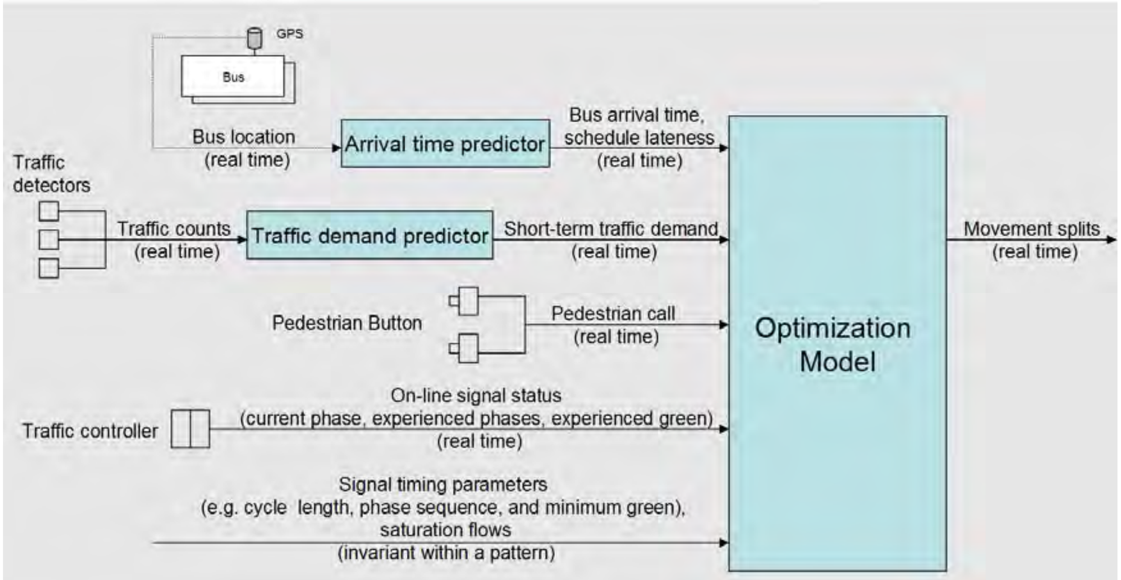

Figure 1-3 Framework of New PRG

The PRG consists of three important modules: a bus arrival predictor, a traffic demand predictor and signal timing optimization models. The bus arrival predictor is essentially the same as that in RTA 65A0026, but has been updated. The predictor uses both historical and real-time transit data to make the prediction. Since traffic patterns may gradually change over time, the historical model has been reformulated to be capable of spontaneously responding to the changes. The updated predictor will be introduced in Chapter 2.

To consider the impacts of TSP operations to buses and other traffic, future traffic demand estimates are a prerequisite, at least for several cycles after the TSP impact. A traffic demand predictor has been developed for that purpose. Based on loop detector data from previous cycles, a moving-average scheme has been adopted for the prediction. Chapter 3 will elaborate on this scheme.

The core of the PRG consists of two optimization models. These take as inputs the predicted bus arrival times, predicted traffic demand, pedestrian presence information and signal status. The outputs of the optimization models are TSP strategies (either early green or green extension) and the corresponding pattern of force-off points. There are two models: one is for early green and the other for green extension. With justifiable assumptions, analytical closed-form expressions of bus delay and traffic delay for each movement relative to either early green or green extension are derived and incorporated into the objective functions of the models. A set of constraints has been developed to guarantee the under-saturation condition, to follow the ring-barrier structure of actuated controllers and to ensure safety. The models are essentially quadratic programming problems and are thus easy to solve. Chapter 4 will describe the model formulations, and present a numerical example to demonstrate the validity of the models.

1.4 Strategies to Address Limited Communication Resources for Large Fleet

ATSP is built on the premise that a signal priority algorithm can anticipate the arrival of a transit vehicle at a traffic signal with sufficient lead time so that the priority treatment can be done before the signal phase at which the bus arrives, in order to 1) maximize the likelihood that the bus will go through the intersection, 2) minimize delay for the transit bus, 3) minimize the impact on other traffic, and 4) ensure pedestrian safety. It is therefore imperative that a prediction algorithm must be able to predict the time-to-arrival (TTA) at the next signalized intersection. One of the main contributions of PATH research on ATSP has been to develop approaches that allow the GPS/Advanced Communication System (ACS) that has already been instrumented on buses to become an integral portion of the ATSP, which facilitates the integrated application of transit ITS technologies. However, there are technical challenges. Because the ACS system was designed for transit operation monitoring and data collection purposes, the polling rate of the ACS is at approximately a two minute interval.

In this study, we formulate and analyze by simulations the problem of establishing communication between transit vehicles and the TMC for requesting signal priority service. The availability of this communication channel is, however, not guaranteed because of the potentially large number of transit vehicles requesting the service at the same time and the limited number of time slots per second available to process these requests. The contention channel has a limited number of time slots for processing the requests. So we need to design an efficient protocol to coordinate the processing of the requests, so that requests with higher priorities are processed first, and the number of dropped requests is minimized.

We have analyzed a simple supply-demand model by considering deterministic scheduled arrival times at signalized intersections. The schedules allow us to estimate an "average" demand profile for signal priority at each intersection. Results show that the queue size and queue waiting time are sensitive to the process rate of the contention channel. We conclude that with a process rate of two requests per second, using the process priority protocol, the contention channel is capable of processing the deterministic demand profile obtained from the scheduled arrival times at intersections. Nonetheless, we should emphasize that the results depend critically on the distribution of the arrival times.

Within this study, we have assumed a deterministic demand profile. The arrival time at each intersection is random and needs to be predicted using a TTA algorithm. The statistical properties of the TTA prediction error distribution for each link connecting two successive intersections need to be included. For a link with a smaller variance, the corresponding set of request processes should have a higher priority. This will serve as another guiding principle in further revising the protocol. A protocol for communication establishment must also consider the communication requirements of the AVL system. The polling schedule should also be included in the analysis, so that the polling channel can be utilized for processing some of the signal priority requests. We hope to cope with all the above unconsidered issues in the next phase of the project.

1.5 Reference

1-1 Zhou, K. et al. Development of Adaptive Transit Signal Priority Systems, University of California, Berkeley: Institute of Transportation Studies, California PATH Program, PATH Research Report UCB-ITS-PRR-2005, August 2005.

2 TIME-TO-ARRIVAL PREDICTION MODEL

The accuracy of bus arrival time prediction is very important for successful TSP operations. Large prediction error could result in heavy interruption in the normal traffic signal control and makes a TSP system unacceptable. The objective of the development of arrival time prediction model was to develop algorithms to accurately predict bus' arrival time at signalized intersections that can meet the requirements of TSP operations.

2.1 Models Developed Under Project RTA 65A0026

Most existing bus arrival time prediction algorithms use historical data based models to mathematically describe the intrinsic relationship between a dependent variable (time to destination from current location) and independent variables such as historical link travel time, upstream schedule derivation and headway distribution, in addition to the current bus location. Although historical data based models might have captured the best average information, the accuracy of arrival time prediction highly relies on the similarity between real-time and historical traffic patterns. Variations of the historical average could cause significant inaccuracy in prediction results.

Under project RTA 65A0026, an adaptive bus arrival time prediction model has been developed for TSP applications, and has been documented in Reference 2-1. This model consists of two parallel components: a historical data based model that aims to capture the best knowledge of the average traffic patterns along the corridor, and a real-time model that aim to describe the prevailing traffic pattern as well as the deviation from the historical patterns. The output of this model, the expected time of arrival (ETA), is a weighted combination between the historical model and the real-time model that integrates the historical traffic patterns and the current traffic pattern in the framework of the algorithm.

Bus GPS data consists of four fields: Coordinated Universal Time (CUT) time, longitude, latitude, and speed over ground. A node-link representation is used to describe a bus route. A node refers to either a bus stop or the stop-bar at an intersection, and a link refers to the road segment connects two adjacent nodes. GPS coordinates (longitude and latitude) of each node were collected and converted into Euclidean coordinates and bus trajectory is projected into a one-dimensional bus path that starts from the first node of the heading direction. Bus trajectory along the route is divided into moving sections and stopped sections as they represent different status of bus operations and need to be treated separately. A stopped section refers to the time when bus stopped either at a bus stop to load/unload passengers or in traffic jam, and a moving section refers to the time between two stopped sections.

Figure 2-1 plots a particular bus trajectory on a time-space diagram. It clearly shows a linear relationship between the travel distance and the travel time for moving sections, with slopes possibly varying among sections. This motivated the use of a constant running speed model to represent the intrinsic relationship between distance-to-next-node (D2N) and time-to-next-node (T2N).

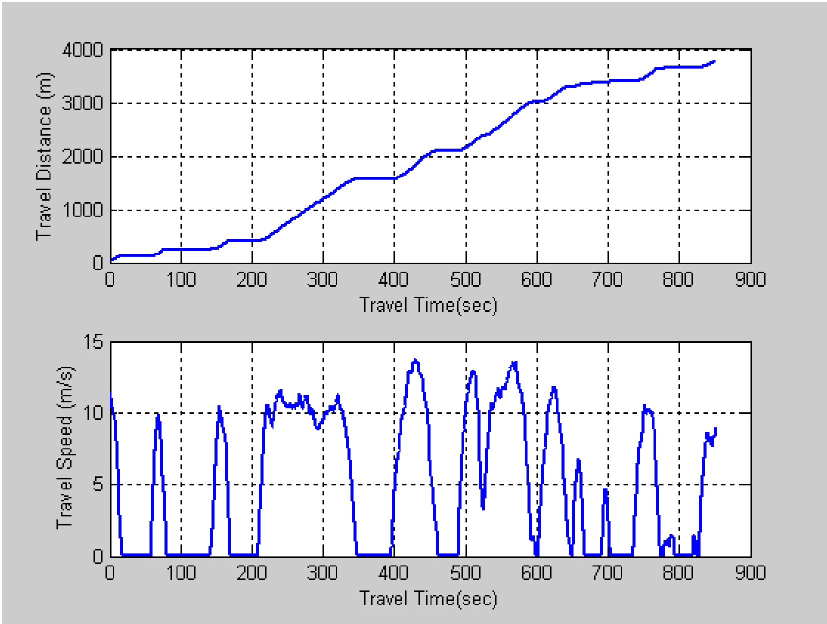

Figure 2-1 Bus Location and Speed as Function of Time

At time t, if the next node is an intersection, ETA is given by

$$\mathrm{ETA} = t + \mathrm{T2N} \qquad \textbf{2-1}$$

where T2N is the estimated time-to-next-node from current bus location and is a weighted combination of historical model and real-time model. In case that there is a downstream bus stop before the intersection, ETA is given by

$$\mathrm{ETA} = t + \mathrm{T2N} + \mathrm{DT} + \mathrm{T2I} \qquad \textbf{2-2}$$

where DT is the estimated dwell time at the bus stop and is calculated as the average dwell time at the bus stop. T2I is the estimated time-to-intersection from the bus stop and relies on historical model only.

2.1.1 Historical Model

Historical data, at certain level, represent the average traffic patterns. The purpose of the historical model is to address the problem: Given a moving section with length *D*, what is the "best" fit for the time, *T(D)*, that a bus would spend on the section?

The following linear equation is used to capture the intrinsic relationship between dependent variable, *T*, and independent variable, *D*.

$$T(D) = \alpha D + \beta \qquad \textbf{2-3}$$

The linear regression method (Least Squares) is then applied to determine the parameters α and β. The estimation error consists of two parts: the measurement error on D and the regression error. The error variance is approximately given by

$$\sigma_{T,H}^2 = \alpha^2 \sigma_d^2 + \sigma_{reg}^2 \qquad \textbf{2-4}$$

where σ_d and σ_{reg} are the standard deviations of the measurement error and the regression error, respectively.

In case that there is a downstream bus stop before the intersection, Equation 2-3 estimates the travel time from the bus stop to the downstream intersection, given the distance between bus stop and the stop-bar of the intersection, i.e., T2I, the last component of Equation 2.2.

When a bus is traveling towards the next node, the expected time-to-next-node, i.e. T2N, needs to be updated as the bus changes its location, i.e. D2N. It is assumed that the bus would maintain its average traveling speed until the arrival at the next node.

The average traveling speed, from Equation 2-3, is given by

$$\bar{v}_H \stackrel{\text{def}}{=} \frac{D}{T(D)} = \frac{D}{\alpha D + \beta} \tag{2-5}$$

with the error variance given by

$$\sigma_{\bar{v},H}^2 = \left(\frac{1}{T(D)}\right)^2 \sigma_d^2 + \left(\frac{D}{T^2(D)}\right)^2 \sigma_{T,H}^2 \tag{2-6}$$

And the expected time-to-next-node is given by

$$T2N_H = \frac{D2N}{\bar{v}_H} \tag{2-7}$$

with the error variance given by

$$\sigma_{T2N,H}^2 = \frac{2}{\bar{v}_H^2}\sigma_d^2 + \left(\frac{D2N}{\bar{v}_H^2}\right)^2 \sigma_{\bar{v},H}^2 \tag{2-8}$$

2.1.2 Real-Time Model

As aforementioned, historical models would have captured the best average information, and the prediction accuracy highly relies on the similarity between real-time and historical traffic patterns. When the bus is far away from the destination point, larger weight needs to be placed on the historical model due to the uncertainties in the downstream traffic conditions; and when the bus is close to its destination, emphasis needs to be placed on the real-time traffic condition to compensate for the variation from the historical average. In the latter case, the bus is likely to move to the destination with its prevailing average running speed rather than the historical average speed.

Since transit vehicles share the right of way with general traffic, bus movement, at certain level, represents the real-time traffic condition. Based on the observation of the linear relationship between travel distance and travel time along moving sections, bus movement along a moving section is modeled as the following

$$d(t) = \bar{v}_{RT}(t)t + b(t) + w_d(t) \overset{\text{def}}{=} H(t)x(t) + w_d(t) \qquad \textbf{2-9}$$

where $H(t) = \begin{bmatrix} t & 1 \end{bmatrix}$ and $x(t) = \begin{bmatrix} \bar{v}_{RT}(t) & b(t) \end{bmatrix}^T$. $w_d(t)$ is the measurement noise and is assumed to be Gaussian white i.e.,

$$\begin{aligned} &E[w_d(t)] = 0 \\ &E[w_d(t)w_d(t+\tau)] = \sigma_d^2\delta(\tau) \end{aligned} \qquad \textbf{2-10}$$

d and t are the travel distance and travel time along the moving section, all measured from the start point of the moving section. $\bar{v}_{RT}(t)$ is the average traveling speed upon time t.

The following Recursive Least-Squares (RLS) filter is used to estimate the unknown parameter *x* from the measurements, and the estimation adapts to the movement of bus.

$$
\begin{aligned}
\hat{x}_{k+1} &= \hat{x}_k + W_{k+1}\left[d_{k+1} - H_{k+1}\hat{x}_k\right] \\
S_{k+1} &= H_{k+1}P_k H_{k+1}^T + \sigma_d^2 \\
W_{k+1} &= P_k H_{k+1}^T S_{k+1}^{-1} \\
P_{k+1} &= P_k - W_{k+1}S_{k+1}W_{k+1}^T
\end{aligned}
\qquad \textbf{2-11}
$$

where W_{k+1} is the filter gain, S_{k+1} and P_{k+1} are the covariance of the residual and the state estimate, respectively.

The expected time-to-next-node is then given by

$$T2N_{RT} = \frac{D2N}{\bar{v}_{RT}} \qquad \textbf{2-12}$$

with the error variance given by

$$\sigma_{T2N,RT}^2 = \frac{2}{\bar{v}_{RT}^2}\sigma_d^2 + \left(\frac{D2N}{\bar{v}_{RT}^2}\right)^2 P(1,1) \qquad \textbf{2-13}$$

2.1.3 Adaptive Model

Equation 2-7 estimates the time-to-next-node based on bus current location and historical data, while Equation 2-12 estimates that based on the real-time bus movement data. The purpose of the adaptive model is to balance the outputs from the historical model and the real-time model therefore to compensate for the variance of real-time traffic pattern from historical traffic pattern and the uncertainties of the downstream traffic condition.

Considering $T2N_{RT}$ as an observation of $T2N_H$, which is assumed to be a Gaussian random variable with the mean and variance given by Equation 2-7 and Equation 2-8, respectively, i.e.,

$$T2N_{RT} = T2N_H + w,\ \ w \sim N(0, \sigma_{T2N,RT}) \qquad \textbf{2-14}$$

The maximum a *posteriori* (MAP) estimate of Equation 2-14 is given by

$$T2N_A = w_H T2N_H + w_{RT} T2N_{RT} \qquad \textbf{2-15}$$

with

$$\begin{aligned} w_H &= \frac{\sigma^2_{T2N,RT}}{\sigma^2_{T2N,RT} + \sigma^2_{T2N,H}} \\ w_{RT} &= \frac{\sigma^2_{T2N,H}}{\sigma^2_{T2N,RT} + \sigma^2_{T2N,H}} \end{aligned} \qquad \textbf{2-16}$$

The corresponding error variance of Equation 2-15 is given by

$$\sigma^2_{T2N,A} = \frac{\sigma^2_{T2N,RT}\sigma^2_{T2N,H}}{\sigma^2_{T2N,RT} + \sigma^2_{T2N,H}} \qquad \textbf{2-17}$$

As can be send from Equation 2-17, the error variance of the adaptive model is smaller than that of the historical model and that of the real-time model, i.e.,

$$\begin{aligned} \sigma_{T2N,A} &< \sigma_{T2N,H} \\ \sigma_{T2N,A} &< \sigma_{T2N,RT} \end{aligned} \qquad \textbf{2-18}$$

2.2 Updates under this Project

Updates have been made under the current project to further improve the performance of bus arrival time prediction. First of all, as GPS location measurement contains error, it is necessarily to have an observer which smoothes the measurement noise, tracks and estimates bus state (location and speed). Secondly, the historical model has been updated to have a better fit with the kinematics of bus acceleration and deceleration movement. As a consequence, the adaptive model has also been revised. Finally, an adaptive mechanism is applied on historical model to have its parameters dynamically updated based on real-time bus movement.

2.2.1 Bus State Observer

As aforementioned, a bus state observer needs to be implemented to smooth the measurement noise and to track and estimate bus state. Figure 2-2 illustrates the framework of updated prediction algorithm.

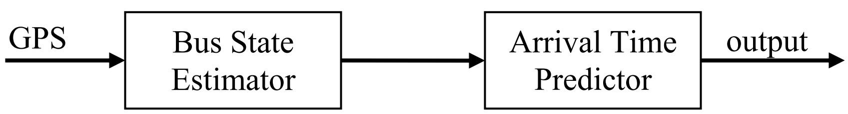

Figure 2-2 Framework of Prediction Algorithm

The following constant speed model is used to describe bus dynamics

$$\begin{aligned} X_{k+1} &= AX_k + w_k \\ y_k &= CX_k + w_k^y \end{aligned} \qquad \textbf{2-19}$$

where $X = [d \quad v]$ is bus state variable with *d* represents the travel distance with respect to the start node and *v* represents the speed. $A = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}$ is the transition matrix (*h* = 1 second is the GPS sampling interval), and $C = [1 \quad 0]$ is the measure matrix. w_k and w_k^y are independent white Gaussian sequences with zero means and covariance matrix Q_k and R_k, respectively.

The standard Kalman filter is applied to the estimated bus state as the follows

$$
\begin{aligned}
S_k &= CP_kC^T + R \\
G_k &= P_kCS_k^{-1} \\
\hat{X}_{k+1} &= A\hat{X}_k + G_k\left[y_{k+1} - CA\hat{X}_k\right] \\
P_{k+1} &= (I - G_kC)(AP_kA^T + Q
\end{aligned}
\tag{2-20}
$$

where S_k is measurement prediction covariance, P_k is the state prediction covariance, and G_k is the filter gain.

2.2.2 Updated Historical Model

When applying linear regression, the first question needs to be answered is whether the linear model fits the underlying physics. Considering a typical bus movement on a moving section, the bus would first accelerate from zero speed with a constant acceleration level A_1 until it reaches a desired constant speed V; it then keeps the speed for certain of time t^* and decelerates to zero speed with a constant deceleration level A_2. Let *D* and *T* be the lengths in distance and in time of the moving section, then

$$
\begin{aligned}
D &= \frac{V^2}{2A_1} + \frac{V^2}{2A_2} + Vt^* \\
T &= \frac{V}{A_1} + \frac{V}{A_2} + t^*
\end{aligned}
\tag{2-21}
$$

Defining the average running speed of this moving section as $\bar{v}_H = \frac{D}{T}$ yields the following

$$
\bar{v}_H = \frac{aD}{D + b}
\tag{2-22}
$$

where *a* and *b* are parameters associated with A_1, A_2 and V.

Although Equation 2-22 is non-linear, it can be transformed to a linear regression problem as

$$\bar{v}_H^{-1} = \alpha + \beta D^{-1} \qquad \textbf{2-23}$$

where $\alpha = \dfrac{1}{a}$ and $\beta = \dfrac{b}{a}$.

The regression models given by Equation 2-3 and Equation 2-23 are applied to northbound, morning peak (0700 to 0900 hrs) bus trips, and Figure 2-3 plots the regression line of Equation 2-3, the previous historical model, and that of Equation 2-23, the updated historical model. When bus is beyond 150 meters upstream to an intersection, the previous historical model under estimates the average speed, which would result in over estimated time-to-intersection.

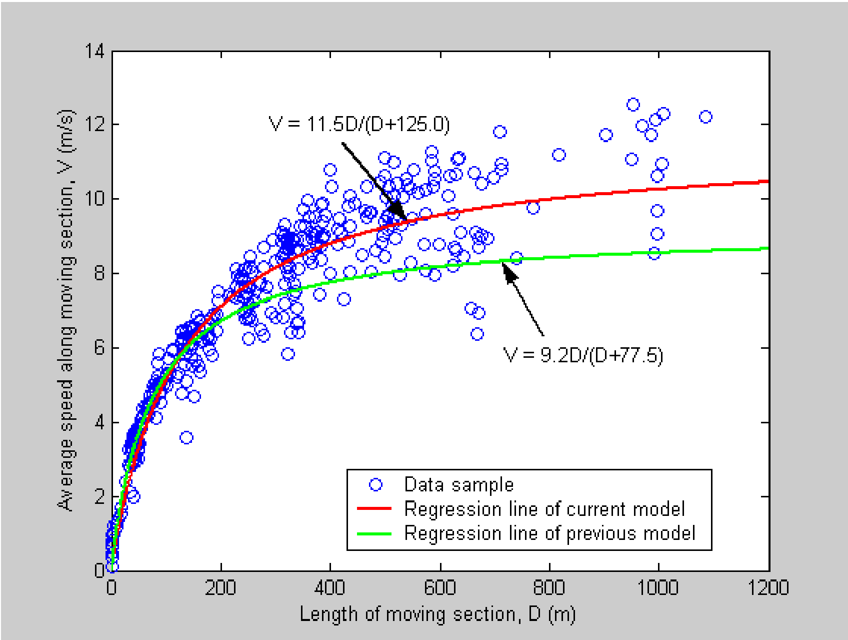

Figure 2-3 Regression Lines of Equation 2-3 and Equation 2-23

A measure to test whether a regression model fits the underlying physics is to check whether the regression residual has a normal distribution. Figure 2-4 and Figure 2-5 plot the normal probability of the residual of Equation 2-23 and that of equation 2-3, respectively. In Figure 2-4, most data samples are close to the straight line indicating that the residual has a near normal distribution, while in Figure 2-4, the data samples do not fit the straight line well implying the residual is not normal distributed.

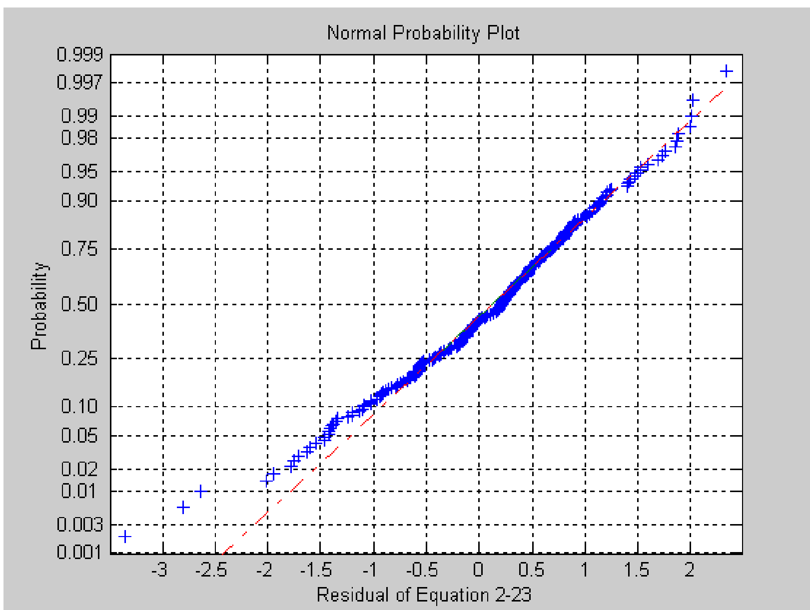

Figure 2-4 Normal Probability Plot of Regression Residual of Equation 2-23

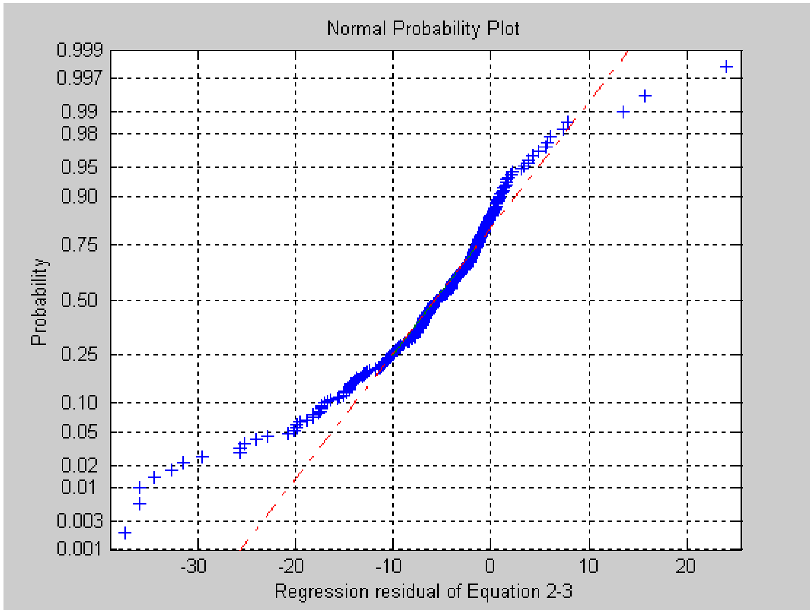

Figure 2-5 Normal Probability Plot of Regression Residual of Equation 2-3

On the other hand, as the goal of the historical model is to estimate bus average speed along a moving section, another measure for the performance of the model is to compare the error mean of speed estimation. Figure 2-6 and Figure 2-7 plots the histogram of estimation error for Equation 2-23 and Equation 2-3, respectively. A normal density function superimposed on the histogram. The plots clearly show that the error of the updated model given by Equation 2-23 is less than that of the previous model given by Equation 2-3. The mean of error is 0.1 for updated model and is 1.0 for the previous model.

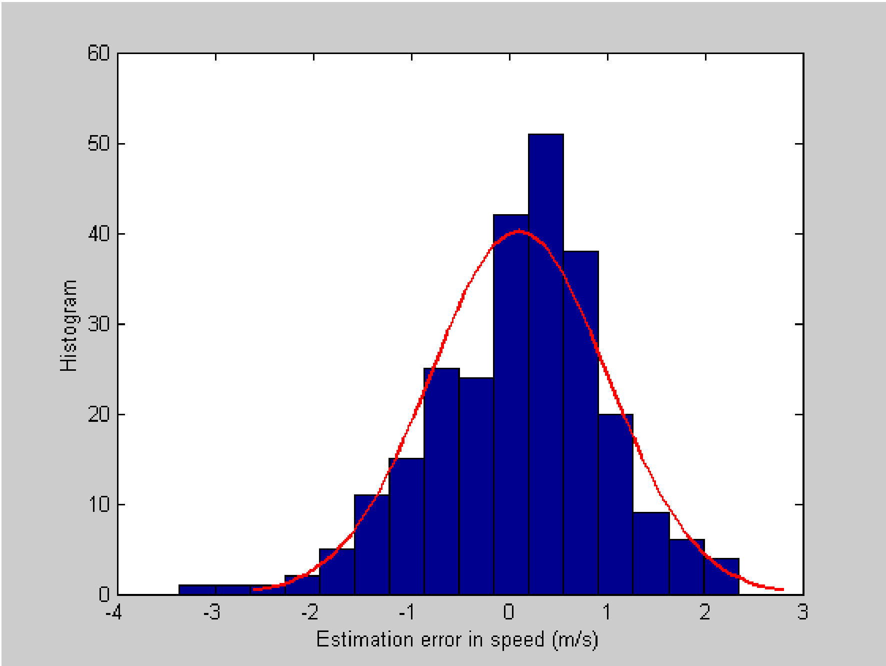

Figure 2-6 Histogram of Speed Estimation Error by Equation 2-23

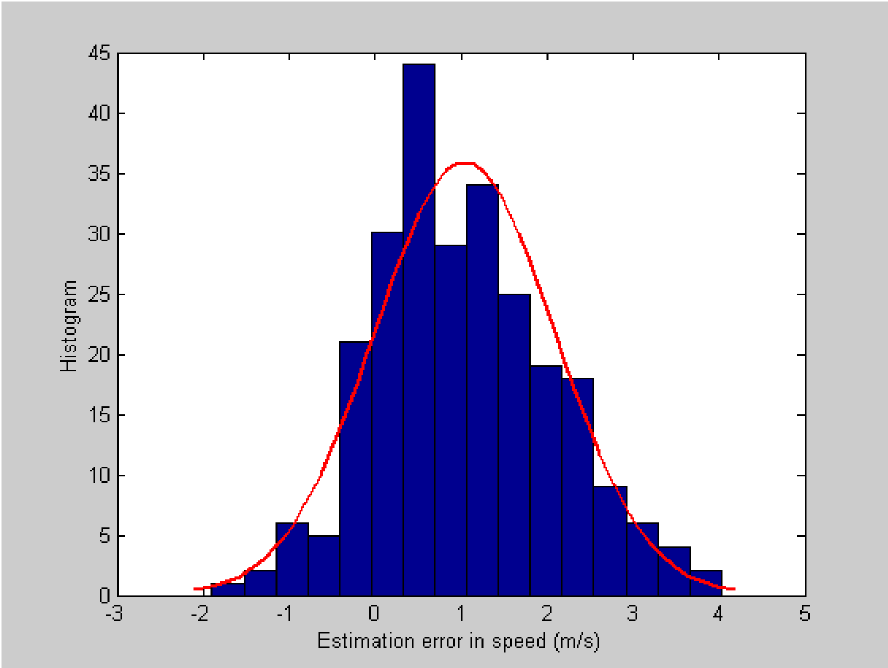

Figure 2-7 Histogram of Speed Estimation Error by Equation 2-3

2.2.3 Updated Adaptive Model

As shown in Equation 2-15, the expected time-to-next-node, $T2N_A$, is a weighted combination between the output of the historical model, $T2N_H$, and the output of the real-time model, $T2N_{RT}$. The algorithm can be simplified as

$$T2N_A = \frac{D2N}{\bar{v}_A} \qquad \textbf{2-24}$$

where $\bar{v}_A$ is the weighted combination of $\bar{v}_H$, which is given by Equation 2-22, and $\bar{v}_{RT}$ given by Equation 2-11, i.e.,

$$\bar{v}_A = \bar{w}_H \bar{v}_H + \bar{w}_{RT} \bar{v}_{RT} \qquad \textbf{2-25}$$

The error variance of $\bar{v}_H$, from Equation 2-22, is given by

$$\sigma^2_{\bar{v},H} = \frac{ab}{(D+b)^2}\sigma^2_d + \sigma^2_{reg} \qquad \textbf{2-26}$$

where σ^2_{reg} is the variance of regression error. The error variance of $\bar{v}_{RT}$, from Equation 2-11, is given by

$$\sigma^2_{\bar{v},RT} = P(1,1) \qquad \textbf{2-27}$$

The resulted weightings between the two models become

$$\bar{w}_H = \frac{\sigma^2_{\bar{v},RT}}{\sigma^2_{\bar{v},RT} + \sigma^2_{\bar{v},H}}$$
$$\bar{w}_{RT} = \frac{\sigma^2_{\bar{v},H}}{\sigma^2_{\bar{v},RT} + \sigma^2_{\bar{v},H}} \qquad \textbf{2-28}$$

which are different from Equation 2-16.

2.2.4 Dynamic Historical Model

The historical model relies on the parameters *a* and *b* to estimate bus average speed. The parameters could change in time-of-day, time-of-month, and time-of-year. It is essential to have a mechanism to dynamically update the parameters to address the variation of the parameters.

Applying the method of Recursive Least-Squares (RLS) to Equation 2-23, i.e.,

$$V_k^{-1} = \alpha_k + \beta_k D_k^{-1} + w_k \overset{\text{def}}{=} H_k x_k + w_k \qquad \textbf{2-29}$$

where D_k and $V_k = \dfrac{D_k}{T_K}$ are observed lengths in distance and time of the moving section that the bus just traveled across. $H_k = \begin{bmatrix}1 & D_k^{-1}\end{bmatrix}$, $x_k = \begin{bmatrix}\alpha_k & \beta_k\end{bmatrix}^T$, and $w_k \sim N(0, \sigma_{reg,k})$ is the regression error. The similar recursive process given by Equation 2-11 is then applied to dynamically update α_k and β_k, based on real-time bus movement. The parameters of the historical model become

$$a_k = \frac{1}{\alpha_k} \qquad b_k = \frac{\beta_k}{\alpha_k} \qquad \textbf{2-30}$$

2.3 Case Study

The performance of updated bus arrival time model is examined by the following case study, using real-world transit data. In this example, the distance between two adjacent signalized intersections is 525 meters and the actual travel time is 41 seconds. After passing the upstream intersection, both of the historical model and the adaptive model predict the arrival time to the downstream intersection.

Figure 2-8 shows the prediction results as well as bus trajectory. The adaptive model which integrates the historical model and the real-time model worked much better than the historical model. When the bus just passed the upstream intersection, the two models performed similar as the adaptive model is in transition mode and has larger estimate variance at the beginning. As the bus moving towards the downstream intersection, the adaptive model provides prediction results that are adaptive to the bus movement and converged fast to the actual arrival time.

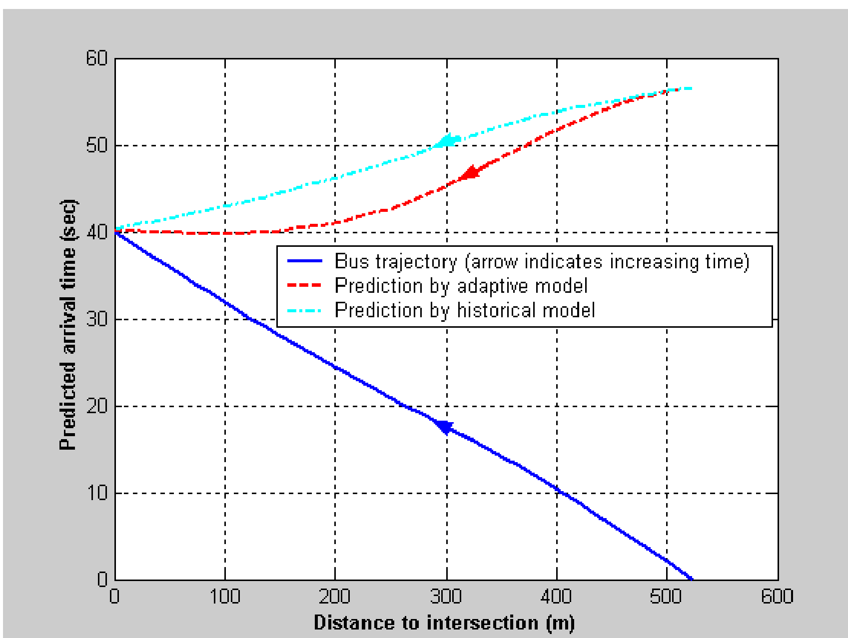

Figure 2-8 Arrival Time Prediction

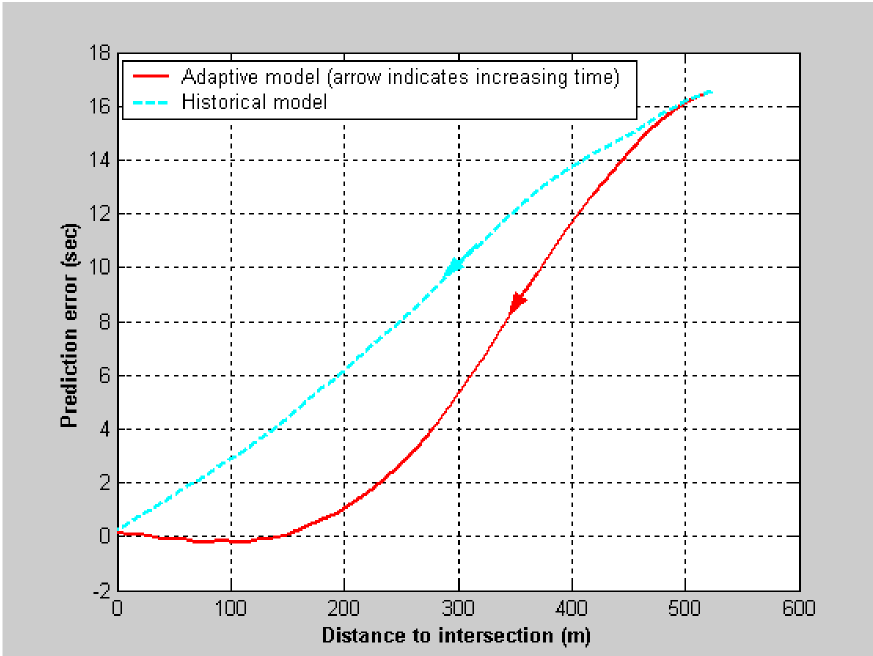

Figure 2-9 Prediction Error

Adjusting the signal timing to facilitate the movement of transit vehicles through signalized intersections poses a strict requirement on the estimate error of the arrival time prediction. According to Caltrans, the estimate error at the point a priority treatment being trigged should be within 10% of the signal cycle length. With the normal signal cycle length varying from 60s to 120s, an +/- 5s error bound would be appropriate to measure the confident level of the arrival time prediction. The adaptive model achieved the requirement at 300m and 24s to go to the intersection, while the historical model achieved this at 165m and 14s to go to the intersection, as shown in Figure 2-9. Noticing that the normal detection range for most current TSP systems is between 50m to 100m, both the historical model and the adaptive model provide much larger flexibilities than conventional models.

2.4 Summary

This chapter reports the updates that have been made on bus arrival time predictor. The predictor developed under project RTA 65A0026 has been validated that can meet the requirement of ATSP operations. In order to further improve its performance, a Kalman filter based bus state observer has been included to smooth the measurement noise and to track bus movement. A kinematics model is applied to model bus' acceleration and deceleration movement along moving sections and it leads to an updated version of the historical model. Side-by-side comparisons were made to show the improved performance. The case study demonstrates the effectiveness of the updated adaptive model, e.g., the prediction error fall into +/- 5 seconds range when the bus is 300 meters (or 24 seconds arrival) to the intersection.

2.5 Reference:

2-1 Zhou, K. et al. Development of Adaptive Transit Signal Priority Systems, University of California, Berkeley: Institute of Transportation Studies, California PATH Program, PATH Research Report UCB-ITS-PRR-2005, August 2005.

Y. Bar-Shalom and T. E. Formann. *Tracking and Data Association*. Academic Press, Orlando, FL, 1988

3 ARRIVAL TRAFFIC FLOW PREDICTION MODEL

In order to balance the trade-off between the needs of normal traffic and the needs of transit vehicles at signalized intersections, the advanced ATSP algorithm described in chapter 4 needs an estimation of arrival traffic flow to calculate the delay on normal traffic. This chapter reports such an effort for predicting traffic demand at signalized intersections.

3.1 Literature Review

The implementation of Intelligent Transportation Systems (ITS) has made available vast quantities of real-time traffic data. The development of Dynamic Traffic Management (DTM) Systems has obtained increasing attention over the past 20 years as a result of growing traffic problems and increasing potential of information technologies. The accuracy of estimates of current and short-term expected traffic situations is crucial to successful DTM applications. The problem of short-term traffic flow prediction is to determine the traffic volume in the next time window, usually in the range from five minutes to half an hour. Let $\{V_t\}$ be a discrete time series of vehicular traffic flow rates at a specific detection point, the univariate short-term traffic flow prediction problem is

$$\hat{V}_{t+k} = f\left(V_t,\ V_{t-1}, V_{t-2}, \cdots\right), \qquad k = 1, 2, 3, \cdots \tag{3-1}$$

where $\hat{V}_{t+k}$ is the prediction of V_{t+k} computed at time *t*. The prediction where $k = 1$ is the single interval or one-step forecast. Likewise, multiple interval forecasts are those where $k > 1$.

In the past, various approaches have been developed to forecast traffic flow, From the standpoint of how information is used, short-term traffic flow forecasting models can be categorized into three groups: 1) those using historical information only, 2) those using real-time information only, and 3) those using both historical and real-time information.

3.1.1 Historical Data Based Algorithms

The basic premise behind historical data based algorithms is that traffic patterns are cyclical. It is assumed that traffic at a given location repeats itself from day to day. In other words, knowledge of typical traffic conditions on Monday at 7:00 a.m. will allow one to predict the conditions on any particular Monday at 7:00 a.m. The prediction is given by

$$\hat{V}_{t+k} = V_{t+k}^{\text{hist}} \qquad \textbf{3-2}$$

where V_{t+k}^{hist} is simply the average flow rate, computed using historical data, at particular time-of-day and day-of-week associated with time $t+k$. The historical data based algorithm is attractive in that it requires no real-time data. However, the model has no way to react to dynamic changes, such as incidents, in the transportation system [1].

To overcome this limitation, historical data based algorithm can be extended to include the real-time data as follows:

$$\hat{V}_{t+k} = \frac{V_{t+k}^{\text{hist}}}{V_t^{\text{hist}}} V_t \qquad \textbf{3-3}$$

where V_t is the flow rate at current time interval, and $\frac{V_{t+k}^{\text{hist}}}{V_t^{\text{hist}}}$ is the projection rate of historical average from time interval t to $t+k$. A major weakness of this methodology is that it implicitly assumes that the projection ration will remain constant [2].

3.1.2 Time Serials Models

Since a greater of traffic data are observed in the form of a time serials which is a collection of observations made sequentially, time serials analysis is a valuable tool for traffic flow prediction. Many of the earlier time serials models used in short-term traffic flow forecasting has been reviewed by Han [4], and most recently, by De Gooijer [5], including moving average method, exponential smoothing method, ARIMA models, and Kalman filtering models.

The moving average method uses the average of n past observations for one-step forecast, i.e.,

$$\hat{V}_{t+1} = \frac{1}{n}\sum_{i=0}^{n-1} V_{t-i} \qquad \textbf{3-4}$$

Note that in such a model the n previous observations are equally weighted, and observations older than n are ignored. The advantage of moving average method is its flexibility in that observations are not forced into any particular patterns. However, it can not generate meaningful multi-period forecast and may become less reliable as the forecasting horizon gets longer.

The underlying model for exponential smoothing methods is given by

$$V_t = \mu_t + w_t, \quad t = 1,2,\cdots \qquad \textbf{3-5}$$

where μ_t represents the time-variant mean term and w_t is a white noise error term. The smoothing equation is

$$L_t = \alpha V_t + (1-\alpha)L_{t-1} \qquad \textbf{3-6}$$

where α denotes a smoothing parameter and L_t is the smoothed level that estimates μ_t. The k-step-ahead prediction is given by

$$\hat{V}_{t+k} = L_t, \quad \mathrm{k} = 1, 2, \cdots \qquad \textbf{3-7}$$

In other words, one forecasts V k-step ahead by using the last available estimated (smoothed) level state, L_t.

The parameter α in general should fall between zero and one. As α approaches zero, the forecast approaches a straight average of past observations, and as α approaches one, the forecast approaches random walk forecast. If parameter α is pre-determined from historical data, the exponential smoothing method is actually a heuristic model but with real-time update. If it is estimated from real-time data, the result is essentially a fitted ARIMA(0,1,1) model [6].

The double exponential smoothing (HOLT's) method is a more general form of simple exponential smoothing method mentioned above. The linear exponential smoothing model is given by the model equation

$$V_t = \mu_t + \beta_t t + w_t \qquad \textbf{3-8}$$

where β_t represents the time-varying slope (trend) term. The smoothing equations are

$$L_t = \alpha V_t + (1-\alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \qquad \textbf{3-9}$$

where T_t is the smoothed trend that estimates β_t, and γ is the trend smoothing weight. The k-step-ahead prediction equation is

$$\hat{V}_{t+k} = L_t + kT_t, \quad \mathrm{k} = 1, 2, \cdots \qquad \textbf{3-10}$$

If parameters α and γ are estimated from real-time data, HOLT's method is fitted ARIMA(0,2,2).

3.2 Methodology

3.2.1 Data Source and Preliminary Analysis

The data used for this development were loop data collected at El Camino Real corridor state highway route 82. Loop data of particular lanes are combined to associate with the signal phases. Advance loop data were used if they are available; otherwise presence loop data were used. The resolution of loop data is in 1 second.

At signalized intersections, traffic flow will be interrupted by traffic control device. As a consequence, using cycle based traffic volume as input to time serials model would be more appropriate than using the average crossing multiple cycles.

When evaluating dynamics systems models for traffic flow forecasting, it is certainly important to consider the accuracy of the model. However, it is equally important to the environment in which the model will be developed, used and maintained.

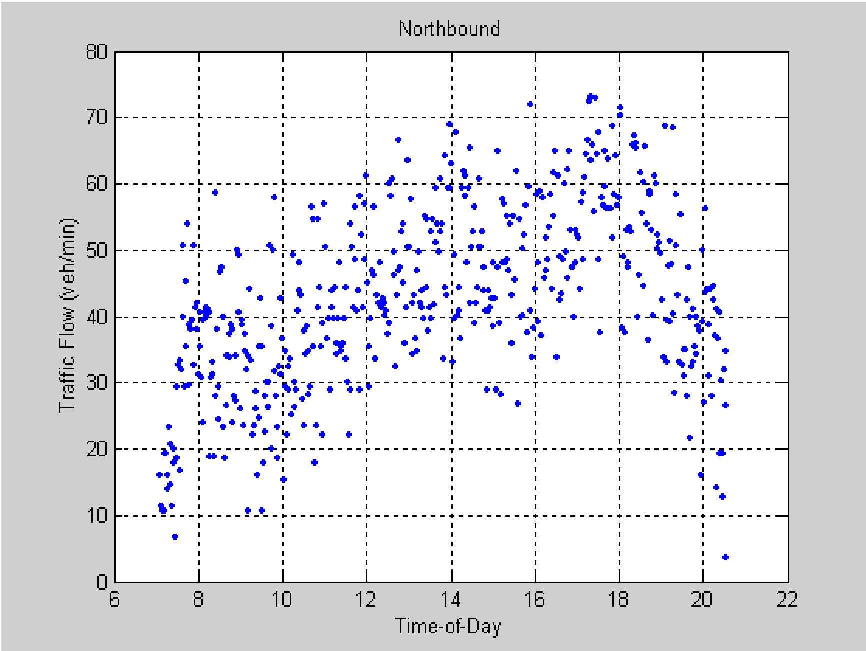

Figure 3-1 Traffic Flow v.s. Time-of-Day (Northbound)

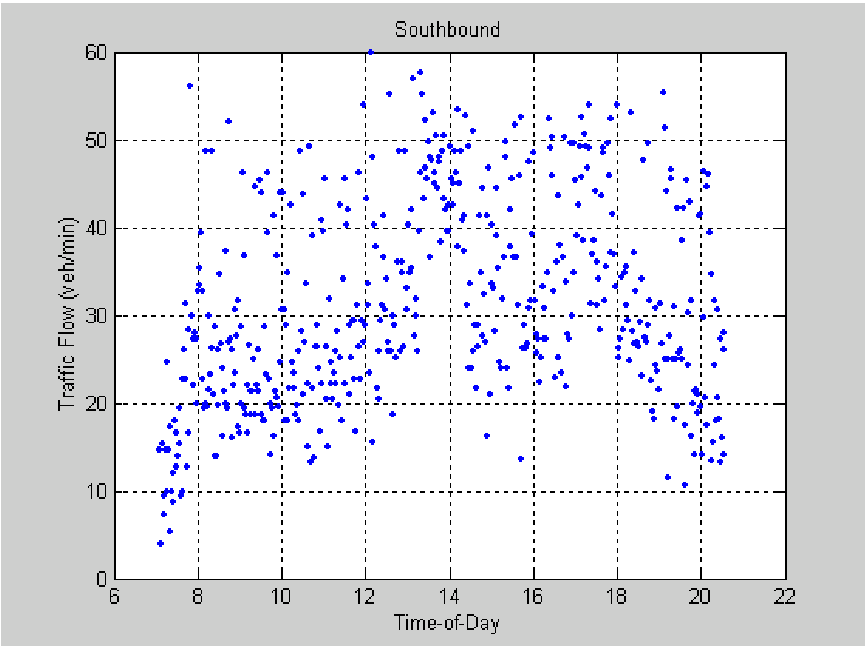

Figure 3-2 Traffic Flow v.s. Time-of-Day (Southbound)

Figure 3-1 and Figure 3-2 plot arrival traffic flow as function of time-of-day for northbound and southbound, respectively. We examined the increment in cycle-based traffic flow, i.e., $\Delta V_j \stackrel{\text{def}}{=} V_j - V_{j-1}$, where V_j is the traffic flow at the j-th signal cycle. The normal probability plots for increment in northbound and southbound are shown in Figure 3-3 and Figure 3-4, respectively.

As shown in Figure 3-3 and Figure 3-4, the data fit the straight line quite well, indicating the increment form a normal distribution. The Jarque-Bera test was performed to test for goodness-of-fit to a normal distribution. The p-value of Jarque-Bera test for northbound increment is 0.98 while the p-value for southbound increment is 0.73. Statistic T-test was also made to test whether the mean of the increment is zero or not. The p-value associated with T-test for northbound increment is 0.96 and the p-value for southbound increment is 0.99. In other words, both increment in northbound and southbound passed the hypothesis test that it has a normal distribution with zero mean. This implies that the cycle-based traffic flow follows the following random walk model:

$$V_j = V_{j-1} + \varepsilon_j \tag{3-11}$$

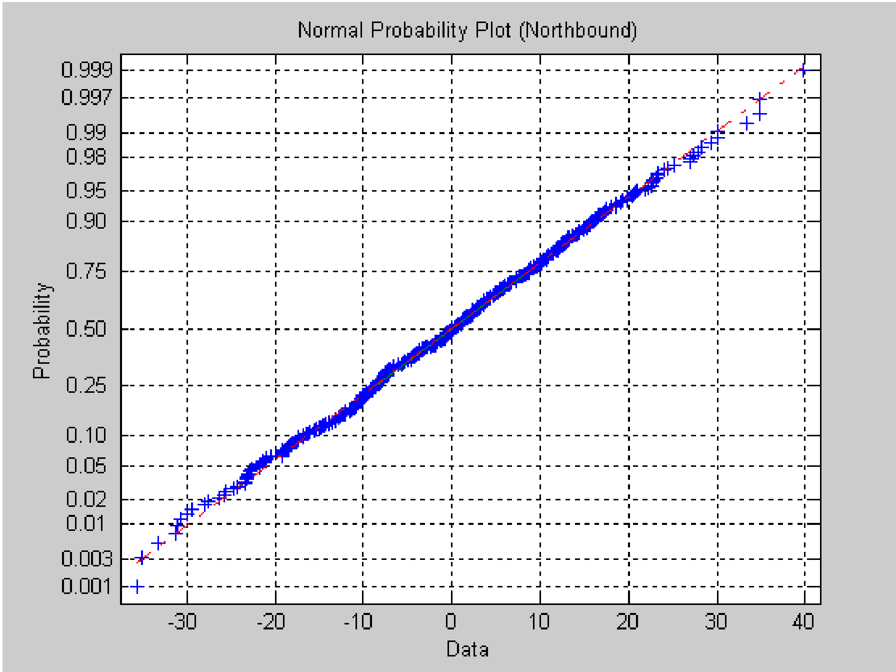

Figure 3-3 Normal Probability Plot for Northbound Increment

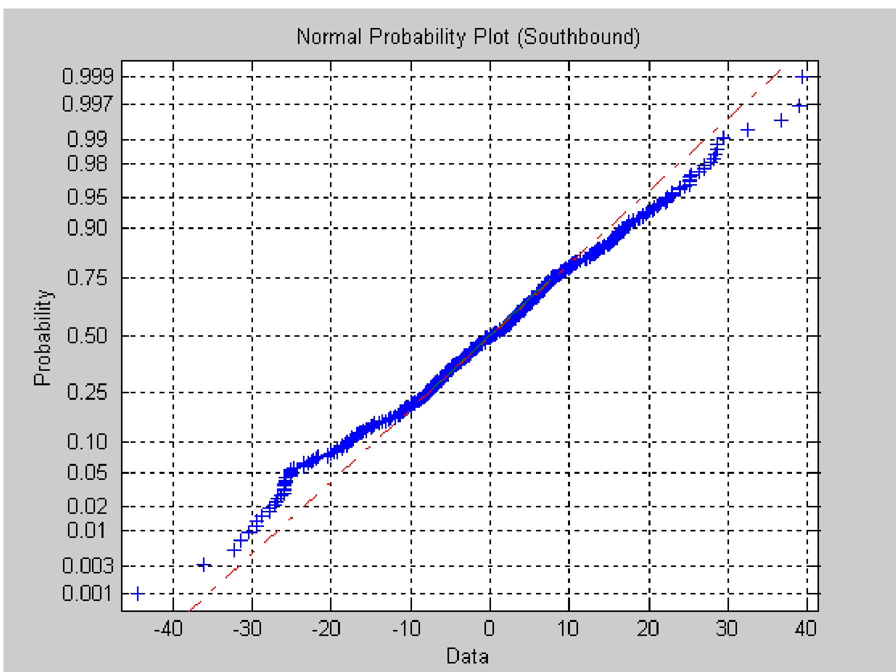

Figure 3-4 Normal Probability Plot for Southbound Increment

3.2.2 *Adaptive Recursive Least-Squares Method*

The following linear model is used to represent Equation 3-11

$$V_j = \sum_{i=1}^{n} w_i V_{j-i} + \varepsilon_j \qquad \textbf{3-12}$$

where n is the window size of past observations. The recursive least-squares (RLS) method as described in chapter 2 was then applied to estimate unknown weightings, w_i from previous and current observations. In other words, the weightings are adaptive to local changes. The parameter n is selected as 7, based on the significance of correlations.

3.3 Case Study and Results

The adaptive RLS method described in previous section is applied to predict cycle-based traffic flow, and the results are illustrated in Figure 3-1 and Figure 3-2 for northbound and southbound, respectively. As shown in those figures, the developed RLS model can capture the deterministic trend from stochastic random walk process.

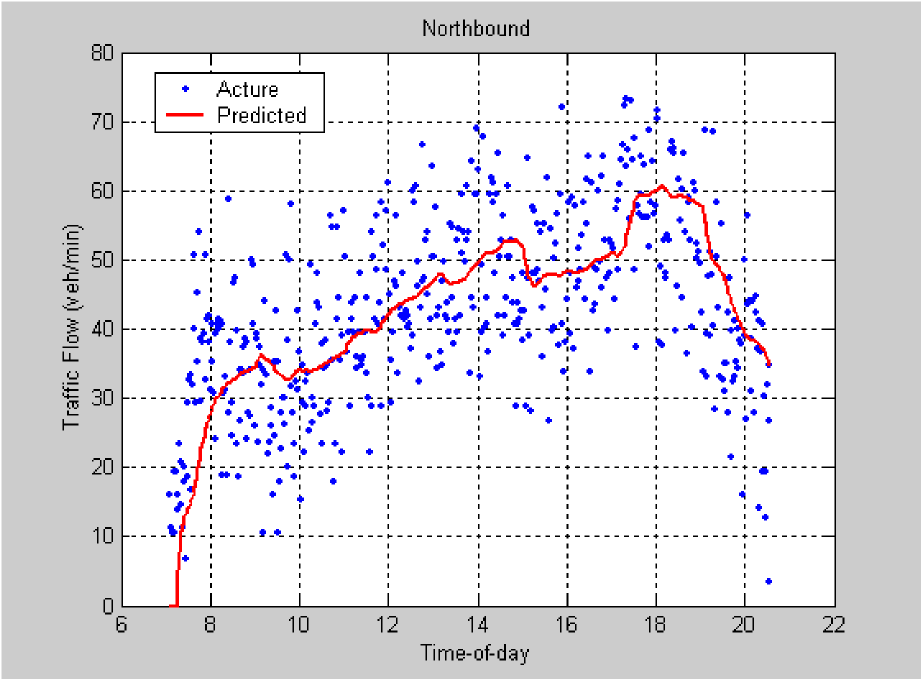

Figure 3-1 Predicted Traffic Flow (Northbound)

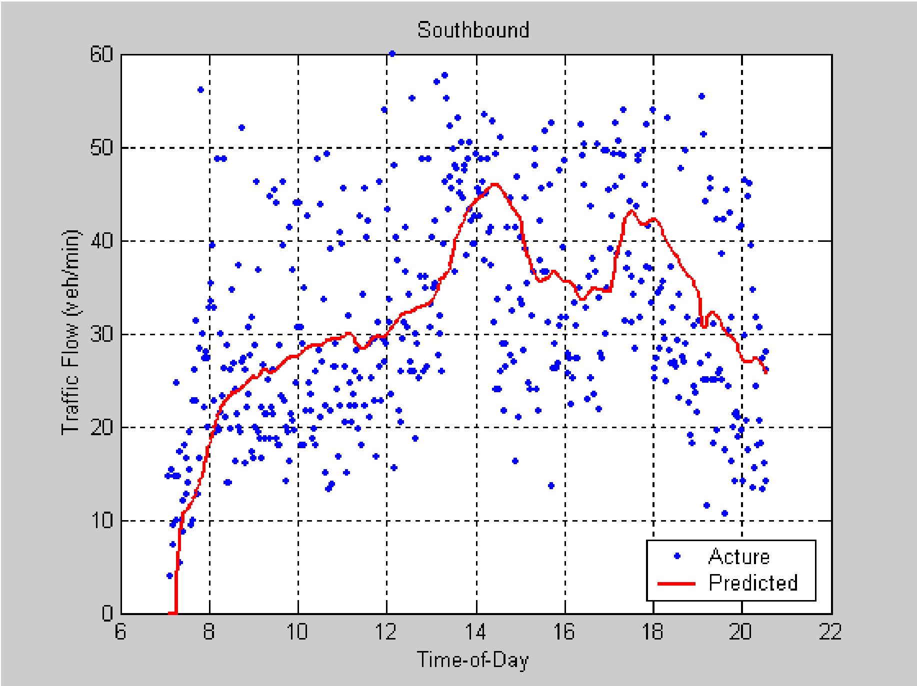

Figure 3-2 Predicted Traffic Flow (Southbound)

3.4 Reference

3-1 Y. Kamarianakis and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. Transportation Research Record 1857, Transportation Research Board, 2003.

3-2 I. Kaysi et al. An Integrated Approach to Vehicle Routing and Congestion Predictions for Real-time Driver Guidance. Transportation Research Records 1408, Transportation Research Board, 1993.

3-3 R. Chrobok et al. Different Methods of Traffic Forecasting Based on Real Data. European J. of Operational Research 155(2004), 558-568.

3-4 C. Han and S. Song. A Review of Some Main Models for Traffic Flow Forecasting. 2003 IEEE Proceedings of Intelligent Transportation Systems, Vol. 1, 216-219.

3-5 J.G. De Gooijer and R.J. Hyndman. 25 Years of IIF Time Serials Forecasting: A Selective Review. Working Paper, Monash University, Australia, May 2005.

3-6 C. Chatfield et al. A New Look at Models for Exponential Smoothing. The Statistician, 50(147-159), 2001.

3-7 R.K. Oswald et al. Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression. Research Report UVACTS-15-13-7, University of Virginia, December, 2001.

4 ADAPTIVE TSP ALGORITHMS ON ACTUATED SIGNAL CONTROLLERS

An adaptive TSP algorithm was developed to manipulate actuated signal control to grant priority to buses. The proposed TSP algorithm with an optimization model attempts to minimize the weighted sum of traffic delay and bus delay at an isolated intersection.

4.1 Model Assumptions

In order to simplify the optimization model, the following assumptions were made about intersection geometry, traffic demand and traffic signal settings.

- The model deals with isolated intersections along a corridor which is coordinated by an actuated system. As defined by National Electrical Manufacturers Association (NEMA) in Figure 4-1, movements 1, 6, 2, and 5 are on the main corridor streets; movements 4, 7, 3, and 8 are on cross streets. Movement 2 or 6 is the sync movement, which actually represents the coordination direction.
- Short-term traffic demand is stationary. Arrivals are uniformly distributed within each time interval.
- In this study, a new definition for signal cycle is used for the ease of calculation. In contrast to the traditionally defined signal cycle which references to the on/off of sync movement, the newly defined cycle starts from the onset of the cross street movements and ends after main street movements.
- The model is activated only when transit vehicles are expected to arrive during red periods.

- The model can change green splits for, at most, three consecutive cycles.
- TSP will not cause residual queues for any movement after the control cycles.
- Upstream streets and left-turn pockets are long enough to accommodate queues.
- All-red phases are not considered in delay calculations.
- Right-turns are not considered in delay calculations.

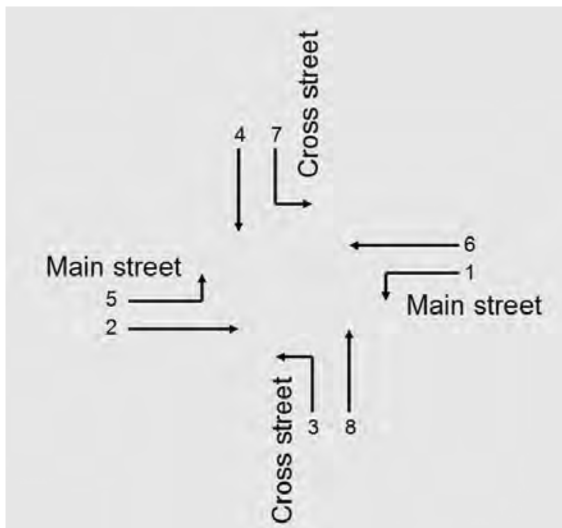

Figure 4-1 Definitions of standard NEMA 8 movements

4.2 Model Inputs and Outputs

As shown in Figure 4-2 Control input/output diagram, the optimization model has five real time inputs. An arrival time predictor (ATP) developed by PATH [9] generates and updates bus arrival time and its schedule lateness. Another real time input is predicted short-term traffic demand, which is obtained by using a moving average method that analyzes the time-series traffic counts from traffic detectors. Moreover, pedestrian call and online signal status come from the traffic signal controller in real time.

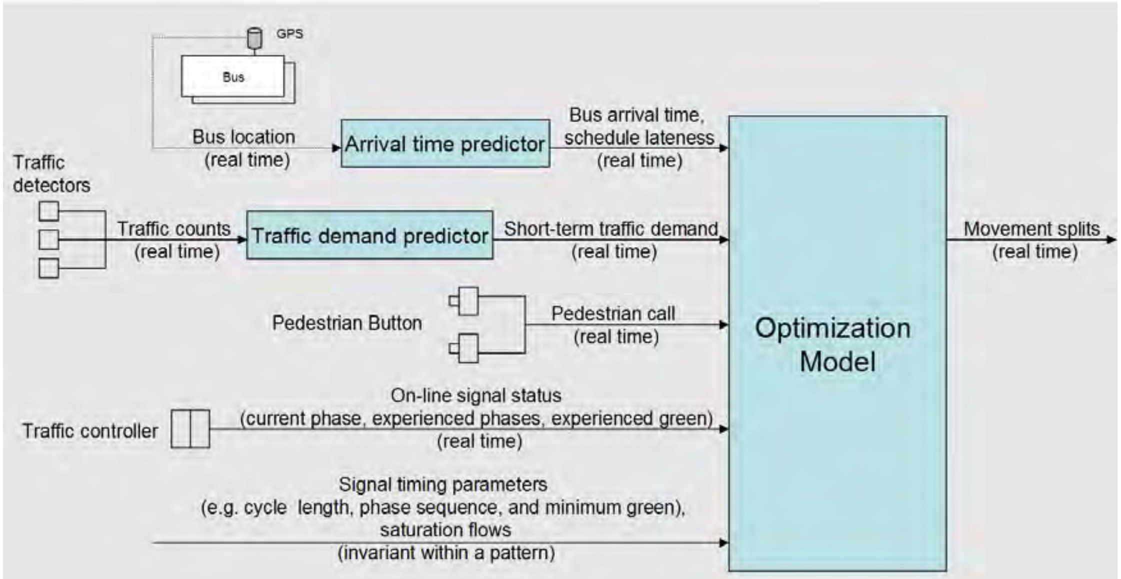

Figure 4-2 Control input/output diagram

In addition to the real-time inputs, the optimization model also has some static inputs such as signal timing parameters (e.g. cycle length, phase sequence, and minimum green) and saturation flows, which are invariant within a pattern typically defined by time of day (TOD), i.e. morning peak, middle of day, afternoon peak, night, etc.

The model outputs are priority requests in the form of movement splits. A recent study [9] validated that the 170E signal controller, a popular model for actuated systems, is capable of performing more adaptively through online updating of timing parameters, such as force-off points, gaps, maximum green, etc. However, an actuated control system cannot work the same as an adaptive system because of two constraints in their control logic. The first one is the cycle length constraint, which requires the duration between the end of the sync movement and the end of the next sync movement, which is the cycle length, to be a constant. The other constraint concerns movement sequence: no movement can be revisited before the cycle end. Essentially, the two constraints aim to keep all signals in coordination and make the control logic simple and applicable to specific type of controllers. Because neither of the two constraints can be overwritten, the proposed model outputs, which consist of adjusted timing parameters, must satisfy the two constraints.

4.3 Optimization Model

The parameters and decision variables used in the optimization model, as well as its objective function and constraints, are introduced and discussed in this section.

According to an aforementioned assumption, the optimization model manipulates movement splits for, at most, three cycles. We define the cycle that covers the predicted bus arrival time as cycle 1. Correspondingly, the previous and following cycles are labeled cycle 0 and cycle 2, respectively. The TSP model confirms accurate prediction results by the end of cycle 0, provides bus priority in cycle 1, and uses cycle 2 as a transition cycle to compensate other traffic for priority impacts. Noted that the signal cycles mentioned here and hereafter are based on the definition in Section 4.1.

Phase sequence can be defined by lead-lag relationships within the four conflicting movement pairs: 1&2, 3&4, 5&6, and 7&8. Lead-lag is a relative relationship, so four binary variables as defined in equation 4-1 can uniquely represent a phase sequence.

$$L_i = \begin{cases} 1, \text{if mov'} i \text{ is lead} \\ 0, \text{if mov'} i \text{ is lag} \end{cases}, \forall i = 1,3,5,7 \qquad \textbf{4-1}$$

For a better understanding of the formula, a numerical example is provided in **Error! Reference source not found.**. In this example, 1 and 7 are leading movements, while 3 and 5 are lagging, thus L_1=1, L_3=0, L_5=0, and L_7=1. Moreover, as defined by actuated control logic, movement 6, which ends earlier than movement 2, is the sync movement. Therefore, the phase sequence ending with sync movement is: 2&5, 4&7, 4&8, 3&8, 1&6, and 2&6, as shown in **Error! Reference source not found.**. It is noted that a phase is a combination of multiple movements such as 2&5.

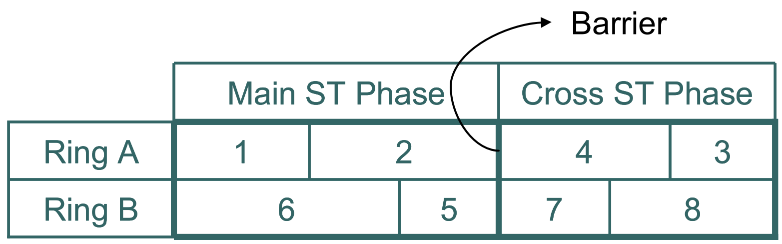

Figure 4-3 Phase sequence for an imaginary intersection

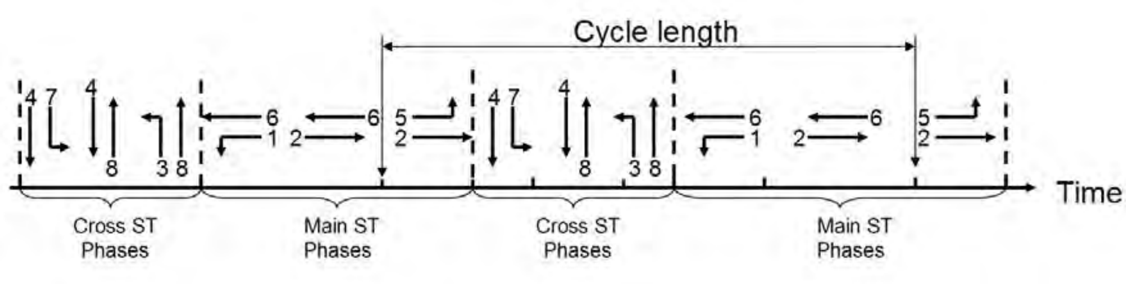

Figure 4-4 Expanded phase sequence for the example

4.3.1 Decision Variables

As depicted in Figure 4-5, cycle 1 and cycle 2 are control cycles when the TSP algorithm manipulates the movement splits. All the movement splits within control cycles are decision variables for the optimization model. In contrast, all other cycles are background cycles with timings that are untouchable by the TSP for the requesting bus.

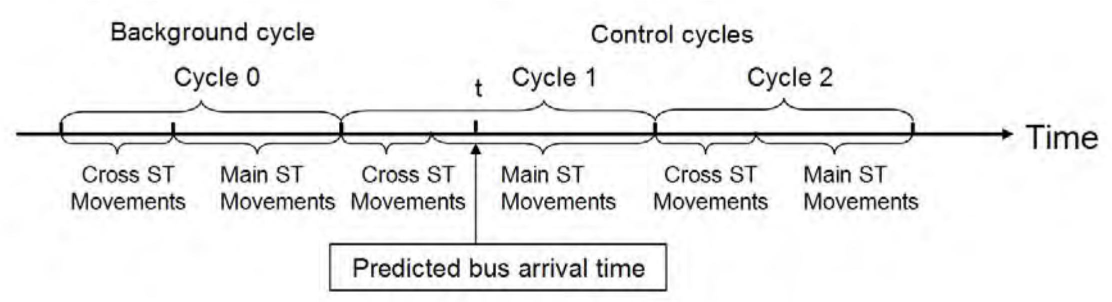

Figure 4-5 Background cycle and control cycles

The green split for movement *i* in cycle *j* is defined as g_{ji}. Similarly, r_{ji} is defined as the red time for movement *i* before green split g_{ji}.

$$g_{ji}, i = 1,2,3,4,5,6,7,8; j = 0,1,2 \qquad \textbf{4-2}$$

4.3.2 Constraints

The proposed optimization model has six sets of constraints: minimum green, cycle length, barrier, under-saturation, red-green relationship and real-time update.

4.3.2.1 <u>Minimum green constraint</u>

The minimum green constraint requires a minimum protected green for each movement. Pedestrian walking is consolidated into this minimum constraint. A pedestrian presence variable is defined by equation 4-3. When the pedestrian button is pushed, the minimum green for the corresponding movement is elongated to the protected "walk" and "flash don't walk" time, as described by equation 4-4.

$$Ped_{ji} = \begin{cases} 1, \text{if ped' button is pushed for mov'} \, i \text{ in cycle } j \\ 0, \quad \text{if no ped' info for mov'} \, i \text{ in cycle } j \end{cases} \quad (i = 1,..,8, j = 0,1,2) \qquad \textbf{4-3}$$

$$g_{ji} \geq (1 - Ped_{ji}) G_i^{\min} +, Ped_{ji} G_i^{Ped} \qquad (i = 1,..,8, j = 0,1,2) \qquad \textbf{4-4}$$

where $G_i^{\min}$ is the minimum green for movement *i*.

4.3.2.2 Cycle length constraint

The cycle length constraint, as described in Section 4.2, is formulated in equation 4-5. The first expression represents lead-lead phase sequential relationship while the second one represents lead-lag or lag-lag relationships. Lead or lag relationships mean the left-turn traffic is released before or after the opposing traffic.

$$\begin{cases} L_1 L_5 (C - \sum_{i=1}^{4} g_{ji}) = L_1 L_5 (C - \sum_{i=5}^{8} g_{ji}) & j = 1,2 \\ (2 - L_1 - L_5)[C - L_5 (g_{j-1,1} - g_{j1}) - L_1 (g_{j-1,5} - g_{j5}) - (1 - L_1)(1 - L_5)(g_{j-1,i} - g_{ji}) \\ - g_{ji} - g_{j,i+1} - g_{j,i+2} - g_{j,i+3}] = 0 & i = 1,5, j = 1,2 \end{cases} \qquad \textbf{4-5}$$

where C is cycle length.

4.3.2.3 Barrier constraint

The barrier constraint, as shown in equation 4-6, means ring A and ring B at the same side of the barrier have the same duration.

$$\begin{aligned} g_{j1} + g_{j2} &= g_{j5} + g_{j6} \\ g_{j3} + g_{j4} &= g_{j7} + g_{j8} \end{aligned} \qquad j=1,2 \qquad \textbf{4-6}$$

4.3.2.4 Under-saturation constraint

The under-saturation constraint, in the form of equation 4-7, is to guarantee no residual queues will be present after the two control cycles.

$$\frac{\lambda_i \sum_{k=j}^{2} (g_{ki} + r_{ki})}{\mu_i \sum_{k=j}^{2} g_{ki}} \leq 1, \qquad i = 1,\ldots,8,\ j = 1,2 \qquad \textbf{4-7}$$

where λ_i is traffic arrival rate for movement *i* and μ_i is saturation flow for movement *i*.

4.3.2.5 Red-green relationship

Without considering the predefined all-red period, a traffic signal must be showing green to some approach while showing red to other approaches. Such red-green relationship forms another constraint for the optimization model, as described in equation 4-8.

$$r_{ji} = \begin{cases} g_{j-1,i+1} + g_{j-1,i+2} + g_{j-1,i+3} - L_i(g_{j-1,i+2} + g_{j-1,i+3} - g_{j,i+2} - g_{j,i+3}) & i = 1,5,\ j = 1,2 \\ g_{j,i-1} + g_{j,i+1} + g_{j,i+2} & i = 2,6,\ j = 1,2 \\ g_{j,i-2} + g_{j-1,i-1} + g_{j,i+1} + L_{i-2}(g_{0,i-2} - g_{1,i-2}) + L_i(g_{0,i+1} - g_{1,i+1}) & i = 3,7,\ j = 1,2 \\ g_{j,i-3} + g_{j-1,i-2} + g_{j,i-1} + L_{i-3}(g_{0,i-3} - g_{1,i-3}) + (1 - L_{i-1})(g_{0,i-1} - g_{1,i-1}) & i = 4,8,\ j = 1,2 \end{cases}$$

4-8

4.3.2.6 Real-time update constraint

A real-time update constraint in the optimization model, as shown in equation 4-9, is to achieve the "adaptive" goal. One major advantage of such ATSP system is that the central control module can make specific decisions based on updated real-time information. If the control module is aware of the execution status of a movement, either skipped or ended, it will not consider the length of this movement $g_{ji}^{\exp}$ as a decision variable any more. While for other statuses, either ongoing or forthcoming, $g_{ji}^{\exp}$ will be another lower bound of the decision variable g_{ji} other than that in equation 4-4.

$$\begin{cases} g_{ji} \geq g_{ji}^{\exp} \\ M_{ji}(M_{ji}-1)g_{ji} \leq M_{ji}(M_{ji}-1)g_{ji}^{\exp} \end{cases} \qquad j=0,1,2; i=1,\ldots,8 \qquad \textbf{4-9}$$

where:

$g_{ji}^{\exp}$: experienced green time for movement *i* in control cycle *j*;

M_{ji} : execution status for movement *i* in control cycle *j*;

$$M_{ji} = \begin{cases} 0, & \text{if the movement is not started yet} \\ 1, & \text{if the movement is ongoing} \\ 2, & \text{if the movement is ended or skipped} \end{cases} \qquad i=1,\ldots,8; j=0,1,2$$

4.3.2.7 <u>The Example</u>

For the numerical example shown in the beginning of this section, cycle length constraint 4-5 and under-saturation constraint 4-7 are (for cycle $j=1,2$):

$$\begin{aligned} & g_{j-1,5} + g_{j6} + g_{j7} + g_{j8} - g_{j5} = C \\ & g_{j,1} + g_{j2} + g_{j3} + g_{j4} + g_{j-1,5} - g_{j5} = C \end{aligned} \qquad \textbf{4-10}$$

$$\begin{aligned} & \frac{\lambda_i(g_{2i}+r_{2i})}{\mu_i g_{2i}} \leq 1, \forall i=1..8 \\ & \frac{\lambda_i(g_{1i}+r_{1i}+g_{2i}+r_{2i})}{\mu_i(g_{1i}+g_{2i})} \leq 1, \forall i=1..8 \end{aligned} \qquad \textbf{4-11}$$

And red-green relationship 4-8 can be used to calculate red time for all movements. For example, we calculate r_{24}. As shown in Figure 4-4, the phases between movement 4 in cycle 1 and in cycle 2 are 3&8, 1&6, 2&6, and 2&5. All the phases are in cycle 1. Therefore, we have:

$$r_{24} = g_{13} + g_{11} + g_{12} \qquad \textbf{4-12}$$

4.3.3 Objective Function

Before discussing the objective functions, we need to determine the actuated signal timing without TSP. In the simplified semi-actuated control logic, green time is terminated when a detected gap in the platoon exceeds a predefined value. If we assume all movements are under-saturated and traffic arrivals are uniformly distributed, then the green splits for this background cycle can be calculated as a linear programming problem:

$$\max_{G_i} G_2 + G_6 \qquad \textbf{4-13}$$

$$\text{Subject to: } \frac{\lambda_i C}{\mu_i G_i} \leq 1, \forall i = 1,..,8$$

$$\sum_{i=1}^{8} G_i = C$$

$$G_1 + G_2 = G_5 + G_6$$

$$G_3 + G_4 = G_7 + G_8$$

where: G_i is the green split for movement *i* in the background cycle.

The value of the objective function is total delay, which consists of traffic delay and weighted bus delay. For traffic delay, we have two scenarios for each movement:

1. No Residual queue in cycle 1
2. Residual queues exist in cycle 1 but not in cycle 2.

Thus traffic delays in cycle 0, 1, and 2 are calculated by equation 4-14, which can be consolidated into formula 4-15.

$$\begin{cases} \text{Scenario I}: \ d_i = \frac{\mu_i}{2}\rho_i r_{0i}^2 + \frac{\mu_i}{2}\rho_i ({r_{1i}}^2 + {r_{2i}}^2) \\ \text{Scenario II}: d_i = \frac{\mu_i}{2}\rho_i r_{0i}^2 + \frac{\mu_i}{2}\rho_i (r_{1i} + r_{2i})^2 - r_{2i}\mu_i g_{1i} \end{cases} \quad \textbf{4-14}$$

$$d_T = \sum_{i=1}^{8} [\frac{\mu_i}{2}\rho_i (r_{1i} + r_{2i})^2 - r_{2i}\mu_i \min(g_{1i}, \rho_i r_{1i}) + \frac{\mu_i}{2}\rho_i r_{0i}^2] \quad \textbf{4-15}$$

where: $\rho_i = \dfrac{\lambda_i}{\mu_i - \lambda_i}$

The ideal signal priority scheme is to manipulate signal timings to allow transit vehicles to pass through without disturbing other traffic. If the optimization model could have long enough lead time, it can easily shift the transit movement back or forth in control cycle 1 and then compensate other movements in control cycle 2. Per an aforementioned assumption, the optimization model is activated only when a transit vehicle is expected to arrive during its red period, i.e., either before or after its normal green. If a transit vehicle is expected to arrive before its normal green in control cycle 1, the optimization model would make the decision before control cycle 1 to reducee green time for the movements before the transit movement. Such a TSP strategy is so-called early green strategy.

Green extension, the other popular TSP strategy, will be executed when a transit vehicle is predicted to arrive right after its normal green phase. There are two scenarios for the green extension strategy: (1) when the transit movement is not the last movement of a cycle, it can be extended within the current cycle; and (2) when the transit movement is the last movement of a cycle, normally movement 2 or 6, the movement cannot be extended theoretically because of the cycle length constraint. However, the extended green can be executed technically by either fixing the controller timer or inserting an extra green period at the end of the previous cycle. Both of these TSP treatments need special TSP functions in the controller software[9]. Moreover, such green extensions would impact existing coordination for main street phases. As a result, traffic engineers typically have some restrictions for this strategy, for example, the extended green cannot be longer than 10% of its cycle length. Therefore, a specific objective function is designed for this special green extension strategy.

Most rapid transit services which need TSP are running along major corridors, thus for the ease of demonstrating the model, we assume buses discussed in the following examples are running on movement 2 and 6. Hereby, we defined a binary variable in equation 4-16 to indicate buses' running directions. It should be noted that the proposed model does not require transit services to be running along a major corridor, the mathematical equations listed below need only minor changes, for example adding a heading (compass direction) variable, to accommodate buses running on other approaches.

$$B = \begin{cases} 1, & \text{if bus is on movement 2} \\ 0, & \text{otherwise bus is on movement 6} \end{cases} \qquad \textbf{4-16}$$

The predicted bus arrival time t_{bus} is referenced to the end of a sync movement or a real clock [9]. To compute bus delay, we converted t_{bus} into T_{bus}, which is referenced to the end of green for the bus approaching direction;

$$T_{bus} = t_{bus} + B(1-L_1)g_{01} + (1-B)(1-L_5)g_{05}. \qquad \textbf{4-17}$$

When implementing TSP, the model shrinks the red time for a bus's approaching direction, from *R'* to *R*, as shown in Figure 4-6. At T_{bus}, the bus is expected to arrive at the intersection to join a standing queue. The number of queued vehicles in front of the bus is N_T. The corresponding queue discharging time is bus delay d_{bus} because the bus leaves the intersection at $T_{bus} + d_{bus}$. The queue disappears at t_q, which can be computed by formula 4-18.

$$t_q = R + B\rho_2 r_{12} + (1-B)\rho_6 r_{16} \qquad \textbf{4-18}$$

where: R is the red time for the bus movement, $R = Br_{12} + (1-B)r_{16}$.

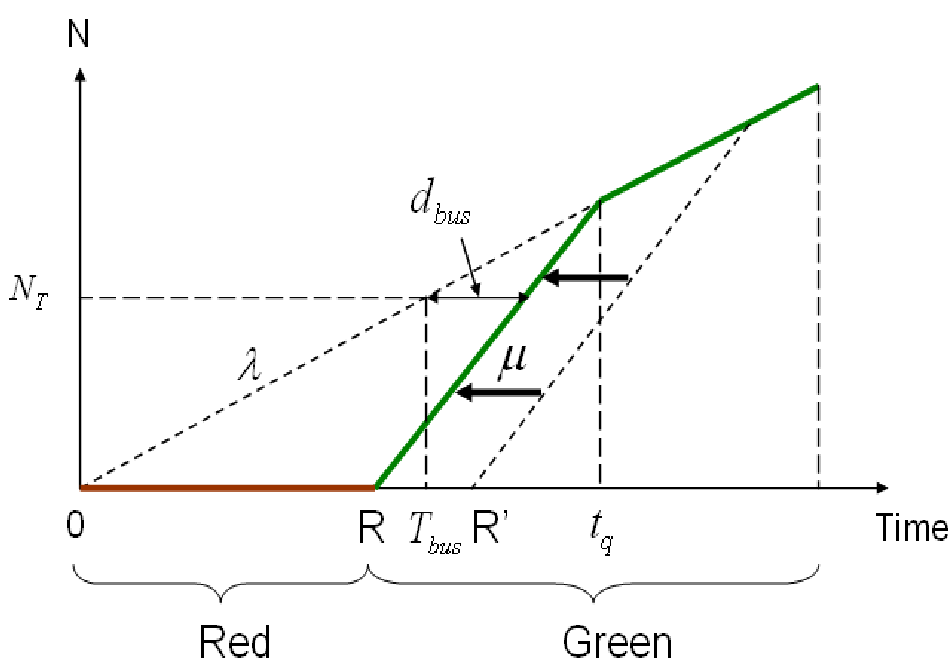

Figure 4-6 N-T curve for bus movement with TSP

The geometry of Figure 4-6 was used to derive the relationship 4-19. Accordingly, bus delay can be obtained by equation 4-20 below.

$$\frac{d_{bus}}{R} = \frac{\max(t_q - T_{bus}, 0)}{t_q} \quad \textbf{4-19}$$

$$d_{bus} = \frac{R}{t_q} \max(t_q - T_{bus}, 0) \quad \textbf{4-20}$$

Therefore, the objective function for normal TSP is:

$$\min d = \sum_{i=1}^{8} [\frac{\mu_i}{2} \rho_i (r_{1i} + r_{2i})^2 - r_{2i}\mu_i \min(g_{1i}, \rho_i r_{1i}) + \frac{\mu_i}{2} \rho_i r_{0i}^2] + w_b \frac{R}{t_q} \max(t_q - T_{bus}, 0) \quad \textbf{4-21}$$

where: w_b is the weighting factor for buses.

For the special green extension strategy, the green movement in cycle 0 is extended until the approaching bus leaves the intersection. Therefore bus delay is zero. So the objective function for the special green extension strategy is:

$$\min d = \sum_{i=1}^{8} [\frac{\mu_i}{2} \rho_i (r_{1i} + r_{2i})^2 - r_{2i}\mu_i \min(g_{1i}, \rho_i r_{1i}) + \frac{\mu_i}{2} \rho_i r_{0i}^2] \quad \textbf{4-22}$$

Notice that the extended green makes the following two changes on signal timings and constraints in cycle 0 and cycle 1.

$$\begin{aligned} g_{02} &= G_2 + BT_{bus} \\ g_{06} &= G_6 + (1-B)T_{bus} \end{aligned} \quad \textbf{4-23}$$

$$\begin{aligned} &L_1 L_5 (C - T_{bus} - \sum_{i=1}^{4} g_{1i}) = L_1 L_5 (C - T_{bus} - \sum_{i=5}^{8} g_{1i}) = 0 \\ &(2 - L_1 - L_5)[C - T_{bus} - L_5(g_{01} - g_{11}) - L_1(g_{11} + g_{05} - g_{15}) - g_{12} - g_{13} - g_{14}] = 0 \\ &(2 - L_1 - L_5)[C - T_{bus} - L_1(g_{05} - g_{15}) - L_5(g_{15} + g_{01} - g_{11}) - g_{16} - g_{17} - g_{18}] = 0 \end{aligned} \quad \textbf{4-24}$$

4.4 Computation Procedure and Online Update

The two objective functions, 4-21 and 4-22, are both quadratic and all of the constraints are linear. However, there are minimum functions in the quadratic terms, thus they are not standard quadratic programming problems. Moreover, general actuated controllers work on 10 hertz, so their working frequency is 10 times per second. As a result, the decision variables in the model are not integers. Therefore, the two optimization problems are not integer programming ones either. By converting the minimum function into a concave linear constraint, we have a standard quadratic programming problem with 25 decision variables. The optimization package provided by MATLAB is then used to solve the problems.

Given a set of traffic signal information and transit movement data, the optimization model can output two sets of movement splits for normal TSP strategy and the aforementioned special green extension strategy, respectively. Meanwhile, there are two optimal values for the two objective functions, both of which are the weighted sum of traffic delay and bus delay. At the end, the strategy with the smaller weighted delay is chosen as the optimal strategy for the coming bus. The final priority request, which is the output to PRS, consists of the corresponding movement splits.

4.5 Case study

An analysis of the numerical example is provided here to illustrate the performance of the proposed model. In this example, the intersection has four lanes on each main street approach, one of which is a left-turn lane. On cross streets, one of the three lanes is a left-turn lane. Table

4-1 lists all basic signal information under a medium-congested scenario whose saturation degree is 0.67. Pedestrian buttons are not pushed in this case.

Table 4-1 Intersection information for the numerical example

Movement	1	2	3	4	5	6	7	8
Minimum Green (sec)	4	6	4	6	4	6	4	6
Demand (veh/hour)	200	1200	200	800	200	1200	200	800
Saturation flow (veh/hour)	1200	5400	1200	3600	1200	5400	1200	3600
Green split (sec)	20	53	20	27	20	53	20	27
Delay (sec/veh)	50	23.8	50	46.7	50	23.8	50	46.7

Suppose a bus is coming along movement 6, its arrival time is predicted by an ATP. The proposed TSP algorithm is activated by the inputs, thereby runs the optimization model and outputs movement splits to the signal controller. Table 4-2 presents the performance of the TSP algorithm under the medium-congested scenario. In the table, the bus movement is movement 6, while others are non-bus movements. Both of the average and total vehicle delays reported are the total for cycles 0, 1, and 2. The first case, with a weighting factor of 1, is the "no TSP" base case for comparison. The objective of this case is to minimize total vehicle delay including bus delay, so its logic is similar to an adaptive signal. Essentially, there is no TSP for buses in this case because their delay is treated the same as that of other traffic.

As weighting factor increases, the coming bus has relatively higher priority over other traffic. Accordingly, the bus delay is reduced and other traffic delay increases. For active rule-based TSP systems, the priority for buses also favors traffic along bus movements [9]. However, it is not that obvious for adaptive TSP because such the system optimally allocates the disturbance of bus priority treatment to all other traffic. The elongated bus movement in cycle 1 would incur a chopped green for this movement in cycle 2 because other movements need to be compensated in the transition cycle. Therefore, average vehicle delays for traffic on both bus movement and non-bus movements are increasing as the weighting factor goes up.

Table 4-2 Performance of the TSP algorithm in a medium-congested scenario

Weighting factor	Average vehicle delay (sec/veh)						Total vehicle delay	
	Bus		Bus movement		Non-bus movements			
	(sec)	Diff'	(sec)	Diff'	(sec)	Diff'	(sec)	Diff'
1 (Ref' no TSP)	10.24	0.00%	19.05	0.00%	37.46	0.00%	15782.41	0.00%
50	4.98	-51.34%	19.27	1.13%	37.65	0.52%	15877.40	0.60%
100	3.12	-69.54%	19.37	1.67%	38.02	1.50%	16018.45	1.50%
150	1.1	-89.24%	19.48	2.24%	38.68	3.25%	16263.63	3.05%
200	0.2	-98.04%	19.60	2.88%	39.07	4.29%	16416.62	4.02%
250	0.14	-98.61%	19.64	3.06%	39.09	4.35%	16428.58	4.09%
300	0.02	-99.85%	19.66	3.19%	39.18	4.60%	16464.55	4.32%
350	0.00	-100%	19.70	3.38%	39.18	4.60%	16469.16	4.35%
400	0.00	-100%	19.70	3.38%	39.18	4.60%	16469.16	4.35%
Actuated signal (no TSP)	23.81	132%	23.81	24.97%	39.79	6.22%	17204.77	9.01%

Furthermore, bus delays are more sensitive to weighting factor changes than other traffic delays are. When the weighting factor is 50, the average bus delay under a different predicted arrival time is reduced by 51%, while traffic delay on the bus movement approach and the non-bus movement approach is increased by only 1.13% and 0.52%, respectively. When the weighting factor is increased to 350, buses experience no delay no matter when they arrive at the signal; while average vehicle delays for other traffic rise 0.65 sec and 1.72 sec only. The proposed adaptive TSP system works well in the medium-congested scenario because it can significantly reduce bus intersection delay without incurring too much extra delay for other traffic.

For the case shown in the last row of Table 4-2, the signal is controlled by an actuated control system. Neither this case nor the reference case provides TSP to buses. Traffic delay for the actuated case is the same as those in Table 3. Average bus delay is calculated using 4-20 and 4-25.

$$\overline{d_{bus}} = \frac{\int_0^{t_q} \frac{R}{t_q}(t_q - t)dt}{C} = \frac{(1+\rho_6)(C - g_{06})^2}{2C} \qquad \textbf{4-25}$$

As shown clearly in Table 4-2, the proposed adaptive TSP system outperforms the TSP function residing within the actuated signal in terms of both bus delay and other traffic delay. The reason is that the typical actuated signals lack an objective function, thus they apply rule-based logic rather than optimization methods to allocate time resource among conflicting movements. Therefore actuated systems do not utilize time resources as optimally as adaptive systems do.

A recent study [10] recommends TSP for medium- or low-congested conditions because active TSP systems always incur significantly delay on non-bus movements in higher congested scenarios. The proposed adaptive TSP system, however, can use the weighting factor to control the impact on other traffic even under conditions with heavily congested traffic. Table 4-3 illustrates the performance of the adaptive TSP algorithm in a scenario with a saturation degree of 0.89. Similar to the results in Table 4-2, the average bus delay reduces dramatically as the weighting factor increases, while the average delay for other traffic shows a much slower increase. However, the delays experienced by other traffic are more significant than those under medium-congested conditions. The reason is that time resources among conflicting traffic are scarcer when traffic is highly congested.

Table 4-3 Performance of the TSP algorithm in a heavily congested scenario

Weighting factor	Average vehicle delay (sec/veh)				Total vehicle delay		GE/EG switch point (sec)
	Bus		Other traffic				
	(sec)	Diff'	(sec)	Diff'	(sec)	Diff'	
1 (Ref')	25.09	0.00%	38.98	0.00%	22475.62	0.00%	No TSP
100	10.95	-56.38%	40.22	3.19%	23190.88	3.18%	22
200	5.45	-78.29%	41.53	6.55%	23932.43	6.48%	31
300	1.69	-93.27%	43.11	10.60%	24834.72	10.50%	36
400	0.96	-96.19%	43.58	11.80%	25102.70	11.69%	38
500	0.31	-98.75%	44.13	13.22%	25421.80	13.11%	40
600	0.27	-98.92%	44.17	13.32%	25444.25	13.21%	40
Actuated signal	36.36	44.91%	40.85	42.83%	23565.10	4.85%	No TSP

If the predicted arrival time falls at the beginning of a cycle, it is better to extend green than to provide early green because the early green strategy needs to shrink all of the other green splits, so more movements are influenced. When the bus arrival time is closer to the end of a cycle, the impact of a green extension is larger than that of early green. Thus, there is a specific point in bus arrival time when the optimal TSP strategy switches from green extension to early green. This point is named GE/EG switch point. As shown in the right most column in Table 4-3, when the weighting factor increases, the bus delay has more weight over the delays of other movements, so the GE/EG switch point moves towards the cycle end.

Figure 4-7 and Figure 4-8 show average bus and other traffic delays versus weighting factor and predicted arrival time in the heavily congested scenario, respectively. Neither of the two surfaces is smooth because delays are nonlinear functions of arrival time. They have breaks when the system decides to switch its strategy from green extension to early green. According to the trends of the surfaces, we can see bus delays decrease as the weighting factor grows and the arrival time increases, while other traffic delays rise as the weighting factor grows. When bus arrival time falls at either end of a cycle, TSP algorithm can easily extend green or do nothing to manipulate TSP requests. When bus arrival time falls in the middle of a cycle, the TSP algorithm has to provide priority which would incur larger delay for other traffic. Therefore, other traffic delays peak when arrival time is at the middle of a cycle. Moreover, the peak value increases with the increase of weighting factor.

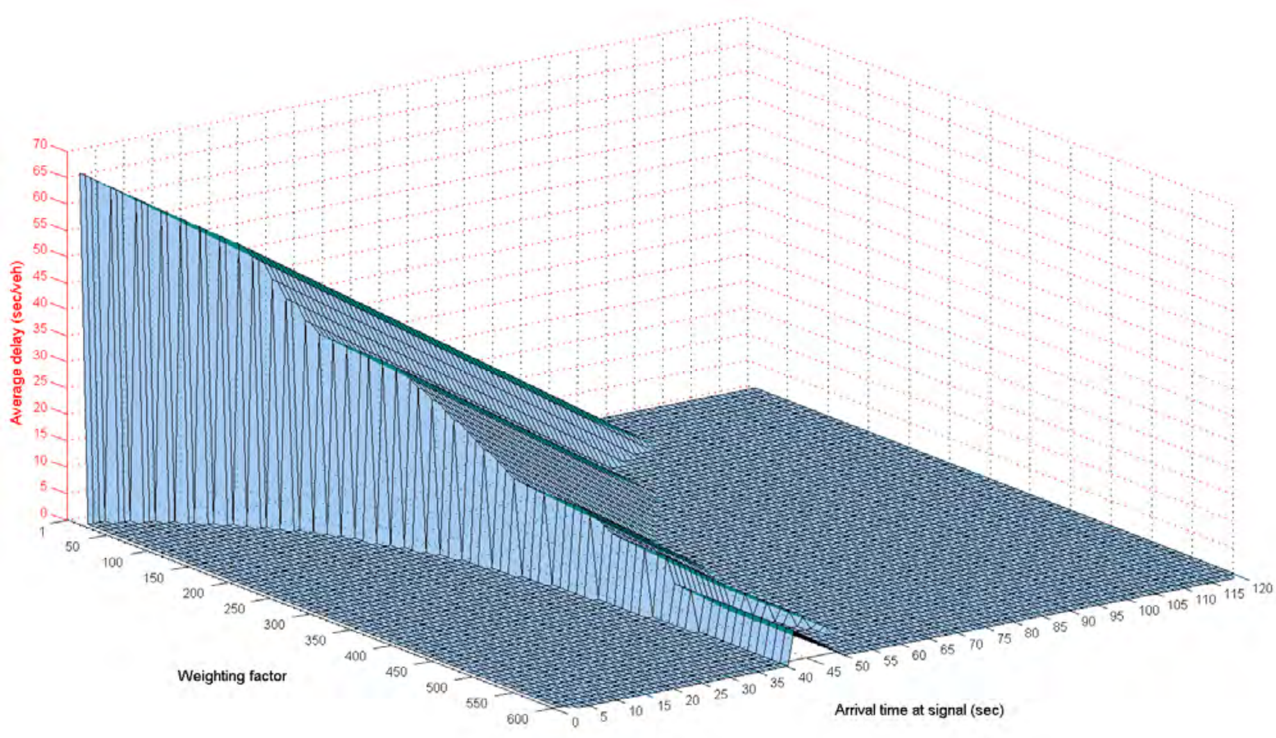

Figure 4-7 Average bus delays in the heavily congested scenario

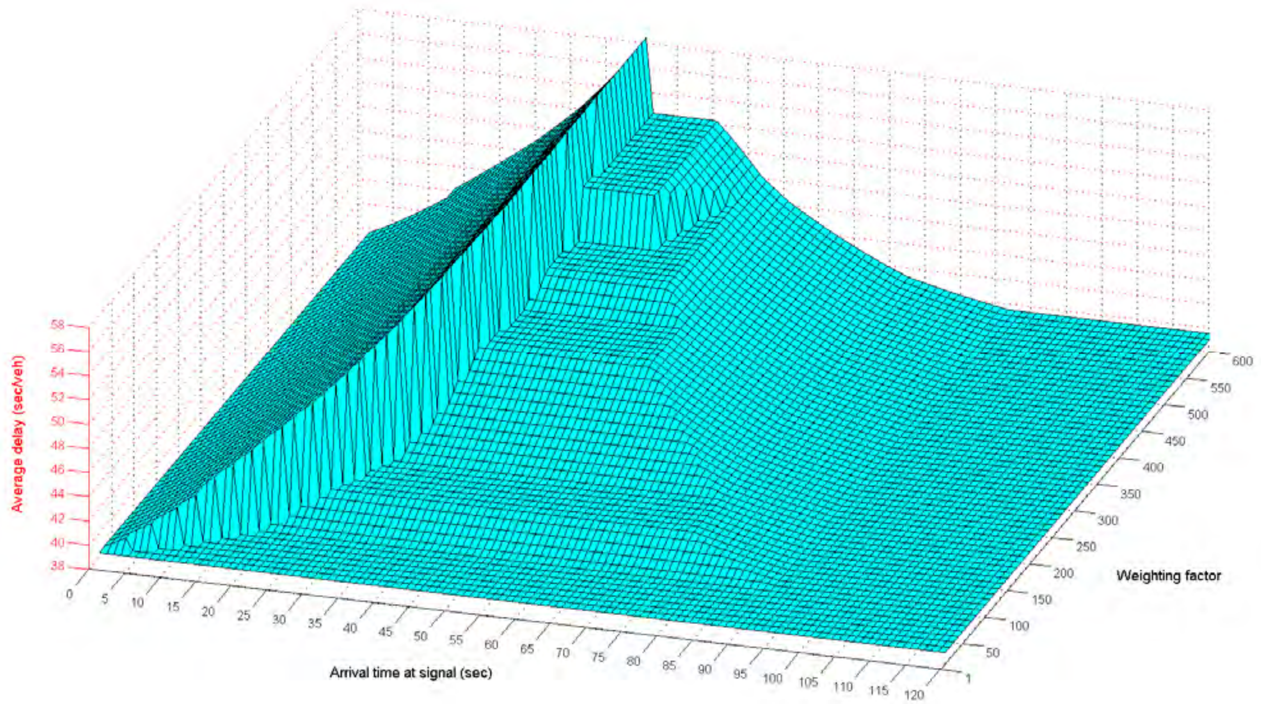

Figure 4-8 Intersection delay for other traffic in the heavily congested scenario

4.6 Discussions

Among the assumptions listed in Section 4.1, some can be relaxed by making minor modifications to the optimization model. For example, when there is no left-turn signal for a leg of an intersection or when the intersection is T-shaped, we can simply add constraints that specify zero as the maximal value for the corresponding split. Moreover, as aforementioned, when buses are not running on movement 2 or 6, we can add a heading (compass direction) variable for each movement indicating where buses are coming from. Note however, in such cases, the early green strategy is the only means to adjust signal timing, which is to shrink red or extend green in cycle 1. Extra green cannot be added at the beginning of cycle 1 because those movements cannot be at the end of a cycle. So the optimization model can be simplified by only searching for optimal early green strategy. As to the assumption of sufficiently long upstream streets and left-turn pockets to accommodate queues, a back of queue constraint can be added to circumscribe maximum queue length.

A strong assumption in this study is that traffic arrivals follow a uniform distribution. However in the real world, the traffic arrival without upstream controls is a Poisson process. If upstream controls exist, vehicles form platoons at upstream intersections. For delay calculations, platoons and their dispersion will make the formulation used in this study much more complicated. To relax this assumption, the delay in objective functions can be modified by equation 4-26 or by Webster formula [12].

$$\overline{w} = \frac{1}{2}\mu\rho R^2 + \frac{\rho(R+G)\Delta}{2(G-\rho R)} \qquad \textbf{4-26}$$

where: Δ is a measure of the variability in traffic arrivals.

In addition, one of the advantages of actuated signal control is its ability to deal with fluctuation arrivals. In such sense, we might have overestimated the delays for actuated signal in Tables 4-2 and 4-3 because we assume no fluctuation in traffic arrivals.

Another real time input, which is the predicted arrival time, also has variance. One option to accommodate the variance is to conduct sensitivity analysis. Such analysis can reveal the impacts of the variance on optimal solutions. Besides, the TSP algorithm can keep updating priority requests as a bus gets closer to the intersection because the prediction error reduces as the bus gets closer.

The proposed adaptive TSP is applicable for Selective Vehicle Detection (SVD) systems that identify specific vehicle types or conditions and also for near-side bus stops if TSP starts only after the bus departs the stop. The only difference is there will be fewer decision variables for the optimization model. Because time is already in the middle of cycle 1 when a bus departs from a near-side bus stop or touches a SVD detector and activates the algorithm, some movements have already passed. Such cases give the traffic control system only limited lead time to borrow “seconds” from other traffic and thus result in timings that are not as efficient as those for far-side bus stops or AVL/GPS systems.

4.7 Conclusion

This chapter reports the development of an adaptive TSP algorithm in closed-loop actuated control systems. The performance of the adaptive TSP algorithm is illustrated in a numerical example under medium congested and heavily congested conditions.

The purpose of the study is to provide transportation authorities with a cost-efficient way to achieve adaptive TSP in the state-of-the-practice traffic control systems. Moreover, it provides a quantitative model to balance the benefit and impact of TSP with other traffic. Given a specific traffic situation, the optimization model can make cost-benefit-analysis for different weighting factors. In future studies, the weighting factor could be a function of factors such as maximum allowed movement delays, maximum additional delay, maximum vehicle delay, longest queues, number of transition cycles, transit headways, or schedule lateness. In other words, the TSP algorithm can work with these specific factors instead of ambiguous weighting factors.

4.8 Reference

1. Baker, R.J. et al. An Overview of Transit Signal Priority, White Paper, Intelligent Society of America, July 11, 2002.
2. Lowrie, P.R. The Sydney Coordinated Adaptive Traffic System (SCATS) – Principles, Methodology, Algorithm. Proceeding of International Conference on Road and Road Traffic Signaling, London, 1982.
3. Hunt, P.B., D.I. Robertson, and R.D. Bretherton. SCOOT, A Traffic Responsive Method of Coordinating Signals. TRRL Laboratory Report 1014. Transportation and Road Research Laboratory, Crowthorne, Berkshire, U.K. 1981.
4. Turksma S. An Overview of UTOPIA-SPOT Development in Northern Europe. Paper Presented in 8th ITS World Congress, 2001, Sydney.
5. Mirchandani, P., L. Head, A. Knyazyan and W. Wu. An Approach towards the Integration of Bus Priority and Traffic Adaptive Signal Control. Preprint CD-ROM of Transportation Research Board 80th Annual Meeting, 2001, Washington, D.C.
6. Smith, B. et al. Data-Driven Methodology for Signal Timing Plan Development: A Computational Approach. Computer-Aided Civil and Infrastructure Engineering, 17, pp.387-395, 2002.
7. Luyanda, F., Gettman, D., Head, L. ACS-Lite Algorithmic Architecture. Report by Federal Highway Administration, 2003.
8. AASHTO/ITE/NEMA. National Transportation Communications for ITS Protocol: Objective Definitions for Signal Control and Prioritization. NTCIP 1211, V01.26, September 2002.
9. Zhou, K., Li, M., Yin, Y., Zhang, W-B, Skabardonis, A.,. Development of Adaptive Transit Signal Priority Systems, University of California, Berkeley: Institute of Transportation Studies, California PATH Program, PATH Research Report UCB-ITS-PRR-2004 (Draft), August 2004.

10. Rakha, H., et al. Sensitivity Analysis of Transit Signal Priority Impacts on Operation of a Signalized Intersection, *Journal of Transportation Engineering*, Vol. 130, No. 6, November/December 2004, pp. 796-804

11. Daganzo, C., Fundamentals of Transportation and Traffic Operations, Pergamon-Elsevier, Oxford, U.K. (1997)

12. Webster, F.V., Traffic Signal Settings. Road Research Technical Paper, No. 39, road Research Laboratory, Ministry of Transport, HMSO, London, U.K. (1958)

5 IMPLEMENTATION OF ADAPTIVE TRANSIT SIGNAL PRIORITY (ATSP) SYSTEM

5.1 System Overview

A prototype ATSP system has been developed and implemented under PATH Project RTA 65A0026. Chapter 1 provides a briefing of the developed prototype system and illustrates the system architecture as shown in Figure 1-1. The system implementation in this project is to upgrade each sub-system of the proposed ATSP system in order to accommodate new technologies and newly developed control algorithms.

5.2 Hardware Upgrade

Figure 5-1 illustrates the ATSP system hardware architecture. Only two hardware sub-systems have been upgraded since the previous project. The first one is the wireless communication means. Cellular digital packet data (CDPD) modems are obsolete, so they are replaced by general packet radio service (GPRS) modems. The other change is that a laptop PC has been set up in the Cabinet at 9th Avenue. The traffic response field master (TRFM) control software, which was loaded in the 170E signal controller, is now installed on the laptop PC. In addition to the normal signal coordination functions and ATSP request functions, the laptop PC can also perform system monitoring, logging and trouble shooting functions.

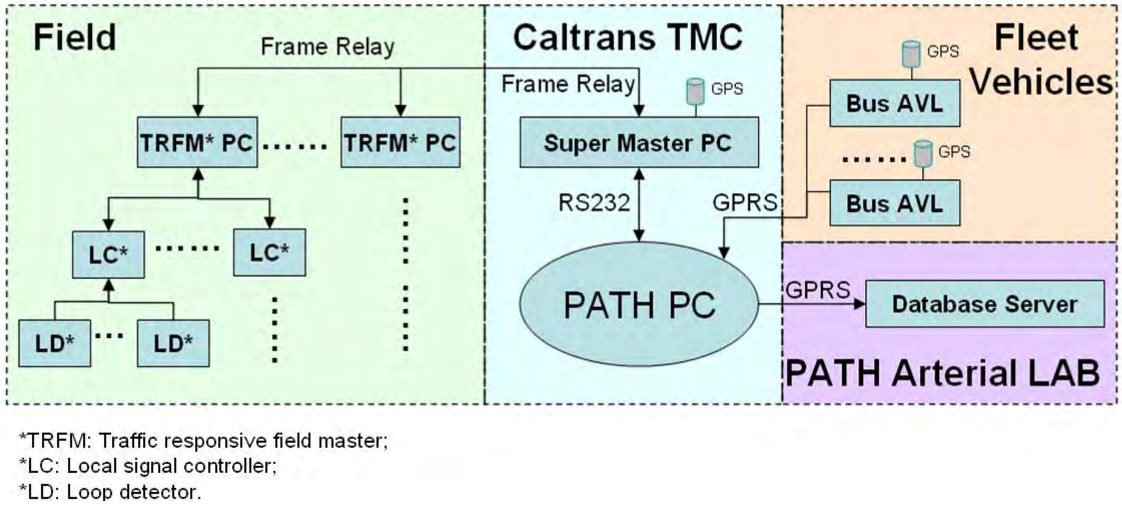

Figure 5-1 ATSP System Hardware Architecture

5.3 Software Upgrade

As shown in Figure 5-2, there are fifteen modules in the ATSP software architecture. Twelve of these modules were developed and upgraded by PATH. Chapter 2 through Chapter 4 have elaborated on time-to-arrival (TTA) predictor, arrival flow predictor, and priority requester (PRG and PRS) respectively.

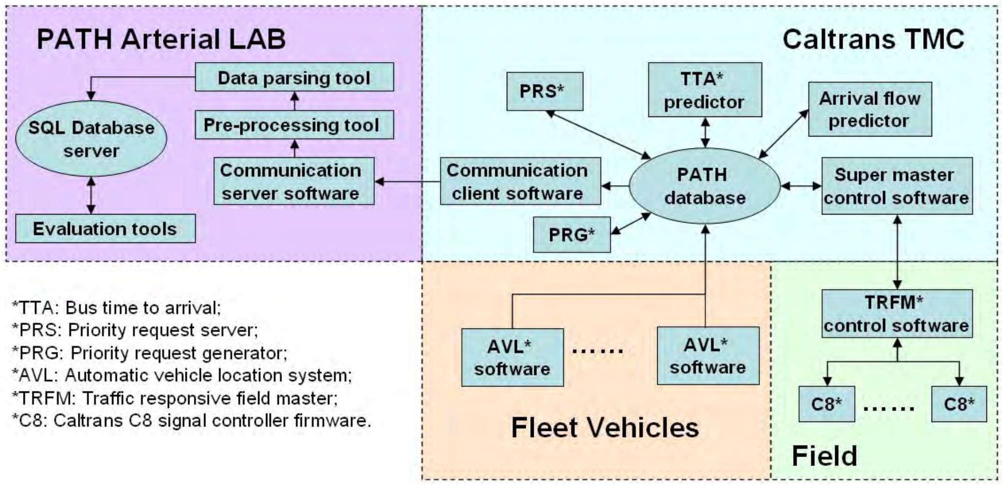

Figure 5-2 ATSP System Software Architecture

Some modules have been updated because of operating system upgrades or required hardware upgrades. The operating system on the PATH computer which is located at the control center in Caltrans District 4 was upgraded from QNX-4 to QNX-6. Accordingly, the PATH database and other software which are hosted by the PATH computer were upgraded to QNX-6 version too. The automatic vehicle location (AVL) communication software and communication client and server program at Caltrans TMC and PATH arterial lab respectively, were updated because of the communication hardware upgrade.

In the previous development project, all the data processing for lab tests and field tests was done by some file-based software modules. The major challenge of developing such file-based programs is to synchronize numerous data files with various file formats, including transit vehicle movement files, traffic signal status files, loop detector output files, and other software components output files. Although the previously developed data analysis tools served that project purpose, they have very limited expansion capabilities for further data analysis . Therefore, the PATH team turned to available database software other than self-developed tools to manage and process the field test data. MySQL, the most popular open source database software, was chosen to serve the data management purpose. After the database was properly set up in MySQL, three software modules, including data pre-processing, data parsing, and evaluation modules, wereen developed using MySQL Application Programming Interface (API) based on those previously developed file-based programs.

Three other modules, including the super master control software, the TRFM control software and Caltrans C8 signal controller firmware, are owned and developed by Caltrans engineers. The super master control software, which acts as the ATSP system monitor and data exchanger between the PATH computer and the TRFM, has not been changed since the previous development project. The other two software modules including TRFM control software and C8 signal controller firmware have been updated according to the changes on the PRG and PRS algorithms. In the previous ATSP system, the TRFM control software was embedded in the signal controller at 25th Ave. As the TRFM PC replaced the 170E controller to perform the field master function, the TRFM software for PC version was developed and installed on the TRFM PC. On the other hand in the previous ATSP system, the C8 signal controller firmware performed the traffic signal transition after the execution of a TSP command. In the current ATSP system, PRG optimizes two signal cycles including the signal transition cycle. Therefore, some changes have been made on the C8 signal controller firmware so that 170E controllers are capable of executeing the transition commands sent by the PRG.

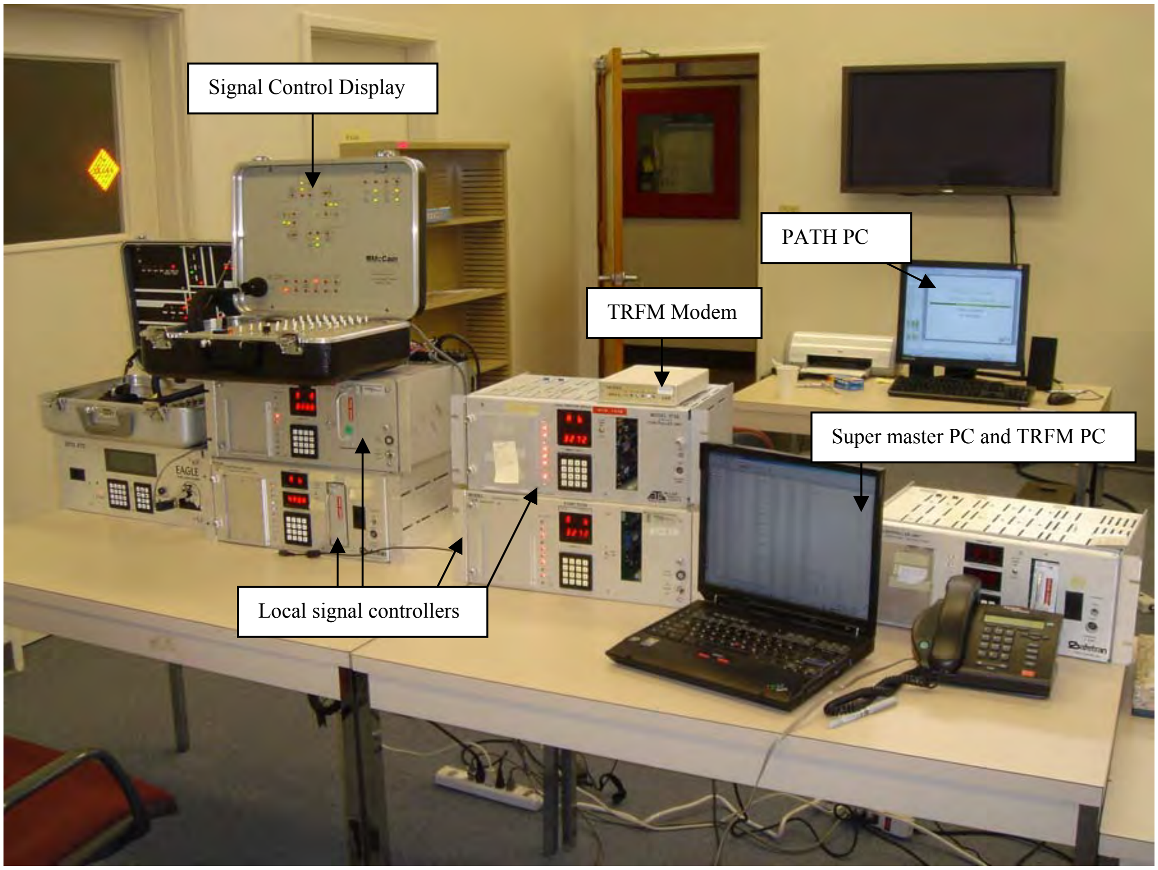

Figure 5-3 Hardware-in-the-loop Simulation Test-bed

With great support from Caltrans District 4, PATH set up a hardware-in-the-loop test-bed in its Transit/Traffic Arterial Lab. As shown in Figure 5-3, the test-bed is actually a compact ATSP system which contains major components of the system including local signal controllers, TRFM modems, TRFM PC, super master PC, and PATH PC. During the simulation run, the PATH PC simulates a bus run and generates a TSP request. Once the TSP request is received by the super master PC, the super master control software might grant the request and send it to the TRFM control software. Finally, the TSP command is sent from the

TRFM control software to the local signal controller. The whole data exchange process is exactly the same as the ATSP system running in the field. Moreover, the performance of the simulated system can be monitored by the signal controller display and software debugging tools on PATH PC, super master PC, and TRFM PC. By using this test-bed, both Caltrans engineers and the PATH team can debug each software component and communication interface. The test-bed is also a perfect tool to demonstrate the feasibility and stability of the ATSP system before its field implementation.

6 FIELD OPERATIONAL TESTS

Following the agreement with Caltrans, a prototype ATSP system was developed and field tested under PATH Project RTA 65A0026 at three intersections from 25th Avenue to 28th Avenue along El Camino Real corridor. The PATH project team drove a probe vehicle, which was developed and equipped by PATH, back and forth along the three TSP-capable intersections for two weeks. The first week was for the existing scenario without TSP, and then the second week was running with the prototype ATSP system.

Based on the experiences from the previous field test, Caltrans and SamTrans were more confident with the updated ATSP system. They decided to expand the testing site from three intersections to seven intersections. Moreover, an active SamTrans bus operator drove a not-in-service SamTrans bus along the testing site for the operational test.

6.1 Testing Description

Figure 6-1 shows the expanded testing site along El Camino Real corridor. Within this site, 20th Avenue is the only intersection which is not TSP capable because its running controller firmware was different from those at the other intersections. Other than 20th Avenue, field data were collected at all the other seven intersections.

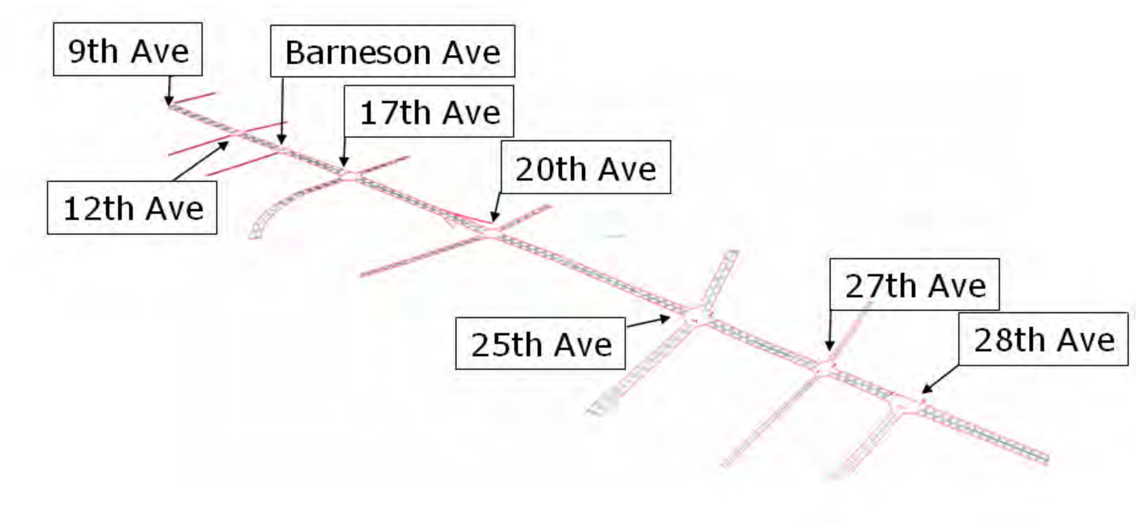

Figure 6-1 Map of Expanded Testing Site along El Camino Real

Through the hierarchical communication links among the super master PC, PATH PC, and TRFM PC, the second-by-second traffic signal data are collected by the PATH PC. This data includes traffic signal phase, interval, pattern plan, local cycle timer and master cycle timer, traffic volume data including cumulative count and occupancy at each loop detector. Meanwhile, all the TSP requests sent from PRG to PRS and then to TRFM are recorded on the PATH PC together with all the input and output logs for each software component for system checking and debugging purposes. The PATH PC saves all the above information in text files and then sends them to the PATH database server every 15 minutes. At the database end, some developed data pre-processing tools and parsing tools automatically check and clean those data and then parse them into a MySQL database.

A SamTrans bus was equipped with the data acquisition system for the field tests. The data acquisition system consists of three major components: a laptop PC, a GPRS modem, and a

GPS receiver. The AVL software, including the data process tools and communication tools, were installed on the laptop PC and automatically load after the PC reboots.

The field tests using SamTrans bus include the following two testing periods each of which is one-week-long:

<u>Test Period 1 (01/23/2006 till 01/27/2006):</u> "Before" data collection for the testing route, during which, signal and bus data were collected but no signal priority request was granted.

<u>Test Period 2 (01/30/2006 till 02/03/2006):</u> "After" data collection for the testing route, during which, all data was collected and signal priority requests were granted if needed.

Within each test period, drivers were requested to drive in three time windows per day. Each of the three time windows represents a typical traffic demand scenario. The three time windows are:

Morning Peak (A.M.):	**7:00 A.M. ~ 9:00 A.M.**
Mid-day (M.D.):	**11:00 A.M. ~ 1:00 P.M.**
Afternoon Peak (P.M.):	**4:00 P.M. ~ 7:00 P.M.**

During each time window, the assigned SamTrans bus operator drove the testing bus back and forth along the testing route on weekdays. Although no passengers were on aboard, drivers were told to follow their routine behaviors such as cruising speed, accelerations and decelerations.

6.2 Summary of Testing Results

As shown in Table 6-1, there are totally 143 effective bus runs in the "before" scenario and 110 effective testing runs in the "after" scenario. All of the effective bus runs were distributed among morning peak (A.M.), mid-day (M.D.), and afternoon peak (P.M.).

Table 6-1 Summary of Effective Bus Runs

	Date		A.M.	M.D.	P.M.	Total
"before" scenario	1/23/2006	Monday	0	0	2	2
	1/24/2006	Tuesday	0	5	12	17
	1/25/2006	Wednesday	10	11	0	21
	1/26/2006	Thursday	12	12	9	33
	1/27/2006	Friday	9	6	6	21
	1/30/2006	Monday	2	3	15	20
	1/31/2006	Tuesday	8	9	12	29
	"before" total		41	46	56	143
"after" scenario	2/1/2006	Wednesday	8	7	8	23
	2/2/2006	Thursday	10	8	13	31
	2/3/2006	Friday	7	7	6	20
	2/6/2006	Monday	4	6	13	23
	2/7/2006	Tuesday	4	0	0	4
	2/9/2006	Thursday	0	9	0	9
	"after" total		33	37	40	110
Total			74	83	96	253

The impacts of TSP implementation are classified into a variety of categories for different stakeholders. The parties or stakeholders whom TSP operations may affect include prioritized buses, passengers on these buses, normal traffic in the same direction as bus operations (normally on the main streets, thus called as major-phase traffic), cross-streets and main-street left-turn traffic (called as minor-phase traffic) and pedestrians. TSP operations may impose different impacts on these parties, which can be further categorized into different measures of effectiveness (MOEs). Based on the data collected in the field test, this section summarizes the bus trip data analysis from the perspective of bus travel time/speed, number of stops at intersections and intersection delay, and summarizes the ATSP incurred delays to different stakeholders.

6.2.1 Summary of Bus Trip Data Analysis

The purpose of trip-based analysis was to qualify the impacts of TSP operations on transit. Although the intersection at 20th Avenue was not TSP capable, the TSP treatments granted upstream would cause a bus arrive earlier and consequently affect the delay at 20th Avenue. Therefore, the trip-based analysis was conducted along the whole testing site, including 20th Avenue.

The measures of effectiveness (MOEs) selected to assess the benefits of TSP are as follows:

- travel time on segment
- number of stops at red
- total intersection delay on segment
- maximum intersection delay on segment
- average delay per stopped intersection
- average traveling speed on segment

During the testing period, the testing bus was cycling around the testing site along EL Camino Real, without following any schedule. Therefore, the bus schedule reliability is not included in the analysis.

In order to include the intersection delay at the two boundary intersections, the segment for northbound trips is defined as the roadway from 31st Avenue to 9th Avenue, which is about 2,933 meters long; and the segment for southbound trips is 2,954 meters long, from 5th Avenue to 28th Avenue.

As aforementioned, the ATSP algorithm provides optimized green splits that facilitate the movement of in-service transit vehicles through signalized intersections while minimizing the negative impacts on normal traffic. The average execution rate of TSP treatments is 17% on both directions, i.e., 1.2 TSP treatment being granted for both per northbound trip and per southbound trip. One third of granted priorities are green extension treatments and the other two third are early green treatments.

Table 6-2 compares the aforementioned MOEs for before-and-after scenario. Statistic t-test was made to test whether the changes are statistically significant or not. The following observations can be made from Table 6-2:

- The operations of ATSP saved bus trip travel time by 13% (51 seconds) at northbound and by 9% (32 seconds) at southbound. The changes are statistically significant.
- ATSP increased bus average traveling speed by 11% (1.5 m/s or 3.4 MPH) at northbound and by 7% (1.1 m/s or 2.4 MPH) at southbound. The changes are statistically significant.

- It saved the bus' total intersection delay by 19% (26 seconds) at northbound and by 14% (18 seconds) at southbound. Moreover, the maximum intersection delay was reduced by 19% (11 seconds) northbound and by 18% (9 seconds) southbound; and the average delay per stopped intersection was reduced by 14% (6 seconds) and 9% (3 seconds) for northbound and southbound, respectively. The changes are statistically significant.
- ATSP operations reduced number of stops due to traffic signals. However, the changes are statistically insignificant because the objective of ATSP is to minimize the weighted bus and normal traffic delay rather than minimizing the number of stops.

Table 6-2 Summary of ATSP impacts on transit operations

MOE	Direction	"before" scenario	"after" scenario	Changes (value)	Changes (%)	*t*-test
Travel time (sec)	Northbound	386.7	335.8	-51.0*	-13.2%*	0.0005*
	Southbound	347.8	315.8	-32.0*	-9.2%*	0.0061*
No. of stops at red	Northbound	3.5	3.4	-0.1	-3.7%	0.2869
	Southbound	3.8	3.5	-0.3	-8.2%	0.0802
Average traveling speed (m/s)	Northbound	13.8	15.3	1.5*	10.8%*	0.0010*
	Southbound	15.4	16.4	1.1*	6.9%*	0.0081*
Total intersection delay (sec)	Northbound	136.2	109.9	-26.3*	-19.3%*	0.0065*
	Southbound	125.6	107.4	-18.1*	-14.4%*	0.0248*
Maximum intersection delay (sec)	Northbound	57.4	46.5	-10.8*	-18.9%*	0.0482*
	Southbound	51.3	42.3	-9.1*	-17.6%*	0.0149*
Average delay per stop (sec)	Northbound	39.6	34.0	-5.6*	-14.1%*	0.0114*
	Southbound	33.4	30.6	-2.8*	-8.5%*	0.0074*

* Significant at 5% level

Note that above time saving on bus was achieved by only 1.2 TSP treatment per each directional trip. As a consequence, the negative impacts on cross street traffic would be considerably small. Moreover, granted priorities would also benefit the mainline traffic, which could cause the impacts of ATSP operations on normal traffic be positive.

Figure 6-2 to Figure 6-4 plot the empirical cumulative distribution functions (CDF) of bus travel time, average traveling speed, and total intersection delay, respectively, for both of "before" scenario and "after" scenario. The CDF plots clearly show how ATSP operations saved bus time and improved bus traveling speed.

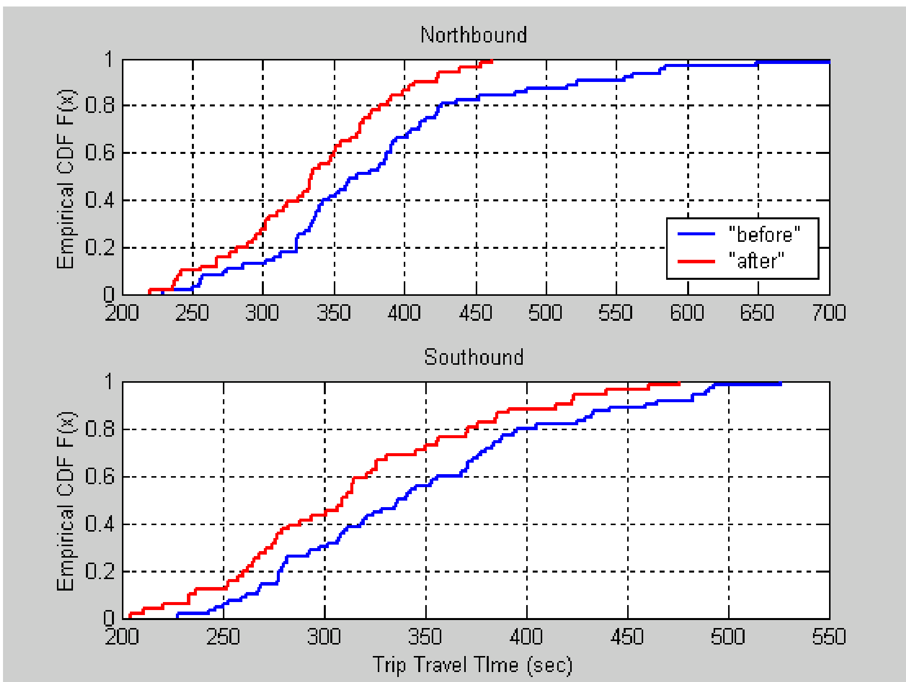

Figure 6-2 CDF Plot of Bus Trip Travel Time

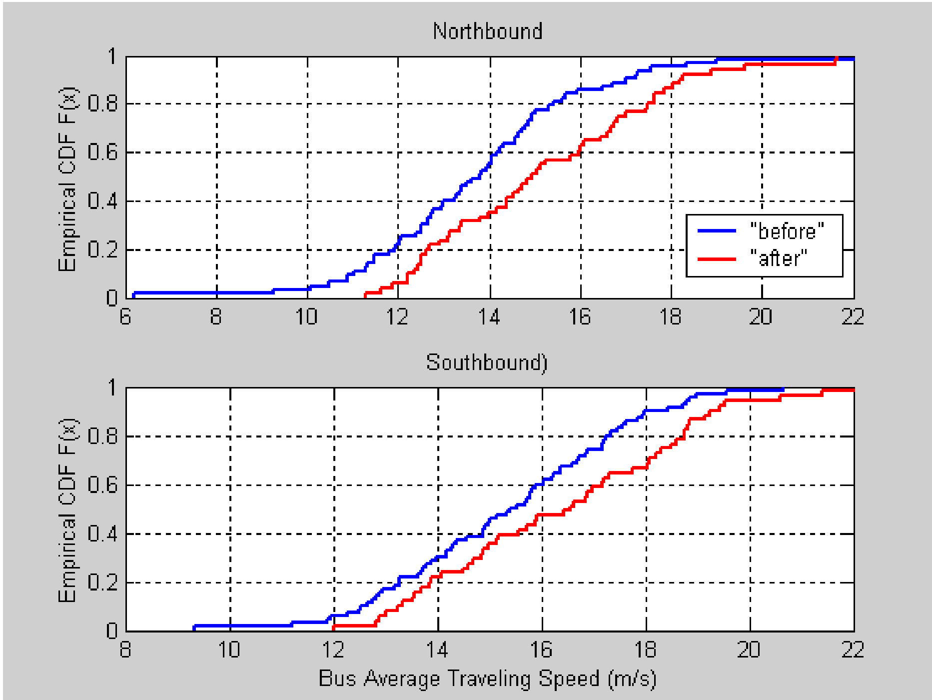

Figure 6-3 CDF Plot of Bus Average Traveling Speed

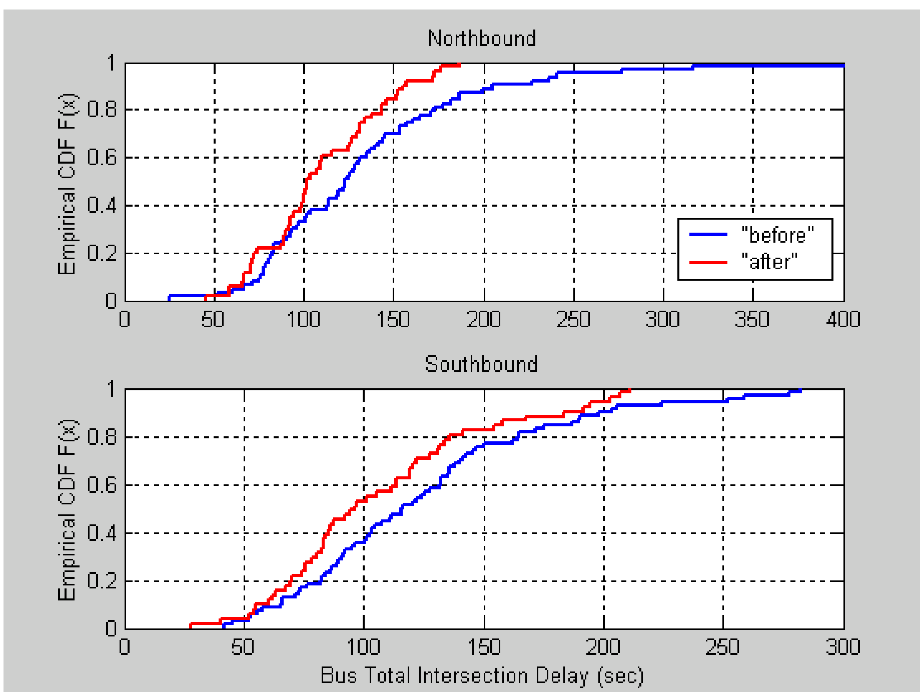

Figure 6-4 CDF Plot of Bus Total Intersection Delay

6.2.2 Summary of Intersection Delay Analysis

At signalized intersections, time is the critical resource that different stakeholders, such as traffic at different approaches, transit, and pedestrians, are competing for. As mentioned in Chapter 4, ATSP is an unconventional strategy to re-assign time resources among these parties. With ATSP, transit vehicles have higher priority to get time resources compared to other parties. The major concern of such treatments is how to better accommodate transit vehicles without significantly impacting other parties. In such sense, intersection delay is one of the most important MOEs to evaluate ATSP impacts on all stakeholders.

Table 6-2 compares intersection delays for an inferred "before" scenario and an "after" scenario. It should be noted that the inferred "before" scenario here is not the real "before" scenario as mentioned in Table 6-1 Summary of Effective Bus Runs The "before" scenario in Table 6-2 is inferred based on the "after" scenario. As discussed in Chapter 4, ATSP re-assign time resources, phase lengths in this case, among different stakeholders. Without ATSP, the traffic signal will follow the original semi-actuated signal control logic and assign a cycle to different phases. A program was developed to mimic the semi-actuated control logic and reverse the phase lengths with TSP in the "after" scenario back to a scenario without TSP. Such an inferred "before" scenario is more comparable with "after" scenario than the real "before" scenario when the TSP evaluation is conducted intersection by intersection and trip by trip. Therefore, all the following evaluation results are comparisons of inferred "before" scenarios and "after" scenarios.

To simplify calculations, it is assumed for the inferred "before" scenario that all phases except coordination phases will be no longer than their green splits if pedestrian buttons are not pushed. It is a reasonable assumption because according to *Highway Capacity Manual* 2000, green splits are calculated based on the approaches' historical demand. Therefore, green splits are best guesses of phase lengths.

For both of the inferred "before" scenario and the "after" scenario, the same assumptions as described in Chapter 4 about the optimization model have been made for delay calculation. Intersection traffic delays for major phases and minor phases and bus delays are retrieved by putting the real phase lengths and traffic demands back into the optimization model. Major phases in this study are bus approaching phases, which are southbound and northbound El Camino Real. Accordingly, minor phases are all phases except major phases. Note that the intersection delays for both "before" and "after" scenarios are calculated for local signal cycles, which consist of one background cycle and two control cycles as defined in Chapter 4, i.e. the adjacent signal cycle before bus arrival, the bus arrival cycle and its following cycle.

As mentioned in Chapter 4, the proposed ATSP system can benefit buses and the traffic traveling along with buses. As expected, Table 6-2 illustrates that traffic intersection delays for major phases are smaller after executing TSP. In contrast, traffic intersection delay for minor phases increased after providing TSP. When calculating the average passenger delay at intersections, the average number of passengers on vehicles other than buses is assumed to be

1.2. While for SamTrans buses, the average onboard passengers including drivers were assumed to be 15 per bus.

At 9th Avenue for the inferred "before" scenario, the average bus intersection delay, the average major phase vehicle delay and the average minor phase vehicle delay are 41.58 seconds, 14.16 seconds and 14.42 seconds, respectively. With the execution of TSP, the average bus delay is sharply reduced by 95% to 1.98 seconds per bus; the average major phase traffic delay is reduced by 81% to 2.70 seconds per vehicle; the minor phase delay is increased by 0.93 seconds per vehicle (6%) to 15.35 seconds per vehicle. The statistic *t*-test results show that the delay savings for buses and also major phase traffic after implementing TSP are significant, while the incurred delay for minor phase traffic is negligible. When evaluating the whole intersection by looking at passenger delays, the average passenger delay for all approaches including buses is reduced by 55% from 15.57 seconds per passenger to 6.98 seconds per passenger. The statistic *t*-test also shows that the ATSP system significantly reduced the average intersection passenger delay at 9th Avenue.

Table 6-3 Summary of ATSP Impacts on Intersection Delay
(second/vehicle or second/passenger)

	Delay (sec/veh or sec/pax)		Inferred "before" scenario				"after" scenario			
			Bus	**Major**	**Minor**	**Pax***	**Bus**	**Major**	**Minor**	**Pax***
9th Ave.	Mean		41.58	14.16	14.42	15.57	1.98	2.70	15.35	6.98
	Standard deviation		19.15	8.50	6.36	5.36	6.86	1.36	4.83	1.88
	Change	sec/veh	N/A	N/A	N/A	N/A	-39.60	-11.46	0.93	-8.59
		%	N/A	N/A	N/A	N/A	-95%	-81%	6%	-55%
		t-test*	N/A	N/A	N/A	N/A	sig't*	sig't*	insig't*	sig't*
17th Ave.	Mean		61.38	33.20	9.61	25.93	28.56	28.48	11.11	22.19
	Standard deviation		19.49	10.71	3.09	7.34	29.64	12.29	3.33	7.60
	Change	sec/veh	N/A	N/A	N/A	N/A	-32.82	-4.72	1.49	-3.74
		%	N/A	N/A	N/A	N/A	-53%	-14%	16%	-14%

		t-test*	N/A	N/A	N/A	N/A	sig’t*	insig’t*	sig’t*	sig’t*
25th Ave.	Mean		51.30	36.67	13.72	27.42	29.09	30.88	15.11	24.19
	Standard deviation		27.65	10.39	2.76	6.54	28.34	8.07	2.37	5.40
	Change	sec/veh	N/A	N/A	N/A	N/A	-22.21	-5.79	1.39	-3.23
		%	N/A	N/A	N/A	N/A	-43%	-16%	10%	-12%
		t-test*	N/A	N/A	N/A	N/A	sig’t*	sig’t*	sig’t*	sig’t*
27th Ave.	Mean		45.35	18.26	16.94	19.13	17.93	11.63	17.15	12.36
	Standard deviation		17.49	8.58	7.00	7.52	21.74	5.29	3.31	4.30
	Change	sec/veh	N/A	N/A	N/A	N/A	-27.42	-6.62	0.21	-6.76
		%	N/A	N/A	N/A	N/A	-60%	-36%	1%	-35%
		t-test*	N/A	N/A	N/A	N/A	sig’t*	sig’t*	insig’t*	sig’t*
28th Ave.	Mean		45.58	18.83	13.24	18.76	14.07	4.95	16.77	7.20
	Standard deviation		15.06	5.94	2.50	4.40	18.81	2.54	4.23	2.73
	Change	sec/veh	N/A	N/A	N/A	N/A	-31.50	-13.89	3.53	-11.56
		%	N/A	N/A	N/A	N/A	-69%	-74%	27%	-62%
		t-test*	N/A	N/A	N/A	N/A	sig’t*	sig’t*	insig’t*	sig’t*

Note:

Pax*: Delay for passengers on buses and other vehicles;

t-test*: Statistic *t*-test to check the delay change is statistically significant or insignificant;

sig’t*: Delay change is statistically significant;

insig’t*: Delay change is statistically insignificant;

17th Avenue and 25th Avenue are two busiest intersections along the testing corridor. As a result, the same amount of green times “borrowed” for buses and major phase traffic from minor phases incurs more delay for minor phase traffic at the two intersections than at 9th Avenue. As described in Chapter 4, the weighting factor is the key to balance the level of priority and the incurred minor phase traffic delay. In the field test, a constant weighting factor was applied for all the seven intersections. As expectated, the TSP optimization model reduces the level of priority to buses at those busy intersections in order to balance the incurred delays for minor phase traffic. As illustrated in Table 6-2 for 17th Avenue, the average bus delay is reduced by 53% from 61.38 seconds to 28.56 seconds; the average major phase traffic delay is reduced by 14% from 33.20 seconds per vehicle to 28.48 seconds per vehicle. Meanwhile, the incurred minor phase traffic delay is only 1.49 seconds per vehicle. The average passenger delay at 17th Avenue is reduced by 14% from 25.93 seconds per passenger to 22.19 seconds per passenger. The statistics *t*-test shows that the passenger delay saving is statistically significant. Similarly for 25th Avenue, TSP saved 43% of average bus delay and 16% of average major phase traffic delay, while only cost an extra 1.39 second per vehicle for minor phase traffic. The average passenger delay saving is 12% which is statistically significant.

27th Avenue and 28th Avenue have less traffic on minor phases than 17th Avenue and 25th Avenue. With TSP at 27th Avenue, the average bus intersection delay is reduced by 60% from 45.35 seconds to 17.93 seconds; the average major phase traffic delay is reduced by 36% from 18.26 seconds per vehicle to 11.63 seconds per vehicle. While the resulting extra delays to minor phase traffic are only 1% and 0.21 seconds per vehicle. In other words, the negative impact of TSP on minor phase traffic is very minor. For average passenger delay, there is a 35% reduction which is also statistically significant. At 28th Avenue, the average bus delay is reduced by 69% from 45.58 seconds to 14.07 seconds. The average delay saving for major phase traffic is 13.89 seconds per vehicle which is 74% of their average delay in the "before" scenario. As a result, the average minor phase traffic delay is increased by 3.53 seconds per vehicle delay which is statistically insignificant. For the whole intersection, the average passenger delay is sharply reduced by 62% from 18.76 seconds per passenger to 7.2 seconds per passenger. The signal control efficiency, from the point of view of per passenger delay, is significantly improved.

One of the major incentives for TSP is that transit vehicles carry more passengers than other vehicles. Giving appropriate priority to transit vehicles might reduce the overall passenger delay. This represents a more appropriate measurement of effectiveness than the overall vehicle delay for traffic signal control at busy intersections. Table 6-3 and Figure 6-2 show the sensitivity analysis results for passenger intersection delay with different number of passengers on board. Note that the bus is empty when the number of passengers, actually the driver in this case, is one.

Table 6-4 Sensitivity analysis of passenger intersection delay (second/passenger)

Scenario	Number of pax	9th Ave	12th Ave	Barneson	17th Ave	25th Ave	27th Ave	28th Ave
Before	1 (empty)	14.35	6.49	16.40	24.37	26.78	18.20	17.94
	5	14.71	6.72	16.62	24.83	26.96	18.47	18.18
	10	15.14	7.00	16.89	25.39	27.19	18.80	18.47
	15	15.57	7.27	17.15	25.93	27.42	19.13	18.76
	20	15.98	7.52	17.41	26.46	27.64	19.44	19.04
After	1 (empty)	7.21	5.44	16.86	21.99	24.06	12.17	6.99
	5	7.14	5.35	16.70	21.99	24.10	12.22	7.05
	10	7.06	5.25	16.51	22.09	24.14	12.29	7.12
	15	6.98	5.15	16.32	22.19	24.19	12.36	7.20
	20	6.90	5.05	16.14	22.29	24.24	12.43	7.27
Change	1 (empty)	-7.13	-1.05	0.46	-2.38	-2.72	-6.04	-10.95
	5	-7.56	-1.37	0.08	-2.84	-2.87	-6.25	-11.13
	10	-8.09	-1.75	-0.38	-3.30	-3.05	-6.51	-11.35
	15	-8.59	-2.12	-0.83	-3.74	-3.23	-6.76	-11.56
	20	-9.08	-2.47	-1.27	-4.17	-3.41	-7.01	-11.77

In Figure 6-2, each dot represents a delay reduction at one of the seven testing intersections for one of the five levels of bus occupancies. The slope of each curve represents the sensitivity of the passenger intersection delay reduction at one intersection. All the curves have positive slopes because the average bus delay savings are positive so that more passengers on buses would result in more passenger delay savings. Among the seven testing intersections, 12th Avenue and Barneson Avenue have higher sensitivities than other intersections because they have relatively smaller traffic volumes so that the number of passengers on buses has more significant impacts on average passenger delay. Moreover for Barneson Avenue, the curve crosses 0% line at about six passengers on board. It means the average passenger intersection delay would be reduced with TSP at Barneson Avenue when there are more than six passengers on the bus. For the other six intersections, the re-optimized signal timing can reduce the average passenger delay even though buses are empty because the existing semi-actuated signal control is less efficient.

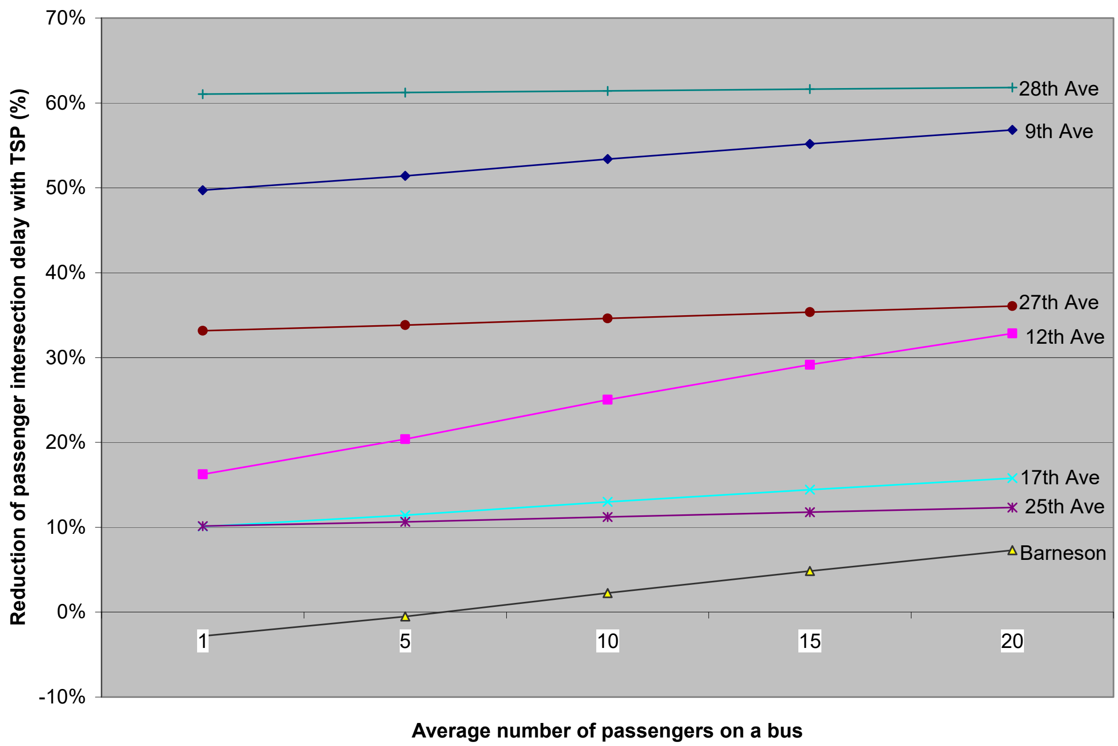

Figure 6-5 Sensitivity analysis of passenger delay reduction (%)

Table 6-4 illustrates the distribution of executed priority requests at the seven testing intersections. For example at 9th Avenue, 12 priority requests are executed among the total 110 testing trips. 5 of them are early green requests; the other 7 are green extension requests. All 12 trips with priority requests are northbound trips. Three of them happened during morning peak; five of them happened during mid-day; the other four trips happened during afternoon peak. The TSP execution rate is about 11% at 9th Avenue. On 70% of trips, buses do not need priority. In other words, there is a 70% chance that a bus will arrive at 9th Avenue during the green period. Note that the rate TSP is not needed is estimated by the detailed analysis on 10 randomly picked trips. The other 19% of priority requests at 9th Avenue are blocked. Among the blocked priority requests, most of them are blocked by protected signal control logic, for example pedestrian call and minimum green requirements. Sometimes when the travel time predictor yields big prediction errors due to the instability of communication or GPS system, TSP requests might be blocked by PRG itself.

12th Avenue and Barneson Avenue have an extremely low execution rate that is lower than 5%. It is mainly caused by the very low traffic demand on their minor phases. Due to the semi-actuated signal control logic, all the green time would be assigned to coordinated phases when there is no call from minor phases. Therefore, buses would have better chances to meet a green signal at the less busy intersections and thus do not need TSP.

The 17th Avenue and 25th Avenue intersections are the busiest. The rates that TSP is not needed are lowest at the two intersections, which are 40% and 20% respectively. 25th Avenue is a wide street connecting two shopping areas with a high frequency of pedestrian button usage. As a result, the block rate at 25th Avenue, at 48%, is the highest among all intersections.

Table 6-5 Distribution of executed priority requests

		9th Ave	12th Ave	Barneson	17th Ave	25th Ave	27th Ave	28th Ave	Sum
Executed priority requests	EG*	5	0	3	31	27	8	6	80
	GE*	7	4	2	15	8	6	0	42
	NB*	12	2	5	22	17	0	5	63
	SB*	0	2	0	24	18	14	1	59
	A.M. *	3	1	1	7	9	3	1	25
	M.D. *	5	2	3	24	21	6	2	63
	P.M. *	4	1	1	15	5	5	3	34
	Total	12	4	5	46	35	14	6	122
% of all*	Executed	10.91%	3.64%	4.55%	41.82%	31.82%	12.73%	5.45%	15.84%
	TSP not needed*	70%	90%	70%	40%	20%	80%	80%	64%
	Blocked*	19.09%	6.36%	25.45%	18.18%	48.18%	7.27%	14.55%	19.87%

Note:

EG*: Executed early green requests;

GE*: Executed green extension requests;

NB*: Northbound trips with executed requests;

SB*: Southbound trips with executed requests;

A.M.*: Trips during morning peak with executed requests;

M.D.*: Trips during mid-day with executed requests;

P.M.*: Trips during afternoon peak with executed requests;

% of all*: Percentage among total 110 testing trips;

TSP not needed*: Trips which TSP requests are not needed;

Blocked*: Trips which TSP requests are blocked.

To sum up, the field operational tests of the ATSP system at seven intersections show very positive results. Bus delay savings and passenger intersection delay savings are statistically significant. The major phase traffic delays are reduced, while the incurred minor phase delays are insignificant.

There were some limitations of the field test. First of all, although a great deal of time was spent to fine-tune the sensitivity of the loop detector system, the real accuracy of each loop detector is unknown. Accordingly, the accuracy of arrival flow prediction is indefinite. Secondly, although SamTrans' bus operators drove a SamTrans bus for the field test, the bus was still not in service and did not stop for passengers. Thus, the testing results might vary from real transit service operations. Therefore, the accuracies of loop detectors need to be measured using other detection means, for example pneumatic road tubes and counters or video sensors, to further verify the effectiveness and impacts of the ATSP system. Moreover, in-service buses are needed to run future field tests.

6.3 References

13. Transportation Research Board's (TRB's) Committee on Highway Capacity and Quality of Service, *Highway Capacity Manual 2000*, 2000.

7 INVESTIGATION OF POTENTIAL COMMUNICATION APPROACHES FOR ATSP

ATSP is built on the premise that the signal priority algorithm anticipates the arrival of a transit vehicle at a traffic signal with sufficient lead time so that the priority treatment can be achieved prior to the signal phase at which the bus will arrive. This is to maximize the likelihood that the bus will go through the intersection on a green phase, to minimize the delay for the transit bus and, at the same time, minimize the impact on the rest of the traffic and ensure pedestrian safety. It is therefore imperative that a prediction algorithm is able to predict the time-to-arrival (TTA) at the next signalized intersection, denoted by $T_A(t)$, with a predetermined lead time and accuracy to support the ATSP strategy. The ATSP concept further recommends that TTA is estimated based on GPS positioning, which can be obtained every second using a the state-of-the-art GPS. The estimation of TTA can be obtained either on the bus or off the bus. Because off-bus TTA estimation will require communication from a bus to an off-bus computer at a rate of 1 Hz, it is deemed more practical with the current communication technology that a on-bus system involving GPS will monitor the bus location and determine the need and timing for a priority call and once the call is needed, a TSP request be sent from the bus to a off-bus TSP requesting computer. This approach will significantly reduce the demand on the bandwidth of a communication channel.

One of the main contributions of the PATH research on ATSP is to develop approaches allowing the GPS/Advanced Communication System (ACS) that has already been instrumented on buses to become an integral portion of the ATSP. This facilitates the integrated application of transit ITS technologies. However, there are technical challenges with this approach. Because the ACS system is designed for transit operation monitoring and data collection purposes, the polling rate of the ACS is typically at approximately 2 minutes intervals. Since buses can travel long distances and possibly cross several intersections within 2 minutes, the low polling rate of current ACS applications makes it impractical for ATSP to receive timely priority requests. However, most of the ACS systems have built in additional bandwidth allowing smaller numbers of buses with higher priorities or that have priority needs (such as emergency calls) to receive more frequent polling. The study summarized below investigates the ATSP bandwidth needs and potential protocol approaches for accommodating TSP requests using existing GPS/ACS systems.

7.1 Analysis of Priority Request Demand

The TSP request demand mainly depends on the number of buses in service on the corridor and the number and spacing of intersections. In reality, the demand is complicated by the stochastic nature of the predicted arrival time at an intersection. Nonetheless, one could first analyze and investigate the priority request demand problem by considering the mean arrival time at each intersection. In this study, we use Samtrans bus services along El Camino Real as a case study. We first consider the situation when transit vehicles arrive at their next intersections on time according to the bus schedules. This implies the process deadlines are pre-determined and the TTA calculated at the current time t, $T_A(t)$, which is the scheduled arrival time minus the current time, can easily be calculated without the need of a TTA prediction algorithm. It is clear that each link connecting two successive intersections has unique traffic conditions, so the TTA prediction distribution will be different for different links.

Our analysis focuses on the bus schedules of 30 SamTrans (http://www.samtrans.com) bus routes running primarily alone the El Camino Real in the San Mateo County of San Francisco Bay Area. The bus routes are: KXN, KXS, MXN, MXS, PXN, PXS, RXN, RXS, 250E, 250W, 251E, 251W, 260E, 260W, 262N, 262S, 271E, 271W, 274E, 274W, 295N, 295S, 296N, 296S, 390N, 390S, 391N, 391S, 397N, 397S. Here “N”, “S”, “E” and “W” denote the north, south, east and west bound routes, respectively. The schedule for each bus route provides the arrival times, with a time resolution in seconds, at specific time points. The time points are a subset of the signalized intersections along the route. The arrival times at other intersections that are not provided in the schedule are obtained by using linear interpolation of the given arrival times at the time points. This linear mean travel time model is indeed the historical model developed in ([1]). The arrival time prediction model reported there is a historical model “continuously” tuned by a real-time adaptive model.

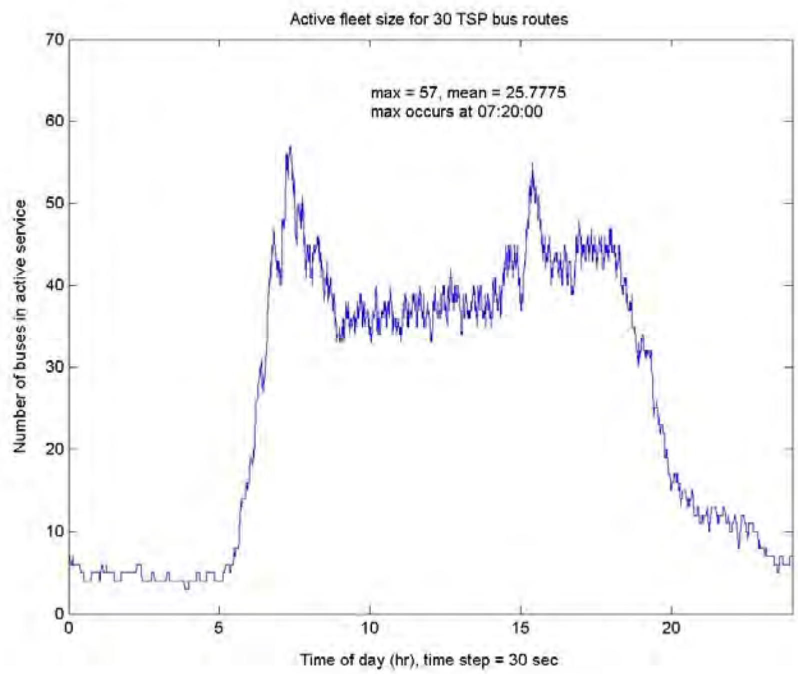

Figure 7-1 Number of buses in active service for 30 SamTrans TSP bus routes

Using the bus schedules, we can analyze the priority request demand in terms of how many vehicles could generate a TSP request at the same time. Figure 7-1 Number of buses in active service for 30 SamTrans TSP bus routes**Error! Reference source not found.**shows the number of active vehicles with respect to the TOD (in hour). We see that during the morning commute hours, as many as $N(t) = 57$ buses could be in service in the network. In principle, the same number of buses could be making priority requests if they all arrive at their next intersections at the same time, that is, with the same process deadline. Given the arrival times at intersections, which are process deadlines, we need to set a rule to determine when a priority request can be made. The rule permits a vehicle to make its initial request when it is at a fixed period, a "look ahead" time, away from its next arrival time. If the initial request is not processed, additional requests are made before the transit vehicle arrives at its next intersection or a pre-specified time limit.

In the demand analysis, we assume that the request times for a Single Request Process per approach (to an intersection) per bus: Let t be the current time and $T_A(t)$ be the TTA computed at the current time, so the process deadline is $t + T_A(t)$. Let T be a required "look-ahead" time. A priority request can be made if $T_A(t) \le T$.

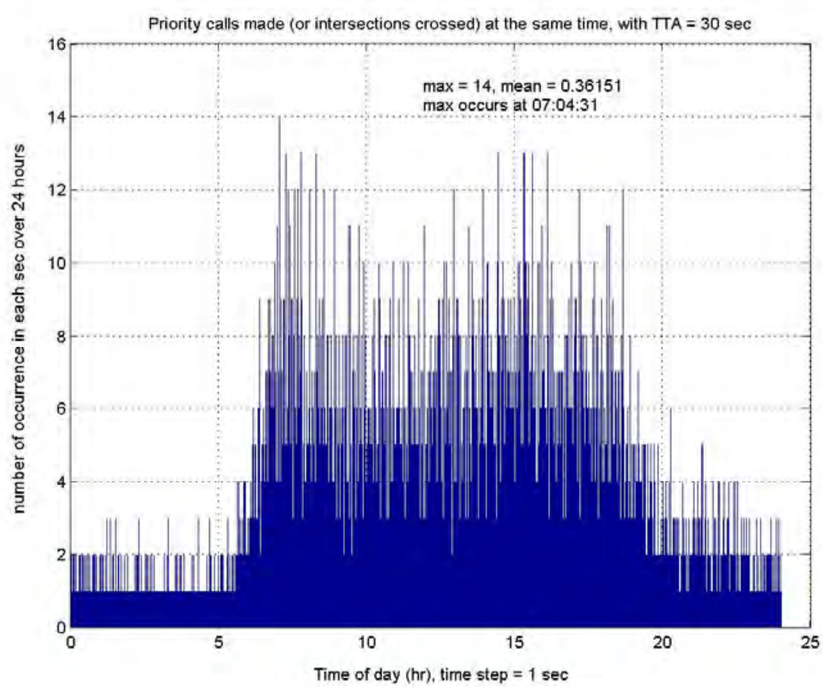

Figure 7-2 Number of calls made in each second over 24 hours

The scheduled arrival times have the format of hh:mm:ss. So if the analysis uses a 1-sec or finer resolution, an initial request is made as soon as $T_A(t) = T$. This requires the simulation time step to be 1 sec or an integer fraction of 1 sec. It is clear that each call has a one-to-one correspondence with a request process. So zero dropped calls means zero dropped request processes. We first analyze the request demand in terms of the number of calls made. If there is only a single request process, the number of calls is exactly the number of intersections that will be crossed. **Error! Reference source not found.** shows the number of calls made in each integer second if calls are made at exactly TTA = 30 sec. This is also the number of intersections that will be crossed exactly 30 sec from the current time. On the average a call will be made in roughly every 67 sec, as shown in the histogram in **Error! Reference source not found.** compiled for the 30 bus routes over 24 hours. This average call time is consistent with our belief that the polling channel alone cannot process all the requests, since the polling frequency could be as slow as once every two minutes. If we change the call time to TTA = 20 sec in **Error! Reference source not found.**, the plot simply shifts to the right by 10 sec. So it is *time-invariant* with respect to TTA. As seen in**Error! Reference source not found.**, there can be as many as 14 simultaneous calls during the morning commute hours. Because of the simulation resolution (1-sec time step), the percentage of time when there are no calls is not transparent from the "dense" plot in **Error! Reference source not found. Error! Reference source not found.-Error! Reference source not found.** shows the hourly distribution of the number of calls made. In the 8^{th} hour, 7-8am, there are zero calla in a second for 55.39% of the time. This percentage increases to 3158/3600 = 87.72% for 5-6am, which is an off-peak hour. If we zoom into **Error! Reference source not found.** and consider only the time window 05:30-05:40, there are indeed many seconds when there is no call. This is shown in **Error! Reference source not found.** with a total of only 36 calls in the 10-minute span and the probability of zero calls in a second is 94.17%. Because of the skewed hourly distribution of the request demand shown in **Error! Reference source not found.**, it will be more meaningful to focus the analysis on the time window from 5am to 10pm. **Error! Reference**

source not found.-Error! Reference source not found. gives the histograms of the number of seconds when *N* calls are simultaneously made for 0500-2000. Note that the probability of zero calls in a second is 66.89%. It would be 74.23% if the entire 24 hour period is considered. These histograms are independent of TTA.

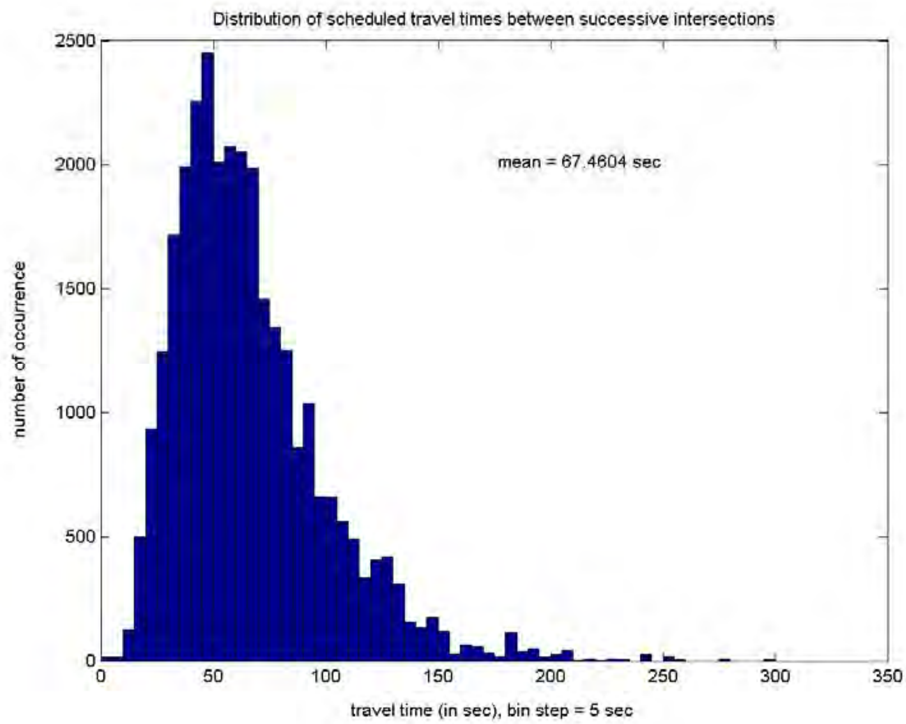

Figure 7-3 Distribution of scheduled travel times between successive intersections

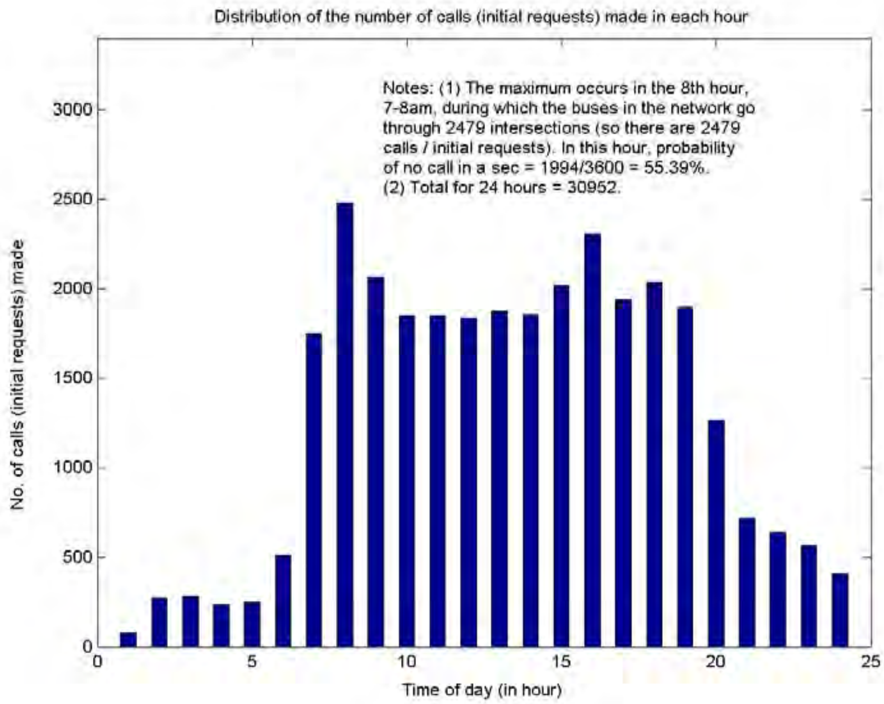

Figure 7-4 Number of calls made in each hour

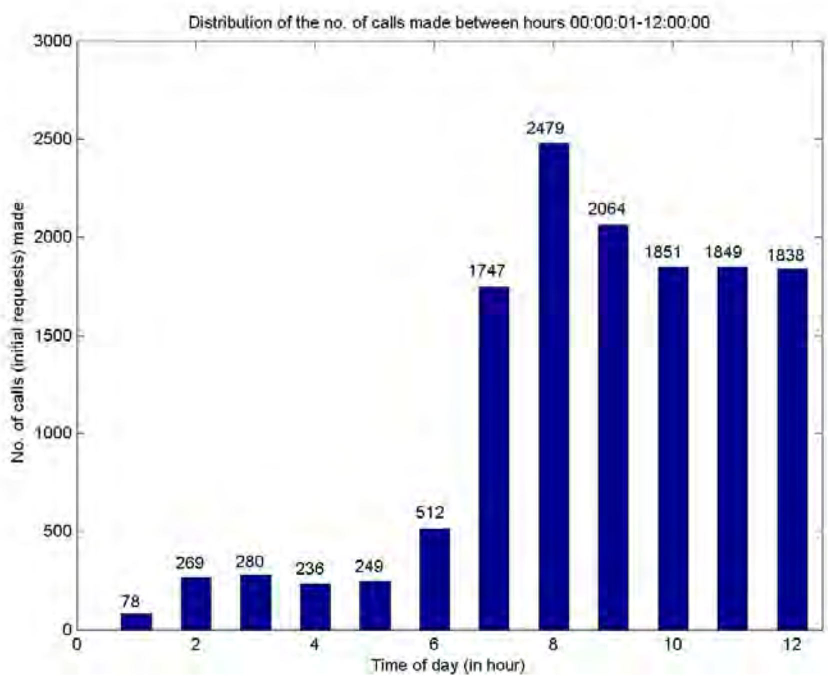

Figure 7-5 Number of calls made in each hour, 1st – 12th hours

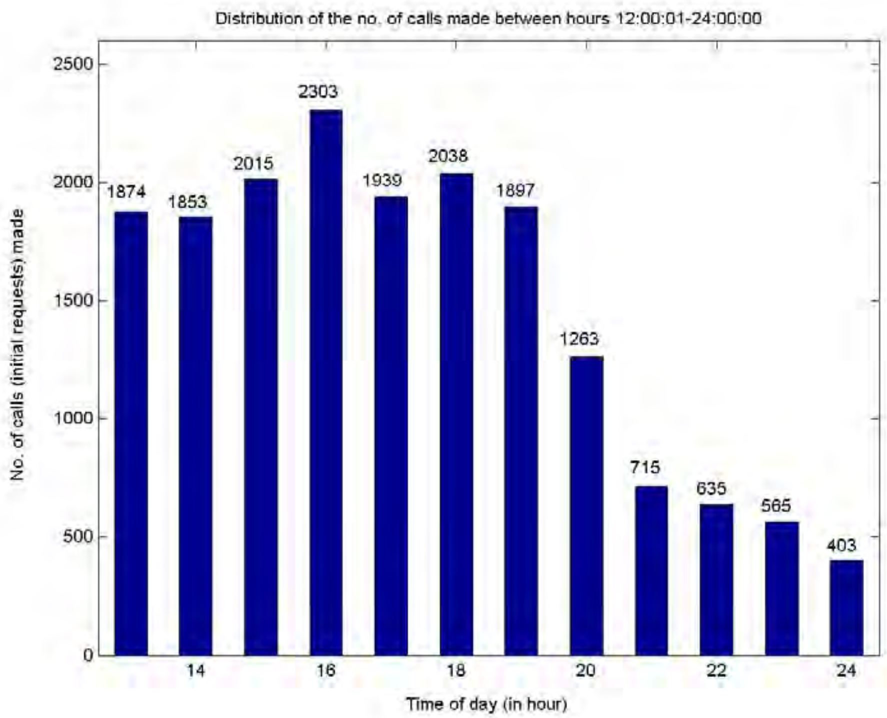

Figure 7-6 Number of calls made in each hour, 13th – 24th hours

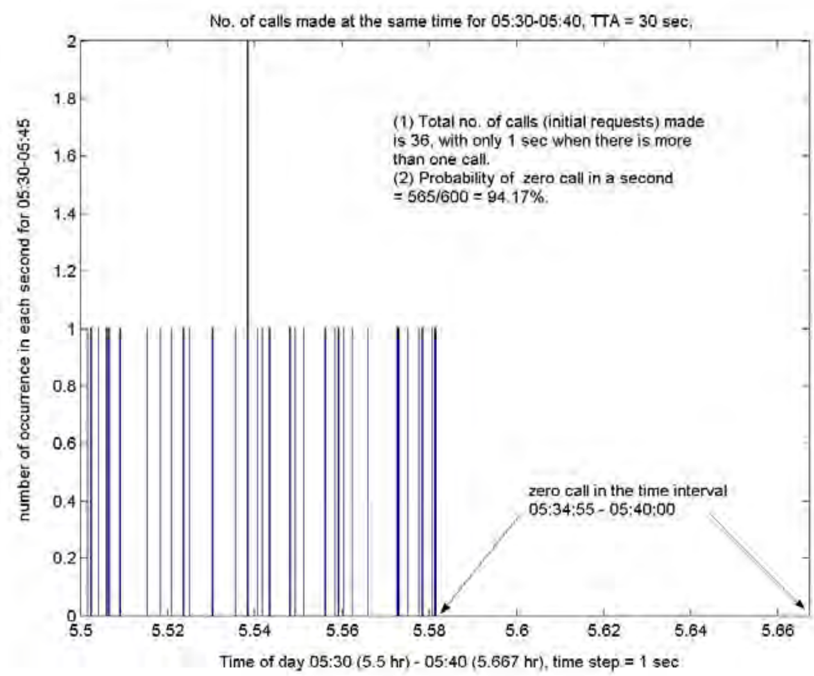

Figure 7-7 Calls made per second for 05:30-05:40 are sparsely distributed

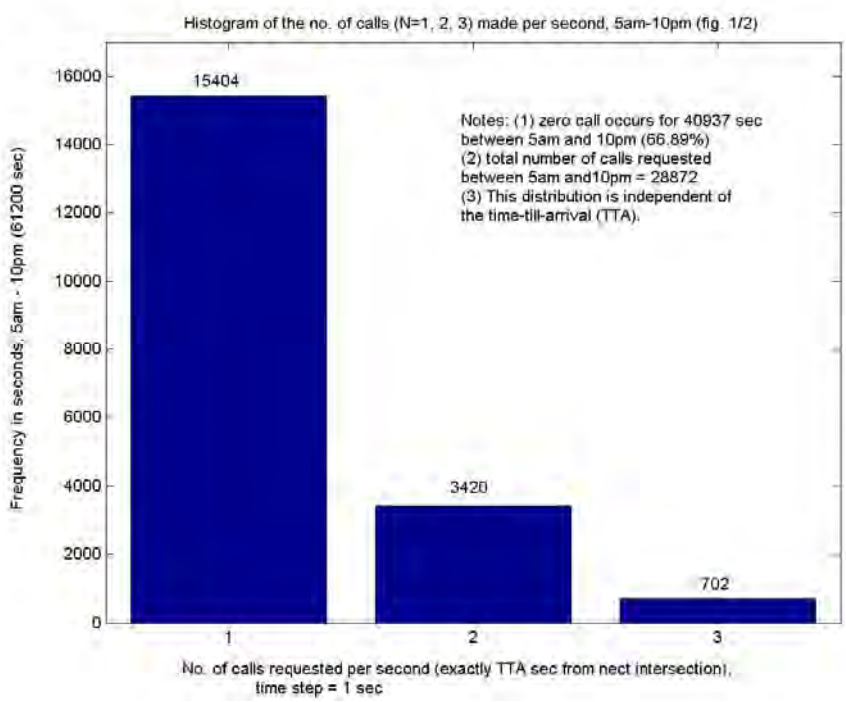

Figure 7-8 Histogram of the number of simultaneous calls for 5am-10pm (part 1/2)

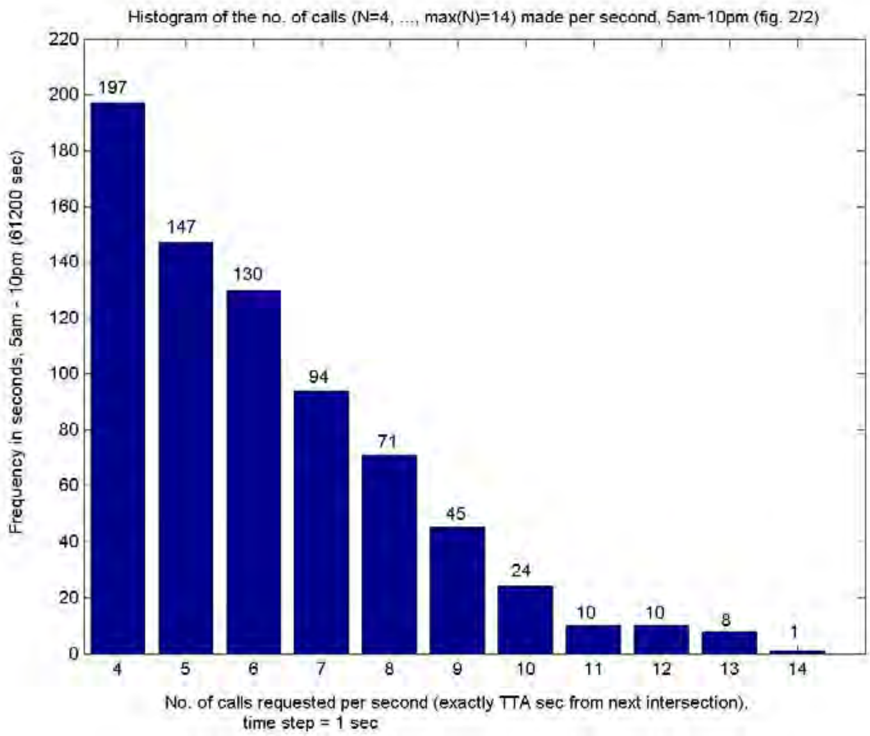

Figure 7-9 Histogram of the number of simultaneous calls for 5am-10pm (part 2/2)

The above supply-demand analysis is based on a set of deterministic arrival schedules (process deadlines) at the intersections. In reality, there is a possibility that the initial request (or call) is not received or not immediately answered. A repeat call will help to ensure that the TSP request is processed before the process deadline. Furthermore, the arrival time at an intersection is random and estimated using a TTA prediction algorithm ([1]). The TTA prediction has different statistical variations for different links connecting successive intersections. Nonetheless, it is typical that the prediction error has a smaller standard deviation and converges to zero as the transit vehicle approaches its next intersection. Since there are statistical uncertainties in the TTA prediction, the predicted arrival time announced at the time of the initial request is likely to be inaccurate, with a fairly large expected error. An unreliable TTA prediction is less useful for determining how a signal cycle should be modified to facilitate the passage of a transit vehicle through a signal-controlled intersection. Therefore, a more accurate predicted arrival time may be needed via the second request process when the vehicle is closer to the intersection and the statistical errors in the predicted arrival time are smaller in order to increase the success rate of TSP execution.

In order to reduce the impact to other traffic, a TSP system is required to notify the traffic controller after the transit vehicle has gone through the intersection so that the normal signal cycle can be restored ([2]). An additional request process is thus needed to establish this communication. This will be called the checkout request process. The request is made as soon as the transit vehicle has passed the intersection and needs to be processed within a specified amount of checkout time. This additional process will increase the overall request demand, and consequently, the queue size and average queue waiting time.

In the following analysis, we consider three types of request processes. A first request process begins when the transit vehicle is farther away from its next intersection. Next a second request process begins when the vehicle is closer to the intersection. Finally, there is a checkout request process that begins when the vehicle passes the intersection. Similar to the single request process, a request is repeated until it is processed before its deadline for each of the three types of request processes.

7.2 Analysis of Alternative Communication Approaches

One of the main contributions of the PATH research on ATSP is to allow use of the existing Advanced Communication System (ACS) based advanced transit management equipment. With ACS, GPS and dedicated communication links have been instrumented on buses. PATH's ATSP concept proposes to allow these GPS and ACS to become an integral portion of the signal priority system. In this integrated application, software modifications will enable the existing on-bus GPS/ACS system to (1) monitor the bus location, (2) determine the need and timing for a priority request and (3) transmit the TSP request through the ACS communication link. In discussion with ACS suppliers, it was determined that the first two functions can be added onto ACS with minimum modification to the existing systems. However, because the ACS system is designed for transit operation monitoring and data archive purposes, the update rate for sampling bus locations is designed to be at approximately 2 minute intervals. As stated before, this sampling rate is insufficient to provide the accuracy and timeliness required for the ATSP scheme. ACS vendors indicated that most of the ACS systems have built-in additional communication bandwidth to enable smaller numbers of buses with higher priorities or buses that have priority needs (such as emergency calls) to receive more frequent polling. One of the primary objectives of this project is to investigate potential approaches that can meet the needs for ATSP using existing GPS based Advanced Communication Systems.

7.2.1 Potential Communication Approaches

Two approaches can potentially enable the existing GPS/ACS to provide more timely priority calls. The first approach is a Dynamic Polling Algorithm (DPA) that polls buses only on a needed basis. In normal transit operation, the transit operations center may not need to receive a bus location update every 2 minutes. A DPA allows the communication bandwidth to be strategically utilized so that only the buses approaching a point of interest (such as intersections or bus stops) are sampled. This can put the communication resources to 'priority' use so that the existing communication bandwidth would satisfy both transit management and signal priority needs. The second approach is to develop the 'Priority Polling' algorithm (PPA) specifically to take advantage of the additional bandwidth for priority vehicles.

Although both approaches have potential to meet ATSP needs using the existing ACS, a few other transit agencies including Samtrans expressed concerns about adopting dynamic polling and are not willing to consider the DPA approach. This left the second approach, i.e., utilization of the extra bandwidth, the only one to be considered under this study. We therefore focused on a supply-demand analysis to investigate the feasibility and protocols of TSP requests using the extra bandwidth available on the existing ACS.

7.2.2 Assumptions for TSP Protocols

In a centralized TSP system, when a transit vehicle makes a request for signal priority to be executed at its next signalized intersection, the vehicle must first establish a communication link with the transit or traffic management centre (TMC). The availability of this communication channel, however, may not be guaranteed because of the potentially large number of transit vehicles requesting priority service at the same time and the limited number of time slots per second (bandwidth limitation) available to process these priority requests. We made the following assumptions: (1) If the channel is busy, the request will not be answered or processed. (2) The request can be repeated until it is processed. (3) There can be more than one repeated request per second in the subsequent seconds. To distinguish the initial request from the subsequent requests, we will refer an initial request to as a call. Obviously, a priority request process, which includes the initial request and the subsequent repeated requests, must be done before its scheduled process deadline. This is either the time when the transit vehicle arrives at its next signalized intersection or a pre-specified time by which the request must be processed. After the process deadline has passed, the transit vehicle cannot make any more requests, and we say the request process is dropped, or simply the call is dropped. The time at which a request for priority is granted, which is also the time when a communication with the TMC is established, is called the process time. The difference between the process time and the time of the initial request is the queue waiting time. The timeline for a single request process is illustrated inFigure 7-10.

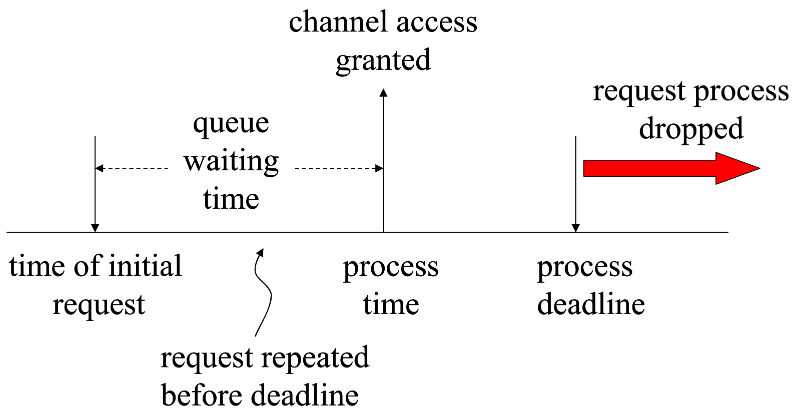

Figure 7-10 Timeline of a single priority request process

The problem is therefore to design a protocol for assigning the communication channel to priority requests made by transit vehicles so that the number of dropped request processes is minimized. We could also have a more general optimization objective that requires some high priority request processes not to be dropped and minimizes the process time for other low priority processes.

The demand for priority service is time-dependent. At any given time-of-day (TOD), this demand depends on the number of transit vehicles in service, and more specifically, the number of these in-service vehicles that will need to have their priority requests processed in, say, T seconds. The demand is also stochastic since the TTA prediction for each link connecting two successive intersections is a random variable. We will analyze and discuss the demand in detail later. On the supply side, the channel bandwidth is limited by the number of requests that it can process per second. This is the process rate. There are two available communication channels: a reserved polling channel and a contention channel. The polling channel is used by the TMC to initiate a communication with a transit vehicle. This operates according to a scheduling protocol which determines the times at which the transit vehicles can communicate with the TMC. Typically a transit vehicle is polled once every two minutes. When it is the turn of a transit vehicle to talk on the polling channel, bus data such as its GPS coordinates, predicted arrival time, on-time performance measure and signal priority request, if needed, will be transmitted. The polling channel will not able to process all the priority requests, because of its slow polling frequency. In order to send their requests, the transit vehicles have to "compete" for access to the contention channel. The access control on the contention channel will be determined by the protocol that is the design objective of this research. In the discussion of our algorithm, we will first focus on channel assignment using only the contention channel, and then modify the protocol to allow the use of the polling channel as well.

7.2.3 Realization of Process Priority

The first requirement in our protocol design is to guarantee that request processes with earlier process deadlines are processed first. This is similar to the Earliest-Deadline-First (EDF) protocol commonly used in real-time operating systems. The EDF protocol is a scheduling principle that places processes in a priority queue. A service deadline is assigned to each process and the protocol then always serves the process with the earliest deadline. Consider the process timeline shown in Figure 7-10.. Let t be the current time (or TOD). Suppose an initial priority request is made when a transit vehicle is T sec from its next signalized intersection, so the process deadline is at time $t+T$. If another initial request from a second vehicle is made at time $t' > t$, its process deadline is at time $t' + T$. If the order of these deadlines is followed, the protocol should make sure that the request process from the first vehicle is processed no later than that from the second vehicle. This can be achieved if the request processes are placed in a "service queue" so that the one with the earliest deadline is always processed first. For a standard TSP system, there is not a server that places request processes in a queue. If a request is made at a time when the contention channel is busy, it is ignored and not processed. Nonetheless, the queue concept can be employed if the probability of processing the request with the earliest deadline is higher than the probability of processing other requests with later deadlines. This can be realized by requiring the number of requests made per second to be higher for the process that has the earliest deadline, thus increasing its probability of establishing communication on the contention channel.

The queue concept is illustrated in Figure 7-10**.** with the number of requests made in subsequent seconds increases arithmetically as $2n$, if none of the requests in the previous seconds is processed. We assume $t = 1$, $t' = 2$, and $T = 4$, so the priority request from vehicle #1 has an earlier process deadline at time $t + T = 5$.

Table 7-1 Realization of a service queue via the probability of channel establishment

	number of requests in each second					
	1st sec	2nd sec	3rd sec	4th sec	5th sec	6th sec
vehicle #1	1	2	4	8	dropped	-
vehicle #2	-	1	2	4	8	dropped

In this example, since there are more requests from vehicle #1 in each second prior to its process deadline, the probability of vehicle #1 gaining access to the contention channel is always higher that of vehicle #2. This is consistent with the desire that the request from vehicle #1 should be processed no later than the request from vehicle #2. One can increase this probability by changing the arithmetic rate or imposing another scheme, such as a geometric rate of 2^n, for the number of requests made in each second in the protocol. As shown inTable 7-1, the waiting time in the queue is the difference between the process time and the time of the initial request. The request process is dropped after the process deadline. Intuitively, an efficient protocol should minimize the number of dropped processes. We could also include a general optimization objective that requires some high priority request processes not to be dropped and minimizes the process time for other low priority processes. The queue size will depend on the channel process rate, the time when an initial request can be made, and the time distribution of the process deadlines. As we will discuss later, the priority is also determined by the accuracy of the prediction TTA in the data message. If a message is more accurate and reliable, it should be sent at a higher priority.

7.2.4 Supply Analysis: Process Rate of the Contention Channel

The capability to process priority requests is limited by the polling frequency of the polling channel and the process rate of the contention channel. From here on, we assume the extra bandwidth, supplied by the contention channel, can process *two requests per second*. The *efficiency* at which requests are processed will depend on (1) the time at which an initial request (or a call) is made relative to the process deadline, (2) the time constraints of other request processes, such as those of a checkout process that will be imposed later, and (3) the process rate of the contention channel. In order to perform this analysis, we need to design a protocol that would maximize the efficiency of this request-process (or demand-supply) system, with the inclusion of the statistical variants of intersection arrival time prediction, the polling schedule, and real-time and/or historical data of an on-time performance measure. We also assume the request process is placed in an "imaginary" queue if the call is not processed immediately. A request process with the earliest process deadline in the queue is processed first. This requirement can be realized by imposing a scheme on the number of repeated requests made in the subsequent seconds in such a way that the probability of establishing a communication on the contention channel increases with earlier process deadlines.

Figure 7-11 shows the histogram of the number of requests processed per second with the initial request (or call) made at TTA = 30 sec. Figure 7-12**.** shows the histogram of the number of calls placed in the queue if they are not processed. In this scenario, there are no dropped calls. The queue waiting time is the time difference between the process time and the time of initial request. The histogram distribution of the waiting time is shown in Figure 7-13.. If we decrease the TTA to make the call time later, there is no change in the queue distribution until TTA drops below 7 seconds, the longest waiting time shown in Figure 7-13. When the TTA falls below 7 seconds, we will start to see dropped calls. The sensitivity to TTA appears to be approximately concave, with no changes if TTA $\geq$ 7 sec. The distribution is clearly sensitive to the process rate. If we decrease it to one request per second, there are still no dropped processes, but the longest waiting time becomes 14 sec. This results in an increase in the average waiting time from 0.86 sec to 1.87 sec. Also the probability that a call is not processed immediately (i.e. with waiting time > 0) increases from 63% to 79%. The waiting time histogram is shown in Figure 7-14..

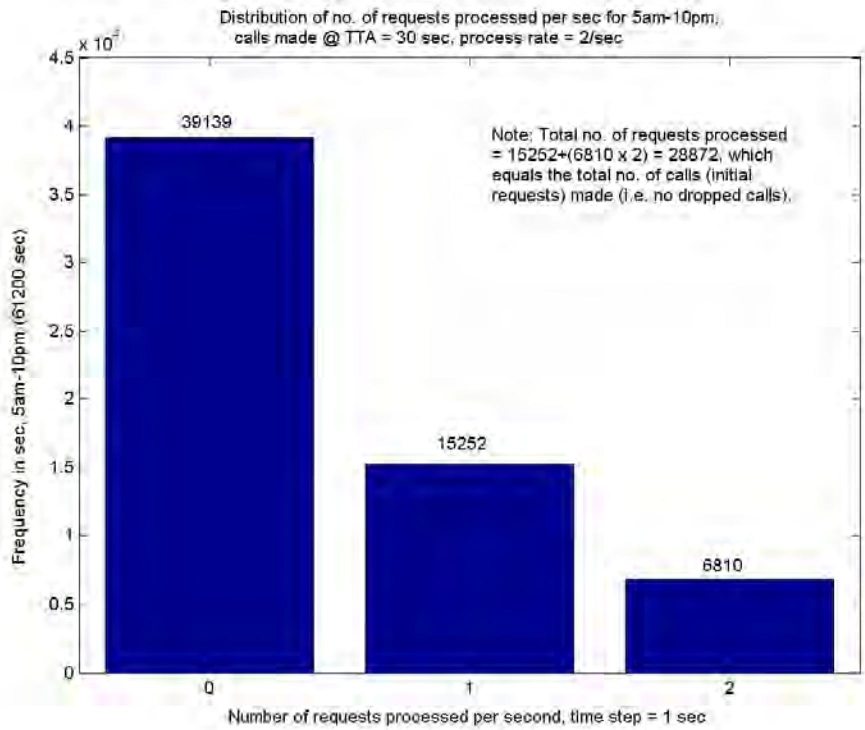

Figure 7-11 Histogram of no. of requests processed per sec, 5am-10pm, TTA = 30 sec

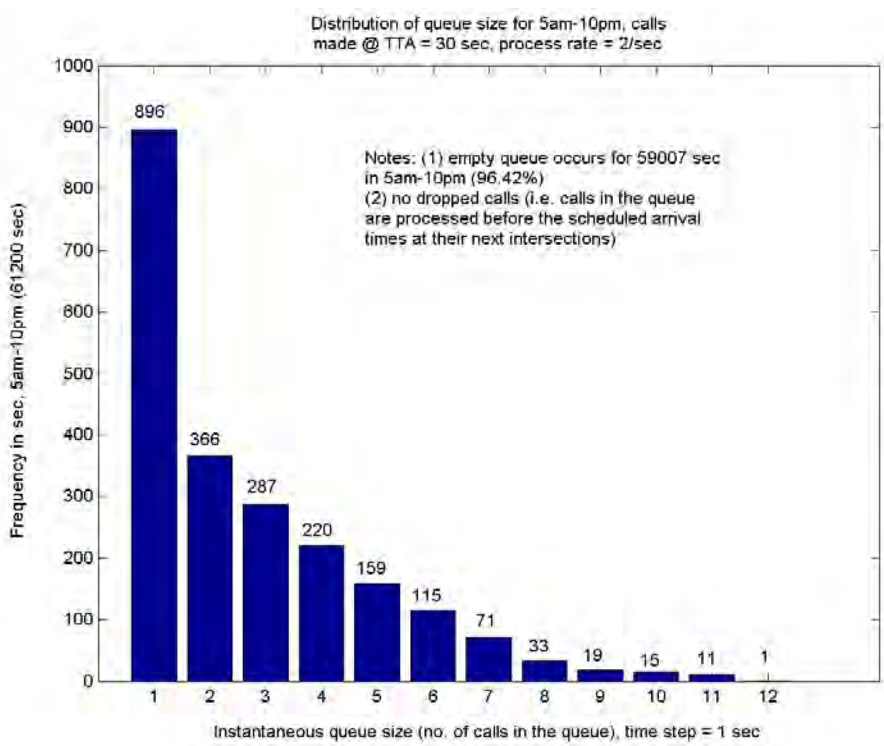

Figure 7-12 Histogram of queue size per sec, 5am-10pm, TTA = 30 sec

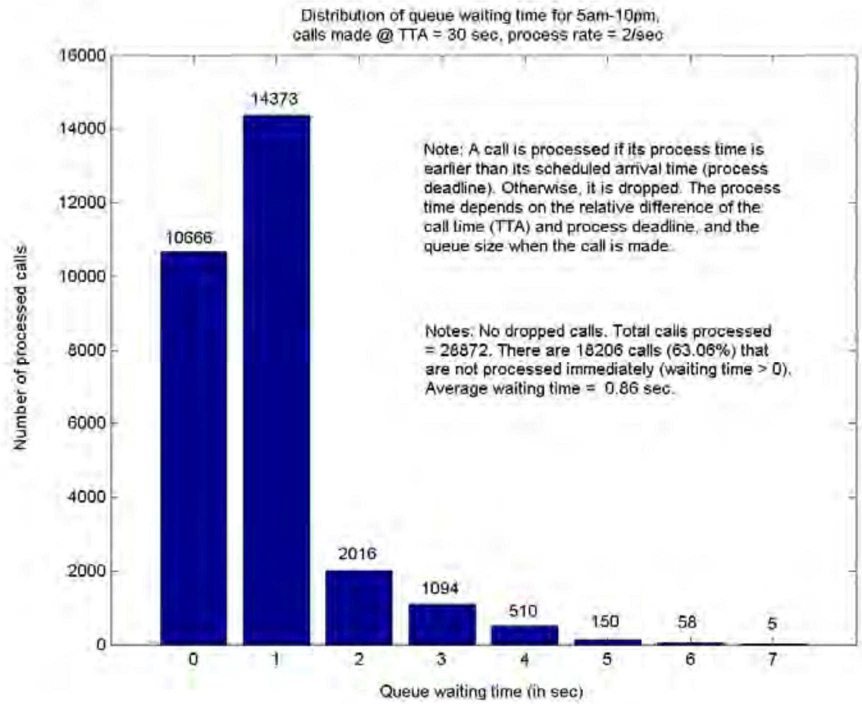

Figure 7-13 Histogram of queue waiting time, 5am-10pm, TTA = 30 sec

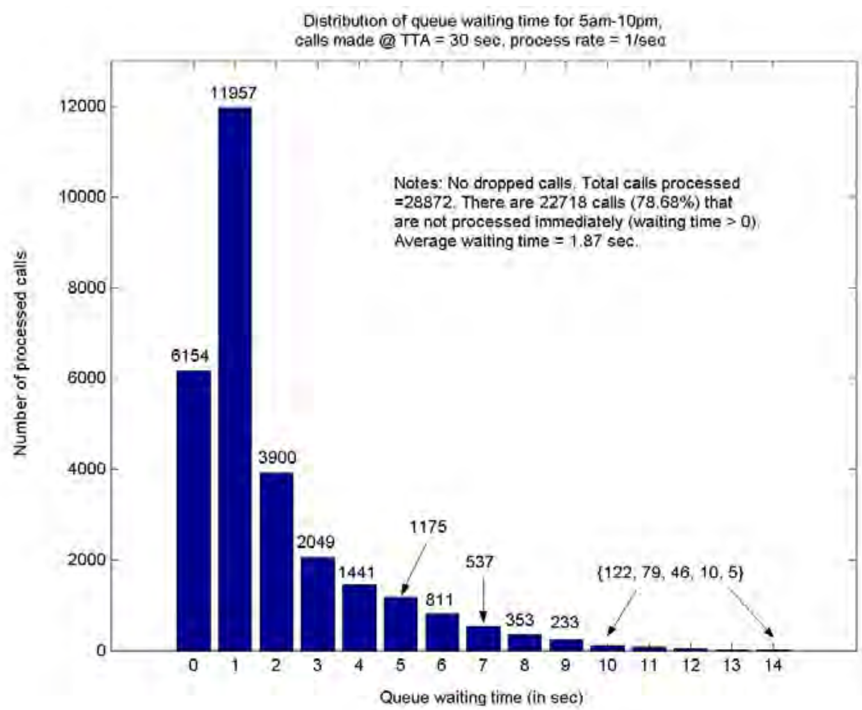

Figure 7-14 Histogram of queue waiting time, rate = 1/sec, 5am-10pm, TTA = 30 sec

Assumption 1 – Request times for First and Second Request Processes: Let t be the current time and $T_A(t)$ be the TTA computed at the current time. Let T_1, T_2 be two fixed "look-ahead" times. For the first request process, a priority request can be made if $T_2 < T_A(t) \le T_1 + T_2$. For the second request process, a priority request can be made if $T_A(t) \le T_2$.

Assumption 2 - Request times for a Checkout Request Process: Let T_{out} be a fixed "checkout" time constant. The initial request for a checkout process is made when a transit vehicle arrives at an intersection at a time $\hat{t}$ (that is, $T_A(\hat{t}) = 0$). A checkout request can be made before the process deadline, which is $\hat{t} + T_{out}$.

With a 1-sec or better simulation resolution, the initial request times (or call times) of the first and second request processes are $\hat{t}_1$ and $\hat{t}_2$, respectively, where $T_A(\hat{t}_1) = T_1 + T_2$ and $T_A(\hat{t}_2) = T_2$, respectively. The first process has a deadline at time $\hat{t}_2 = \hat{t}_1 + T_1$, while the second process has a deadline at time $\hat{t} = \hat{t}_2 + T_2$. The timelines of the three processes are illustrated in Figure 7-15..

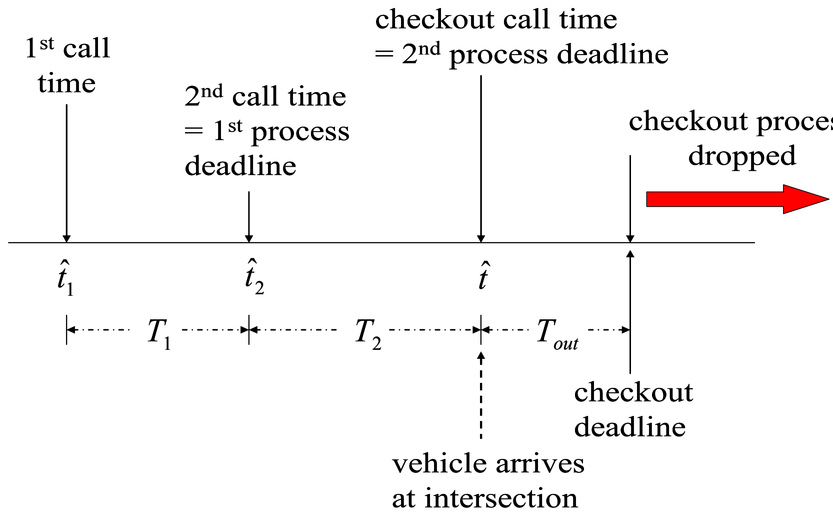

Figure 7-15 Timelines of multiple priority request processes

Since the predicted arrival time included in the data message in a first priority call is less accurate than that in a second priority call, it would be reasonable to require that a second request call be processed earlier than a first request call if they are made at the same time. Also it is desirable to assign a high process priority to a checkout call because of the need to restore a signal cycle to its normal state quickly and smoothly, with minimal delays on the cross street. By using the queue concept discussed in Section 7.2, these requirements can be realized by increasing the probability of communication establishment on the contention channel. We also require that if two processes are of the same type (either first, second or checkout call), the one with the earlier deadline should be processed first.

<u>Assumption 3 – Process Priority of Request Processes</u>: If the three types of request processes are initiated at the *same* time, the checkout call will be processed first, followed by the second call and then the first call, regardless of their process deadlines. This means the probability of

establishing communication on the contention channel is the highest for the checkout call, followed by those for the second and first calls. Two processes of the same type are processed in order of their deadlines.

We illustrate the requirements in Assumption 3 in Table 7-2. The probability of channel establishment is the highest for the checkout process. If the arithmetic rates for the first, second, and checkout processes are $k_1 n, k_2 n, k_3 n$, where $k_1 < k_2 < k_3$, then the probability for the checkout process is $k_3 /(k_1 + k_2 + k_3)$. This probability increases with k_3. So we can make k_3 large if it is highly desirable to first process a checkout call. Other schemes such as geometric rates of $2^{n-1}, 2^n, 2^{n+1}$, can also be used.

Table 7-2 Realization of process priority via probability of channel establishment

	Number of requests per second			
	1st sec	2nd sec	3rd sec	nth sec
veh. #1, checkout call	1	4	8	4(*n-1*)
veh. #2, second call	1	3	6	3(*n-1*)
veh. #3, first call	1	2	4	2(*n-1*)
veh. #4, first call	-	1	2	2(*n-2*)

7.2.5 Analysis of Multiple Request Processes

In this section we analyze a supply-demand scenario that involves multiple request processes. The simulation analysis uses the bus schedules, so that the arrival times at intersections are deterministic and pre-specified. The process rate is two per second. A transit vehicle makes its first call when the TTA $T_A(\hat{t}_1) = T_1 + T_2 = 40$ sec, and makes a second call when the TTA $T_A(\hat{t}_2) = T_2 = 20$ sec. The checkout process is initiated as the vehicle arrives at the intersection and needs to be completed in T_{out} = 10 sec. In this scenario, there are 827 dropped first calls (or request processes), but no dropped second or checkout calls. The high number of dropped first calls is a result of the process priority imposed in Rule 4. If we reverse the priority so that the first call will be processed first when the three types of request processes are initiated at the same time, then all the dropped calls are checkout calls. As shown in Figure 7-16, there are 36,884 seconds (60.27% of the period) when the channel is processing at its full capacity from 5am-10pm (61,200 sec). The probability of a dropped call during 5am-10pm is 0.955%, but it should be noted the dropped calls mainly occur during the peak hours, 7-8am and 3-4pm. The dropped calls in these hours account for 96% of the total dropped calls. At those hours, the probabilities that a call is dropped are 21% and 11%, respectively. The distributions are shown inFigure 7-17 and Figure 7-18. If the first call is made earlier at $T_A(\hat{t}_1)$ = 50 sec, the number of dropped first calls decreases to 763, a reduction of ~7.74%.

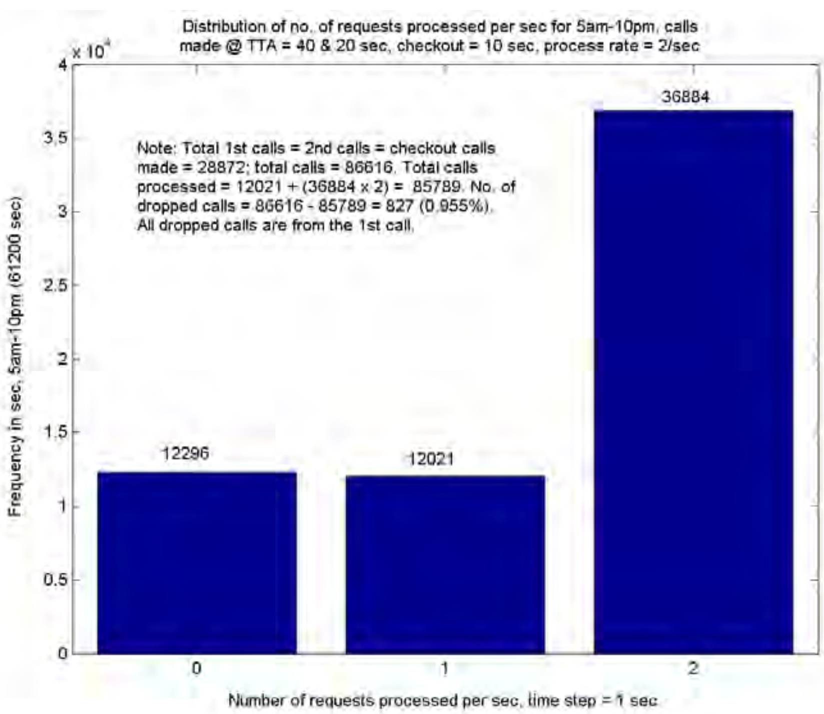

Figure 7-16 Histogram of no. of requests processed per sec; multiple requests {40, 20, 10}

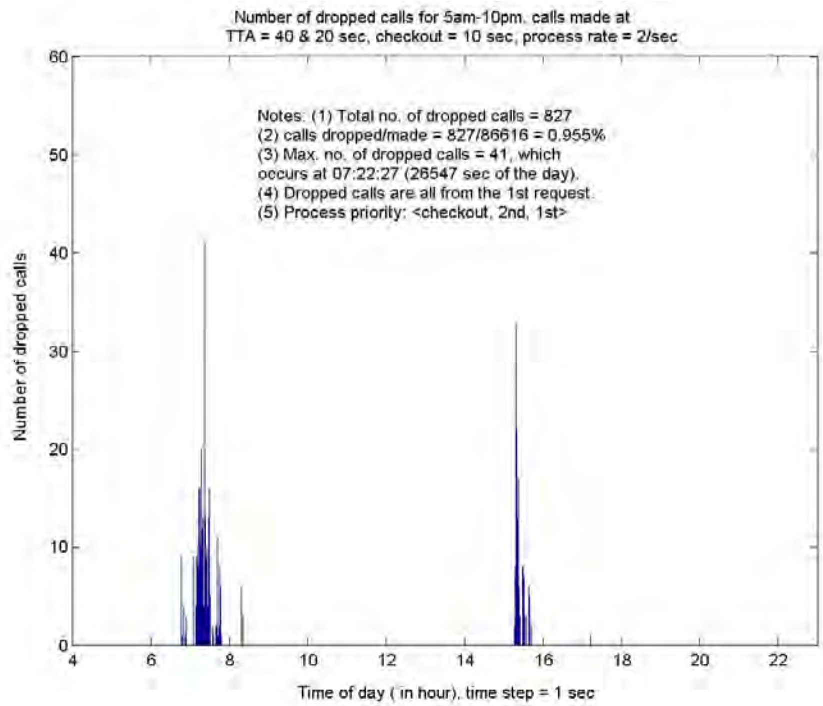

Figure 7-17 Number of dropped calls in 5am-10pm; multiple requests {40, 20, 10}

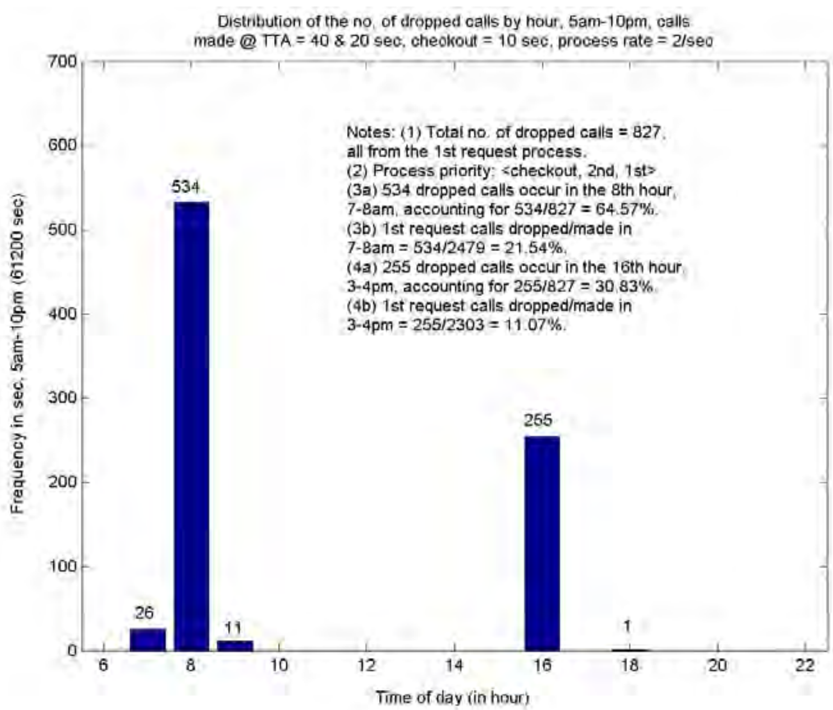

Figure 7-18 Distribution of no. of dropped calls by hour; multiple requests {40, 20, 10}

Figure 7-19 shows the time plot of the queue size, or the number of unprocessed calls, for the time period 5am-10pm. There can be as many as 64 unprocessed calls in the queue. The probability that this queue is non-empty is 39.31%. However, if we zoom into the two peak hours, 7-8am and 3-4pm, the probabilities are very high at 76% and 68%, respectively that there are requests in the queue. The hourly distribution of the frequency of a non-empty queue is shown inFigure 7-20. Since all the dropped calls are from the first request process, we also plot the number of first calls in the queue in Figure 7-21 **.**. There are no first calls in the queue for ~78% of the time in 5am-10pm, with the queue being empty ~60% of the time (seeFigure 7-20). It is interesting to note that the maximum first call queue size (61) occurs at 07:22:26 and the maximum number of dropped first calls (41) occurs at 07:22:27 (seeFigure 7-17), the very next second. During the two peak hours, roughly 65-70% of the calls in the queue are first calls. The average queue size is 1.477 and the average number of first calls in the queue is 0.837. This is consistent with the observations that the queue is empty for ~60% of the time (seeFigure 7-21) and most unprocessed calls are first calls. The histogram of the number of first calls in the queue is shown in Figure 7-22.. The distribution is skewed with a small variance.

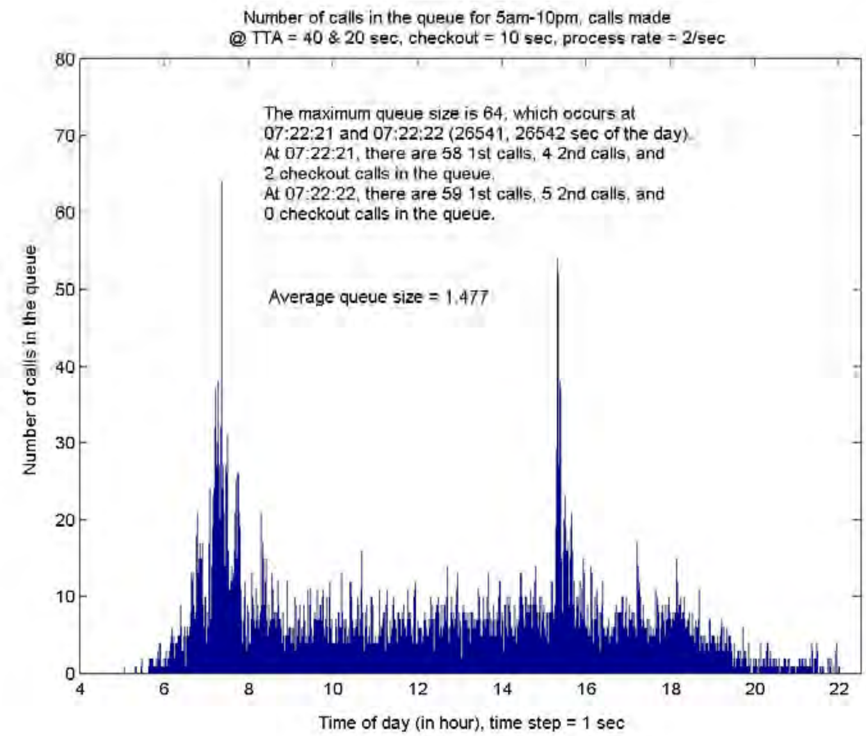

Figure 7-19 Number of calls in the queue; multiple requests {40, 20, 10}

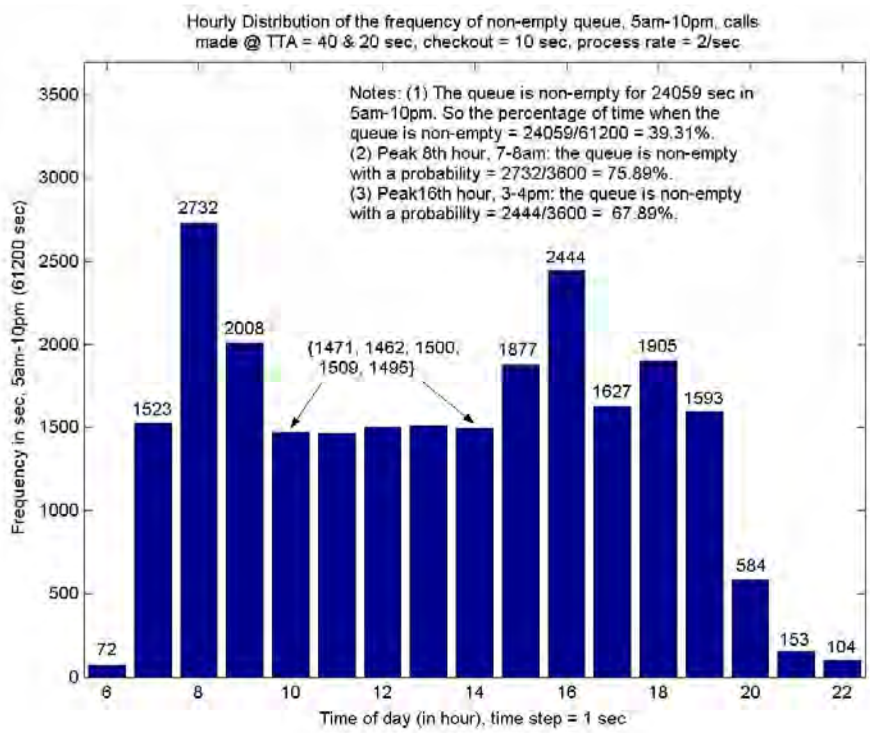

Figure 7-20 Distribution of non-empty queue by hour; multiple requests {40, 20, 10}

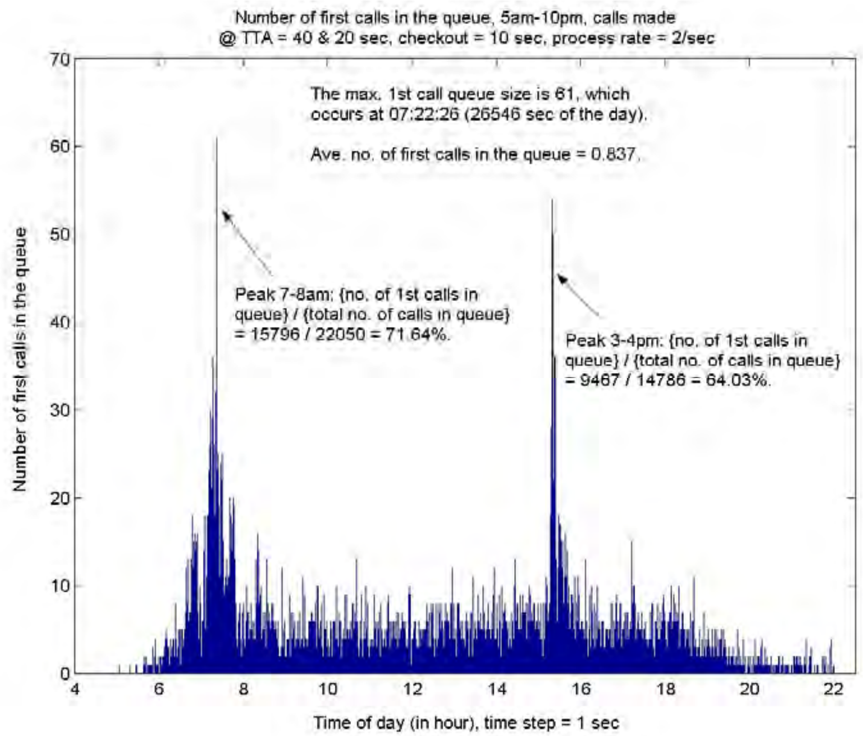

Figure 7-21 Number of first calls in the queue; multiple requests {40, 20, 10}

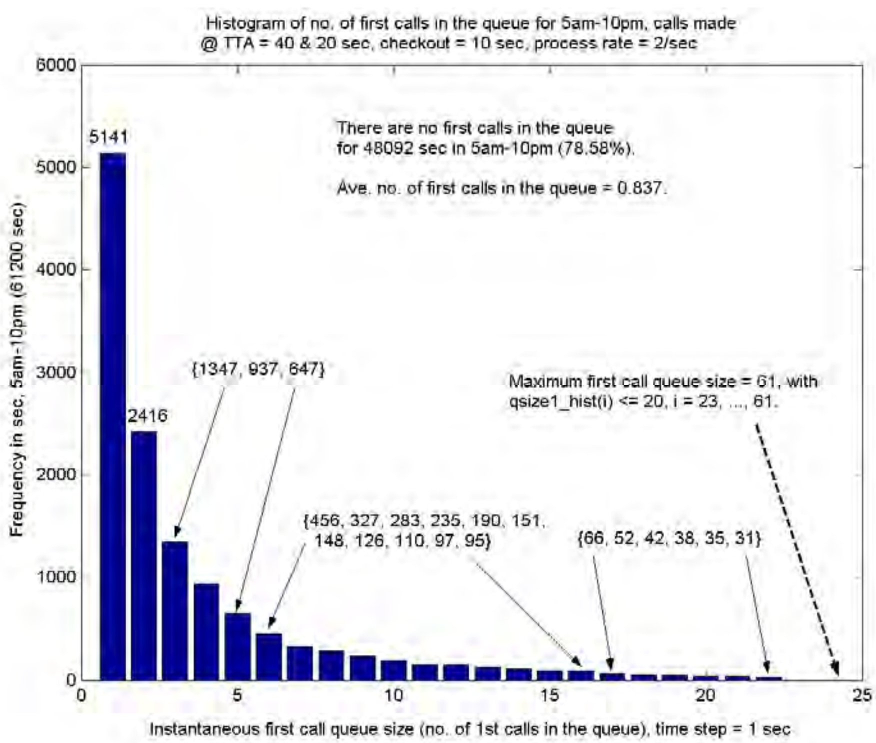

Figure 7-22 Histogram of first call queue size; multiple requests {40, 20, 10}

The queue waiting times for the processed first, second and checkout calls are shown in Figure 7-23, Figure 7-24, and Figure 7-25. As expected, the probability that a type of call is not processed immediately (with queue waiting time > 0) decreases with the priority of the type of call. The same is true for the average waiting time. The average queue waiting time for processing a first call is 3.17 sec, but this average decreases to 1.25 sec for a checkout call which has the highest priority.

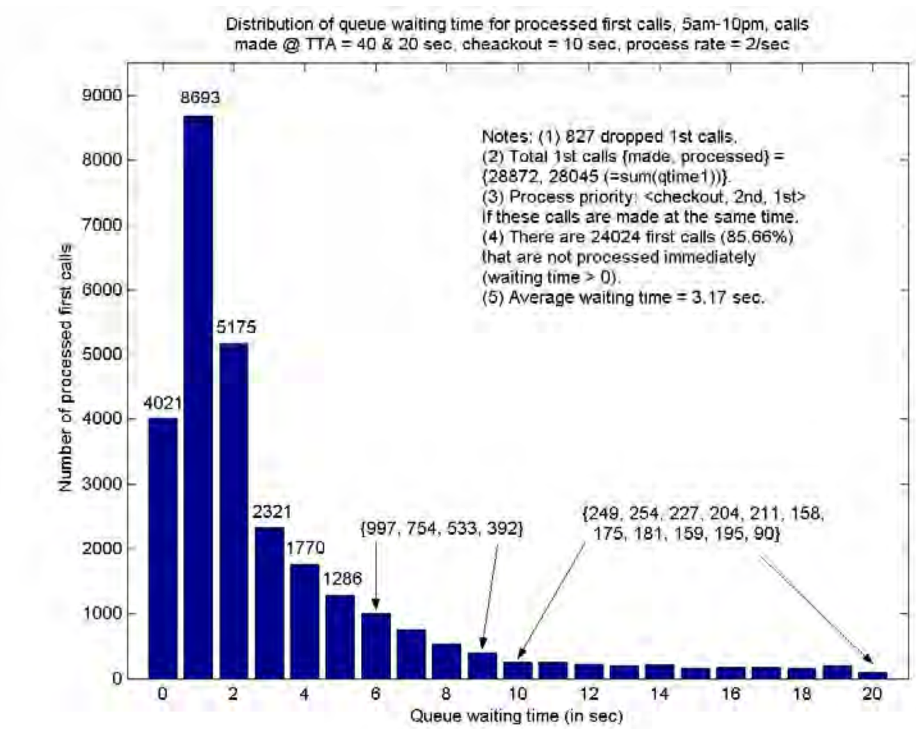

Figure 7-23 Histogram of first call waiting time; multiple requests {40, 20, 10}

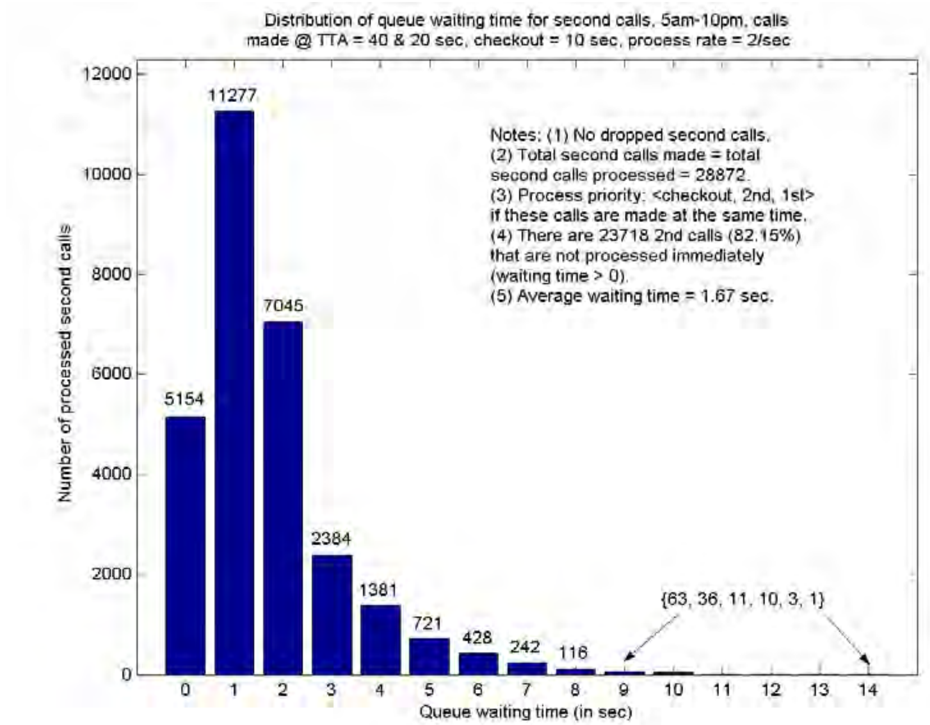

Figure 7-24 Histogram of second call waiting time; multiple requests {40, 20, 10}

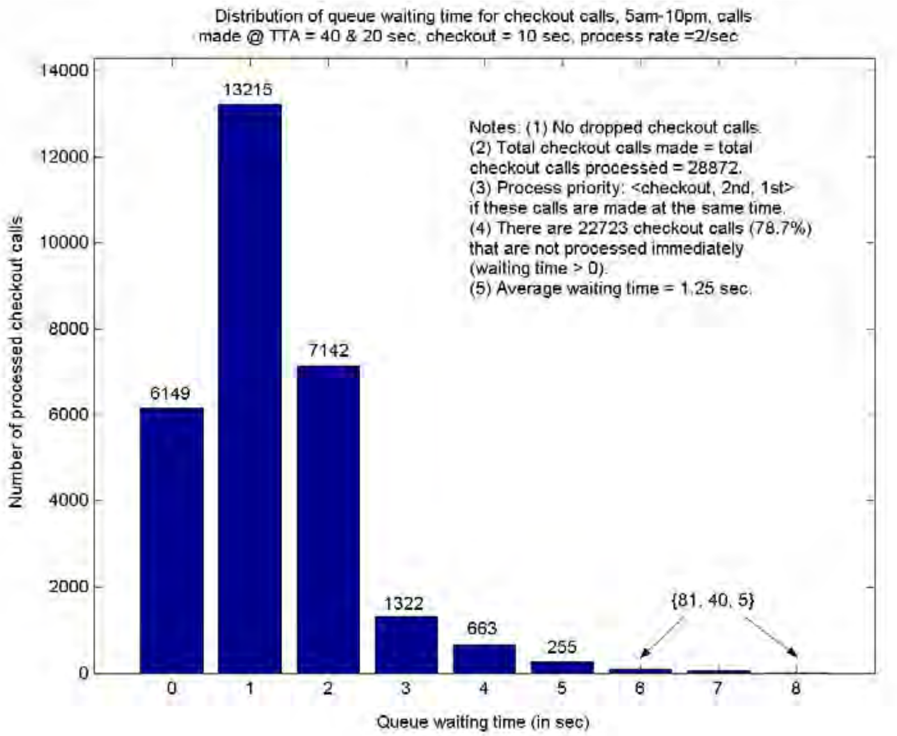

Figure 7-25 Histogram of checkout call waiting time; multiple requests {40, 20, 10}

7.2.6 Taking into Consideration On-Time Performance Measures

When a transit vehicle is in the vicinity[1] of a signalized intersection, the AVL system will compare the actual arrival time with its scheduled arrival time and report whether the vehicle's on-time performance is either "late", "normal" or "early" as defined by the fleet operator. For a given bus route, let the scheduled and actual arrival times at the i^{th} intersection be $T_s(i)$ and $T_a(i)$, respectively. The time difference $\delta T(i) := T_s(i) - T_a(i)$ is a measure of the on-time performance of the transit vehicle at the i^{th} intersection. For SamTrans operations, a transit vehicle is early at the i^{th} intersection if $\delta T(i) > 59$ sec. It is *late* if $\delta T(i) <$ -359 sec (behind schedule by more than 5 min. 59 sec). Otherwise, its performance at the intersection is

[1] The exact radius of this area also depends on the statistical variance of the GPS error in the AVL data.

reported as normal. These performance data are recorded and stored in a database. They are transmitted to the TMC using the polling channel when the transit vehicle is polled. The data for each route can be grouped by the times-of-day (peak, off-peak hours) and by seasons (school-in-session months, summer months, November-December holiday season). The statistics of these historical data can be used to analyze and understand how much of the request demand could be reduced based on the high probability that some transit vehicles will not need signal priority at some intersections, at certain times-of-day and in certain seasons.

We simulate a scenario when transit vehicles are on time for 20% of the time during 5am-10pm, and consequently, do not need to request signal priority 20% of the time. In the simulation, a random number uniformly distributed on [0, 1] is generated for each scheduled arrival time. If it is less than 0.2, there is no need to request signal priority for the chosen scheduled arrival time. Figure 7-26 shows the histogram of the number of requests processed per second for a sample run. In this run, 5837 out of 28872 scheduled arrival times (20.22%) were randomly and uniformly chosen not to request priority. from 5am-10pm. The number of dropped calls is 44, which is a significant reduction compared to that depicted in Figure 7-16 (where there are 827 dropped first calls) when on-time performance is not considered. All of the dropped calls occur during the two peaks hours, 7-8am and 3-4pm. Figure 7-27 shows that the ratio of first calls dropped over those made is only 0.35% for 7-8am. It was 21.54% if on-time performance was not considered (seeFigure 7-18). The histogram for the first call queue size is shown in Figure 7-28. If we compare this to Figure 7-22, the queue length is obviously much shorter, and the average number of first calls in the queue is also smaller. As expected, the average waiting time for processing a first call is shorter and a higher percentage of them are processed immediately. The waiting time distribution for processed first calls is shown in Figure 7-29. These calculations are for one sample run. We could simulate more sample runs to obtain a better approximation of an "average" scenario and account for the variations in the scheduled travel times between successive intersections (seeFigure 7-5). One can employ a variance reduction simulation technique such as the Monte Carlo Method to accomplish this.

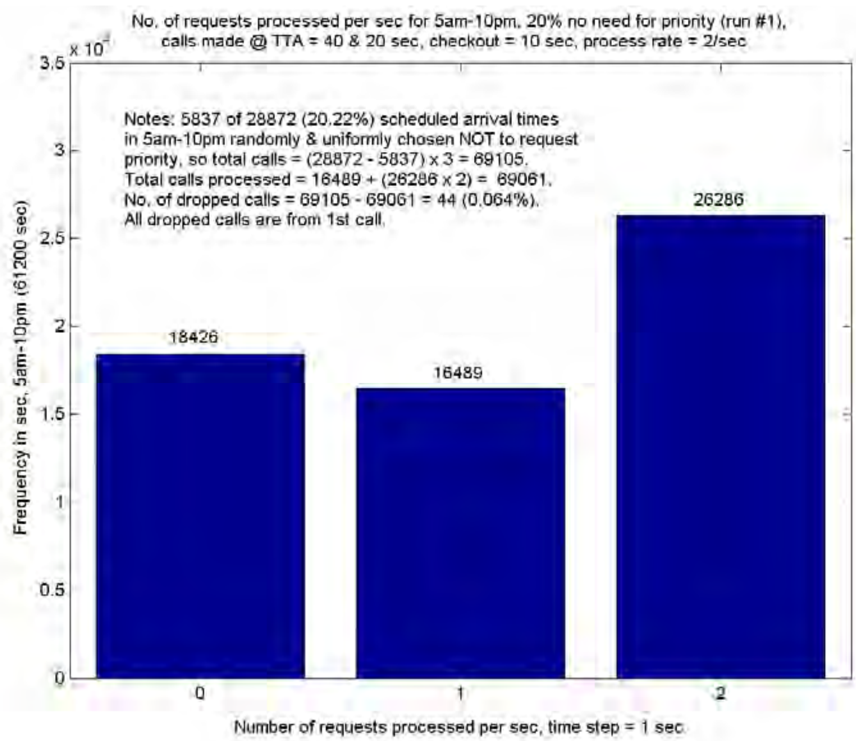

Figure 7-26 Histogram of no. of requests processed; 20% no need for priority

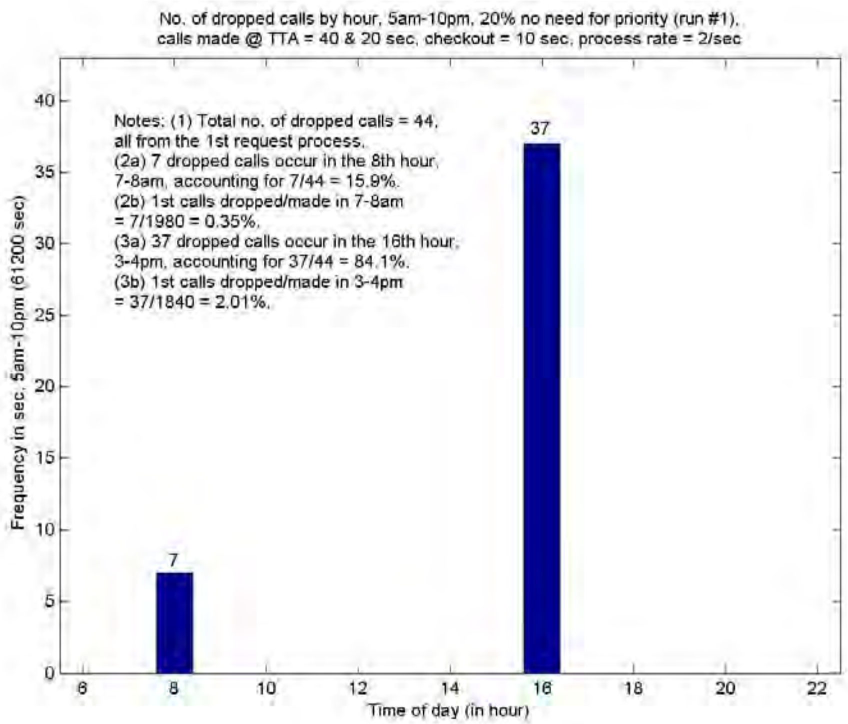

Figure 7-27 Distribution of no. of dropped calls by hour; 20% no need for priority

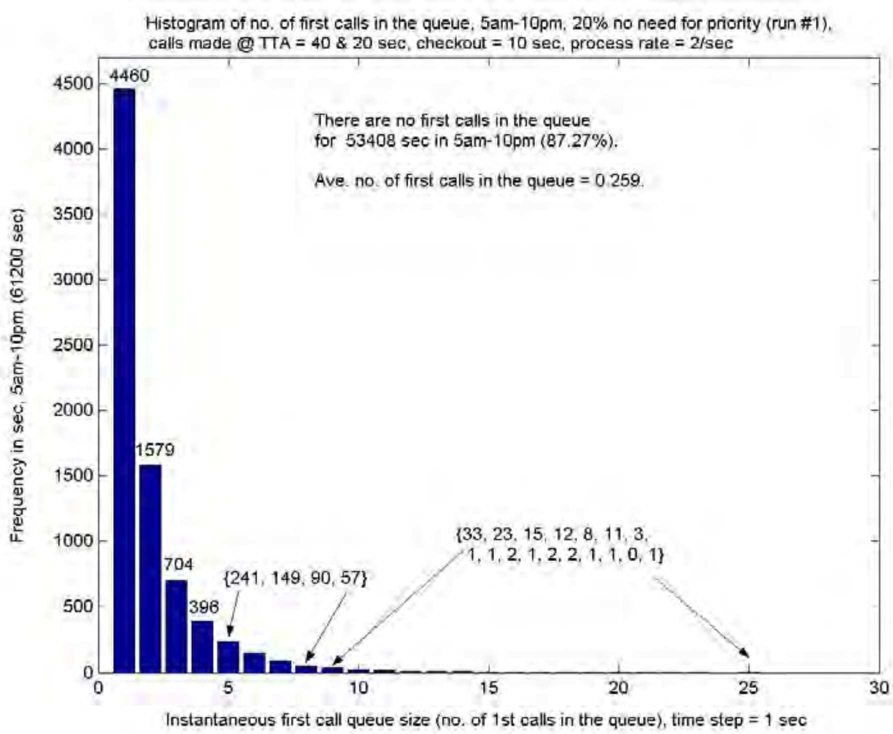

Figure 7-28 Histogram of first call queue size; 20% no need for priority

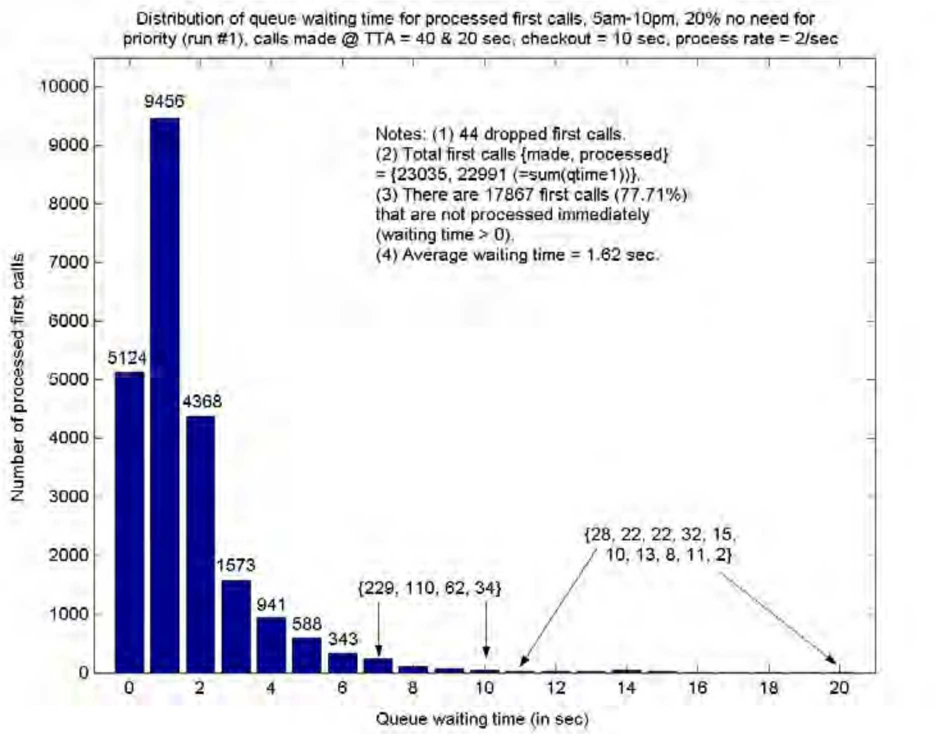

Figure 7-29 Histogram of first call waiting time; 20% no need for priority

We propose that it is highly probable that transit vehicles are on time during off-peak hours and should not need to request signal priority and that priority is needed only during peak hours. If this hypothesis is true, the request demand would be even lower than the 20% on-time performance scenario we just simulated, making it possible to process all the requests with only a small number of dropped calls. More extensive analysis will be performed when on-time performance historical data from SamTrans are made available.

7.3 Concluding Remarks and Future Work

In this report we formulate and analyze by simulation the problems of establishing communication between transit vehicles and TMC for requesting signal priority service. The availability of communication channels is, however, not guaranteed because of the potentially large number of transit vehicles requesting the service at the same time and the limited number of time slots per second available to process these requests. The communication establishment is formulated as a supply-demand problem. The demand for priority service depends on the time it is requested and also the projected arrival time at the next signalized intersection. The arrival time is stochastic and needs to be predicted using a TTA algorithm. On the supply side, the requests are processed by a polling channel and a contention channel. The polling channel cannot process all the simultaneous requests because of its slow polling frequency. The contention channel has a limited number of time slots for processing the requests. So we need to design an efficient protocol to coordinate the processing of the requests, so that requests with higher priorities are processed first and the number of dropped request processes is minimized.

We have analyzed a simple supply-demand model by considering deterministic scheduled arrival times at signalized intersections. The schedules allow us to estimate an "average" demand profile for signal priority at each intersection. We begin with a demand model where there is only one request process (or call) that is initiated when transit vehicles are at a fixed time away from their next intersections. As part of the protocol, we have set a rule requiring a request with the earliest deadline to be processed first. This requirement is realized by imposing a scheme on the number of repeated requests in the subsequent seconds if the initial request is not processed, so that the probability of establishing communication with the TMC increases with increasing priority of the request process. The process priority implies that unprocessed request processes are placed in a “virtual” queue. This is discussed in Section 7.2. This concept of priority is extended to cases when there are multiple request processes from a transit vehicle: (i) a first request process begins when a transit vehicle is farther away from its next intersection, (ii) a second request process begins when the same vehicle is closer to the intersection, and (iii) a checkout process that begins when the vehicle passes the intersection. If the three types of request processes are initiated simultaneously, the checkout process has the highest priority, followed by the second request and then the first request. This process priority protocol is discussed in Section 7.5.

We simulated a scenario when each transit vehicle makes only one request process for signal priority at its next intersection. Results show that the queue size and queue waiting time are sensitive to the process rate of the contention channel. The sensitivity to the call time (time of the initial request) is approximately concave. In the event there are no dropped calls, the queue is almost insensitive to the call time as long as the difference between the scheduled arrival time and the call time exceeds the longest queue time. We also simulated a scenario when there are three request processes. The results show that as long as there are at least 15 sec to process the second calls and at least 9 sec to process the checkout calls, all the dropped calls are from the first request process. The drops primarily (~95%) occur during the two peak hours, 7-8am and 3-4pm. We conclude that with a process rate of two requests per second and the process priority protocol, the contention channel is capable of processing the deterministic demand profile obtained from the scheduled arrival times at intersections. Nonetheless, we should emphasize that the results depend critically on the distribution of the arrival times. If the schedules are revised or if we include additional bus schedules, the results would be quite different.

We also simulated a scenario when on-time performance measure is included. We assumed that signal priority is not needed for 20% of the time from 5am-10pm. Simulations show that the number of dropped first calls in the three-process scenario is reduced significantly, from 827 to only 44. This reduction implies that the contention channel has the capability to process requests from a larger network of transit vehicles if necessary. We expect that the on-time performance measures will be different for peak and off-peak hours, with the need for signal priority being much lower in off-peak hours. If historical on-time performance data are available in the near future, the protocol could be refined to reflect these differences in peak and off-peak hours. We expect the number of dropped calls will be further reduced.

So far we have assumed a deterministic demand profile. The arrival time at each intersection is random and needs to be predicted using a TTA algorithm. The statistical properties of the TTA prediction error distribution for each link connecting two successive intersections need to be included. For a link with a smaller variance, the corresponding set of request processes should have a higher priority. This will serve as another guiding principle in further revising the protocol. A protocol for communication establishment must also consider the communication requirements of the AVL system. The polling schedule should also be included in the analysis so that the polling channel can be utilized for processing some of the signal priority requests. These are also not covered in this report. We hope to cope with all the above unconsidered issues in the next phase of the project.

7.4 References

[1] Chin-Woo Tan, Sungsu Park, Kun Zhou, Hongchao Liu, Peter Lau, Meng Li, and Wei-Bin Zhang, "Prediction of Transit Vehicle Arrival Times at Signalised Intersections for Signal Priority Control", *Proceedings of the 9th International IEEE Conference on Intelligent Transportation Systems* (ITSC), Toronto, Canada, 17-20 September, 2006.

[2] Meng Li, Yafeng Yin, Kun Zhou, Wei-Bin Zhang, Hongchao Liu, and Chin-Woo Tan, "Adaptive Transit Signal Priority on Actuated Signalized Corridors", *Proceedings of the 84th Transportation Research Board Annual Meeting*, Washington, D.C., 9-13 January, 2005.

8 CONCLUSION AND NEXT STEPS

8.1 Conclusions

Although prevailing active TSP systems are efficient in granting priority to buses, they might incur noticeable delays to the minor-phase traffic, raising concerns among traffic engineers and thus impeding the wide-scale acceptance and deployment of TSP systems. The development and implementation of a prototype ATSP system have been conducted under PATH Project RTA 65A0026. This report has summarized a continuous research to address the issues that still remain for eventual deployment of the ATSP system.

In the improved ATSP system, the TTA predictor has added an observer to smooth out measurement noises and applying an adaptive mechanism on historical model to have its parameters dynamically updated based on real-time bus movement. An arrival time flow prediction model, based on an adaptive recursive least-squares method, has been developed to provide PRG an estimation short-term traffic arrival flow, so that a more advanced PRG that is able to adaptively and optimally select either early green or green extension and determine the corresponding signal timing strategies for the TSP operation. The objective that guides the decision is to make a tradeoff between bus intersection delays and other traffic delay. The level of the tradeoff can be adjusted via a weighting factor and should be determined through negotiations among the stakeholders on how much preference the transit operation should be given. Some case studies, hardware-in-the-loop simulation tests, and a field operational field test have been conduced to demonstrate and verify the validity of the proposed adaptive TSP system.

The results of field operational test also show very positive results. The reductions of bus trip travel time 11%, average speed 9%, and total intersection delay 17% are statistically significant. The major phase traffic delays are reduced, while the incurred minor phase delays are statistically insignificant.

According to the results of the field operational test, this project successfully addressed some issues that remained for eventual deployment of the ATSP system. First of all, a more advanced PRG has been developed in order to make an explicit and graceful tradeoff between bus delay savings and traffic delays. Second, the limitations of existing commercial GPS/AVL systems has been studied. Some initial communication protocols have been discussed. However, some aforementioned field testing results still have not reached our expectations due to the limit number of bus runs and the short testing periods. Neither the cases from “before” scenario nor the cases from the “after” scenario are enough to be statistically representative. As a result, the before-and-after scenarios are not statistically comparative which might add some biases to the results. Therefore, the extensive filed operational tests and continuous research can further verify and improve the ATSP system.

8.2 Next Steps

The next steps of the research are to conduct a field operational test with SamTrans Bus Route 390/391 along the El Camino Real corridor at the San Mateo County. The field operational test could serve three purposes: 1) through the test, the ATSP system can be further improved and refined to be suitable for large-scale implementations. 2) Field data analysis and interviews with personnel from SamTrans and city traffic authorities would be able to fully reveal the impacts of the system; and 3) demonstration of the system would facilitate the widespread deployment of the ATSP concept and thus lead to more significant benefits.

The following critical issues will need to be addressed in the follow-up study:

8.2.1 Form a stakeholder advisory committee to facilitate implementation and evaluation of the ATSP system

The stakeholder advisory committee could consist of members from Caltrans, SamTrans, traffic authorities of cities along the El Camino Real corridor and the regional metropolitan planning organization. The committee can provide a channel to understand the technical and institutional concerns from different stakeholders, to engage the stakeholders in discussions that could lead to a more well-accepted TSP system, and to obtain feedback and comments for evaluating the system's performance. Direct and on-going communications between the implementing team and the stakeholders will facilitate the implementation, operation and evaluation of the ATSP system and is very critical for the success of the field operational test.

8.2.2 Investigate a robust and efficient system architecture suitable for a large-scale implementation and identify the corresponding communications links and means

A robust and efficient system architecture is the basis for a successful implementation of the ATSP system. The efficiency and reliability of the architecture of the prototype system should be examined and refined wherever appropriate. Items requiring examination include the locations of elements of PRG and PRS, identifying data transfers necessary to perform logical functions, determining communications means and the associated message latency and building up system redundancy. The architecture should be made suitable for the large-scale field operational test. This includes consideration of field conditions, customization requirements, budget constraints and more importantly, the infrastructure in place (and/or in planning), such as traffic control system, transit management system, equipped transit ITS technologies and communication links.

8.2.3 Work with vendors of bus advanced communications system (ACS) to make the ACS capable of generating priority calls

As described in Chapter 7, a dynamic polling algorithm has been developed to allow the existing GPS/ACS to support ATSP data needs. To implement the algorithm, it will be necessary to design the interfaces with the ACS system, offer recommendations and standards for requirements for TSP and GPS/ACS that will facilitate the integrated application of

GPS/ACS and signal priority technologies. Finally the implementation will require close work with the vendor for the SamTrans ACS system, in order to modify the ACS system to comply with the above recommendations.

8.2.4 Develop a more advanced priority request sever to manage and prioritize the requests generated from multiple buses

A PRS manages and prioritizes priority requests generated by PRG, and sends one or more service requests to signal controllers for execution. For a large-scale implementation, PRS plays an important role in achieving system-wide benefits. This task will improve the current PRS, which is heuristic, by integrating the PRS and the PRG. More specifically, the priority timing strategy will be determined in context of the needs of multiple buses, thereby resulting in a balance between the delay of individual bus (the "latest" bus would have the highest weight) and the system-wide traffic delay.

8.2.5 Implement and integrate the ATSP system along El Camino Real with multiple intersections and multiple buses

The equipment installation could be undertaken by engineers from SamTrans and Caltrans District 4, with assistance from PATH. Additional work will involve calibrating the ATSP system for the corridor and for each intersection. The components that need calibration include the bus arrival time predictor, the arrival flow pattern predictor and the TSP algorithm. Other primary tasks include system integration, system inspection and fine-tuning.

8.2.6 System evaluation and demonstration

It would be extremely beneficial to the transit community if the field operational test can provide a comprehensive, detailed and objective evaluation. The evaluation would attempt to capture every aspect of the impacts of TSP treatments on prioritized buses, major-phase traffic, minor-phase traffic and pedestrians. "Before and after" data analysis should be performed, using a detailed, integrated data set. It will be important to include transit operational data from the GPS/ACS system, traffic signal status, TSP event logs from the traffic control system and traffic condition data from detectors. The evaluation of detailed field data could be complemented through interviews with members of the stakeholder committee and the general passengers before and after the field operational test.