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

Field Operational Tests of Adaptive Transit Signal Priority Systems

Meng Li, et al.

California PATH Research Report
UCB-ITS-PRR-2010-35

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 6400

June 2010

ISSN 1055-1425

Field Operational Tests of Adaptive Transit Signal Priority Systems

Final Report for Task Order 6400

Prepared by:
California PATH
University of California, Berkeley
and
California Department of Transportation
in collaboration with
San Mateo County Transit District

June 2010

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 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
Wei-Bin Zhang
Yue Li
Guoyuan Wu
Fanping Bu

Division of Research and Innovation, Caltrans:
Sonja Sun

Division of Traffic Operations, Caltrans:
Kai Leung

District 4, Caltrans:
James Lau

STATE OF CALIFORNIA DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV. 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA08-0948		

4. TITLE AND SUBTITLE Field Operational Tests of Adaptive Transit Signal Priority Systems	5. REPORT DATE June 2010 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Meng Li, Kun Zhou, Wei-Bin Zhang, Yue Li, Guoyuan Wu, Fanping Bu, Sonja Sun, Kai Leung, and James Lau	8. PERFORMING ORGANIZATION REPORT NO. UCB-ITS-PRR-2010-35
9. PERFORMING ORGANIZATION NAME AND ADDRESS California PATH Program University of California, Berkeley 1357 .46th St. Richmond, CA 94804	10. WORK UNIT NUMBER 193 11. CONTRACT OR GRANT NUMBER 65A0208 TO6400
12. SPONSORING AGENCY AND ADDRESS California Department of Transportation Division of Research and Innovation P.O. Box 942873, MS 83 Sacramento, CA 94273-0001	13. TYPE OF REPORT AND PERIOD COVERED 9/1/2005 – 9/30/2008 14. SPONSORING AGENCY CODE

15. SUPPLEMENTAL NOTES

16. ABSTRACT

An Adaptive Transit Signal Priority (ATSP) system is built upon and integrated with the existing Transit Automated Vehicle Location/Advanced communications system (AVL/ACS) to provide transit buses with needed priority by adaptively determining an optimum signal timing strategy based on status of both the concerned transit bus and the traffic conditions. PATH has conducted a series of studies on ATSP in collaboration with the California Department of Transportation (Caltrans) and San Mateo County Transit District (SamTrans) since 2000, involving the development and implementation of a prototype ATSP system and studies of the deployment issues. This report summarizes a continuous research effort for refining the ATSP system through field operational tests in order to make it readily deployable.

17. KEY WORDS	18. DISTRIBUTION STATEMENT	
Transit Signal Priority, Signal Optimization	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. PRICE
Unclassified		

Reproduction of completed page authorized

DISCLAIMER STATEMENT

Abstract

Adaptive Transit Signal Priority (ATSP) system is built upon and integrated with the existing Transit Automated Vehicle Location/Advanced communications system (AVL/ACS) to provide transit buses with needed priority by adaptively determining an optimum signal timing strategy based on status of both the concerned transit bus and the traffic conditions. PATH has conducted a series of studies on ATSP in collaboration with the California Department of Transportation (Caltrans) and San Mateo County Transit District (SamTrans) since 2000, involving the development and implementation of a prototype ATSP system and studies of the deployment issues. This report summarizes a continuous research effort for refining the ATSP system through field operational tests in order to make it readily deployable.

Executive Summary

Transit Signal Priority (TSP) has been identified as a critical technology for deployment of Bus Rapid Transit (BRT) systems and for improvement of traditional transit services. Although prior applications of TSP for BRT and conventional transit applications have demonstrated positive effects in improving transit travel times and reliability, thereby increasing the attractiveness of public transportation, deployment of TSP has been a challenge in the United States because of the concern from traffic operation authorities that frequent TSP operations could deteriorate the overall performance of signal control that was optimized for traffic flow, and the capital cost for instrumentation of the TSP system.

California PATH has developed the concept of Adaptive Transit Signal Priority System (ATSP), intending to overcome the above issues. PATH has conducted a series of studies on ATSP in collaboration with the California Department of Transportation (Caltrans) and San Mateo County Transit District (SamTrans) since 2000. The development and implementation of a prototype ATSP system have been conducted under the project RTA64A0026 “Implementation of ITS Technologies for BRT”. The improved ATSP system, which also addresses issues toward deployment, has been concluded under PATH Project Task Order 5404 “Toward Deployment of Adaptive Transit Signal Priority Systems”. This report summarizes a continuous research effort to address Caltrans’ call for a field operational test to further evaluate/validate the ATSP system, refine the system and transform the system from a prototype to a readily deployable system.

ATSP System Overview

An ATSP system is built upon and integrated with the existing Transit Automated Vehicle Location/Advanced communications system (AVL/ACS) to provide transit buses with needed priority by adaptively determine an optimum signal timing strategy based on status of both the concerned transit bus and the traffic conditions. The ATSP architectural design follows National Transportation Communications for ITS Protocol (NTCIP) scenario 3, in which the Priority Request Generator (PRG) and Priority Request Server (PRS) are located at the Traffic Management Center (TMC). Such system setup has been emulated and tested in the field under this study. This architecture allows the usage of the existing advanced communications system (ACS) for ATSP without needing additional communication links, therefore enables a most cost effective deployment of TSP.

Priority Request Generator (PRG): A PRG has been 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.

Different than the traditional TSP, the ATSP contains a bus arrival time predictor that can estimate the bus arrival at the intersection some seconds away. The arrival time prediction (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.

By knowing an accurate bus arrival time at a longer 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".

Priority Request Server: The PRS processes and prioritizes all priority requests from different PRGs and generates service requests to be granted by the local signal controller. The PRS keeps monitoring current signal status, which is updated every second by each local signal controller. Based on pre-determined prioritization criteria, the PRS selects the highest priority of service as the potential service request and becomes a validated service request if the Time to Serve (TTS) conditions are met. Upon the selection of a priority service request, the PRS compares it with the current serving request to determine whether it conflicts with the on-going service and determines whether to process the priority request. After priority is generated, the PRS keeps monitoring traffic signal status and priority service status and sends out a clear service message upon the termination of the priority service phase. The ATSP prioritization requests criteria that can be designed based on the principles and constraints agreed upon with traffic control agencies. The essence of these principles is to enhance the intersection safety and to reduce the disturbance of providing transit priority treatments to the underlying signal control strategy.

The protocols for applying advanced communications system (ACS) for ATSP: A prior study investigated the feasibility of using existing ACS as a communication backbone for facilitating TSP requests. Further investigation on the feasibility for the El Camino Real corridor application is simulated by using the collected field data. This analysis shows that it is viable to implement the proposed TSP request schedule using the existing ACS communications system.

Field Testing System

Working with California Department of Transportation (Caltrans) and San Mateo Transit District (SamTrans), the field operational tests were conducted along the central part of Bus Routes 390 and 391 on El Camino Real (State Route 82), from 42nd Ave in the City of San Mateo to Rosedale Ave in Burlingame. The test section is about 6.6 miles long, covering 35 signalized intersections. The existing traffic control systems are operated under the Caltrans C-8 Traffic Signal Control Software together with Model 170E Control System hardware. The systems provide coordinated semi-actuated traffic control with multiple times-of-day plans. The control software was upgraded with the enhancement to support the ATSP operation.

Bus Routes 390 and 391 share the 27 northbound and 28 southbound bus-stops along the section, with three bus-stops as time-points on both directions. Fifteen 390/391 buses operating in revenue service have been installed with the automatic vehicle location (AVL) systems. A suit of software was developed to emulate the central communications scheme for SamTrans' AVL/ACS system with lower a pooling rate. Laboratory testing was conducted to evaluate and calibrate the proposed communications protocols. It was concluded that the additional prediction error (±1.5 seconds) due to the discrete communications protocols is not significant. The incurred increment or decrement of bus delay is less than 1.5 seconds. Due to multiple traffic phases, the incurred changes on each traffic phase duration will be less than 1 second. Moreover, the additional prediction error only impacts the EG treatment, which represents 58.5% of all priority cases. The average impacts on both bus delay savings and traffic phase durations are less than 1 second and very minor. Therefore, the proposed communications protocol using ACS can support the ATSP system without a significant compromise on system performance.

In order to enable ATSP to accommodate a larger size of transit fleets and a greater number of intersections, the ATSP functional elements, such as bus arrival time prediction (ATP), PRG and PRS, were modularized into a layered structure, with the upper layer dealing with the coordination and data exchange and the lower layer handling separate buses and individual intersections. For the data analysis, we also developed a comprehensive transit and traffic database under MySQL (an open source database software). Software tools have also been developed to pre-process and clean the data, save the data into the database, monitor bus movements and ATSP operations in real-time, and calculate the measures of effectiveness (MOEs).

Summary of the Findings through Field Testing Results

After a six-week "before" and a six-week "after" field tests, the results of field operational testing show very positive results. Field operational tests have convincingly demonstrated the effectiveness of ATSP. The findings of ATSP benefits are summarized as follows:

- Bus delays at intersections was reduced by 18.5% for northbound trips and 32.0% for southbound trips;
- Number of stops for red was reduced by 9.5% (northbound) and 16.8% (southbound);
- Average waiting time per stop for red was reduced by 10.5% (northbound) and 18.5% (northbound);
- Average bus running time was reduced by 4.9% (northbound) and 5.6% (southbound);
- All of the above changes are statistically significant at the 5 percent significance level.
- ATSP has no statistically significant impacts on traffic intersection delays.

The FOT along the 35-signal arterial boulevard achieved very positive results, demonstrating that bus service can be significantly improved while the impacts on traffic intersection delays are not statistically significant. This evaluation study also demonstrated that TSP can offer greater benefit for headway-based bus service. For schedule-based bus service, a re-design of the bus schedule to take advantage of TSP would achieve more significant benefits.

The field operational test addresses the deployment issues of the ATSP system and demonstrated that the ATSP system in the way it has been designed is practical for deployment from a number of perspectives:

- The proposed ATSP system architecture is suitable for the large-scale implementation.
- The PRG and PRS algorithms, including bus arrival prediction can meet the goals of reducing bus travel time and minimizing impact to traffic.
- The existing ACS as a communications tool can be used as the backbone to facilitate TSP requests.

The Next Steps

We are working with Samtrans and Caltrans to expand the field testing site to include more TSP capable intersections. "Before" and "after" data collection will be conducted. The evaluation will include a systematic, comprehensive, and objective assessment of 1) technical/system performance, 2) traveler response, 3) emission reduction for buses due to the time saving obtained from TSP and estimation of emission reduction of automobiles due to mode shift because of the introduction of TSP and 4) institutional issues.

After the successful completion of the FOT, one of the immediate next steps is to assist transit agencies, which have already deployed AVL/ACS and have desires of implementing integrating ATSP with their AVL/ACS system, to assess possibilities of deployment of ATSP. The interface with a typical AVL/ACS system will be developed, so the central ATSP elements of PRG and PRS can directly communicate with the transit ACS system.

Future research can be conducted in a number of for further improvement of the deployability of ATSP, including:

- The ATSP system will build upon GPS-based bus AVL systems. GPS signals can be blocked by tall buildings, trees or overpasses, which may degrade the TSP performance at certain geographic locations. We plan to develop means to maintain the ATSP system functional when GPS signals are not reliable.
- We will explore possibility of applying Dedicated Short Range Communications System (DSRC). DSRC is a low-cost wide bandwidth communications infrastructure that is currently developed under the National and California Vehicle-to-Infrastructure Integration (VII) program and is expected to be massively deployed for automobile applications in the near term. DSRC-based ATSP system will be an alternative to the GPS-based system for no-GPS bus systems.
- The predicted bus arrival time can be used for other transit applications, such as prediction of bus arrival time at stations/stops for dynamic passenger information (DPI) and for real-time integrated transit and traffic information. Integrating ATSP with DPI and other features will support integrated application of ITS technologies.

Table of Contents

1 Introduction

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 of active TSP systems often adopt ad-hoc or heuristic TSP control logic, which are not adjustable in a real-time manner. For example, early green (EG) 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 a 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 a wide range of closed-loop signal control systems, including 170E controllers. This may have potential for wide-scale implementation given an existing installed technology base.

California PATH in collaboration with the California Department of Transportation Caltrans (Caltrans) and the San Mateo Transit District (SamTrans) have conducted research and have developed an Adaptive TSP (ATSP) system which utilizes the existing Global Positioning System (GPS) based bus Advanced Communications System (ACS) already deployed by many transit agencies and a signal priority scheme that can provide trade-offs between savings of bus travel time and delays for general traffic therefore to achieve optimum performance for both buses and traffic. The ATSP concept and a prototype ATSP system have been developed under a research agreement between Caltrans and PATH: RTA 65A0026 entitled "Development of Adaptive Transit Signal Priority Systems" (Ying et al. 2004). The improved ATSP system was implemented through a collaboration with Caltrans Headquarters and District 4 under a separate contract between Caltrans and PATH under Task Order 5404"Toward Deployment of Adaptive Transit Signal Priority Systems" (Li et al. 2007). This project also addressed ATSP deployment issues.

This report summarizes a continuous research effort to address Caltrans' call for a field operational test to further evaluate/validate the ATSP system, refine the system and transform the system from a prototype to a readily deployable system.

1.1 Overview of the TSP Technologies

TSP is an operational strategy that facilitates the movements of in-service transit vehicles through signalized intersections (Baker et al. 2004). TSP can be implemented in a variety of ways. A typical TSP system includes four functional applications: vehicle detection, priority request generation (PRG), priority request server (PRS), and TSP control strategies (Smith and Hemily 2005). Priority treatments include passive priority, green truncation, green extension, actual transit phase, phase insertion/rotation, and adaptive/real-time control.

When designing and implementing a TSP system, two fundamental questions need to be carefully assessed: how TSP functionalities are utilized with the existing infrastructures and what impacts the priority implementation will have on general traffic. Although prior deployments of TSP do demonstrate positive effects on reducing transit intersection delay, shortening transit trip time and improving service schedule adherence, concerns have been frequently raised that TSP operations may interrupt the normal operation of signal control and thereby increase delays to general traffic, the capital cost for outfitting the TSP system, and that how the deployed TSP system adapts to advances in technology.

In conventional TSP systems, the priority request is just a flag indicating the presence of a transit vehicle, and the signal controller grants priority based on pre-determined logic. This could result in providing non-necessary priority if the bus is too far away from the intersection and inevitably cause noticeable delay to the non-prioritized traffic, or providing not-enough priority if the bus is too close to the intersection. Besides, the TSP strategies solely consider the transit needs and thus it could cause unbalanced impacts on different movements.

Since the late 1990s, the development of Automatic Vehicle Location (AVL) and Advanced Communications Systems (ACS) to better monitor and control operations has become prevalent throughout the U.S. transit industry. GPS-based AVL is already widely used by U.S. transit agencies. By 2004, AVL technology had been deployed in 54% of all transit buses among respondents to the 2004 Intelligent Transportation Systems deployment survey (TCRP Synthesis 73).

In addition to the widely deployed AVL/ACS, many transit agencies have installed a separate system for TSP, despite the fact that these two systems have common key features such as vehicle positioning and communications. The resulted disintegrated system is a significant waste of resources, has less adaptability to advances in technology, and is costly to upgrade.

TSP involves coordinated improvements in a transit system's infrastructure, equipment, operations, and technologies in order to give preferential treatments to transit vehicles at signalized intersections. According to the Intelligent Transportation Society of America (ITS America), a TSP system may consist of three major components: the Transit Vehicle Detection/Priority Request System that generates a request for priority; the Traffic Signal Control System that receives and processes the request for priority at the intersection(s); and a Communications System that links the Vehicle Detection System with the Traffic Signal Control System, possibly through a Transit/Traffic Management Center (TMC).

1.1.1 Transit Vehicle Detection System

The transit vehicle detection system is critical for implementing TSP. It falls into two main categories: Selective Vehicle Detection (SVD) and AVL. SVD communicates with roadside equipment or inductive loops via an On Bus Unit (OBU) to provide transit vehicle information, while the AVL system often refers to the GPS which is able to monitor the continuous movements of transit vehicles.

The effectiveness of TSP relies on, to a considerable extent, transit vehicle detection and location methods. Three main families of vehicle detection methods have been identified: point detectors, zone detectors and vehicle movement detectors.

Point detectors are one of the most common forms of detection used for TSP. Vehicles are equipped with "electronic registration plates" or "tags", which send a message including the vehicle identification (ID) to a specialized loop detector able to read this message when the vehicle crosses it. In this case, a vehicle is detected at a precise location at a given time.

Zone detectors sense the presence of a vehicle on the approach to an intersection. An infrared, radar or radio frequency emitter is mounted on the vehicle, and the messages are received by a beacon when the vehicle occupies the area covered by this beacon; messages can include complementary information such as vehicle ID, vehicle classification, and vehicle priority level. GPS based AVL and short range communication systems have been recently applied to replace vehicle-based emitters to perform as zone detectors. Recent advances have provided additional information about the location in the zone that could be used in a fashion similar to vehicle movement detection.

Vehicle movement detectors, contract to point detectors, monitor a vehicle's movement through an area. With the improved ability to predict the arrival of the transit vehicle at an intersection, the TSP system should operate more efficiently. Vehicle movement detectors, such as GPS/AVL, are emerging as the most favorable detection systems for TSP, with different possible localization techniques.

In order to eliminate the unnecessary priority time granted for a transit vehicle, most active TSP applications require "check-in" and "check-out" detection. When point detection is used, two detectors are required for each link, one placed at the upstream of the intersection serving as the "check-in" detection to initiate a priority call and one placed at the stop line serving as the "check-out" detection to confirm that the bus has passed through the intersection. When zone detection is used, the presence call is true when the vehicle is still in the detection zone and can be assumed to have exited the zone when the call is false. In addition a priority timeout strategy is usually required to cancel a priority in case that the "check-out" detection failed.

1.1.2 Commonly Used Transit Priority Strategies

Transit and general traffic can have competing interests. The objective of TSP strategies is the reduction of delay for transit vehicles at signalized intersections and the improvement of schedule reliability, while the objective of traffic signal control strategies is to minimize the average delay of all persons and vehicles. With recent advancements in ITS, there is more capability to support transit priority in traffic signal systems, and the TSP system, if implemented properly, can provide improvements for each party. The literature on TSP strategies falls into two general categories: passive and active.

Passive priority is defined as the use of static signal settings to reduce delay for transit vehicles and operates regardless of whether a transit vehicle is present or not. Therefore it does not require a vehicle detection system. The strategies include phase splitting, progression/coordination to favor transit vehicle movements, increasing the priority phase split, or queue jumps. Passive priority strategies are attractive for applications where the transit operations are predictable, transit frequencies are high, and traffic volumes are low, due to the easy implementation and low costs. However they typically make the intersection operate less efficiently overall, especially when the transit frequency is not very high, because the signal setting will be sub-optimal when transit vehicles are not present, which will be the case a large majority of the time.

Active strategies address the limitations of passive strategies by altering signal settings dynamically upon the detection of a transit vehicle or receiving a request for priority by the Transit Vehicle Detection/Priority Request System. This strategy offers a more efficient signal operation by responding to the transit call only when necessary but requires devices to detect the transit vehicles upstream of the intersection and advanced signal controller to employ strategies for granting priority. While there are a variety of treatments of adjusting the normal signal operation to benefit transit vehicles, which may include: early green, green extension, actuated transit phase, phase insertion, and phase rotation, our review found that Early Green and Green Extension are the two dominate methods in the operational systems used in the past and on-going signal priority projects.

The early green strategy shortens the green time of proceeding opposing phases, without violating the minimum green time, pedestrian time, or clearance intervals, to expedite the return to green for the priority movement phase. The green extension strategy is similar to the early green strategy in the sense that the opposing phases are shortened after the priority phase was extended. Both methods intend to facilitate the passage of the transit vehicle in the most efficient manner, depending upon the detection time of the upstream TSP-equipped vehicle within the signal cycle.

Table 1.1 summarizes examples of TSP deployments and their benefits [Irene Li, et al.]. Note that these systems are evaluated using different measures of effectiveness and evaluation methods, therefore the results are not necessarily directly comparable.

Table 1.1 Selected TSP implementations

Location	Transit type	No. of intersection	SP strategy	Benefit	impact
Portland Tualatin Vallley Hwy	bus	10	early green green extension	bus travel time saving 1.4 to 6.4% avg.bus signal delay reduction 20%	N/A
Seattle Rainer at Genesee	bus	1	early green green extension	50% reduction of signal related stops 57% redduction in avg. signal delay 13.5% decrease in intersection person delay 35% reduction in bus travel time variability through intersection	avg. intersection delay did not change to traffic side street effects were insignificant
Europe	bus	five cases	various	10 seconds/intersection reduction in signal delay 40-80% potential reduction in transit signal delay 6-42% reduction in transit travel times in England and France 1 to 2 year payback period for installation of transit priority systems	0.3-2.5% increase in auto travel time
Portland Powell Blvd.	bus	4	early green green extension	5-8% bus travel time reduction bus person delay-general decrease	N/A
Chicago Cermak Rd	bus	15	early green green extension	7-20 reduction in bustravel time depending on time of day schedule adherence improved reduced number of buses needed to operate the service passenger satisfaction increased	8.2 second/veh avg. increase in cross-street delay
Los Angeles Wilshire& Ventura Blvds	bus	211	early green green extension actuated transit phase	8% reduction in running time 35 decrease in bus delay at signals	

1.1.3 Shortcomings of the Existing TSP systems

A common understanding of TSP operations is that frequent priority requests and services at traffic signals interrupt normal signal operations and creates delays to general traffic. The shortcomings of existing TSP control strategies are obvious from this point of view as their performance, to a different level, relies on the means of transit vehicle detection. Nearside optical detection, inductive loops, and RF tag/receiver are point or zone detection systems, which sense the presence of transit vehicles at fixed locations or within a limited area. The detection range influences TSP operations. Short detection range would result in late calls that have limited lead time for Early Green treatment and could miss the potential Green Extension treatment. Large detection range, on the other hand, would lead to less predictability in a transit vehicle's arrival at the intersection due to the uncertainties of bus movements after detection and consequently less efficient TSP operations. Most importantly, there is no optimal detection range that is suitable for any traffic conditions. Under the TSP strategy using point or zone detection, the priority operation is initiated simultaneously upon the detection of a bus regardless of its necessity, which may lead to either false priority calls or insufficient time in the signal cycle to grant enough priority service. False calls are those priority requests that are granted but actually not needed or those failed to discharge the bus during the prioritized interval. In the former case, the bus solicits a non-necessary Early Green service call when it can actually traverse through the intersection within the normal phase interval, which brings no benefits to buses but disrupts non-transit vehicles. The latter refers to the priority requests for Green Extension treatment but failed to discharge the bus within the maximum extension period, which often occurs in the case when bus detectors are located too far from the intersection. In addition, the current TSP systems usually deploy simple methods to shorten non-transit phases to provide Early Green treatments regardless of the real time traffic demand, for example, shorten each phase by a fixed and predetermined ratio, and consequently impact the general purpose traffic.

1.2 Adaptive Transit Signal Priority

AVL based TSP system has the potential to overcome the shortcomings of conventional TSP. The real time bus movement data from GPS can be used to estimate bus location as well as predict bus arrival time to bus stops and intersections. The prediction of transit's arrival at an intersection can help select the optimal time point to trigger the traffic signal controller for priority service.

The addition of real-time traffic information (density, volume) will also be of significant importance in improving signal priority control algorithms. A few of adaptive traffic signal control algorithms have been improved by the manufacturers to embed TSP functions. Two of the most promising adaptive prioritization algorithms are SCOOT (Split Cycle Offset Optimization Technique) version 3.1 and OPAC (Optimized Policies for Adaptive Control). The SCOOT kernel software allows for buses to be detected either by selective vehicle detectors (i.e., bus loops and bus-bone transponders) or AVL systems, Where SCOOT is given a bus identifier as part of the bus detection, it can match this detection with a previous detection of the same bus. This is generally possible with an AVL system; it is also possible in principle with selective vehicle detection systems, but because of data transmission restrictions, the bus identifier may not be transmitted to SCOOT and only a single bit indicates the presence of a bus. The signal timings are optimized to benefit the buses by providing either green extension or recall to an associated phase. Two alternatives exist for extensions: central extension and local extension. Central extension uses the centralized SCOOT processing to determine the priority, while local extension grants the extension locally by the signal controller to avoid the communication delay between the SCOOT central computer and the local controller. Reported bus priority field trials using SCOOT showed to buses with no significant negative impacts to general purpose traffic. In the 10-intersection Camden SCOOT area of London, 22% average bus delay saving per intersection was measured and 70% in light volumes using both extension and recall (Bretherton, et al). OPAC is an on-line signal timing optimization algorithm that optimizes traffic flow (as common signal control) as well as minimizes person delay at intersections by weighting different kinds of vehicles.

Real-time strategies, different with active TSP, attempt to provide transit priority based on optimization of selected objectives. The objective of such integrated active signal/TSP algorithm is usually to optimize traffic-signal-phasing for a model network where real-time information on passenger counts and schedule adherence is available. However, implementing transit priority within existing adaptive signal control strategies has certain flaws (Duerr, et al). For example, adaptive control systems normally consider network-wide effects in the optimization while providing transit priority involves local concern and may lead to conflicts between the local objectives of providing transit priority and the network objective of optimizing traffic flow. Furthermore, most adaptive control systems use macroscopic models of traffic flow and have difficulty capturing the details of transit vehicle movements. Strategies to integrate TSP with adaptive corridor control are being investigated by several institutes in the United States and have not been implemented in practice (Mirchandani, et al.).

Currently under investigation by PATH, adaptive transit signal priority system is designed for corridors without sophisticated network-oriented traffic control systems for the same optimization objectives to accommodate buses and at the same time minimizing the impact on general traffic. This ATSP concept has the following key features:

- It utilizes the existing transit AVL/ACS system to track, predict bus arrival time at signalized intersections, and generate priority requests. All buses instrumented with AVL/ACS are TSP capable, without requiring any additional on-board equipment.
- It utilizes the existing traffic control infrastructure to estimate the traffic needs, i.e., the green time usages and traffic demands on all approaches.
- It dynamically makes the tradeoff between the transit needs and the traffic needs to provide benefits to transit vehicles while minimizing the impact on general traffic. The priority control strategy adapts to bus movements, real-time traffic conditions and signal phasing and timing.

The studies on ATSP have shown the potential for handling the common processing and optimization of signal timing with consideration of transit priority operation.

2. Development of ATSP System

To provide the context of the FOT in the final project report, descriptions of the ATSP architecture and some of the key functional elements of the ATSP system are adopted in this report.

2.1 System Architecture for ATSP

The ATSP system architecture has been defined following the NTCIP standard, *Object Definitions for Signal Control and Prioritization (SCP)*.

2.1.1 NTCIP1211

National Transportation Communications for ITS Protocols (NTCIP) was developed to establish standards for use in implementing transit signal priority (TSP) applications within traffic signal systems. NTCIP SCP Concept of Operation describes the concept of granting priority while maintaining coordination with adjacent intersections. The functionality identified is intended to work in conjunction with the signal coordination object definitions and functions as defined in NTCIP 1202 – Object Definitions for Actuated Signal Controllers. NTCIP 1211 includes a number of signal timing parameters that would modify normal coordination parameters to allow implementation of a priority strategy. The strategies and timing parameters are under the control of the traffic signal system operator.

NTCIP 1211 also defines a Management Information Base (MIB) or data dictionary of parameter controls and status information for SCP related to:

1. Generating and monitoring the status of a request for priority from a source to a logical entity referred to as the Priority Request Server
2. Passing a prioritized list of priority requests to a controller and monitoring the status of the controller responses
3. A set of configuration parameters to manage the process of receiving and responding to priority requests

The NTCIP 1211 SCP Concept of Operations is comprised of two primary elements for transit signal priority, including the Priority Request Generator (PRG) and a Priority Request Server (PRS). The standardization occurs at the interface of these processes and represents the objects developed by NTCIP 1211. The two primary interfaces are (1) between PRG and PRS and (2) between PRS and the traffic signal controller coordinator, which implements special coordination operation. The NTCIP 1211 standard addresses the four likely signal control priority scenarios and uses cases to provide a logical architecture for implementation of TSP, including:

- Scenario 1 – Fleet Vehicle Priority Request Through Fleet and Traffic Management Centers
- Scenario 2 – Fleet Management Priority Request Through Traffic Management Centers
- Scenario 3 – Traffic Management Priority Request
- Scenario 4 – Fleet Vehicle Priority Request

A study sponsored by the Federal Transit Administration (FTA) and conducted by PATH and Kittelson and Associates recommended to FTA an additional scenario involving traffic signal priority request through traffic signals [Li, et al].

2.1.2 ATSP Architecture

ATSP architecture is flexible and could be implemented using any of the Scenarios 1 through 3, where traffic signals are regionally coordinated (including corridor level coordination). The choice of a viable ATSP architecture is primarily based on the characteristics of the available communication systems among transit vehicles, the fleet management, and TMC.

We have chosen to use NTCIP 1211 Scenario 3 as ATSP architecture, which locates the elements of PRG and PRS in the TMC. The architectural design is shown in Figure 2-1. There are a number of advantages for choosing this architecture. First of all, the amount of data from the traffic signal control system is much more than that from the transit vehicles or the fleet management center. Traffic data consists of the traffic signal status, the local and master coordinated controller clocks, the traffic demand information from multiple inductive loop detectors, and the pedestrian button status. In contrast, the transit vehicle data only includes either the AVL data or the processed travel time prediction information. Secondly, the number of intersections in the transit geographical network is much larger than the number of transit vehicles in a fleet. Thirdly, the traffic data conveys more dynamic information of the traffic situation and thus requires a higher data resolution than the transit data. Finally, once a TSP service request is generated by PRS, it requires being reached at the local signal controllers in a timely fashion. The communication latency for a PRS request is much more sensitive than that for the transit vehicle data.

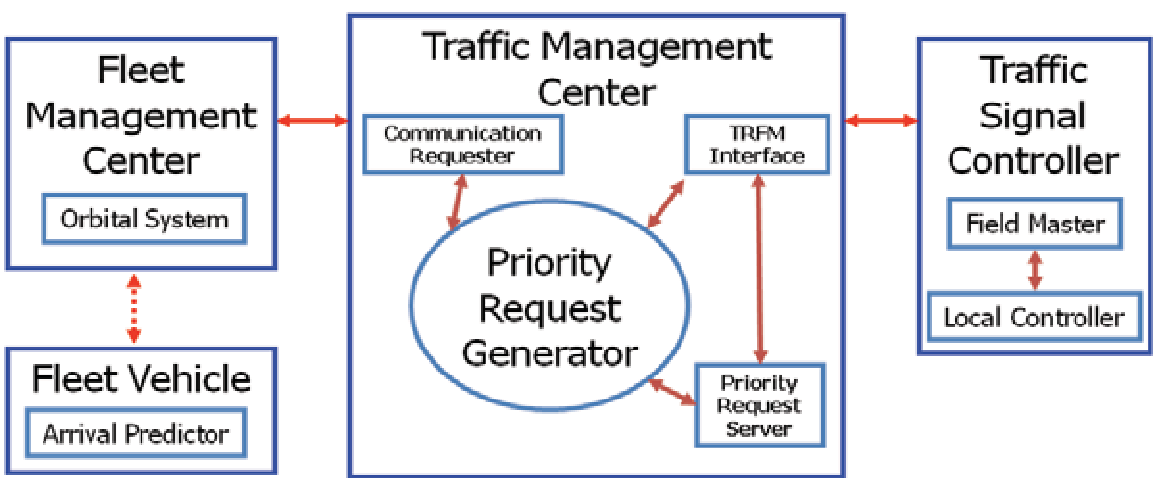


Figure 2-1 System Architecture of the Large-scale ATSP Implementation

The ATSP architecture made one variation from the NTCIP 1211 Scenario 3, involving distributing the transit arrival time predictor (ATP) to each fleet vehicle instead of at the central system in the TMC or the fleet management center. The relocation of ATP can effectively relieve the data communication load between each vehicle and the fleet management center. Under this architecture, the historical vehicle movement data is locally stored at the onboard AVL computer and only processed data based on a larger data set of raw data will be transmitted. This arrangement will allow the existing low bandwidth transit communications system to function as a communication link for ATSP and will not require any additional hardware equipment.

It is noted that most implementations of TSP or ATSP systems are motivated and planned by transit agencies or local transportation planning agencies but typically not traffic management organizations. Hence, ATSP implementation with central intelligence at the TMC will require good coordination and communication among the multiplicity of jurisdictions and project partners.

2.2 Priority Request Generator (PRG)

A PRG has been 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. Different than the traditional TSP, the ATSP contains a bus arrival time predictor that can estimate the bus arrival at the intersection some seconds away. By knowing accurate bus arrival time at a longer 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”.

2.2.1 Bus arrival time predictor

The prediction algorithm we propose uses real-time bus location and bus wheel speed information, together with historical AVL data to predict the arrival time of a bus at the next traffic light. The GPS data gives the *absolute position* of a bus and its corresponding coordinated universal time (UTC). For the type of GPS receiver installed on each of the instrumented buses, the location data have an error distribution that is approximately Gaussian with a zero mean and standard deviation of 15 meters. With the recent advancement in technology, precise positioning is possible using fairly expensive GPS receivers at reference locations (for example, in a differential GPS setup) providing corrections and relative positioning data for receivers. In an urban environment, however, the accuracy of even expensive GPS receivers suffers tremendously when satellite signals are frequently blocked or reflected (for example, due to multi-path effects). Thus GPS alone will not be sufficient for navigation, positioning, and time dissemination.

The GPS data arrive at a rate of about 1 sample per second (1 Hz). In between GPS updates, we rely on the bus wheel speed to estimate the *relative position* of the bus with respect to the last GPS coordinate update. An inertial navigation system (INS) consisting of a distributed array of accelerometers, and possibly, a few gyroscopes, could also be used to provide the location estimates between GPS updates. However, unless an expensive unit is used, the INS solution diverges quickly, especially in situations when GPS data are blocked or not available. The bus wheel speed is readily available and is a cheaper alternative to an INS. For the signal priority application, bus wheel speed is selected to complement the AVL system. The accuracy of the GPS data is also affected by the uncertainties in communication delays and data packet losses. The bus wheel speed data are used to compensate for these drawbacks to obtain more accurate bus positions. The *current* bus location indicates how far the bus is from the next intersection, but is insufficient to predict the time-until-arrival (TUA). We also need real-time downstream traffic conditions. The downstream bus wheel speed and GPS data reveal some information of the downstream conditions imposed on the bus movement.

We have developed a historical model that predicts the time-until-arrival based solely on historical data, as described in Appendix A.2. We then developed an adaptive recursive least squares prediction model in Appendix A.3 that adaptively adjusts its filter gain based on real-time AVL data that are "continuously" made available. The predictions of the bus traveling a distance *d* from the beginning of the section with a length $D = \hat{D}$ from the historical model and the adaptive model are molded together in a weighted combination to form a real-time TUA prediction algorithm.

The historical model developed in Appendix A.2 predicts the TUA in a single calculation and the historical TUA given by $\hat{t}_{gH}(d)$ is computed using a *constant* speed. On the hand, the adaptive model developed in Appendix A.3 adaptively adjusts some parameters using the available past and present real-time data, and the adaptive TUA given by $\hat{t}_{gA}(d)$ is computed using a speed that *adapts* to the real-time traffic conditions. So we can think of $\hat{t}_{gA}$ as an updated measurement of $\hat{t}_{gH}$ as the bus travels towards the end node of the drive section. Given the measurement $\hat{t}_{gA}$, what is the "best a posteriori estimate" of the historical TUA? This "best estimate" will be the predicted TUA. To generalize the TUA prediction problem, we can formulate it as a parameter estimation problem. Consider a single measurement t_m of the unknown parameter t_{gH} in the presence of an additive Gaussian measurement noise w_{tg}, where $w_{tg} \sim N(0, \sigma_{tgA}^2)$ and σ_{tgA}^2 is taken to be exactly the right-hand side of (a-21). That is:

$$t_m = t_{gH} + w_{tg} \qquad (2.1)$$

Note that the process w_{tg} is different from w_{tgA} for the adaptive model in (a-21). That process has a variance that is only approximately given by the right-hand side of (a-21). The problem is to find the "best a posteriori estimate" of t_{gH} when $t_m = \hat{t}_{gA}$. The *prior* information about t_{gH} is that it is Gaussian with mean $\hat{t}_{gH}$ and variance σ_{tgH}^2 given in (a-26) and (a-25), respectively. We assume that t_m and t_{gH} are independent. The *posteriori* probability density function (pdf) given the measurement, t_m, is $p(t_{gH} \mid t_m)$. The *maximum a posteriori* (MAP) *estimator* (a-23) is a realization of t_{gH} that maximizes the posteriori pdf. It is defined as:

$$\text{MAP estimator} := \mathit{Arg\ Max}\ p(t_{gH} \mid t_m) \qquad (2.2)$$

This estimator, which depends on the measurements t_m, and through them on the realization of t_{gH}, is a random variable. It can be shown that the posteriori pdf of t_{gH} is also Gaussian (a-23) with mean and variance given by:

$$\mu(t_m) = \frac{\sigma_{tgA}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2}\hat{t}_{gH} + \frac{\sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2}t_m \qquad (2.3a)$$

$$\sigma_{tg}^2 = \frac{\sigma_{tgA}^2 \sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \tag{2.3b}$$

The mean $\mu(t_m)$ of the posteriori pdf is therefore the MAP estimator since a Gaussian distribution has the maximum at its mean. The MAP estimator molds the historical TUA and the measurement t_m. If the measurement is $t_m = \hat{t}_{gA}$, we get:

$$\hat{t}_g = \mu(\hat{t}_{gA}) = \frac{\sigma_{tgA}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2}\hat{t}_{gH} + \frac{\sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2}\hat{t}_{gA} \tag{2.4}$$

This is our TUA prediction algorithm. We note that the MAP estimator is a *weighted combination* of the historical TUA and the adaptive TUA, and the weightings of the prior mean and the measurement are inversely proportional to their variances.

2.2.2 Time to Initiate Priority

The bus arrival time predictor can provide accurate bus arrival information far away from intersections. However, since buses do not have the right of way like emergency vehicles to overtake other vehicles in the traffic stream, they have to join a queue, wait in the queue, and then discharge with the queue. Therefore, the actual time for a bus to enter the intersection is equal to the summation of the predicted bus arrival time and the queue discharging time. In fact, delay to buses at intersections can be decomposed into two separate components, the delay incurred by waiting in the queue, and the delay incurred by signal control measures (Lin). In order to minimize bus intersection delays, TSP operation is supposed to discharge queues right before the bus arrival.

A model is developed to estimate queue discharging time based on the assumption of “point queue” or “vertical queue” that all queuing vehicles stop behind the stop-bar without a physical dimension. The number of queuing vehicles or length of queue can be estimated based on the real-time traffic counts from loop detectors. For a distributed closed-loop traffic control system, there are two typical loop detector layouts: one is that both advance and presence loop detectors are placed on the main street and the other is that only advance loops are available. There are two queue estimation models for the two general detector assignments.

- Model 1- Queue estimation model for detector assignment 1 (as shown in Figure 2.2)
  Since both cumulative arrival and departure counts can be readily obtained, the number of queuing vehicles is simply equal to the difference of these two counts (with the assumption that queue does not back up to the advance loops).
- Model 2- Queue estimation model for detector assignment 2 (as shown in Figure 2.3)

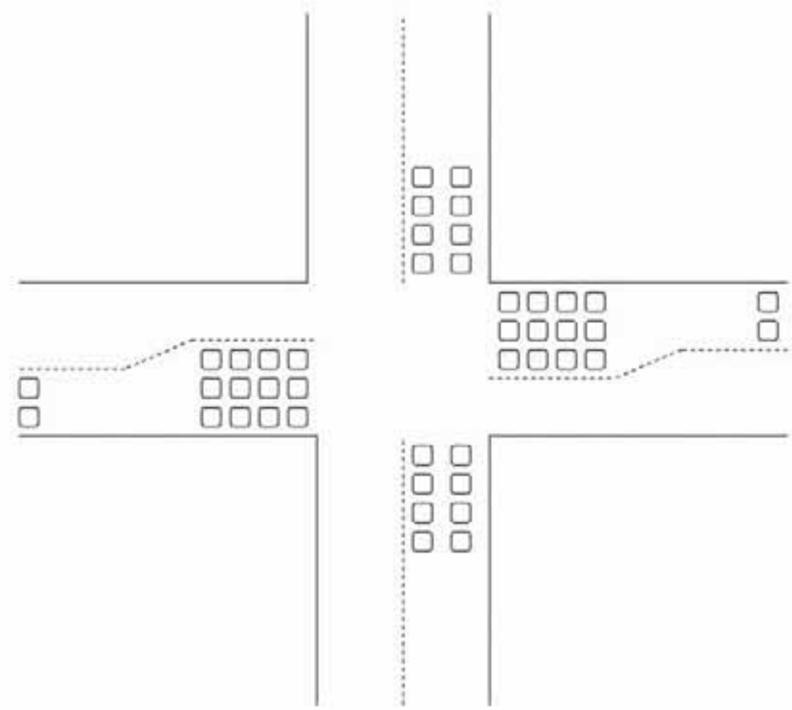


FIGURE 2.2 Detector Assignment 1 for closed-loop signal control system

However, for the latter case, the true departure count is unknown. Three assumptions are made as follows with which the departure count is estimated: 1) all intersections are under-saturated and thus no residual queues would be left until the next cycle; 2) queue would not back up over the advance loops; 3) vehicles in the "point queue" are discharged by the saturation flow. Consequently, the departure count can be expressed as follows:

$$D(t_1) = \begin{cases} \min(A(t_1), D(t_0) + s \times (t_1 - t_0)), & \text{when green is on} \\ D(t_0), & \text{otherwise} \end{cases} \quad (2.5)$$

where $D(t_1)$ is the cumulative departure count at time t_1; $A(t_1)$ is the cumulative arrival count at time t_1; s is the saturation flow and t_0 is start of green of the current cycle. Straightforwardly, the difference of the counted arrival and the estimated departure counts gives the queue length, and the corresponding queue discharging time equal to the number of queuing vehicles divided by the saturation flow.

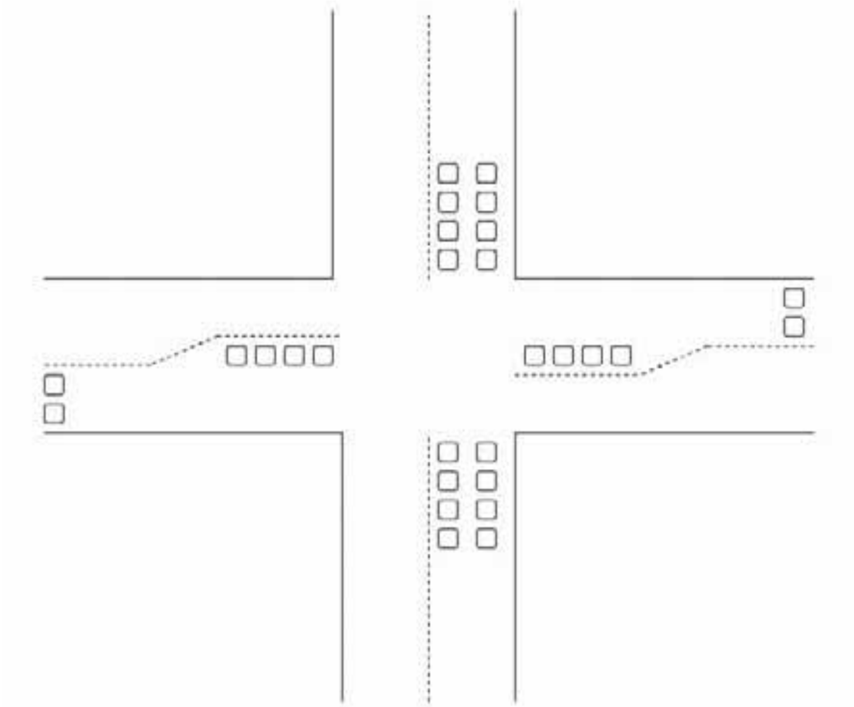


FIGURE 2.3 Detector Assignment 2 for closed-loop signal control system

In order to test the validity of estimation model 2, we performed a comparison between the two estimation models based on real signal and detector data. Figures 2.4 and 2.5 show the estimations by the two models based on one-hour field data from an afternoon. According to the estimations for two directions, the proposed estimation model 2, based on signal timing data and imperfect detector data, is also valid. If we regard the estimations by model 1 as true values, Figure 2.6 shows the estimation error for southbound and northbound queues. The mean estimation error for model 2 is 2 vehicles for both northbound and southbound queues. The standard error of the estimation is also 2 vehicles for two directions. For our study site, there are three lanes on main street, thus the mean error for discharging time is less than 1 second which is accurate enough compared with the AVL system error for bus travel time.

Accurately knowing bus arrival time from a long distance away, it is desirable to initiate TSP operation and adjust signal timing as early as possible so that the negative impacts to minor-phase traffic can be distributed and minimized across multiple cycles. However, in order to maintain the coordination in the actuated signal control, in the current study 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.

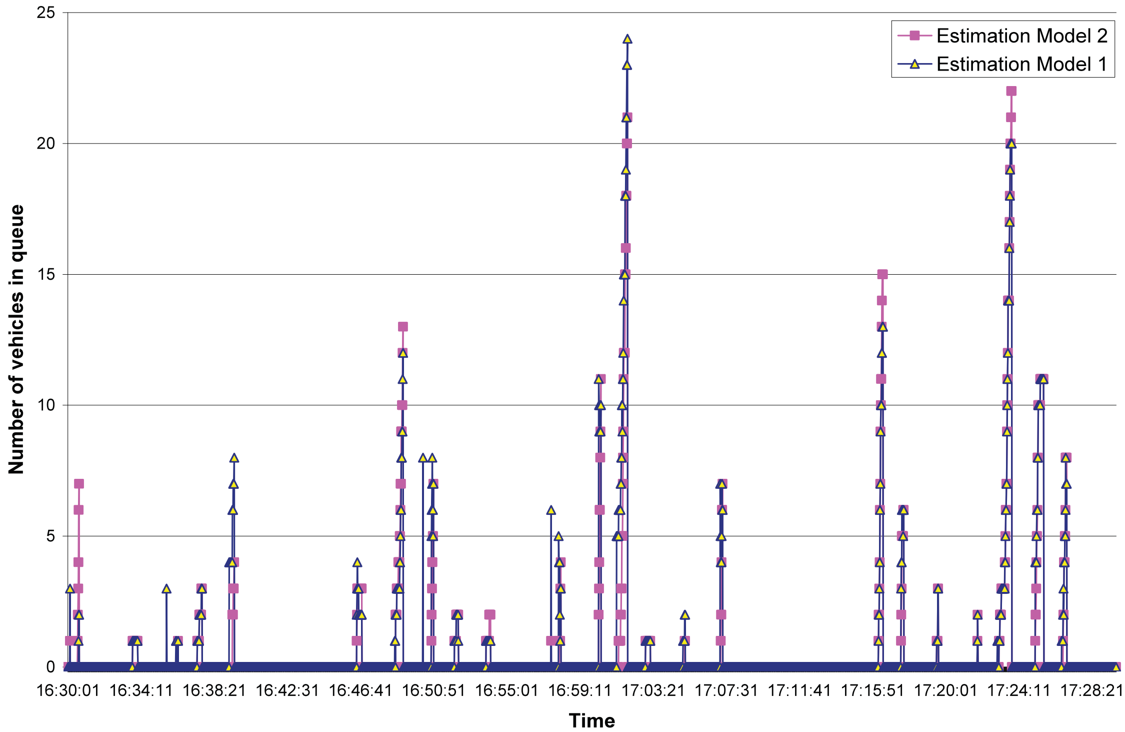


FIGURE 2.4 Comparison of queue estimation models for northbound queues

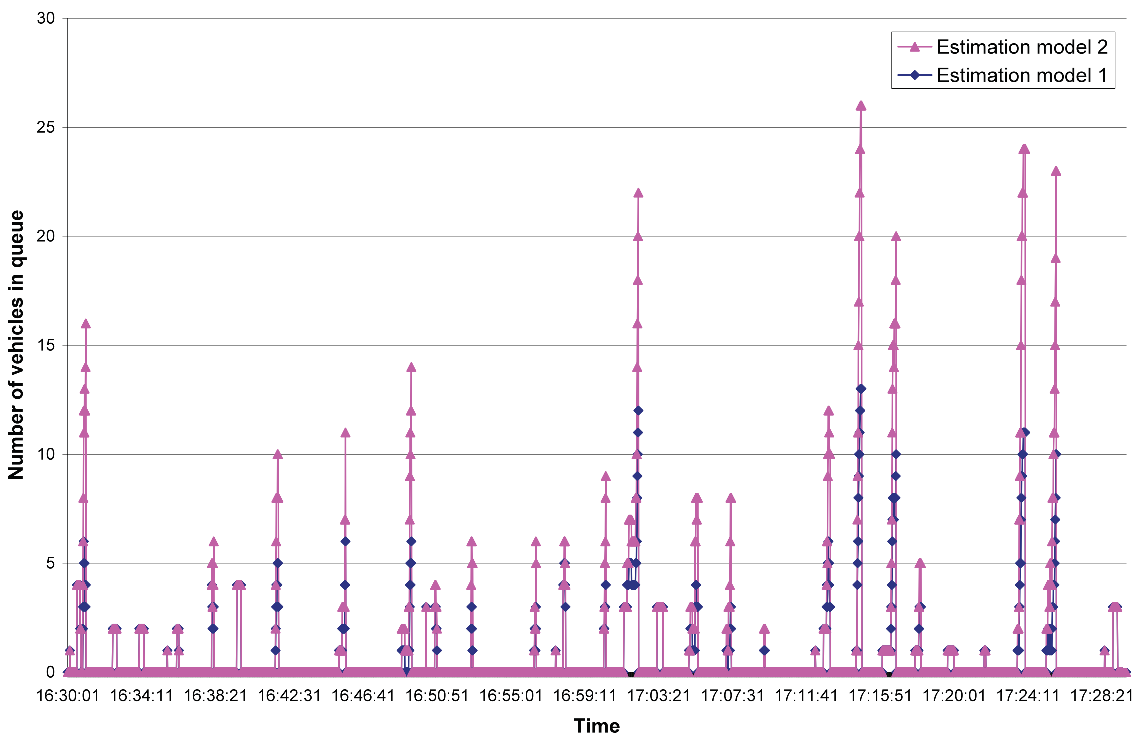


FIGURE 2.5 Comparison of queue estimation models for southbound queues

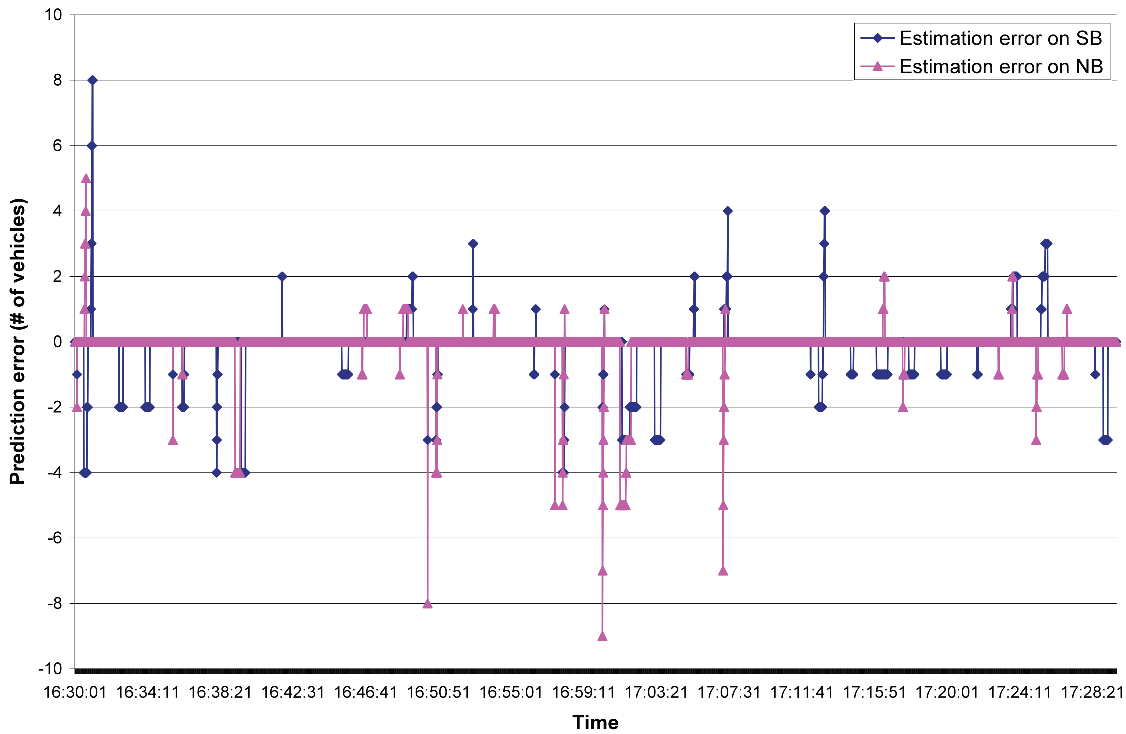


FIGURE 2.6 Estimation errors for southbound and northbound queues

To make an acceptable estimation, detector system error is another important factor. Figure 2.7 and Figure 2.8 plot northbound and southbound departure-arrival curves for 15 minute periods. After field calibrations, the cumulative system errors between departure and arrival are small especially for northbound queues. The errors for northbound during signal patterns 1 through 3, are all less than 0.1%. For southbound during the three signal patterns, the system errors are approximately 6%. As discussed above, all intersections are assumed to be under-saturated, so resetting queue lengths to 0 when the indicator changes to red could avoid cumulative system errors.

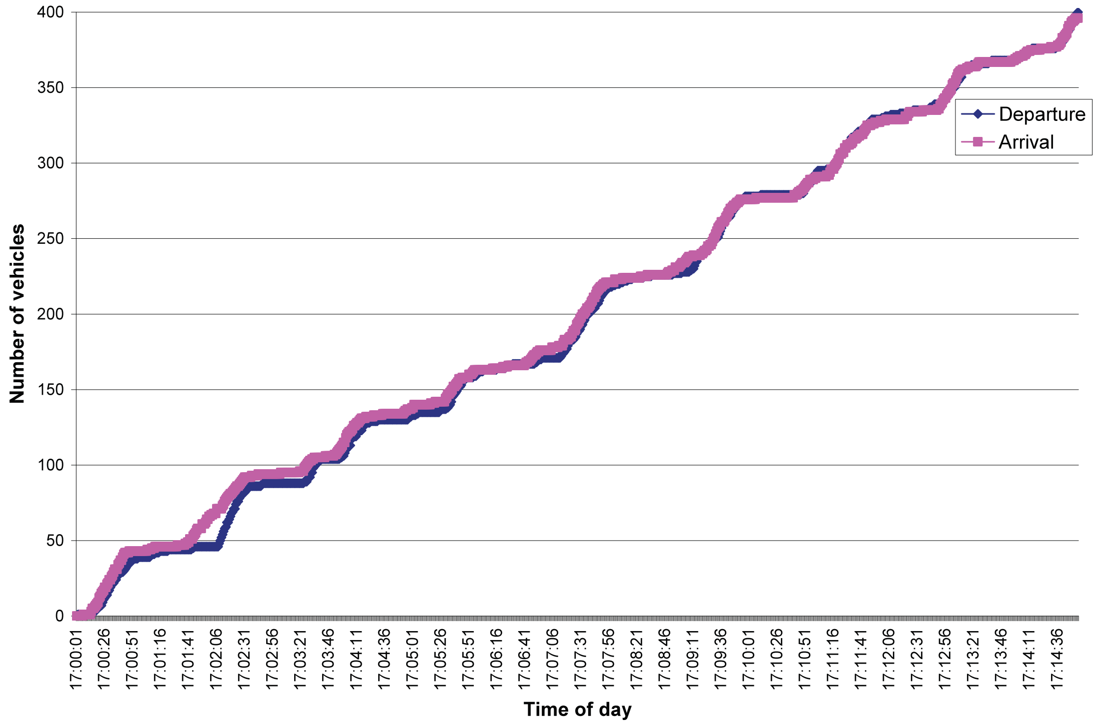


FIGURE 2.7 Northbound Departure-Arrival curves

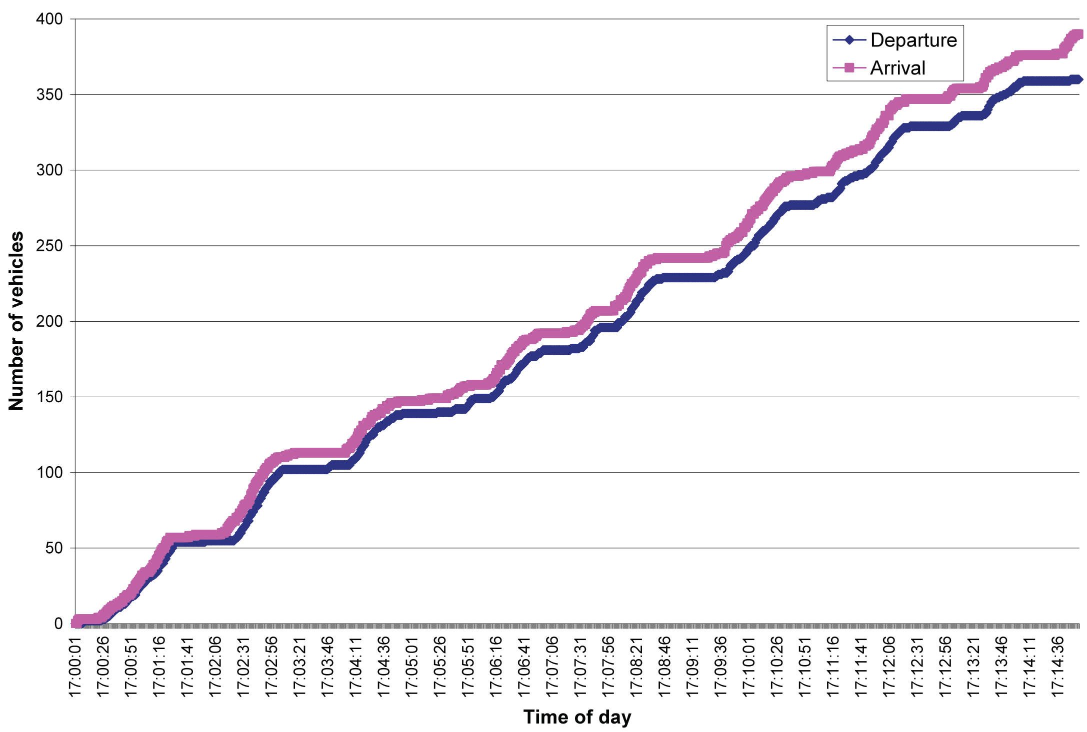


FIGURE 2.8 Southbound Departure-Arrival curves

2.2.3 Type of TSP Treatments

In the developed adaptive TSP system, there are two candidate priority treatments: 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. If the predicted time for bus to enter the intersection is located 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, which is determined with the following considerations:

- Because the cycle length is fixed for each control pattern in the actuated system, the longer the time window is, the less time the rest phases might share. At the bottom line, the time window cannot be so long that the residual time in current cycle is not enough for all pedestrian calls;
- Once one phase is extended, additional vehicles may arrive on other approaches and will cause these phases to be longer, which, in turn, will increase the next green time of the original direction, etc. Increased delays may persist for many subsequent cycles (Newell).
- The 170E signal controller, upon which the adaptive TSP system is built, achieves green extension by adjusting the local cycle clock. After completing green extension, the controller needs four times longer time for clock transition or recovery. All phases during clock transition would be impacted. Therefore, to minimize the above impacts, the time window for green extension cannot be long. In this study it is preset as 10 seconds and the resulting maximum signal recovery time will be less than 40 seconds, which is about a half to one third of a normal signal cycle length.

2.2.3.1 Level of Granted Priority

The TSP algorithm also needs to determine how much priority is appropriate for an approaching bus. It is rather simple for the case of green extension where 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 passed the intersection.

To provide early green treatment, the TSP algorithm will re-allocate the available green time among the remaining phases (at the current time instant) before the bus phase. The available green time is defined as:

$$t_{avail} = \max(0, t_{arrival} - t_{discharge} - t_{guarantee}) \qquad (2\text{-}2)$$

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 pushed. Note that this 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. Otherwise, a method is needed to allocate the available green time among these phases. In fact, the allocation is very often called for in the actual operation. With predicted traffic arrival patterns as input, an optimization model could be developed to allocate the available time so that the negative impacts can be minimized. This model is being developed in the second phase of the project. In this paper, we report another simple alternative, a statistical method to achieve the same purpose.

In viewing that the overall traffic demand pattern is stable at a specific location during a certain period of time, we assume that the phase durations under actuated control system follow their own stochastic distributions respectively during that period of time, which can be statically estimated using the historical signal status data. Note that the ranges of the variables for these distributions should be between the corresponding minimum greens and maximum greens. Figure 2.9 shows sample distributions for phases 1, 5 and 8 for a real intersection. With these phase duration distributions, we determine a probability level such that:

$$t_{avail} = \sum_{i=1}^{n} cdf_i^{-1}(L) \qquad (2\text{-}3)$$

where *n* is the number of the minor phases before the bus phase; cdf_i is the cumulative density function for phase *i*, and *L* is a probability level. Therefore $cdf_i^{-1}(L)$ yields the time that corresponds to the probability level *L,* and it is the time that we allocate to phase *i* in addition to the guaranteed green time. In this sense, *L* can be viewed as the confidence level that we have for each phase to fully accommodate the approaching traffic with the allocated green time plus minimum green time. As illustrated in Figure 2.9, with the uniform confidence level of 0.6, the new green times, denoted as NG1, NG5, and NG8 can be assigned to phases 1, 5 and 8 respectively.

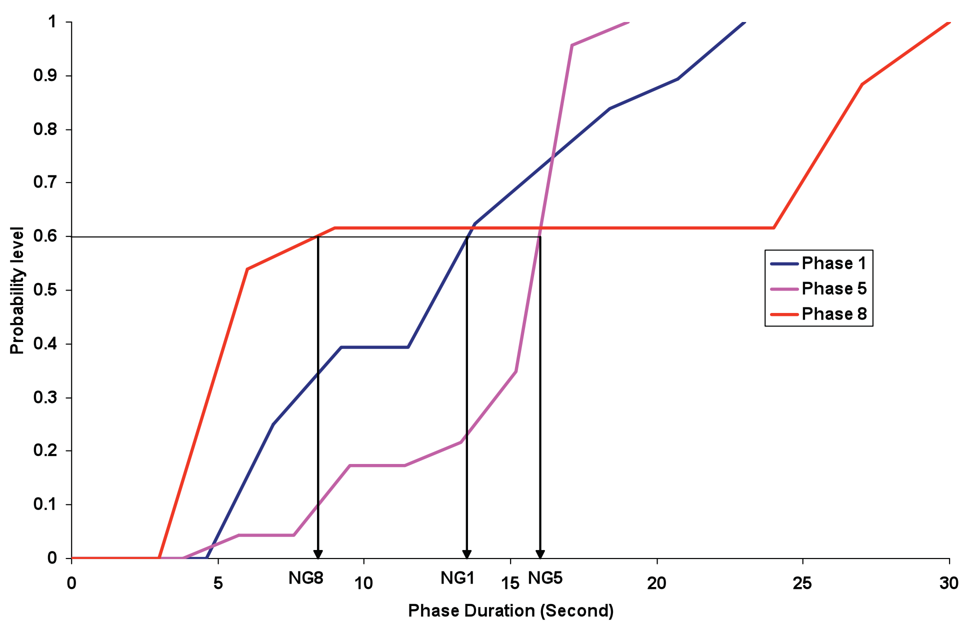


FIGURE 2.9 Green Time Allocation for Early Green Treatment

There are two reasons to support using Equation 2-3. Firstly, it is a fair allocation and evenly distributes the impacts among multiple phases. Second, the way of allocation can ensure that minimum percentages of vehicles are blocked by 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 3 may implicitly lead to much less impact onthe minor-phase traffic.

In summary, Figure 2.10 presents a simplified flow chart for the proposed PRG algorithm. Be aware that the algorithm will initiate a priority request for one approaching bus. If there are more than one priority request from different buses, the PRS will manage and prioritize these requests based on current signal status and buses movement status.

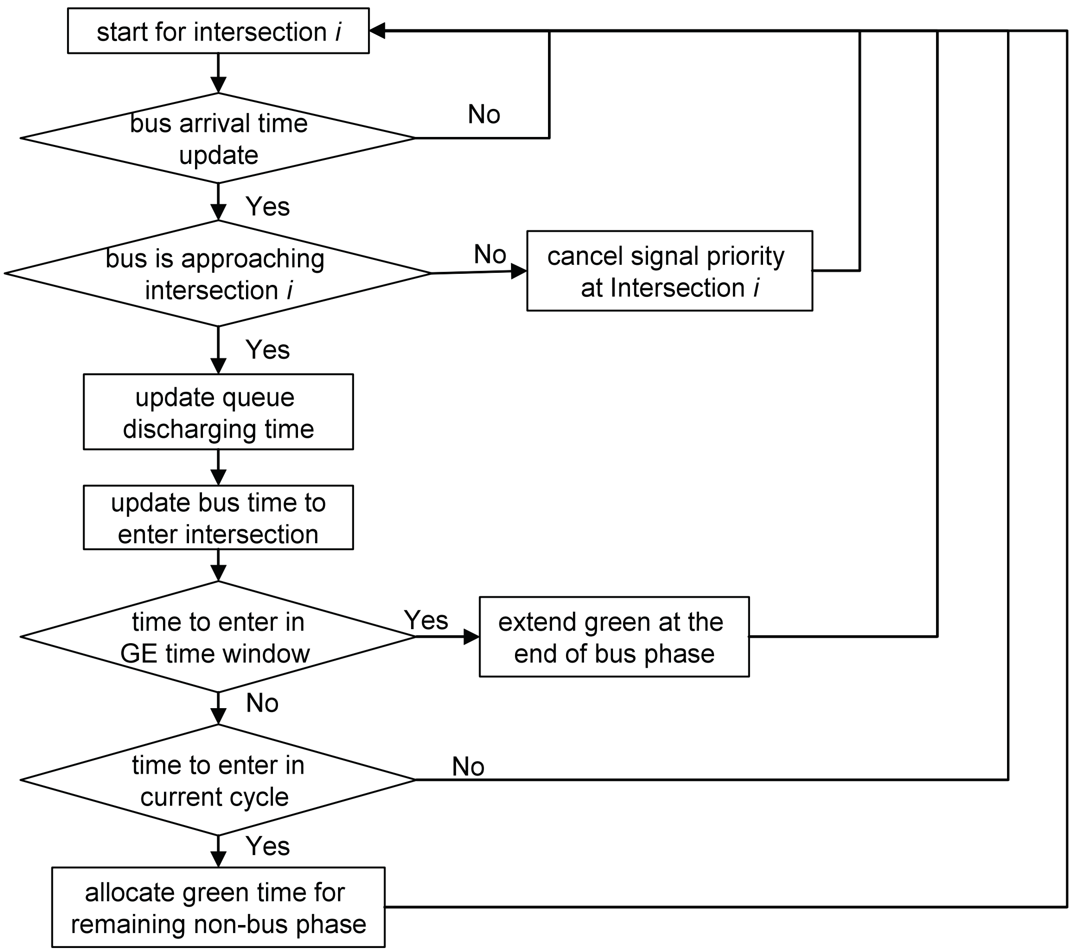


FIGURE 2.10 Flow Chart of PRG Algorithm

2.3 ATSP Priority Request Server (PRS)

The ATSP PRS is developed to manipulate a traffic signal to grant priority to buses to minimize bus intersection delays, and at the same time, limit negative impacts on the minor-phase traffic and ensure pedestrian safety.

In order to keep balance among the three objectives, the ATSP PRS should provide bus priority exactly when needed by buses up to a certain time savings limit. If the priority is overwhelming, TSP becomes preemption and the incurred impacts on the rest of traffic would be too costly. On the other hand, if the priority for buses is too little to see the benefits, TSP is pointless. Therefore, decisions of the ATSP PRS include time to initiate signal priority, type of priority treatments, and level of priority to grant. These decisions will be made online and updated in a real-time manner.

The PRS processes and prioritizes all priority requests from different PRGs and generates service requests to be granted by the local signal controller. It maintains two tables, the Priority Request Table and the Priority Service Table. In the two tables, Request Type corresponds to types of TSP treatments; Request Force Off Point corresponds to the level of priority; Request Time Of Estimated Departure is the estimated time when the bus would pass the stop-bar. Upon receipt of a “priority request” message, the PRS checks the contents of the message to ensure that all fields are present and contain valid values. The wrong message will be ignored and the information of a validated message is stored in the appropriate entry in the Priority Request Table.

The PRS keeps monitoring current signal status which is updated every one second by each local signal controller. Upon receipt of a signal status update from a signal controller, the PRS scans the Priority Request Table entries for all serviceable requests that called for this particular signal and current phase. Based on pre-determined prioritization criteria, the one with the highest priority of service is selected as the potential service request and becomes a validated service request if the Time to Serve (TTS) conditions are met. If the signal is currently serving a transit priority, a further decision needs to be made either to overwrite the on-going service or update it with the new information. Upon the selection of the service request, the PRS stores the service request information in the appropriate entry in the Priority Service Table and sends out a “service request” message to the corresponding signal controller. Figure 2.11 shows a simplified flow chart for the proposed PRS algorithm.

2.3.1 Prioritization Criteria

The simplest prioritization criterion would be First Come First Serve (FCFS). The limitation of FCFS is obvious as it did not consider the overall transit delay at the intersection. For example, with phase 2 & 5 currently on, bus no. 1 on phase 6 approach called for an Early Green before bus no. 2 on phase 2 approach called a Green Extension. Under FCFS, the signal controller will first grant the Early Green treatments for bus no. 1 followed by possible Early Green treatments for bus no. 2. Although two buses could pass through the intersection earlier than without priority treatments, the undesirable intersection delay for bus no. 2 leads to less efficient TSP operations. Trying to minimize the overall transit intersection delay, the following criteria were determined for priority prioritization:

- Green Extension has higher priority than Early Green.
- Among the requests which have the same Priority Request Type, the one that has larger Priority Request Time Of Estimated Departure has higher priority.

Under above prioritization criteria, an on-going Early Green treatment could be overwritten by a Green Extension treatment if conditions are met; however, the PRS will keep serving a Green Extension until it needs to be cleared.

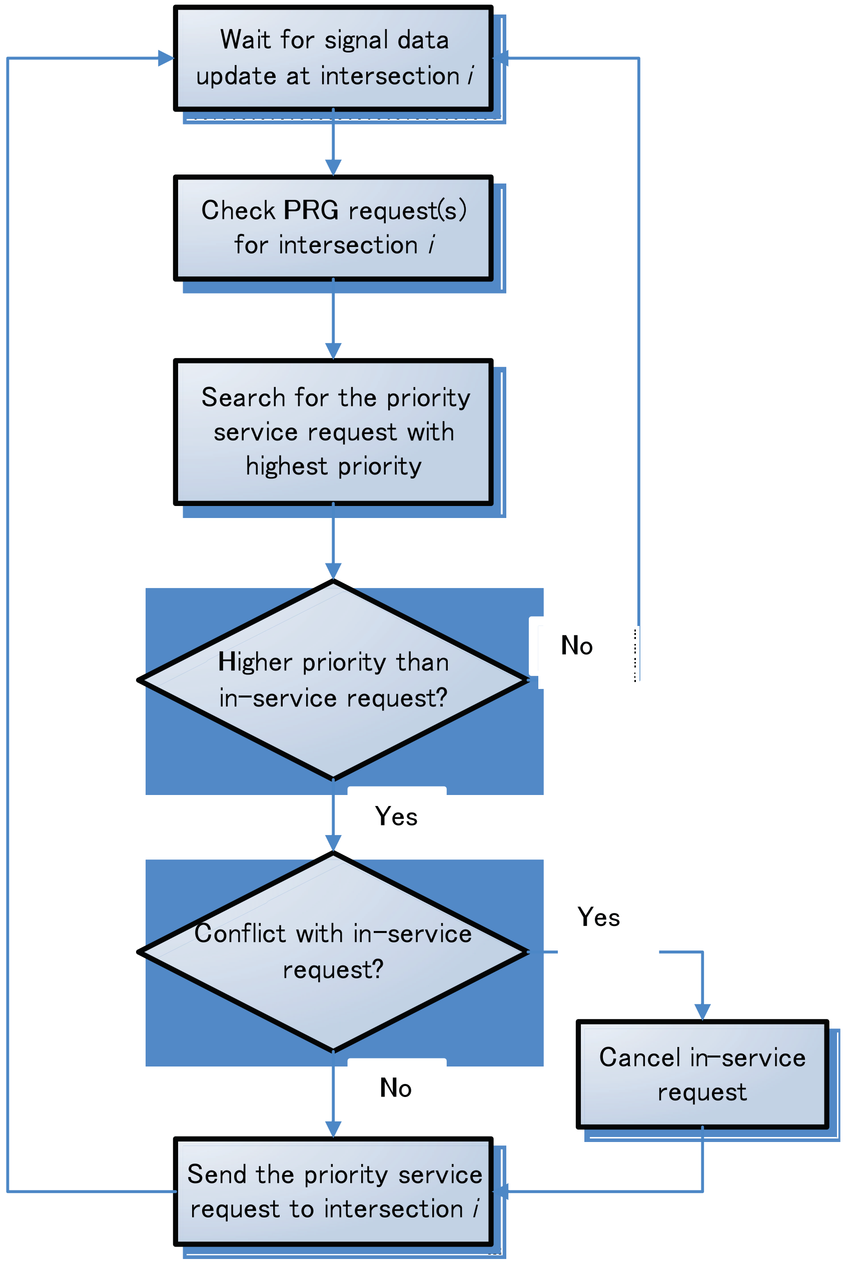


FIGURE 2.11 Flow Chart of PRS Algorithm

2.3.2 Priority Service Request

Priority requests can be granted based on agreements between the traffic and transit agencies. In principle, ATSP shall enhance intersection safety and reduce the disturbance of providing transit priority treatments to the underlying signal control strategy:

- Local controller shall not grant transit priority in the presence of high priority preemption and signal transition.
- Local controller shall not grant transit priority which would shorten the guaranteed minimum green interval (e.g., vehicle minimum green interval, pedestrian intervals), or would cause signal phase or phase interval skipping.
- Local controller shall only grant one priority treatment at a time.

The transit priority requestor is responsible to generate service requests which obey the above principles. The PRS realizes these principles by adding the following constraints on the time to select the service request:

- No service request is allowed, except the “clear service” request, when the signal is in the preemption or transition process.
- An Early Green priority request is only validated as a service request when the Request Phase is in its green interval and its guaranteed minimum green interval has timed out.

2.3.3 Update Service Request

Upon receipt of a validated PRG request, the PRS determines whether it is an initial request, an update request, or a cancel priority request. An initial request is stored in the Priority Request Table as described earlier and an update request overwrites the information that has been stored, while a cancel priority request causes the corresponding data entry removed from the Priority Request Table.

Upon the selection of a validated service request, the PRS compares it with the current serving request stored in the Priority Service Table to determine whether it conflicts with the on-going service and takes one of the following possible actions:

- If there is no validated service request available and the signal is not serving any priority, no further processing takes place.
- If there is no validated service request available while the signal is currently serving a priority, the service needs to be cleared as it was canceled by PRG. The PRS removes the data entry from the Priority Service Table and sends out a cancel service message.
- If there is a validated service request while the signal is not serving any priority, the PRS sends out an initial service request message and stores the information in the appropriate Priority Service Table entry.
- Otherwise the PRS serves the one, which has higher priority of service between the validated service request and the on-going service,.
    - If the latter has higher priority, no further processing takes place.
    - If the former has higher priority and does not conflict with the on-going service, the PRS sends out an update service request message and overwrites the new service information in the Priority Service Table entry.
    - If the former has higher priority and conflicts with the on-going service, the latter needs to be cleared to serve the new request. The PRS sends out a clear service message for the on-going service, removes the corresponding data entries from the Priority Request Table and the Priority Service Table, and sends out an initial service request message for the new service and updates the information in the Priority Service Table.

2.3.4 Clear Priority Service

The PRS keeps monitoring traffic signal status and priority service status and sends out a clear service message upon the termination of the Priority Service Phase. Both corresponding entries are removed from the Priority Request Table and the Priority Service Table. In addition, when the signal just passed the yield point, the PRS automatically clears all Priority Request Table entries, which are entered before the yield point to avoid the potential residuals for the next signal cycle.

2.4 Communications Protocols Using Existing Transit Advanced Communications System (ACS)

In the previous ATSP testing, General Packet Radio System (GPRS) modems were used to send transit AVL data on a second by second basis. The existing transit AVL/ACS system has a much lower polling rate, typically at 60-120 seconds per sample. In an urban area, a bus can travel across several intersections within one ACS sampling cycle. Therefore, it is not practical to use existing ACS' regular polling cycles to carry signal priority-related information for generating priority requests. Means for inserting additional polling at predetermined time to arrival prior to arriving at an intersection need to be explored in order to serve signal priority purpose.

2.4.1 Proposed Communications Protocols

The current transit ACS technologies use center-polling communication protocol. Under this protocol, the center performs carrier sense and collision detection to make sure that the carrier is available and issues a polling command. Each node (bus) can only communicate with the center when polled. No multiple access is allowed. ACS is designed in such a way that the constant communication needs only takes 60-70% of the channel capacity, which leaves additional capacity for meeting other application needs.

Based on the existing ACS communications protocol, the TSP request protocol is designed, which include the following steps:

1) Each bus continuously estimates Time-To-Arrival (TTA);
2) At each regular polling opportunity, predicted TTA is carried to the center;
3) The TMC receives TTA information from all buses and establishes a TSP polling schedule based on the time required when predicted TTA of each bus reaches a predetermined value τ (e.g., 40 second) and whether this bus is behind schedule. Only the late buses will be granted with TSP polling;
4) The TMC issues a 'TSP polling' for a late bus with estimated TTA=τ;
5) Upon receiving the polling, the polled bus will send an updated TTA;
6) The TMC determines if the updated TTA is smaller than τ, otherwise issues another TSP polling at the estimated TTA= τ.

The bus repeats step (5) when receiving polling. It is estimated that this repetition only occurs once. This polling scheme fully relies on the existing ACS communications protocol. Because of the randomness of the arrival time at each intersection, the predicted TTA at a distance to an intersection can have errors. The TSP polling process has accounts for the TTA prediction errors.

2.4.2 Analysis of Priority Call Frequency

To evaluate the feasibility of the applicability of the TSP polling schedule for the bus operation along El Camino Real (ECR), we processed field data and investigated statistics on schedule deviation and the number of late/early buses running within a user-defined time window.

Samtrans operates 15 bus lines along part of ECR, each of which contains 2 opposite directions, either East/West or South/North bound, including: 250 East, 250 West, 251 East, 251 West, 260 East, 260 West, 262 South, 262 North, 271 East, 271 West, 274 East, 274 West, 295 South, 295 North, 296 South, 296 North, 390 South, 390 North, 391 South, 391 North, 397 South, 397 North, KX South, KX North, MX South, MX North, PX South, PX North, RX South, and RX North.

We have obtained operations data from Samtrans covering a duration of 6 months between Jan 1st, 2006 to Mar 31st, 2006 and Aug 1st, 2006 to Oct 31st, 2006. The availability of data varies from line to line. One possible reason is that different lines operate according to different schedules. For example, if we examine the data of the first time period, i.e. Jan 1st, 2006 through Mar 31st, 2006, then we will see that the data of Line 250 include the following segments: from Jan 15th to Jan 22nd, from Jan 28th to Feb 25th. However, if we look into the data of Line 251, there are much more but smaller segments: Jan 16th through Jan 21st, Jan 28th, Jan 30th through Feb 4th, Feb 6th through Feb 11th, Feb 13th through Feb 18th, Feb 20th through Feb 25th, and Feb 27th through March 1st, since Line 251 does not runon Sunday.

To explore the capacities of channels available we developed a MATLAB program to conduct the simulation based on real operation data. The following analysis provides an example using field data collected on Jan 16th, 2006. The following assumptions are used:

a) Only late buses are served with TSP calls;
b) The route between time-points is a straight line;
c) The travel time between signalized intersections is 60 seconds;
d) The geometry of all blocks are homogeneous;
e) There is only one check-in call requested at each signalized intersection;
f) The request initiation time is 40 seconds before the arrival at intersection

Figure 2.12 shows the data analysis results on the actual arrival time deviation distribution for 15 lines for both directions. Detailed deviation distribution for each line is provided in Appendix B. In Figure 2.12, a negative number on the horizontal axis means the time in which the bus arrives earlier than the schedule while a positive one is how much the bus deviates from the schedule. Note that, prior to the arrival time deviation distribution analysis, we preprocessed the original data by removing those buses with arrival deviation greater than 10 minutes early and 30 minutes late. Furthermore, for most of the 15 lines, the arrival time deviations follow a normal-like distribution. However, the distributions of the MX, PX and RX Lines are less likely normal. One possible reason for this is that the sample sizes are not large enough for these lines.

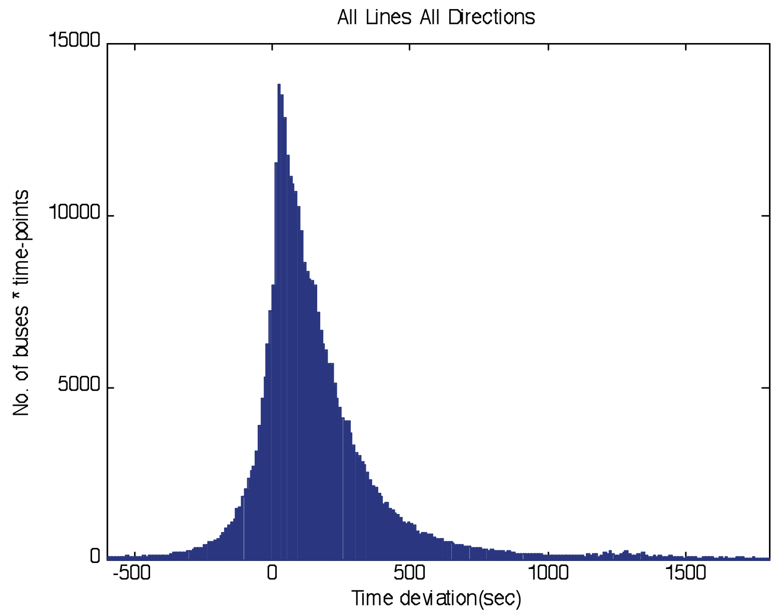


Figure 2.12 Arrival Time Deviation Distribution for 15 Lines (Both Bounds)

Based on the actual arrival time deviation distribution of each line, we have a good understanding on the actual arrival times at time-points of buses in the statistical sense and on how well does each line perform, i.e. schedule adherence. Many transit properties use 5 minutes as a threshold for a 'late bus' definition. When using this definition, from Figure 2.12, the late rate of buses at time-points is 18.48%. However, for the purpose of the signal priority we define the threshold for late buses differently. This arrival time deviation threshold prescribes the number of late buses at each time step that the communications system will need to serve. Therefore, it directly determines the capacity needs for the communications channels.

In one experiment, we set the number of late buses accounting for 20% of the totally operating buses. Then Table 2.1 lists the corresponding arrival time deviation threshold line by line.

We aggregate the data in two different time windows: one hour and 15 minutes, and that obtained for the late buses that may require TSP as illustrated in Figure 2.13 (hourly time window) and Figure 2.14 (15 minutes interval).

Through the analysis of these two sets of data, we found that the general trends of the data processed with an hourly interval and with a 15 minute interval are very similar. This indicates that when a bus is late, it is likely that the lateness will last for a while. Because of this character, as the field data only provide time adherence at time points, we can safely assume that the late bus at the time point is generally late at intersections. From Figure 2.13 and Figure 2.14, it is observed that most of the buses are in operation from 07:00 am to 06:00 pm and the late buses occur often around noon.

Using the same set of data, as we specified the time window as occurring from 10:00 am to 10:05 am, we further investigated the number of late buses and the corresponding late rate using second-by-second data. From Figure 2.15, we can see that the number of buses in operation remains almost the same within this time duration (5 minute interval), as do the number of late buses and the corresponding late rate.

Table. 2.1 Arrival Time Deviation Threshold for All Lines and Each Line

Line Name	Direction	Deviation Threshold (sec)
All Lines	All Directions	278
250 Line	East	210
	West	230
251 Line	East	306
	West	216
260 Line	East	280
	West	284
262 Line	South	274
	North	288
271 Line	East	244
	West	206
274 Line	East	200
	West	190
295 Line	South	528
	North	324
296 Line	South	326
	North	260
390 Line	South	278
	North	294
391 Line	South	286
	North	280
397 Line	South	294
	North	262
KX Line	South	354
	North	326
MX Line	South	108
	North	214
PX Line	South	638
	North	226
RX Line	South	562
	North	190

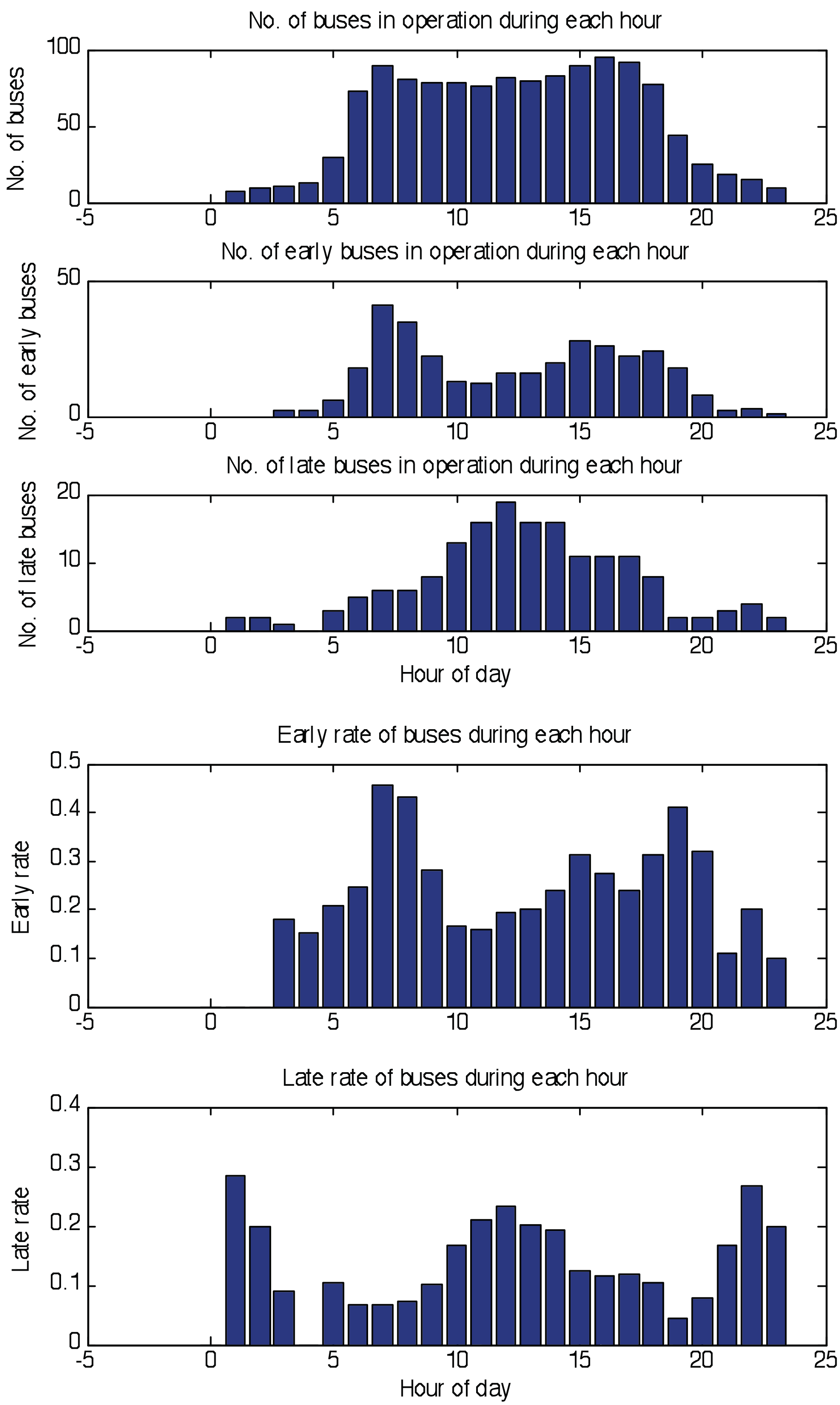


Figure 2.13 All Lines and All Directions Hourly Data

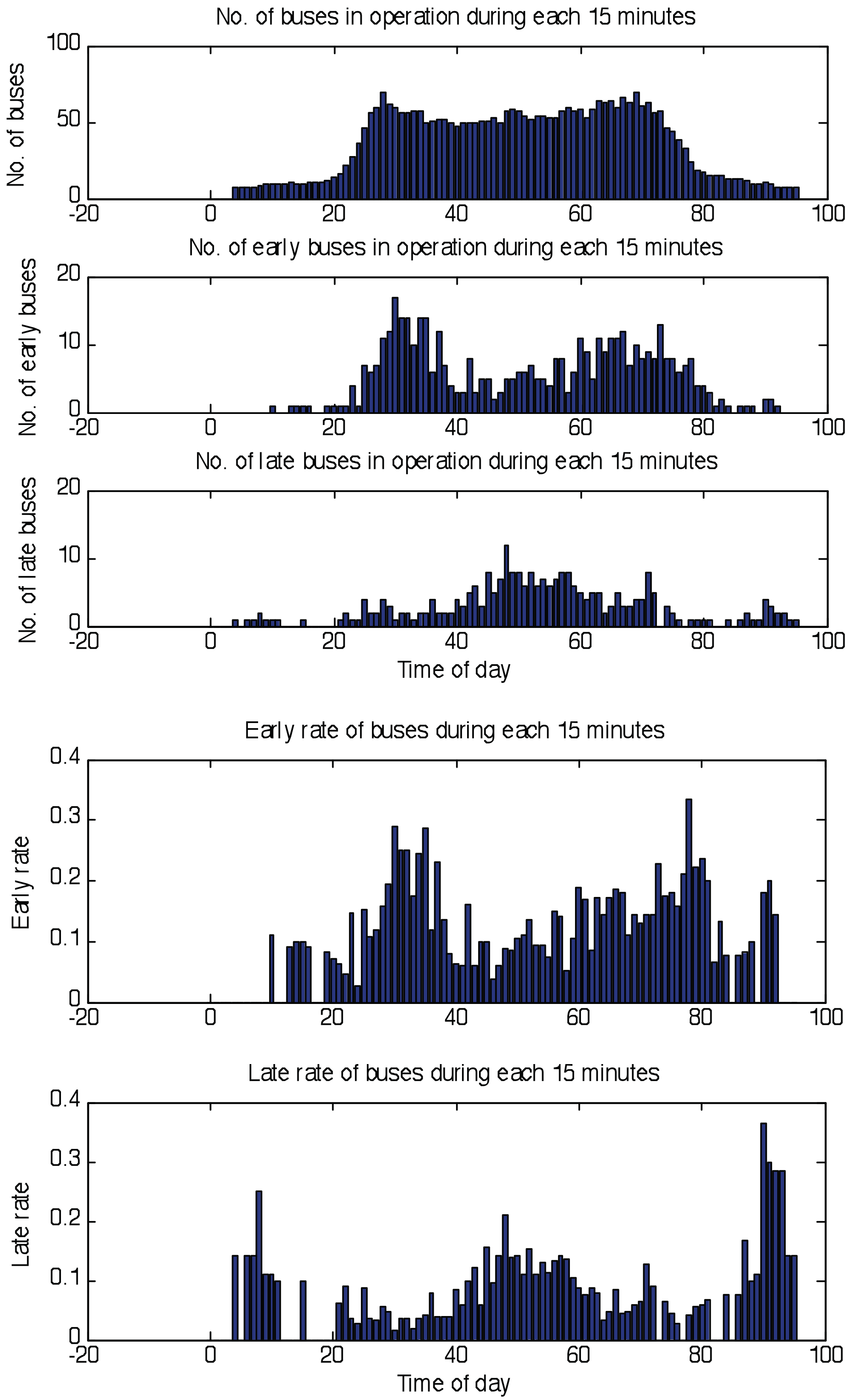


Figure 2.14 All Lines and Directions Data per 15 Minutes

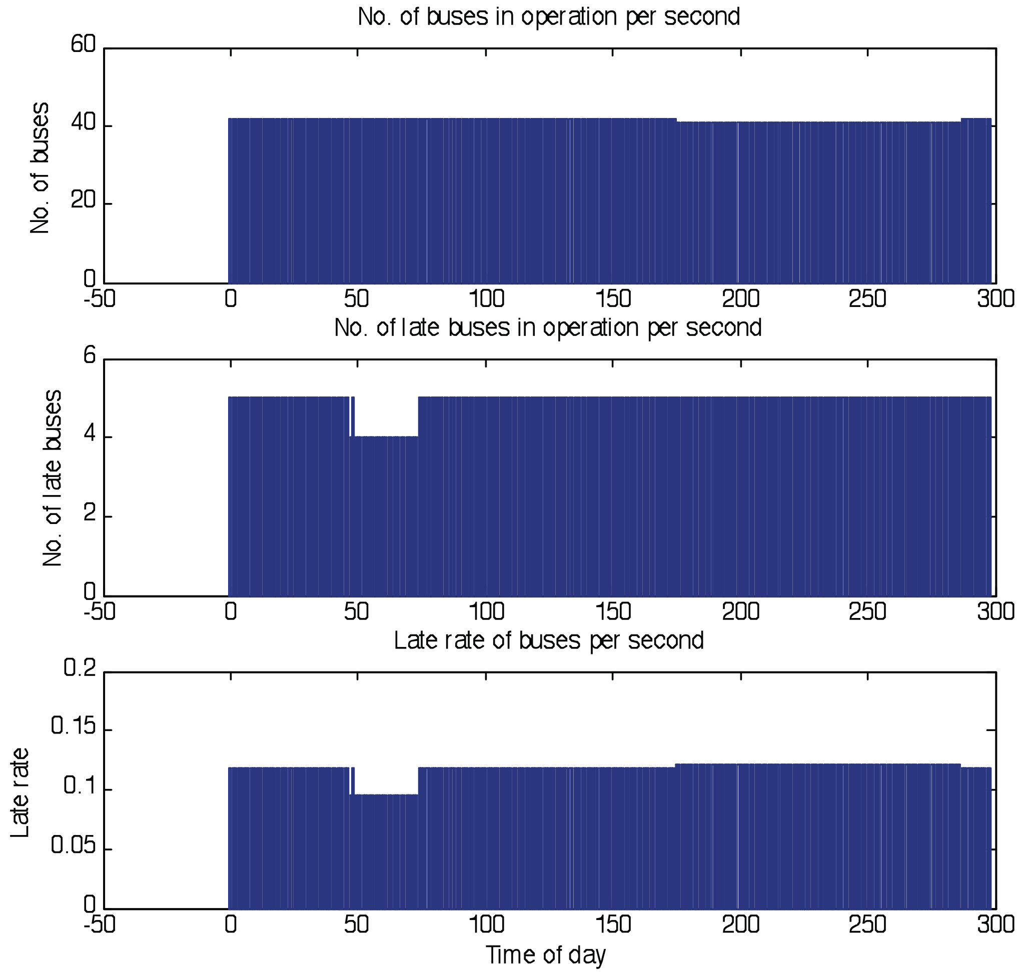


Figure 2.15 All Lines and Directions Data per Second

Assuming that all late buses will need to make one TSP call, Figure 2.16 shows the number of calls that would be generated, second by second, in a typical day of operation. Further analysis is conducted to see if a collision can occur during the tested time period. The analysis shows that over 76.51% of time, there are no TSP calls. About 14.24% of the time, only one TSP call is generated. Less than 10% of time, more than one TSP call occurs. Figure 2.17 provides further details of the TSP call histogram for each time step. If we assume that each request generated can be served within a 20 second time window or otherwise, it will be timed out, we can see that cumulatively there would be around 20 calls to be dropped. (see Figure 2.18 for the number of calls generated and the number of calls being processed and Figure 2.19 for the number of calls being dropped, both at the 20 second interval). However, if we increase this time window to 40 seconds, the dropout rate becomes zero (see Figure 2.20 for the number of calls generated and the number of calls being processed and Figure 2.21 for the number of calls being dropped, both at the 40 second time interval).

This analysis shows that it is viable to implement the proposed TSP request schedule using the existing ACS communications system.

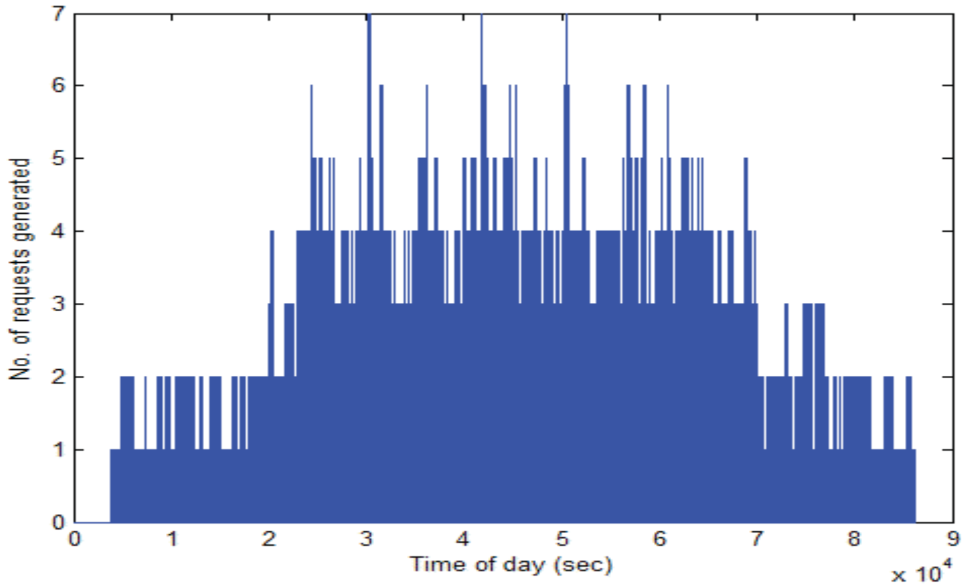


Figure 2.16 Simulation Result of TSP Calls on Jan 16th, 2006

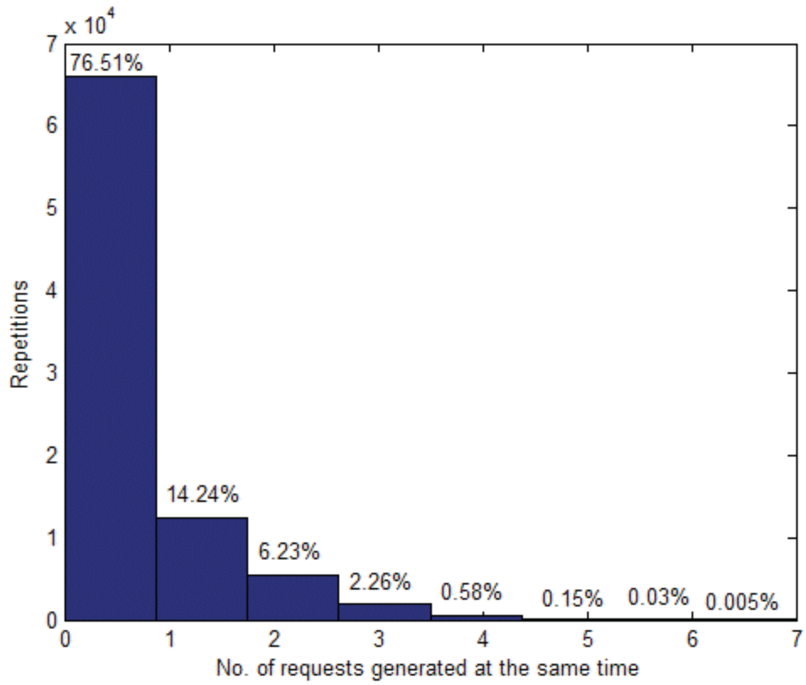


Figure 2.17 Histogram on the Frequency of Calls at Each Time Step

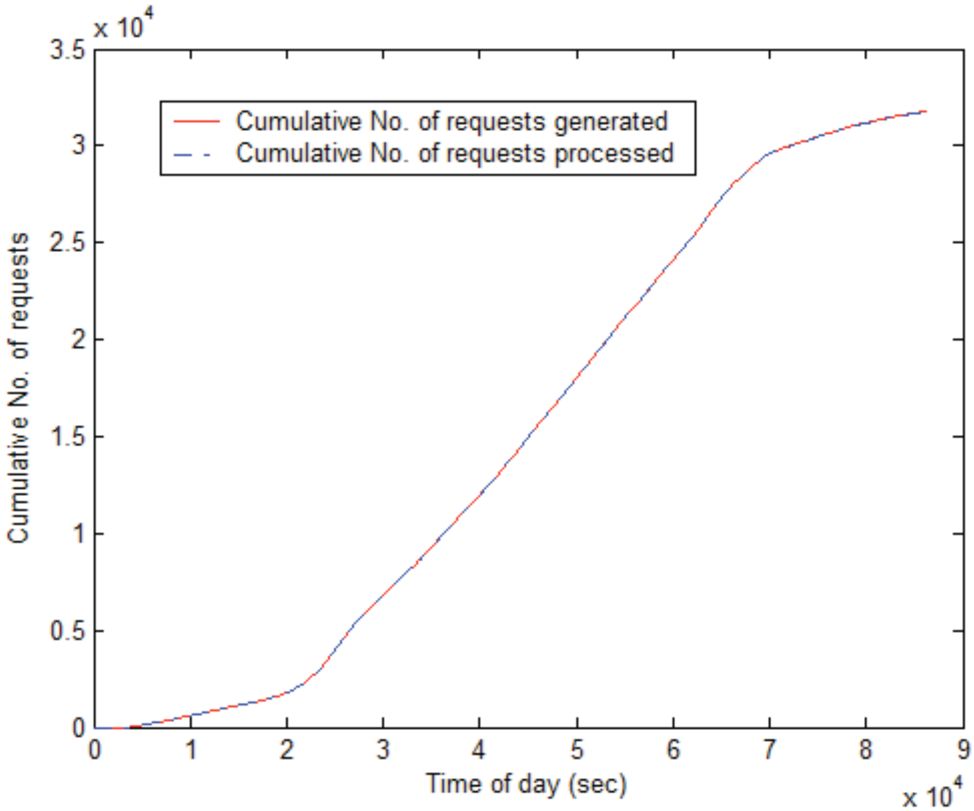


Figure 2.18 Simulation Results on Requests Generated and Processed

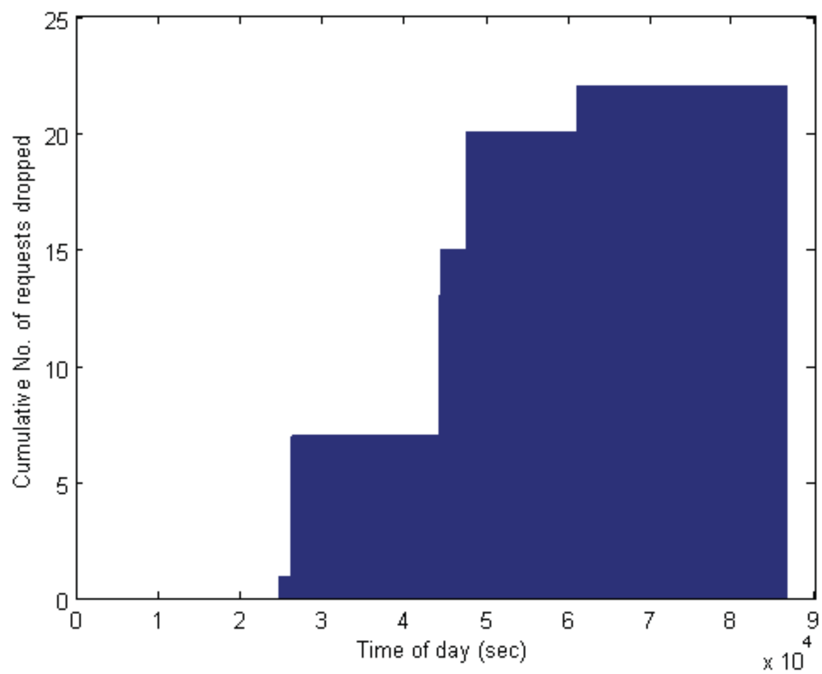


Figure 2.19 Simulation Results on Dropped Requests

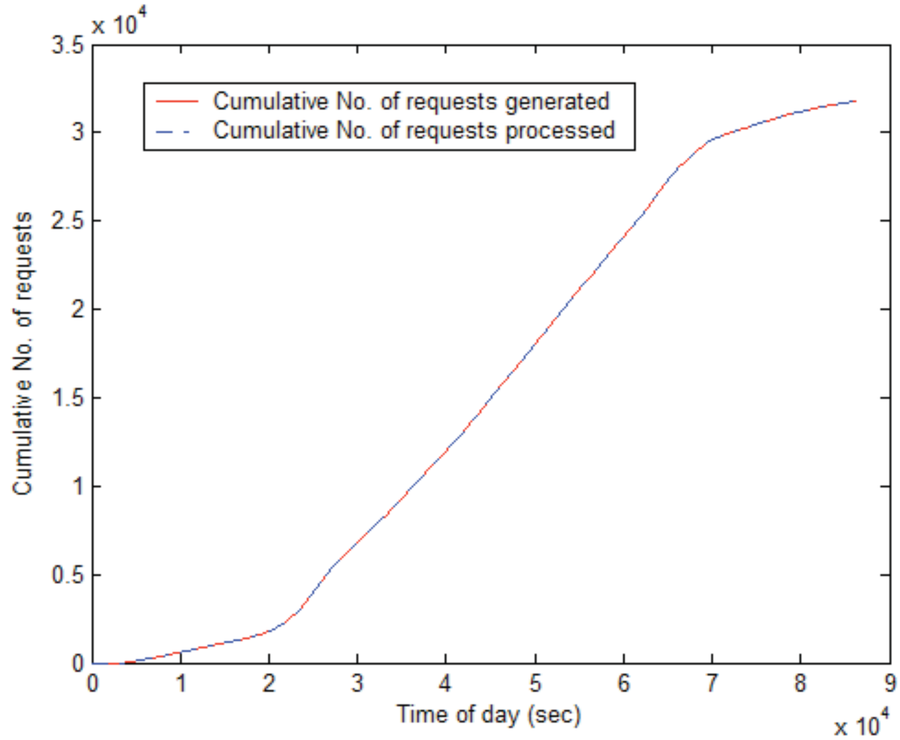


Figure 2.20 Simulation Results on Requests Generated and Processed

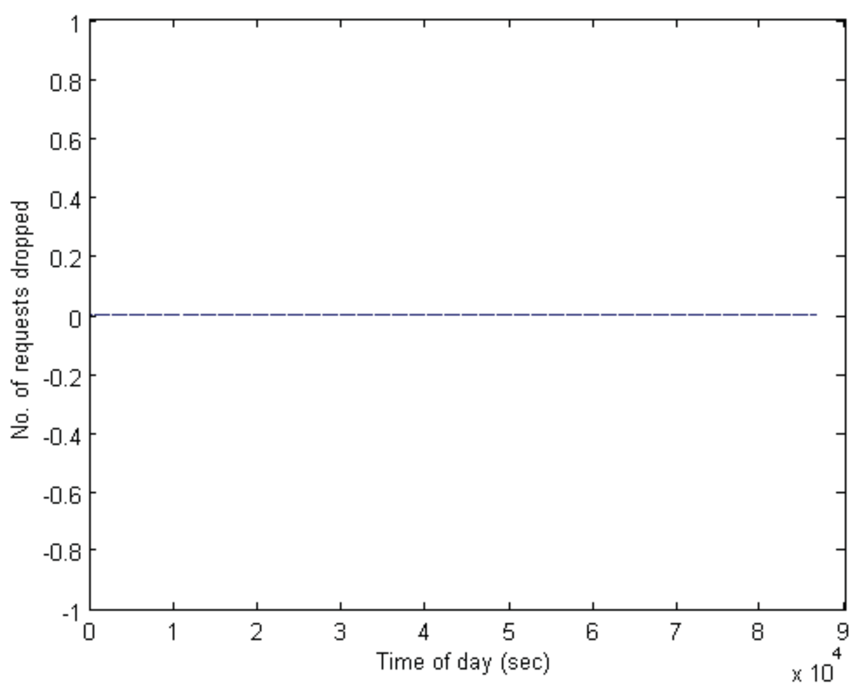


Figure 2.21 Simulation Results on Dropped Requests

3. Field Operational Test System

The field testing was carried out along a 6.6 mile segment of El Camino Real, including 35 signalized intersections, controlled by three field masters. Working with Caltrans, traffic signal controllers have been modified to work with the ATSP system. Fifteen transit buses were instrumented with testing GPS-enabled cellphones. Below provides a description of the testing corridor and the ATSP system.

3.1 The Field Test Corridor

3.1.1 Corridor Description

California State Route (SR) 82 runs from US Route 101 at Blossom Hill Road in San Jose to I-280 in San Francisco along the San Francisco Peninsula. This 52-mile-long SR forms part of El Camino Real but is almost always referred to by local residents as “El Camino”. It serves as an alternative to HW 101 and I-280 between San Jose and San Francisco and traverses through 18 cities in the counties of Santa Clara and San Mateo.

Four transit services are available along the El Camino Real corridor: Bay Area Rapid Transit (BART), SamTrans in San Mateo County, VTA in Santa Clara County, and Caltrain in both counties. These four transit operators provide a means to connect with each other through coordinated schedules, fare collection, and transfer systems, and sometimes through sharing of bus stops, transit centers, and parking facilities.

Three transit stations along the El Camino Real corridor are designed as regional transit hubs. The Millbrae BART/Caltrain inter-modal station is the primary transit hub in San Mateo County. The Palo Alto and Mountain View Caltrain stations are also transit hubs, combining rail and off-street bus loading.

The SamTrans service area covers 446 square miles and carries over 14.9 million passengers annually through San Mateo County, with routes connecting into San Francisco and Santa Clara Counties. Routes 390 and 391 are the backbone of the SamTrans fifty-four fixed routes bus network. Together they account for one quarter of the total ridership. Bus Route 390 provides intercity service from the Palo Alto Caltrain station to the Daly City BART station, while bus Route 391 provides intercity service from the Redwood City Caltrain station to San Francisco, via El Camino Real. These two bus routes run with half-hour headways during the peak hours. They connect all five of San Mateo BART stations (Daly City, Colma, South San Francisco, San Bruno and Millbrae), and connect nine Caltrain stations along El Camino Real (Millbrae, Broadway, Hillsdale, Belmont, San Carlos, Redwood City, Atherton, Menlo Park and Palo Alto).

Working with Caltrans and SamTrans, the section for field operational tests was identified as the central section of Routes 390 and 391, from 42nd Ave in the City of San Mateo to Rosedale Ave in Burlingame. This section is about 6.6 miles long and covers 35 signalized intersections. Bus Routes 390 and 391 share the 27 northbound and 28 southbound bus-stops along the section, with 3 bus-stops as time-points in both directions. Figure 3.1 below shows the overlay of bus routes 390 and 391 with the ATSP field operational tests section, and Table 3.1 lists the 35 signalized intersections along the study section.

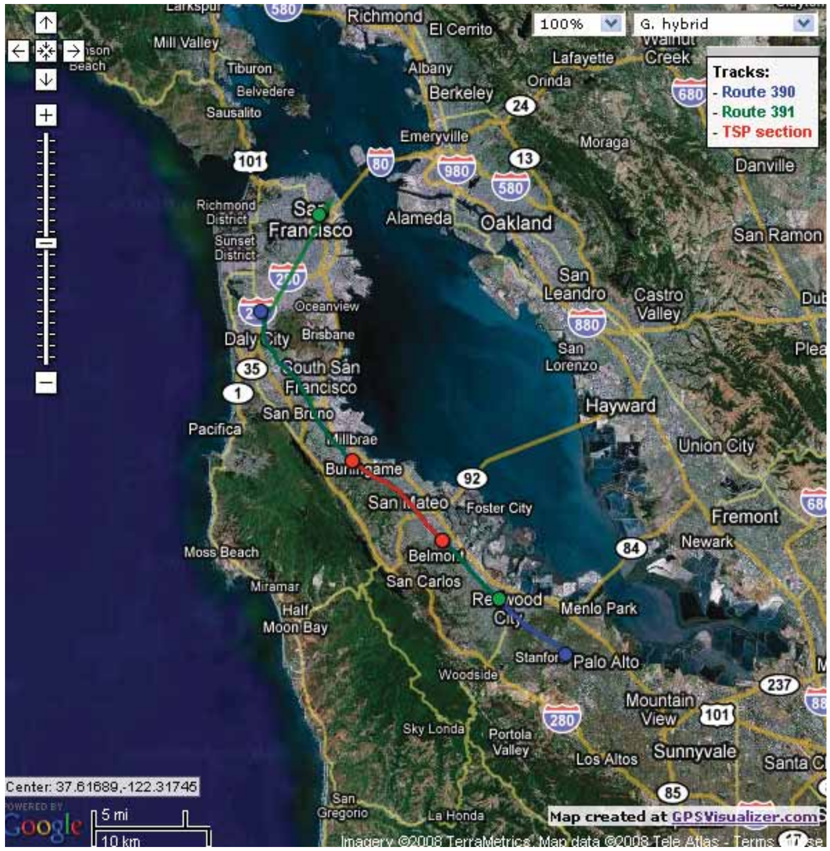


Figure 3.1 Map of SamTrans Bus Routes 390/391 and TSP Section
(Blue line – Route 390; Green line – Route 391; Red line – study section)

3.1.2 Traffic Signal Control Systems

Out of the 35 signalized intersections, the intersection at Hillsdale Blvd is controlled by the City of San Mateo and does not support TSP. Caltrans District 4 utilizes two systems to control these 35 intersections: the Burlingame system and the San Mateo system. The 9 intersections of the Burlingame system are under one field master controller, while the 26 intersections of the San Mateo system are under two field master controllers.

The existing traffic control systems are operated under the Caltrans C-8 Traffic Signal Control software together with Model 170E Control System hardware. The systems provide coordinated semi-actuated traffic control with multiple time-of-day plans. The control software was upgraded with the enhancement to support the ATSP operation. As the intersection at 20th Ave in San Mateo requires special lead-lag control, it does not support TSP.

Table 3.1 35 Signalized Intersections for Testing

Intersection	Field Master Number	City
Ray/Rosedale	FM#6	Burlingame
Oxford		
Hillside Dr		
Lincoln Ave		
Broadway		
Carmelita Ave		
Sanchez Ave		
Oak Grove		
Floribunda Ave		
Chapin Ave		
Burlingame Ave		
Howard Ave		
Bayswater Ave/Cypress Ave		
Peninsula Ave		
Bellevue Ave	FM#7	San Mateo
Poplar Ave		
Tilton Ave/El Cerrito Ave		
Baldwin Ave/Baywood Ave		
Crystal Spring Rd		
2nd Ave		
3rd Ave		
4th Ave		
5th Ave		
9th Ave	FM#8	
12th Ave/Hobart Ave		
Barneson Ave		
17th Ave/Bovet Ave		
20th Ave		
25th Ave		
27th Ave		
28th Ave		
31st Ave		
37th Ave		
31st Ave		
42nd Ave		

3.1.3 Test Vehicles

The 15 test vehicles are in-service SamTrans buses that operate on Routes 390 and 391. These buses were equipped with the cell phone-based AVL system (Figure 3.2), which record and send second-by-second GPS data back to the PATH server.

Figure 3.2 Cell phone based AVL system

3.2 Modularization of ATSP Testing System

The ATSP functional elements, including ATP, PRG and PRS, were modularized into a layered structure, with the upper layer dealing with the coordination and data exchanging and the lower layer handling individual buses and individual intersections.

ATSP is capable of being integrated with a wide range of signal control types. In order to address the deployment issue in California where the less advanced 170E controllers are widely spread, PATH and Caltrans implemented the prototype ATSP system upon the closed-loop signal control systems with 170E controllers. Figure 3.3 demonstrates the system architecture of the prototype ATSP system.

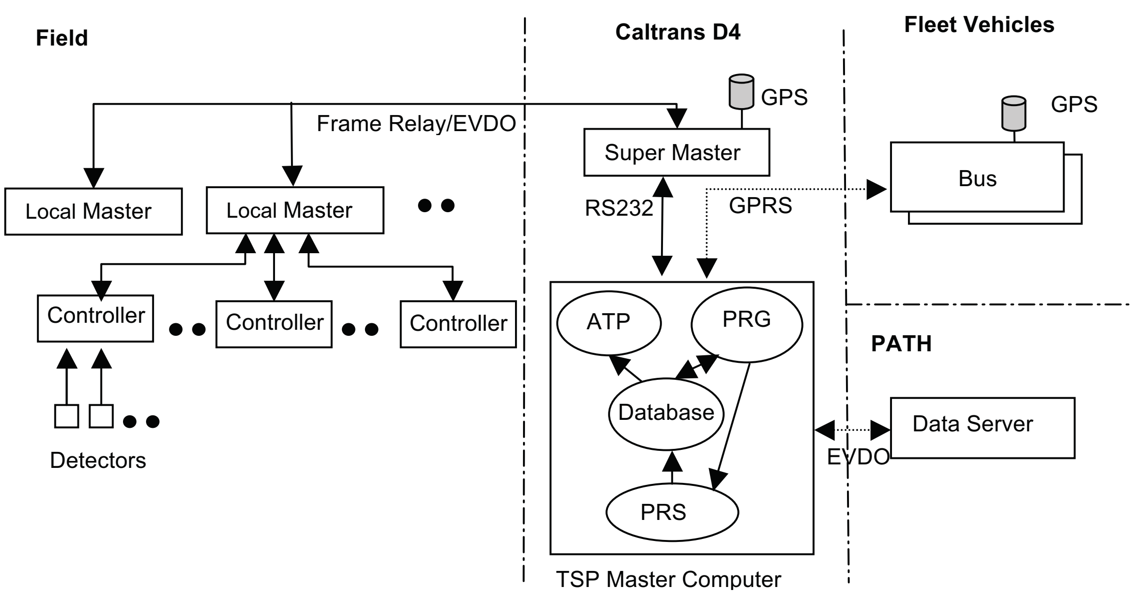


Figure 3.3 System Architecture of the Prototype ATSP System

In the prototype system, both the PRG and PRS are installed on the TSP master computer and physically located at Caltrans District 4 TMC. All the inputs from traffic signal controllers and transit vehicles are directly delivered to the central TSP master computer. For the proof-of-concept testing or a small-scale field operational testing, the system design is adequate because of the limited number of testing transit vehicles and thus limited communications requirements on the transit side.

3.3 Emulation of Proposed Communications Protocols

In order to test whether and how ACS can meet the signal priority requirements and support ATSP operation, PATH developed a testing system to emulate the AVL/ACS communications system with TSP priority request polling calls. A suite of software was designed to emulate the central communications scheme for SamTrans' AVL/ACS system.

3.3.1 Emulation Programs

Two separate channel management programs emulate the regular polling and the contention polling schemes separately. The existing AVL/ACS system maintains a regular polling list and rotates the polling from the top to the bottom of the list within two minutes. Once a second, a bus receives a polling request from the communications center and responses with bus information. The turnaround time is within a second. Typically, there are about one hundred buses on the polling list. It is why the updating frequency for each bus is about one and a half to two minutes. In consideration of a worse situation, the emulation program for regular channel maintains a two-minute updating frequency for all the TSP enabled buses, which are Routes 390 and 391 for the SamTrans case. As shown in Figure 3.4, the onboard system on each bus continuously updates the predicted bus arrival time at the next signalized intersection based on bus movement information. Such predicted information is stored in the onboard AVL/ACS system. When the emulation program is ready to initiate the regular poll for the bus, the latest predicted bus arrival time to the next intersection together with some other trajectory information is transmitted to the communications center. It is noted that the system parameters, e.g. channel bandwidth, number of buses on polling list, etc., can be specified for various system specifications.

The emulation program for the contention polling channel is similar to that for the regular polling channel. As shown in Figure 3.5, the contention polling program has two major functions. One is to wait for new requests generated by the poll request generator. On the other hand, the polling program maintains a request list and organizes the not-yet served polling requests based on time sequence or priority. Given the specified channel bandwidth, the program implements the contention poll by reading the latest bus arrival information.

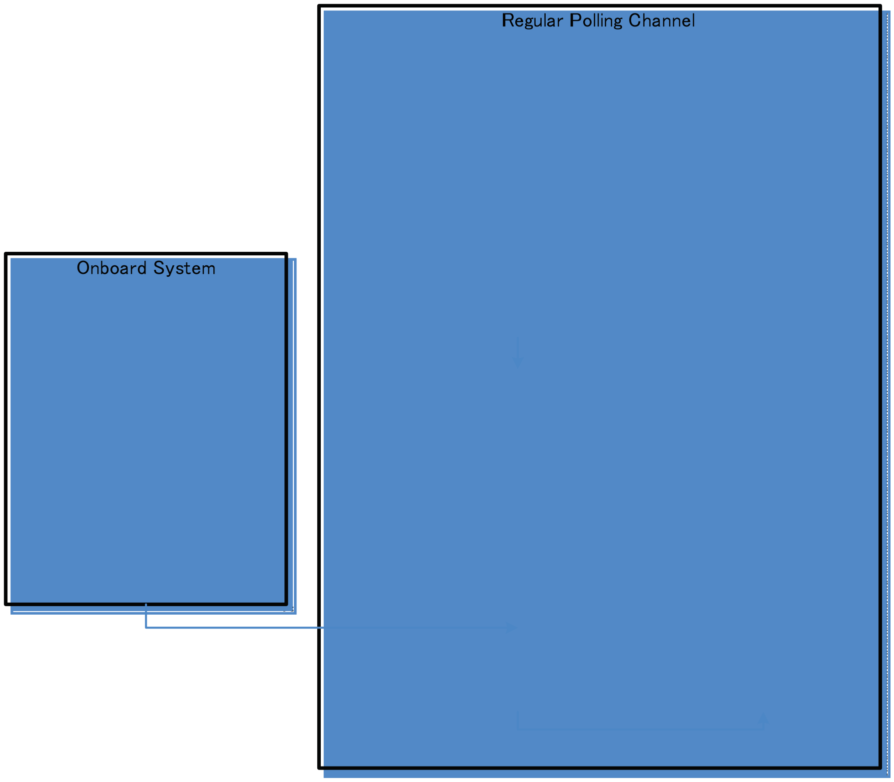


Figure 3.4 Flow chart for the emulation program for regular polling channel

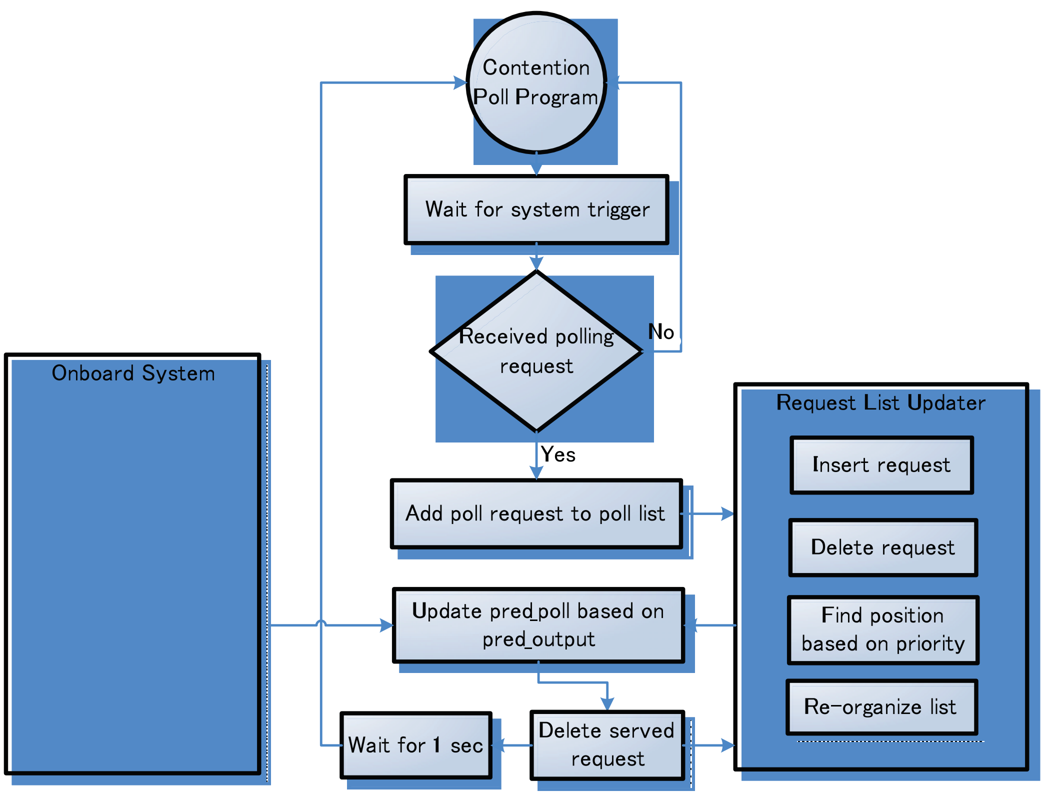


Figure 3.5 Flow chart for the emulation program for contention polling channel

In general, the capacity of the regular polling channel is not sufficient to support ATSP operation. The regular channel in the existing ACS/AVL system can only support a bus fleet with the polling rotation time up to one and a half to two minutes. In other words, each bus will receive a regular poll and get updated every one and a half to two minutes. Such a low data updating frequency cannot support ATSP operation because the average travel time between two adjacent intersections is within one minute. Between two regular polls, a bus might have already gone through two signalized intersections. Therefore, it is necessary to develop a communications protocol and interface between ATSP and the existing AVL/ACS system to fully utilize the existing communications link as shown in Figure 3.4 and Figure 3.5.

Although the 1Hz communication updating rate is used in the previous PATH ATSP system, it is not necessary to have continuous coverage on bus movement to support the PATH's ATSP system. According to the previous testing results, the ATSP system utilizes the predicted bus arrival time only when the bus is approaching and close to the intersection. With this idea, the communications protocol can be designed to request the AVL/ACS system to poll the designated bus when it approaches the intersection.

The objective of the TSP polling generation program (PGP) is to utilize the existing communications channels in the AVL/ACS system to provide the "most" needed information for the ATSP operation. According to previous field testing, the arrival time predictor (ATP) is to predict the bus arrival time at the immediate downstream intersection without consideration of traffic queuing delay at the intersection stop-bar. It is the priority request generator (PRG) which considers the traffic queuing delay based on historical or real-time traffic detection information together with the traffic signal timing and real-time traffic operation information. Generally speaking, buses' trajectories are smoother when they are cruising together with platoons before reaching the back of queues at intersections. It would be desirable for PRG to poll the approaching bus during the cruising period. Moreover, the closer the bus gets to the intersection in the cruising period the more accurate the ATP's prediction results could be. Therefore, the contention poll should be scheduled at the end of the cruising period.

Figure 3.6 illustrates the flow chart for the proposed TSP PGP. The PGP consists of two major components: data updater and poll generator. The data updater waits for the three types of data updating triggers within the data hub system. They are triggers for traffic signal operation data, PRG requests data, and polled bus data. With the latest information, the data updater is able to check the status of the ATSP system and communications system. It is to guarantee polling requests would not be generated if the ATSP system is not functioning, e.g. if the traffic signal is running free or the communication link to the field controllers is broken.

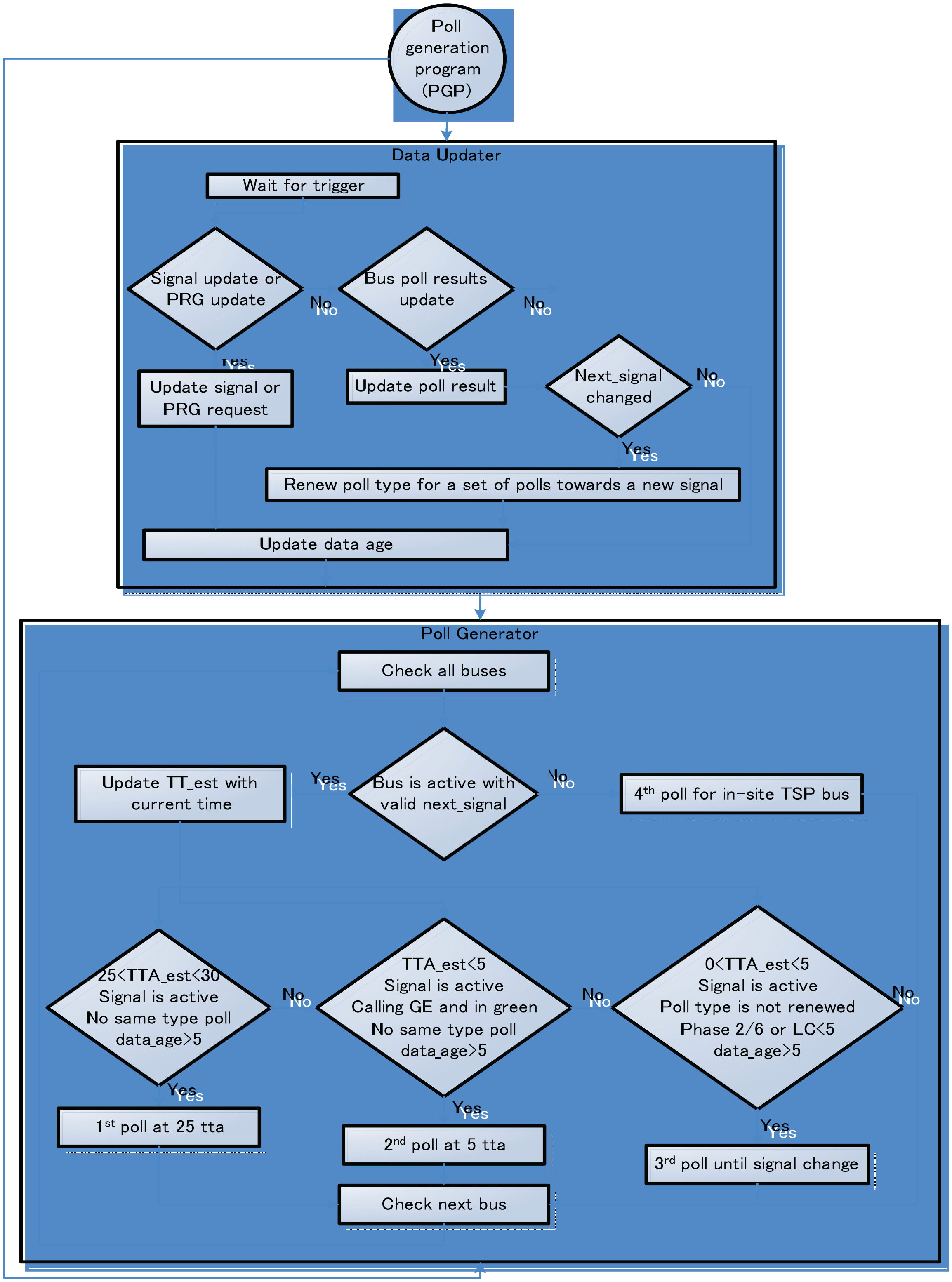


Figure 3.6 Flow chart for the TSP poll generation program (PGP)

On the other hand, the poll generator manages when to initiate the polling request. In general, four types of polls are generated by the PGP. The first poll is when a bus is approaching an intersection but not yet at the back of the traffic queue. According to the prediction data from the previous field testing, the impact of the traffic queue can be as far as six hundred and sixty feet (two hundred meters). Thus, we designed the first polling request to be twenty-five seconds away from the downstream intersection. A timer starts counting down the predicted arrival time after the previous update of bus information, either through a regular periodic poll or contention poll. When predicted time to arrival is counted as 25 seconds, a polling request will be initiated for the latest bus arrival information. It is normally when the priority request is proposed. The second type of poll is designed for the green extension case. Once PRG initiates an extended green request, a polling request will be initiated when the green phase for bus direction is still on and the bus will arrive at the downstream signal within five seconds. Such a poll is designed to check whether the bus is still in need of green extension. The third type of poll is to check when the bus is cleared from the intersection, so the unnecessary priority can be canceled and the predicted bus arrival time can started counting down. The third such poll will be initiated every five seconds when the predicted bus arrival time is within five seconds and the bus phase is green. The fourth type of poll is for those buses with an active GPS update from the regular poll but not yet in the designated site. The polling frequency for the fourth poll is thirty seconds, which is much larger than two minutes for the periodic regular poll. This type of poll is designed for the bus, which just started a fresh service run.

3.3.2 Laboratory Testing of Proposed Communications Protocols

The laboratory testing is designed to evaluate and calibrate the proposed communication protocols before the implementation in the field. Two simulation programs are designed for the laboratory testing. One simulation program is to generate bus trajectories which follow simple kinematic rules and traffic signal operation. The other program is to generate traffic operation information, e.g. current phase, signal color, local and master clock timers, etc., based on designated signal timing tables and the local computer clock.

As illustrated by Figure 3.7, a three-signal arterial is selected as the imaginary test site for the laboratory testing. According to the previous ATSP field testing, the selected three signalized intersections, i.e. 9th Ave, 12th Ave and Barneson Ave, have relatively high probability of requesting TSP treatment due to high side street traffic and a long red for bus phases. The simulated buses run back and forth along this site with a specified turnaround time at each end of the site. The simulated signal operation assumes all phases can reach their force-off points. For 9th Ave (#1), 12th Ave (#2), and Barneson Ave (#3), the morning peak pattern #1 with cycle length 90 seconds is selected. The green durations are 64 seconds (#1), 75 seconds (#2), and 75 seconds (#3), respectively. The coordination offsets are 0 second (#1), 2 seconds (#2) and 49 seconds (#3).

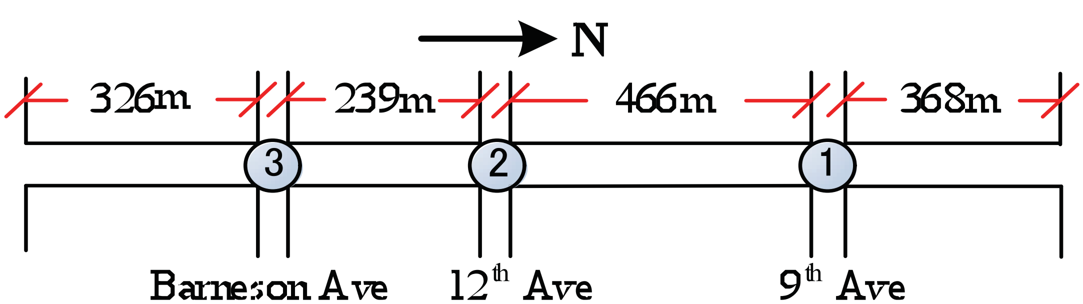


Figure 3.7 Imaginary testing network for the laboratory test

Given the proposed design of PGP as shown in Figure 3.6, three testing scenarios have been conducted. Figure 3.8 illustrates the polling results for the first scenario with only one active bus. The high spikes at about the 45-second level correspond to the trajectory along the longest block (466-meter-long) between 9th Ave and 12th Ave. The poll at the 45th second time-to-arrival (TTA) is the last check-out poll (3rd type poll). The next poll along the trip is the 1st type poll at about the 25th second TTA. Then the following poll is the 2nd type poll when the bus is about five seconds away from the downstream signal. When the bus is within five seconds from the downstream intersection, some 3rd type polls are initiated until the bus is cleared from the intersection.

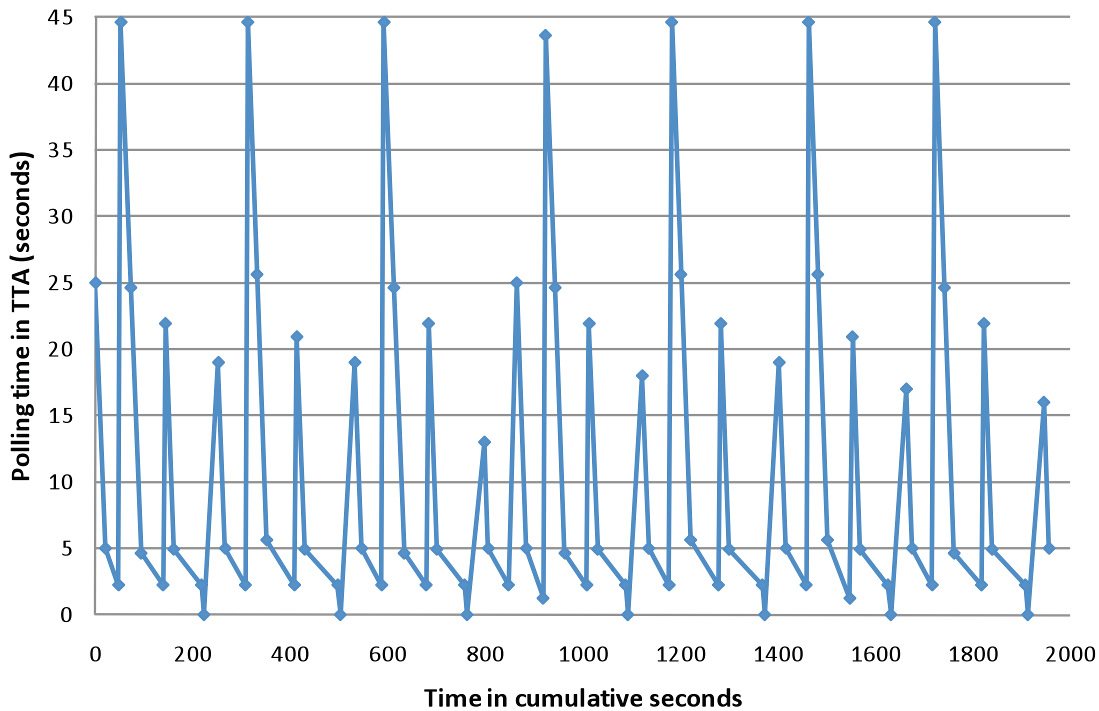


Figure 3.8 Polling request times for single bus scenario

For other blocks with shorter distances, the 1st type polls might not be initiated because the last check out poll is within five seconds of the 25th second TTA. The assumption is that the data with age less than five seconds is still fresh and accurate enough. When buses are at or close to intersections, the polling pattern is similar with those for the trip between 9th Ave and 12th Ave. Most of them are 3rd type polls to update bus check-out information.

Figure 3.9 illustrates the distribution of polling latency for the single bus scenario. The latency here is defined by the time difference between the requested time and service time. Because the bandwidth for the contention channel is 1Hz, the theoretical longest turnaround time is one second. The single bus scenario does not generate more than one polling request within five seconds. There is no conflict on occupying the contention channel. Therefore, the length of polling latency is essentially the gap between the time when the polling request is initiated and the time when the contention channel is open. The maximum latency shown in Figure 4.10 is 1 second. The shape of the distribution is similar to that of a uniform distribution. The average polling frequency is 0.48 second, while the median is 0.51 second.

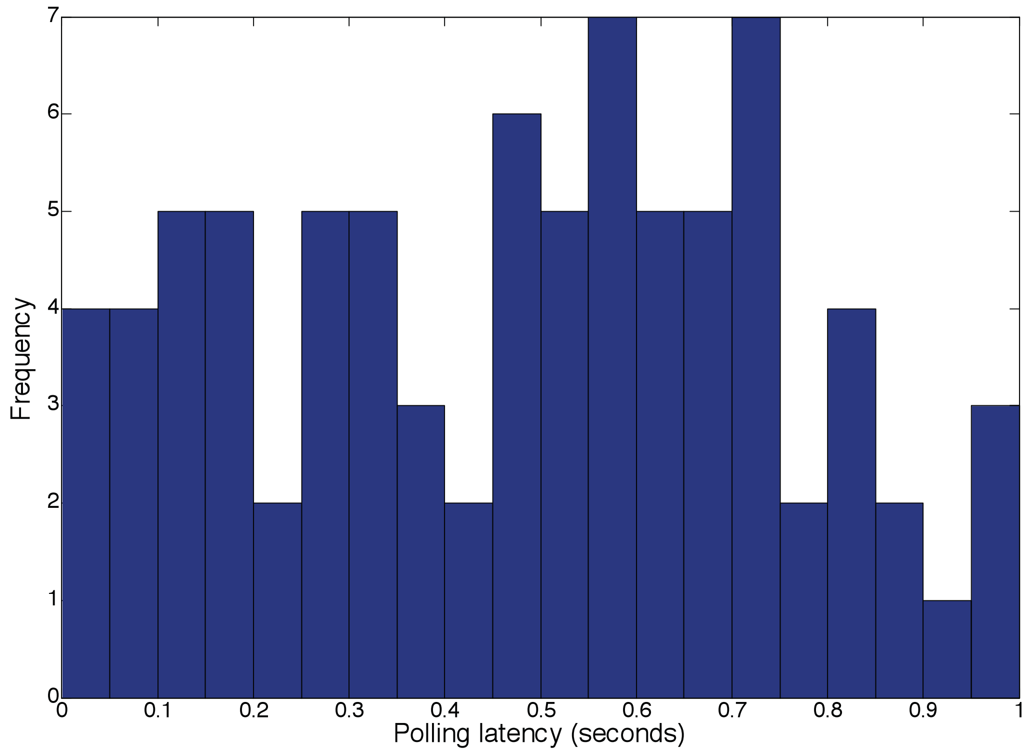


Figure 3.9 Histogram of polling latency for single bus scenario

The second testing scenario is the case with four active buses in site. It is for the scenario within the ATSP field testing site considering the service frequency for SamTrans Route 390/391 with two buses per hour per direction. As illustrated in Figure 3.10, the frequency of the contention poll in the four-bus scenario is much higher than that for the single bus scenario. The polling requests are distributed at three layers: 45seconds when a bus is cleared from an upstream intersection, 25 seconds for the 1st type poll, and within 5seconds for the 3rd type poll. The number of the 3rd type poll is larger than all other types of polls because each bus would generate at least two check-out polls, which are before and after it cleared an intersection.

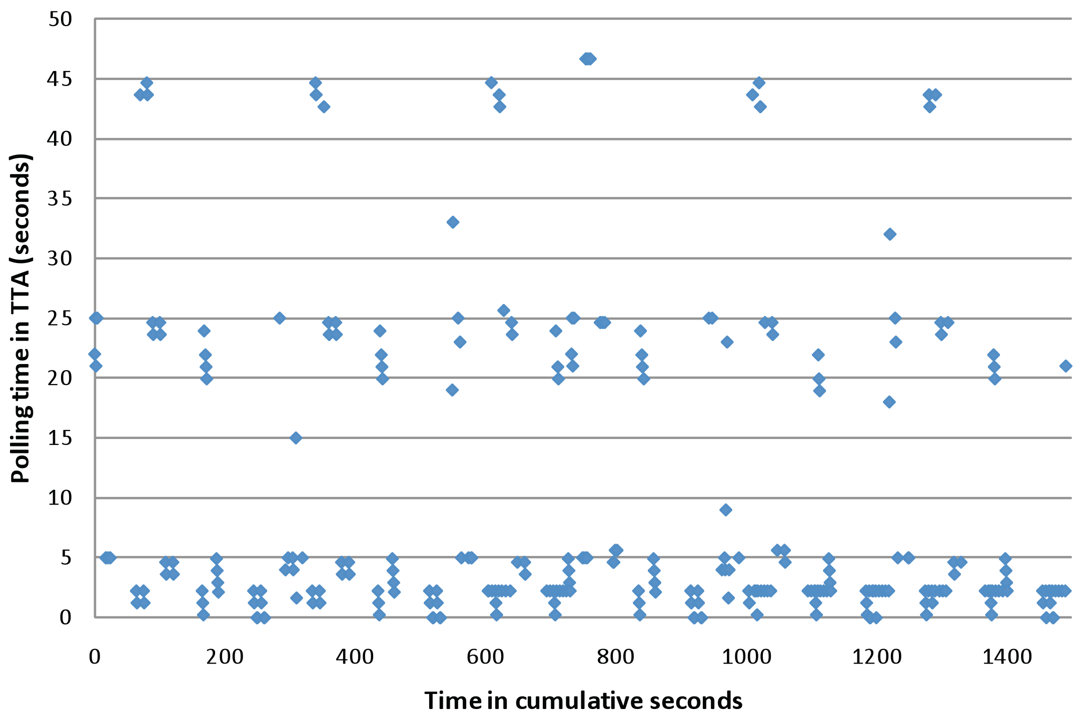


Figure 3.10 Polling request times for four-bus scenario

Figure 3.11 presents the histogram of the polling latency for the four-bus scenario. In this scenario, the worst situation is when four buses generate polling requests at the same time and the channel will open in the next second. Thus, the longest polling latency can be five seconds. However, it is an event with very low probability. As shown in Figure 3.12, the shape of the distribution is not symmetric but heavily weighted on the left with a long tail on the right. In this test with 437 sample polling requests, the mean and median of the polling latency are 0.87 and 0.67 second, respectively. The longest polling latency is 3.79 seconds. Among the 437 sample requests, 40 (9.2%) are 1st type; 99 (22.7%) are 2nd type; 280 (64.1%) are 3rd type; 208 (47.6%) are at signal #1; 115 (26.3%) are at signal #2; and 89 (20.4%) are at signal #3.

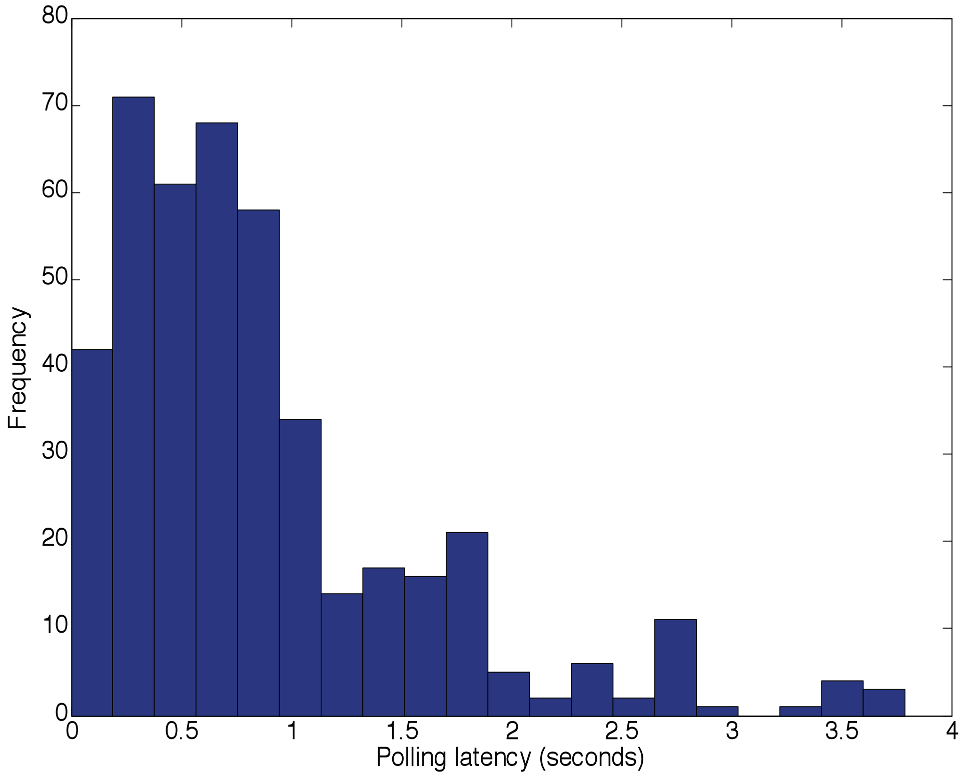


Figure 3.11 Histogram of polling latency for four-bus scenario

The last scenario is designed for the whole fleet of routes 390/391 with two buses per hour per direction. In total, there are eight active buses simultaneously serving the two routes. As illustrated by Figure 4.12, there are more polling requests competing for requesting the contention channel. It is noted that there are some consecutive polling requests between 10-second and 20-second TTA. Those requests are the mix of 3rd type polls and 1st type poll, which are competing for the 1Hz contention channel.

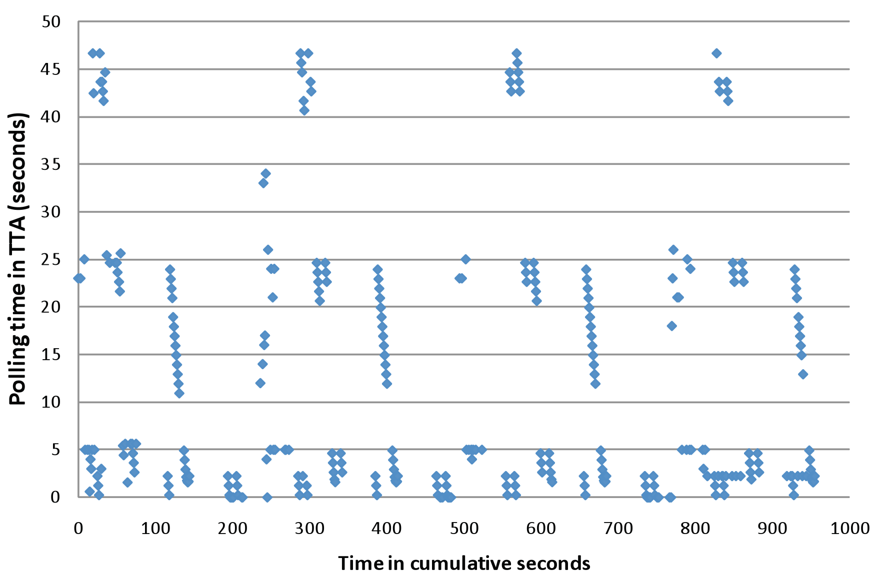


Figure 3.12 Polling request times for eight-bus scenario

Figure 3.13 shows the distribution of polling latency for the eight-bus scenario. Although the longest possible latency is nine seconds, the longest observed latency for this test is eight seconds. Similarly with the four-bus scenario, the shape of the distribution for the polling latency is positively skewed with a long right tail. The mean and median of the polling latency is 2.24 and 1.55 seconds, respectively. For the total of 462 sample requests, 86 (19.7%) are 1st type; 112 (25.6%) are 2nd type; 245 (56.1%) are 3rd type; 144 (31.2%) are at signal #1; 159 (34.4%) are at signal #2; and 125 (27.1%) are at signal #3. In the eight-bus scenario, the probability of a polling latency shorter than five seconds is 88.5%.

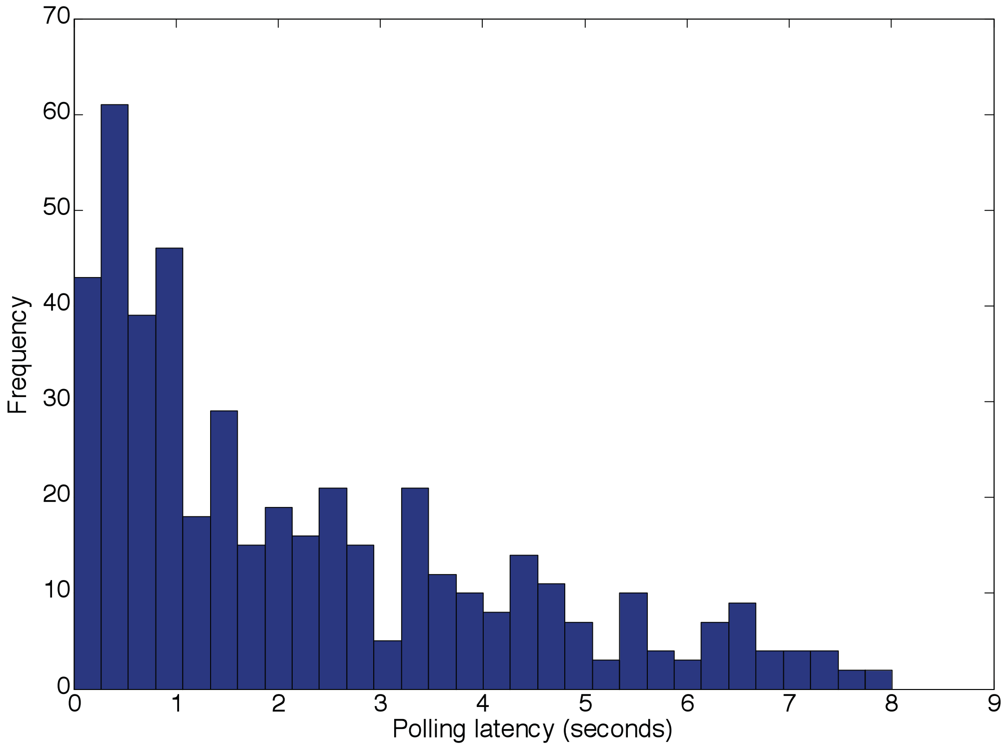


Figure 3.13 Histogram of polling latency for eight-bus scenario

3.3.3 TSP Performance under Proposed Communications Protocols

Under the proposed communications protocols and existing AVL/ACS system, the performance of the ATSP system will be different from that under the continuous wireless communications systems, e.g. GPRS. The magnitude of the performance difference depends on a couple of factors. The first factor is the priority treatments, i.e. early green (EG) or green extension (GE). For the GE treatment, there is almost no performance difference because the 2nd and 3rd type polls guaranteeing the bus would receive extended green before it checks out from the intersection. In contrast, the early green treatment would receive more of an impact because it relies on the prediction when buses are approaching intersections. The other factor is the sign of the additional prediction error due to discrete information updates. When the additional prediction error is positive or the actual arrival time is earlier than the predicted arrival time, the priority would not be enough, thus the benefits of TSP on buses would be less. But the impact on other traffic would be less too. For the negative prediction error, the priority would be more than what the ATSP system designed. It would also lead to additional impact on other traffic.

Under the proposed communications protocols, the impact on TSP performance is mainly from the 1st type poll for early green treatment. The previous field testing shows that the average number of EG treatment is 58.5% of all TSP treatments. Such EG treatments relies on the prediction error at the 1st type poll. According to the eight-bus laboratory testing scenario, the probability of the 1st poll to be initiated between 20-second and 25-second TTA is 88.5%. According to the previous field testing, Figure 3.15 and Figure 3.14 illustrate the difference of prediction error in the time window from TTA25 to TTA20. The shapes of the distributions are very thin with standard deviations as short as 1.37 and 1.67 seconds, respectively. The one sigma windows from the medians of the distributions (μ±σ) cover 78.3% and 83.9% of the whole distributions. In other words, there is an 80% chance that the additional prediction errors due to the discrete communications protocol are within ±1.5 seconds. The additional prediction error is relatively small compared with the average prediction error due to other factors, e.g. equipment accuracy, traffic dynamics, traffic disturbance, etc.

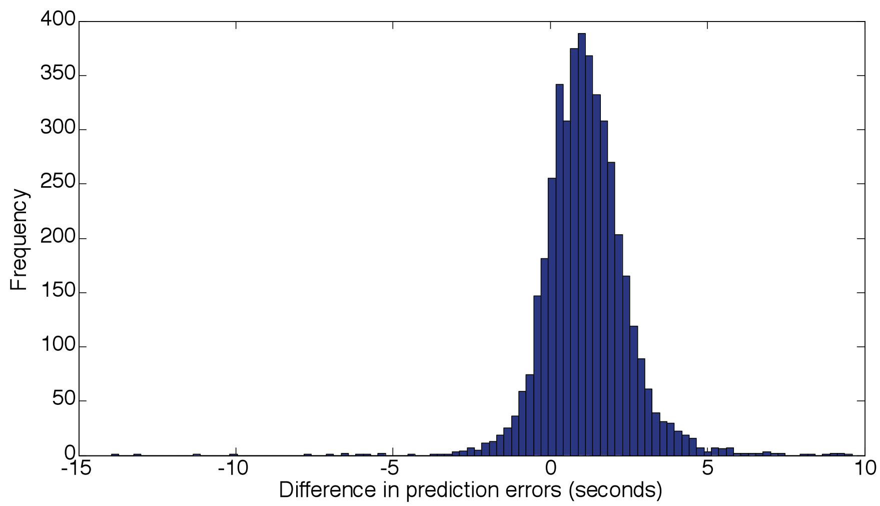


Figure 3.14 Histogram of difference in prediction errors between TTA-25 and TTA-21 for northbound buses)

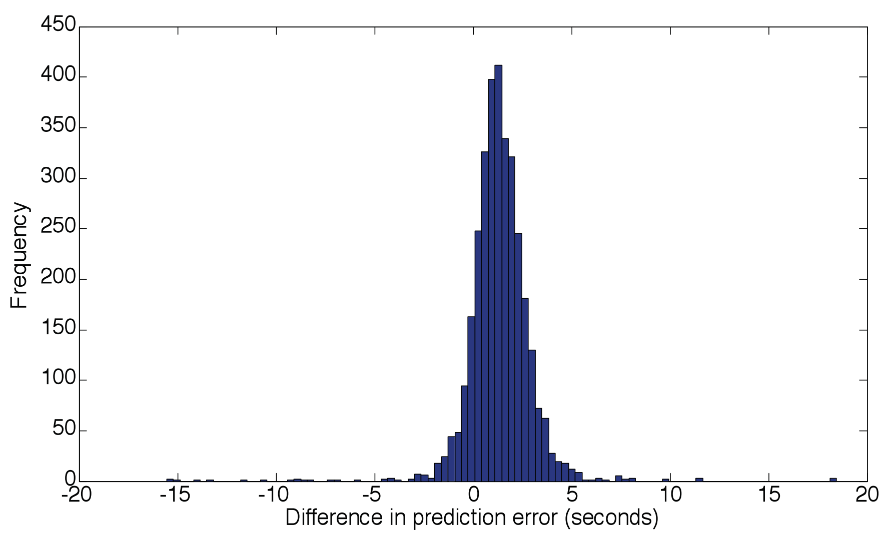


Figure 3.15 Histogram of difference in prediction errors (between TTA-25 and TTA-21 for southbound buses)

In conclusion, the additional prediction error (±1.5 seconds) due to the discrete communications protocols is not significant. The incurred increment or decrement of bus delay is less than 1.5 seconds. Due to multiple traffic phases, the incurred changes on each traffic phase duration will be less than 1 second, which is also very small. Moreover, the additional prediction error only impacts the EG treatment, which represents 58.5% of all priority cases. The average impacts on both bus delay savings and traffic phase durations are less than 1 second and again very small.

The results of this emulation study indicate that the ACS communications channel, combined with newly designed protocol is able to support the ATSP system without significant compromise on system performance.

3.4 Data

High quality data are essential to quantitatively evaluate the effectiveness of ATSP. Second-by-second traffic data (signal phasing and timing, loop counts) are collected 24 hours a day, 7 days a week, by utilizing the existing Frame Relay wired communications link between the field and the Traffic Management Center at District 4. The data are then forwarded to the PATH data server at the Parsons Transit and Traffic Laboratory via EVDO wireless communications.

The PATH team developed cell phone-based AVL portable devices that were installed on 15 SamTrans buses that participated in the field operational tests. Second-by-second GPS location data are transmitted in real-time to the PATH data server through GPRS wireless communications.

In previous projects, transit data and traffic data were saved in data files. When conducting the evaluation, it was very time consuming to synchronize TSP events with bus movements and traffic counts among different data files. Since then, the PATH team has developed a comprehensive transit and traffic database under MySQL. Software tools have also been developed to pre-process and clean the data, save the data into a database, monitor bus movements and ATSP operations in real-time, and to calculate the measures of effectiveness (MOEs).

4. Test Results and System Performance Evaluation

The ATSP system was field tested at 35 intersections along El Camino Real. The test vehicles are SamTrans in-service buses that operate on Routes 390 and 391. The details of the field test and the analysis of the testing data are summarized below.

4.1 The Description of the Field Test

Transit vehicles spend an average of 15 percent of their trip time waiting at traffic signals (Baker et al. 2004). TSP aims to reduce bus intersection delay and thereby lead to shortened trip time and potentially reduced number of stops at red and faster average running speed. This section presents the results of trip-based analysis to evaluate the benefits of ATSP on transit operation. The MOEs used include: bus trip intersection delay, number of stops at signalized intersections, bus trip travel time, and average trip running speed.

Based on the second-by-second data sent back from the AVL system, the performance of the ATSP system was analyzed and evaluated. This section provides a general overview of the testing result, including general information on the bus operation and the ATSP system.

Table 4.1 is a summary of the bus operations. Figure 5.1 shows the average number of TSP calls per bus trip. As Figure 4.1 demonstrates, a bus traveling northbound on average received 10.02 priority treatments, while a bus traveling southbound received 7.48 treatments. The details of the TSP call information at each intersection are given in Table 5.2.

Table 4.1 Daily bus operation statistics

	Minimum	Maximum	Average
Number of operating buses	8	11	9.6
Number of NB trips	11	15	12.2
Number of SB trips	9	14	11.1

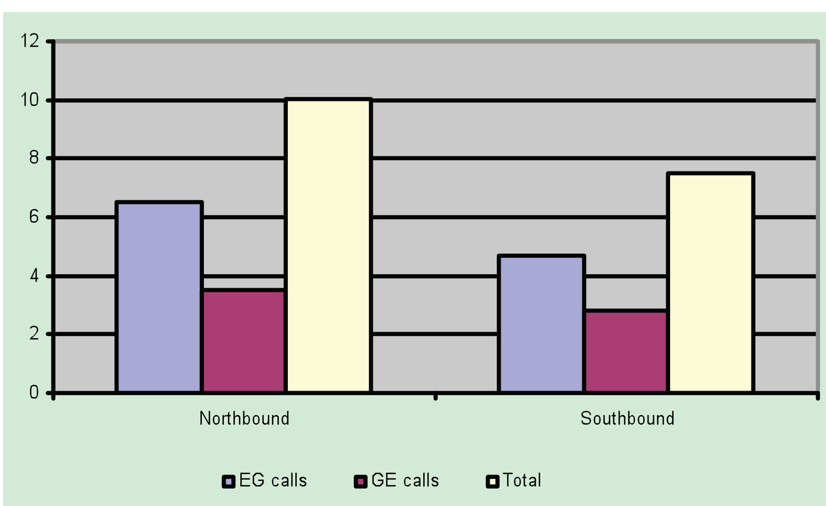


Figure 4.1 Number of TSP Calls

Table 4.2 Detailed TSP call information

Field Master	Northbound		Southbound	
	Average EG calls per trip	Average GE calls per trip	Average EG calls per trip	Average GE calls per trip
Ray/Rosedale	0.22	0.17	0.24	0.15
Oxford	0.31	0.27	0.00	0.02
Hillside Dr	0.11	0.08	0.34	0.27
Lincoln Ave	0.32	0.31	0.05	0.04
Broadway	0.00	0.03	0.21	0.18
Carmelita Ave	0.17	0.10	0.14	0.12
Sanchez Ave	0.29	0.31	0.17	0.17
Oak Grove	0.07	0.15	0.35	0.26
Floribunda Ave	0.32	0.05	0.02	0.04
Chapin Ave	0.04	0.07	0.07	0.10
Burlingame Ave	0.15	0.18	0.07	0.12
Howard Ave	0.21	0.03	0.14	0.10
Bayswater Ave/Cypress Ave	0.29	0.23	0.18	0.03
Peninsula Ave	0.25	0.08	0.10	0.14
Burlinggame, FM#6	**2.76**	**2.06**	**2.09**	**1.75**
Bellevue Ave	0.09	0.03	0.16	0.11
Poplar Ave	0.00	0.00	0.00	0.00
Tilton Ave/El Cerrito Ave	0.64	0.07	0.01	0.00
Baldwin Ave/Baywood Ave	0.12	0.07	0.33	0.08
Crystal Spring Rd	0.03	0.02	0.11	0.05
2nd Ave	0.04	0.09	0.00	0.00
3rd Ave	0.02	0.06	0.20	0.08
4th Ave	0.00	0.01	0.03	0.02
5th Ave	0.35	0.13	0.02	0.00
San Mateo, FM #7	**1.29**	**0.49**	**0.86**	**0.34**
9th Ave	0.52	0.24	0.01	0.00
12th Ave/Hobart Ave	0.22	0.21	0.37	0.33
Barneson Ave	0.23	0.07	0.16	0.14
17th Ave/Bovet Ave	0.27	0.00	0.33	0.08
20th Ave	0.00	0.00	0.00	0.00
25th Ave	0.00	0.00	0.00	0.00
27th Ave	0.00	0.00	0.26	0.04
28th Ave	0.40	0.22	0.07	0.03
31st Ave	0.42	0.02	0.34	0.09
37th Ave	0.00	0.00	0.01	0.00
31st Ave	0.00	0.00	0.00	0.01
42nd Ave	0.41	0.18	0.18	0.00
San Mateo, FM#9	**2.46**	**0.95**	**1.72**	**0.71**
Total	**6.51**	**3.51**	**4.67**	**2.81**

Transit data were collected with TSP on and off. One-way bus trips were retrieved from archived data. A sample refers to a one-way bus trip on the study section. Routes 390 and 391 provide bus service from 4 or 5 am to midnight. Only those samples on weekdays and with the start and end trip time between 7am and 8pm, when the signals are under coordinated traffic control, were included in this study. Table 4.3 summarizes the sample size for the study.

Table 4.3 Bus Trip Sample Size

	Without TSP	With TSP
Northbound Trips	140	136
Southbound Trips	234	154

4.2 The Benefits to Transit

4.2.1 Intersection Stopped Delay

The change in delay, more specifically the waiting time, at individual intersections is a direct measure of benefits for transit. Figure 4.2 and Figure 4.3 show a plot of the empirical cumulative distribution function (CDF) for total intersection delay per trip, respectively for northbound and southbound trips. For northbound trips, the average of total intersection delay was 4.9 minutes without TSP, and it reduced to 4.0 minutes with TSP (18.5% reduction). For southbound trips, it was 4.1 minutes without TSP and reduced to 2.8 minutes with TSP (32.0% reduction). In order to confirm the changes have statistical significance, t-tests were performed at a 5-percent significance level assuming identical average intersection delay for both directions. The small p-values of the t-test, as shown in Table 4.4, indicate that the changes are statistically significant.

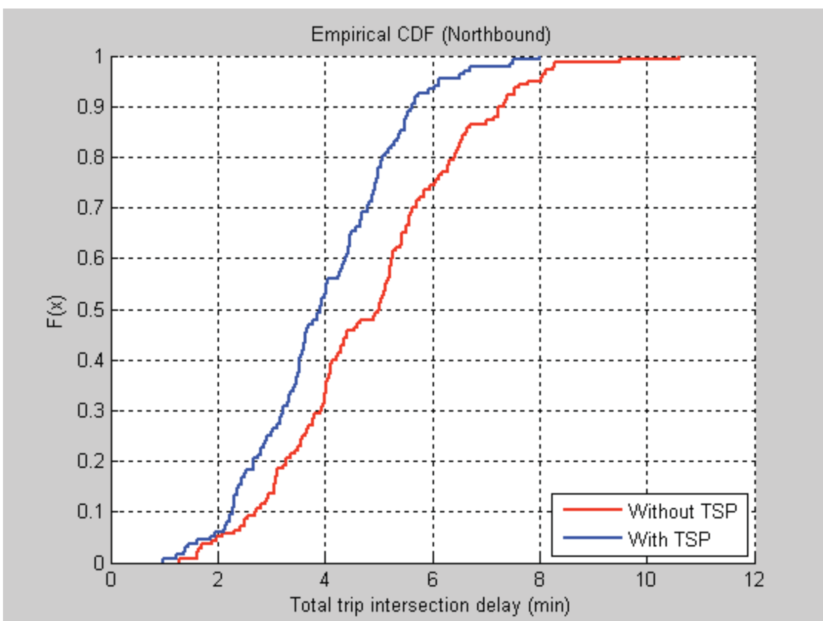


Figure 4.2 Comparison of Northbound Total Intersection Delay

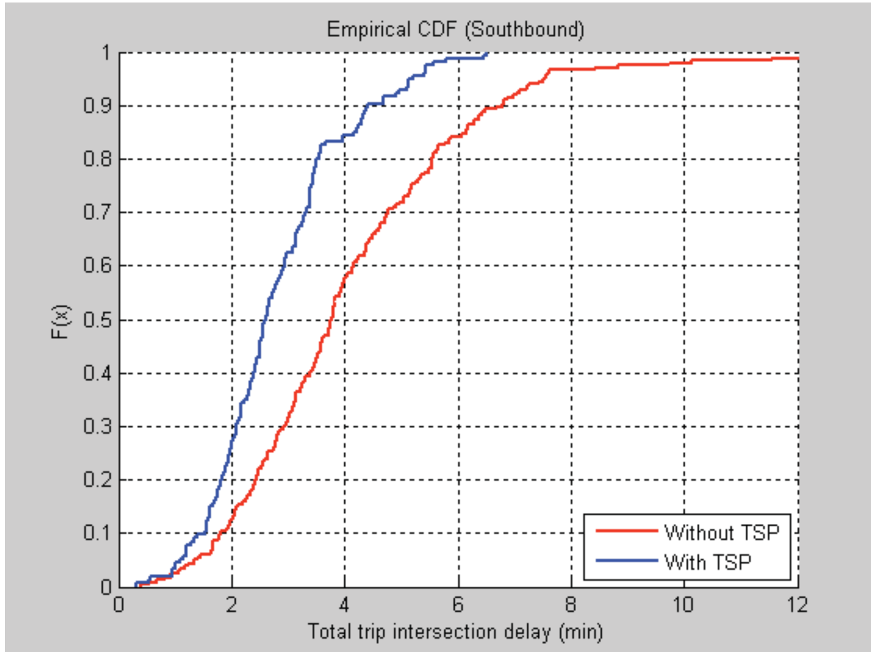


Figure 4.3 Comparison of Southbound Total Intersection Delay

Table 4.4 Total Intersection Delay Comparison

Direction	Without TSP (minutes)	With TSP (minutes)	Change		p-value of t-test
			Value	%	
Northbound	4.9	4.0	-0.9	-18.5%	3.47E-6
Southbound	4.1	2.8	-1.3	-32.0%	5.67E-11

4.2.2 Number of Stops at Red

The reduction in number of stops at red not only decreases bus waiting time at intersections but also leads to reduced fuel consumption, tail-pipe emissions, and lowered vehicle wear and tear due to braking.Figure 4.4 and Figure 4.5 shows CDF plots for the number of stops at red, respectively for northbound and southbound trips. With TSP enabled the number of stops in both directions reduced by 1 (9.5% reduction on northbound and 16.8% reduction on southbound), as shown in Table 4.5, T-test showed the reductions are statistically significant.

Table 4.5 Number of Stops Comparison

Direction	Without TSP	With TSP	Change		p-value of t-test
			Value	%	
Northbound	8.1	7.3	-0.8	-9.5%	0.008
Southbound	6.1	5.1	-1.0	-16.8%	1.87E-6

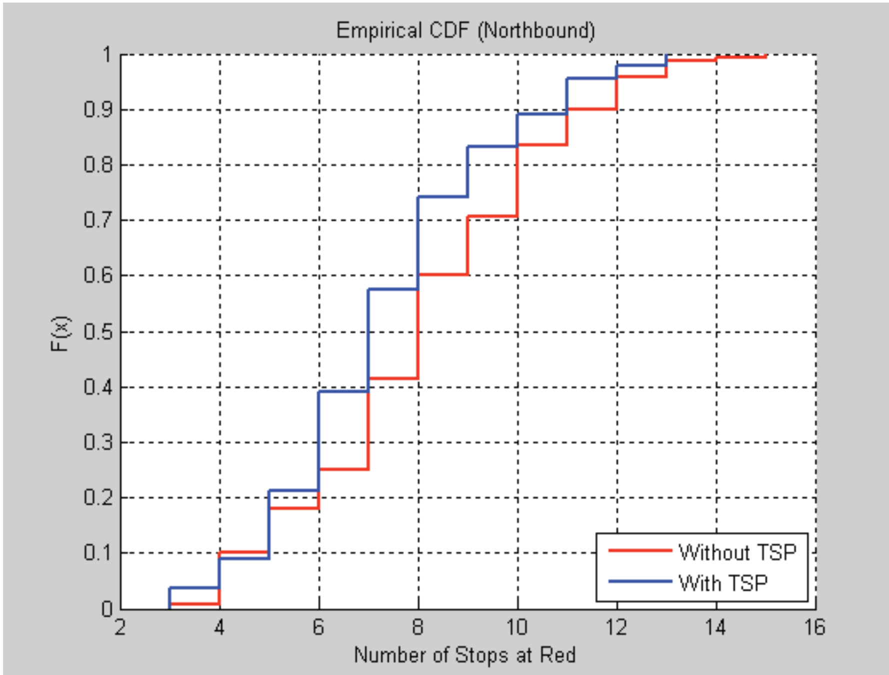


Figure 5.4 Comparison of Northbound Number of Stops at Red

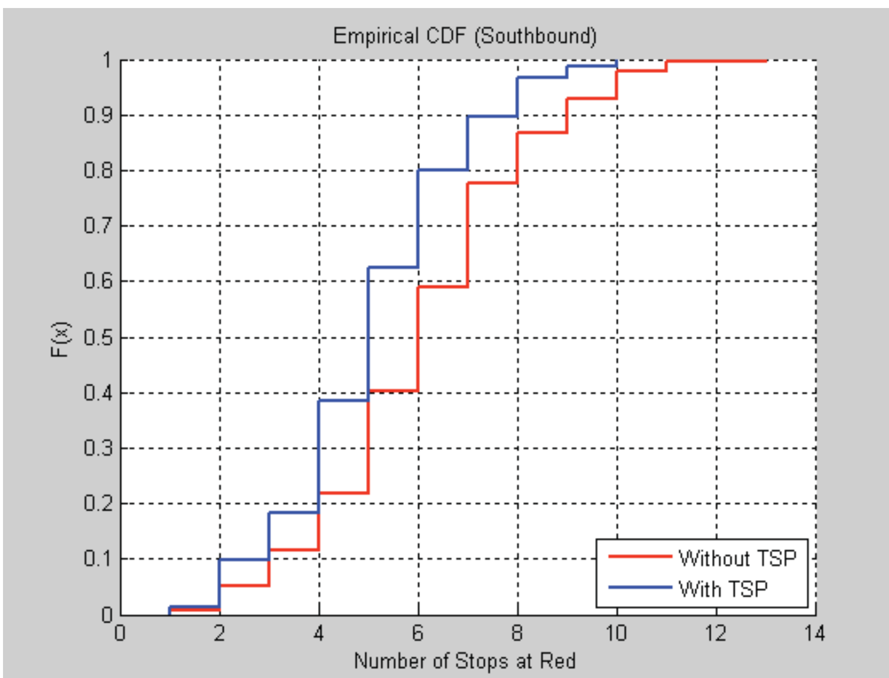


Figure 4.5 Comparison of Southbound Number of Stops at Red

Signal priority is not preemption so buses have to stop for red sometimes while TSP can reduce the waiting time for each of the stops. Table 4.6 compares the average waiting time given stops for both directions. For northbound trips, the average waiting time changed from 36.9 seconds without TSP to 33.1 seconds with TSP (3.9% reduction); while for southbound trips it changed from 41.9 seconds without TSP to 34.1 seconds with TSP (18.5% reduction), as shown in Table 4.6. The changes in both directions are statistically significant.

Table 4.6 Average Waiting Time per Stop Comparison

Direction	Without TSP (seconds)	With TSP (seconds)	Change		p-value of t-test
			Value	%	
Northbound	36.9	33.1	-3.9	-10.5%	0.0017
Southbound	41.9	34.1	-7.7	-18.5%	0.0002

4.2.3 Trip Travel Time

Figure 4.6 and Figure 4.77 shows the CDF plot for bus trip travel time, respectively for northbound and southbound trips. The impacts of TSP on bus trip travel time as well as the results of t-tests are summarized in Table 4.7.

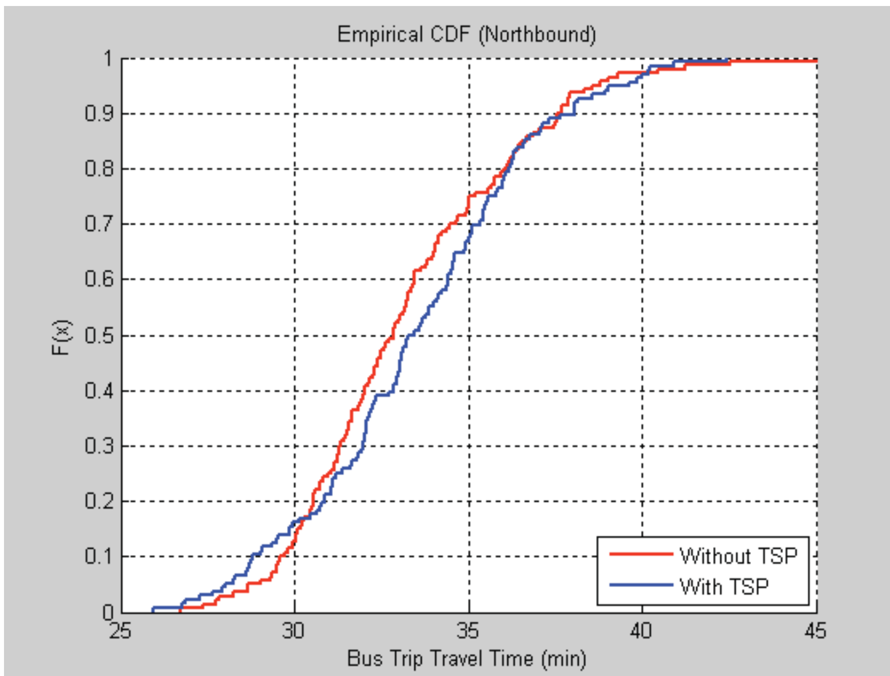


Figure 4.6 Comparison of Northbound Travel Time

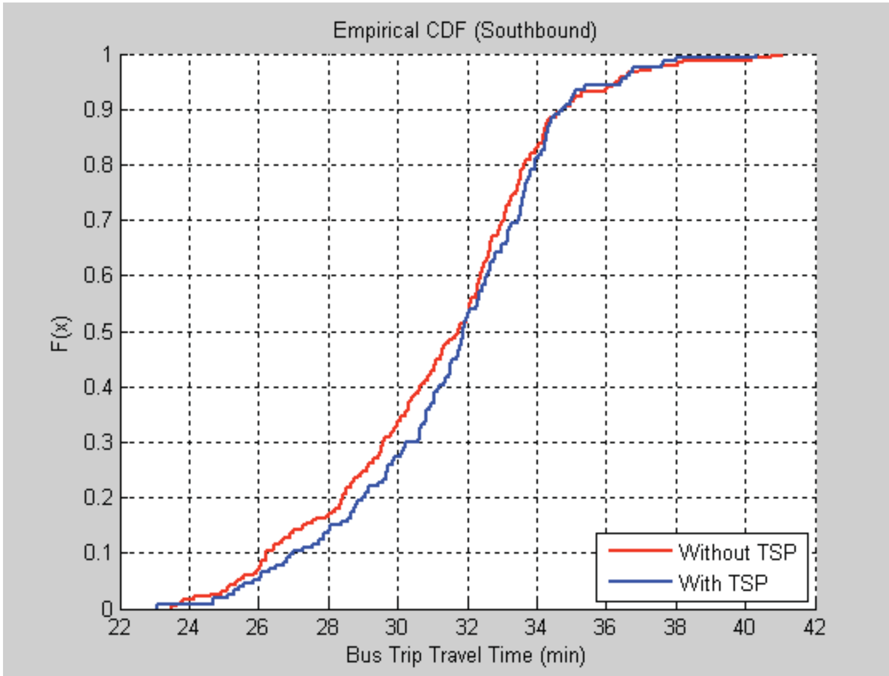


Figure 4.7 Comparison of Southbound Travel Time

Table 4.7 Travel Time Comparison

Direction	Without TSP (minutes)	With TSP (minutes)	Change Value	Change %	p-value of t-test
Northbound	33.3	33.5	0.2	0.7%	0.5424
Southbound	31.2	31.6	0.4	1.2%	0.2733

Despite the TSP benefits we have seen on bus total intersection delay and number of stops for red presented in the previous sections, statistically, there is no change in bus trip travel time in both directions for with and without TSP cases. One reason for that might be because bus services on routes 390 and 391 are schedule-based and bus operators are very cognizant of their timing on routes and thereby they stopped longer at bus-stops or made non-passenger stops. To demonstrate this bus-dwell-time behavior, Figure 4.8 and Figure 4.9 show the CDF plot for dwell time at 3 time-points on route, respectively for northbound and southbound trips.

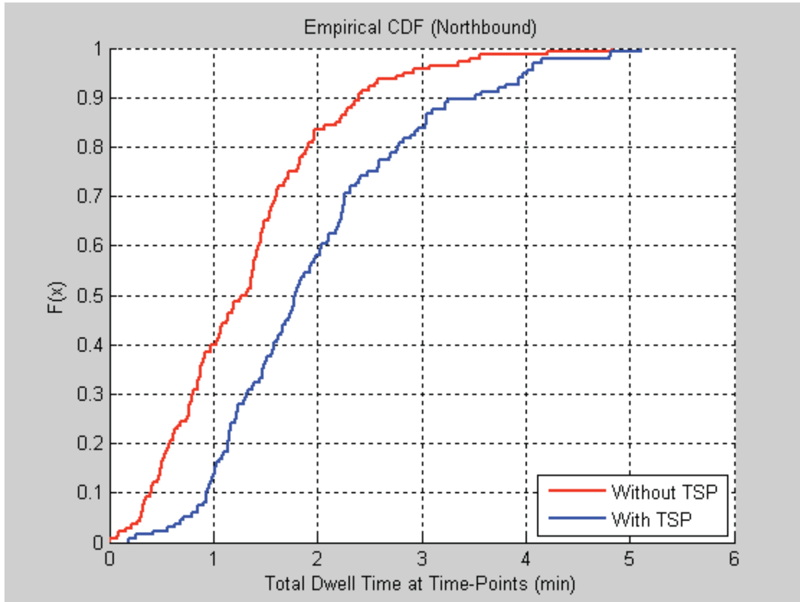


Figure 4.8 Comparison of Northbound Dwell Time at Time-Points

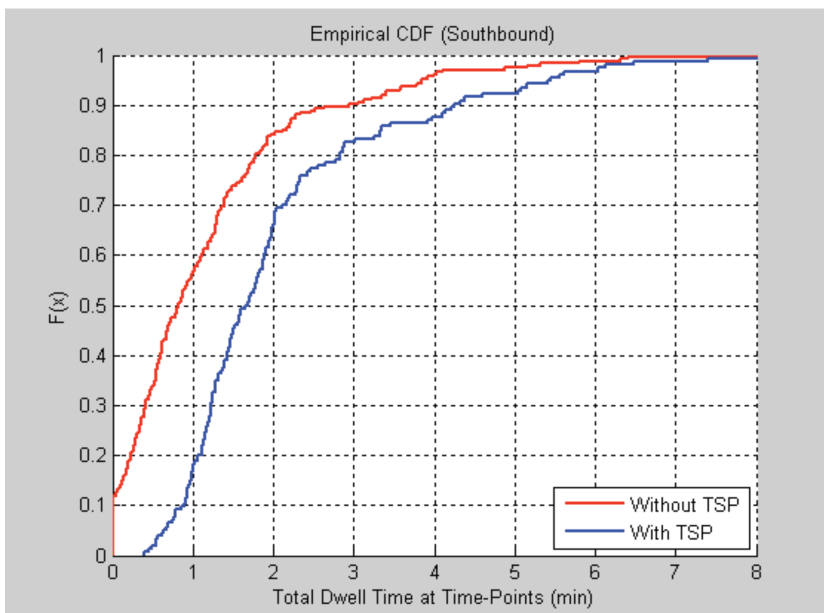


Figure 4.9 Comparison of Southbound Dwell Time at Time-Points

With TSP on, in both directions, bus dwell time at 3 time-points increased significantly when compared with the cases without TSP (48.7% on northbound and 75.5% on southbound). T-test indicates the changes are statistically significant, as shown in Table 4.8**Table 4.**.

Table 4.8 Time-Points Dwell Time Comparison

Direction	Without TSP (minutes)	With TSP (minutes)	Change		p-value of t-test
			Value	%	
Northbound	1.3	2.0	0.7	48.7%	1.32E-8
Southbound	1.2	2.1	0.9	75.5%	9.95E-10

The observed bus-dwell-time behavior change not only occurred at time-points but also at normal bus-stops.

Figure Figure 4.10 and Figure 4.11 graphically demonstrated how the number of stopped bus-stops per trip changed for both directions, when comparing TSP on cases with TSP off cases.

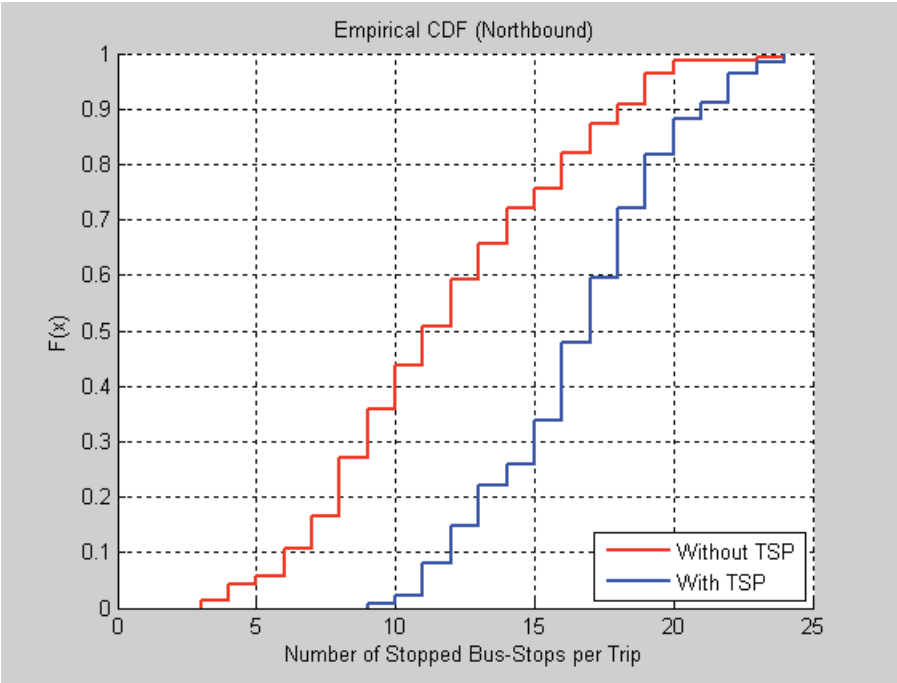


Figure 4.10 Comparison of Northbound Number of Stopped Bus-Stops

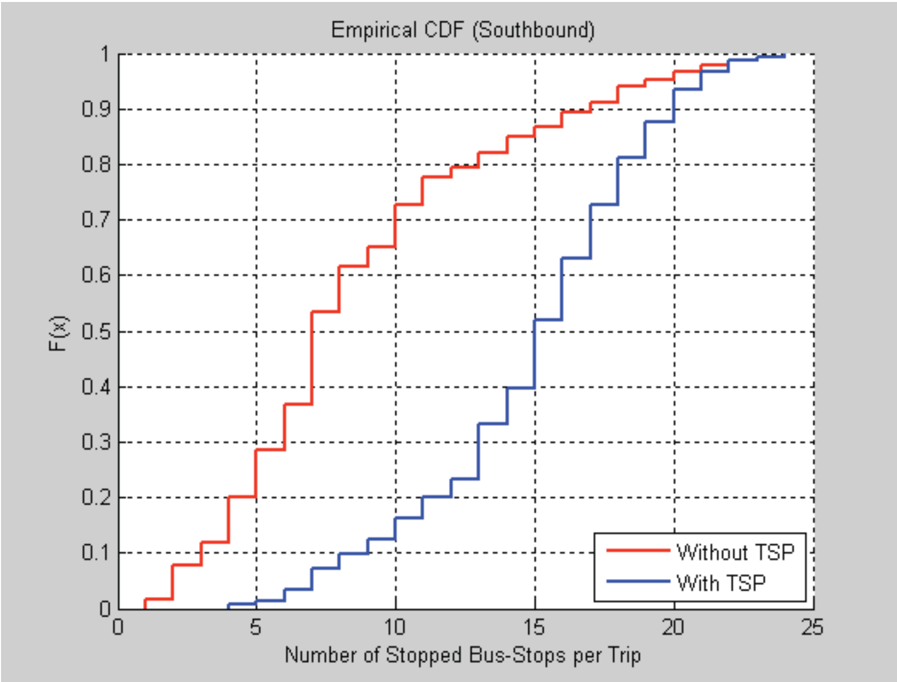


Figure 4.11 Comparison of Southbound Number of Stopped Bus-Stops

Table 4.9 below summarizes bus dwell-time behavior for TSP on and off cases. It clearly shows that total bus dwell-time as well as the number of stopped bus-stops significantly increased when TSP is enabled, while the average dwell time per stop remained about the same.

Table 4.9 Bus-Dwell-Time Behavior Comparison

Direction		Without TSP	With TSP	Change		p-value of t-test
				Value	%	
Northbound	Number of Stopped Bus-Stops	11.8	16.6	4.8	40.5	3.06E-20
	Average Dwell Time per Stop (sec)	31.0	30.0	-1.0	-3.12	0.2258
	Total Dwell Time per Trip (min)	6.0	8.3	2.2	37.2	1.86E-13
Southbound	Number of Stopped Bus-Stops	8.7	14.9	6.2	71.6	1.18e-30
	Average Dwell Time per Stop (sec)	38.2	33.4	-4.8	-12.5	0.0117
	Total Dwell Time per Trip (min)	4.9	8.0	3.1	63.2	6.24E-26

The changes in total bus dwell-time are graphically shown in Figure 4.12 and Figure 4.13 below.

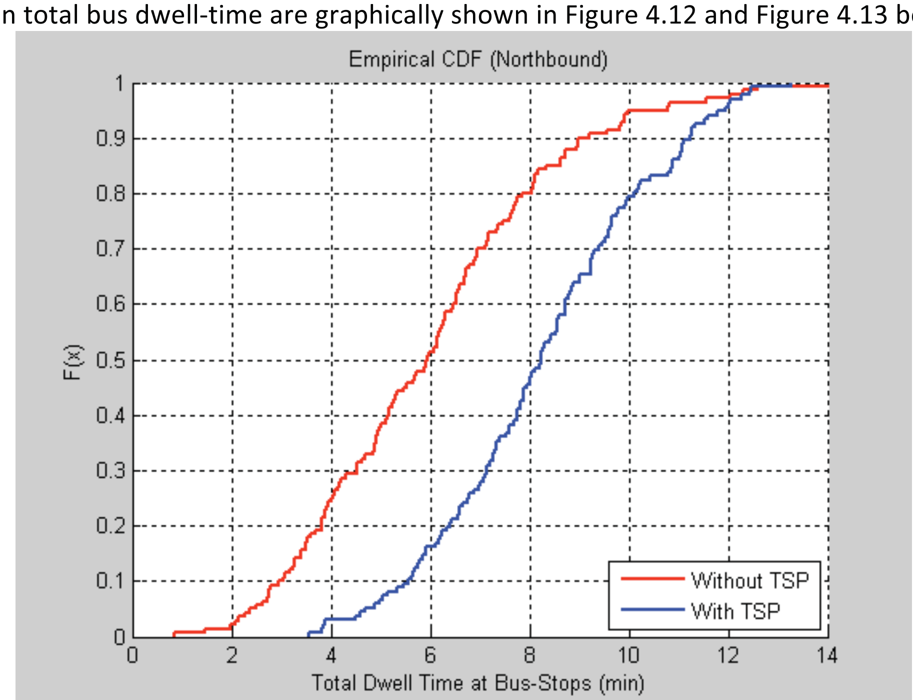


Figure 5.12 Comparison of Northbound Total Dwell Time

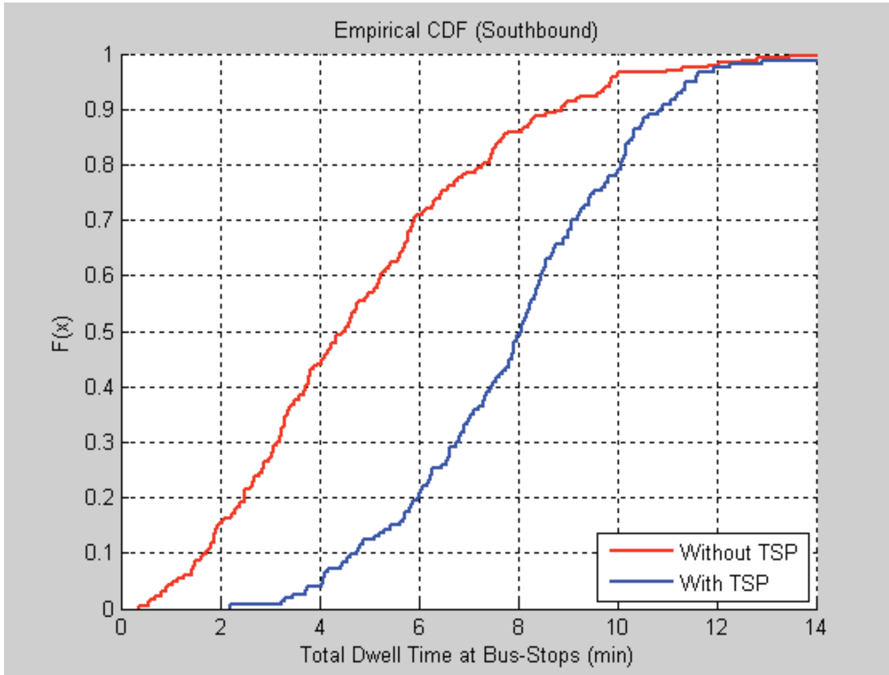


Figure 4.13 Comparison of Southbound Total Dwell Time

In order to make a meaningful comparison of bus trip travel time, the aforementioned bus dwell-time behavior change must be considered. Figure 4.4 and Figure 5 shows the CDF plot for bus trip travel time, excluding the dwell time, respectively for northbound and southbound trips.

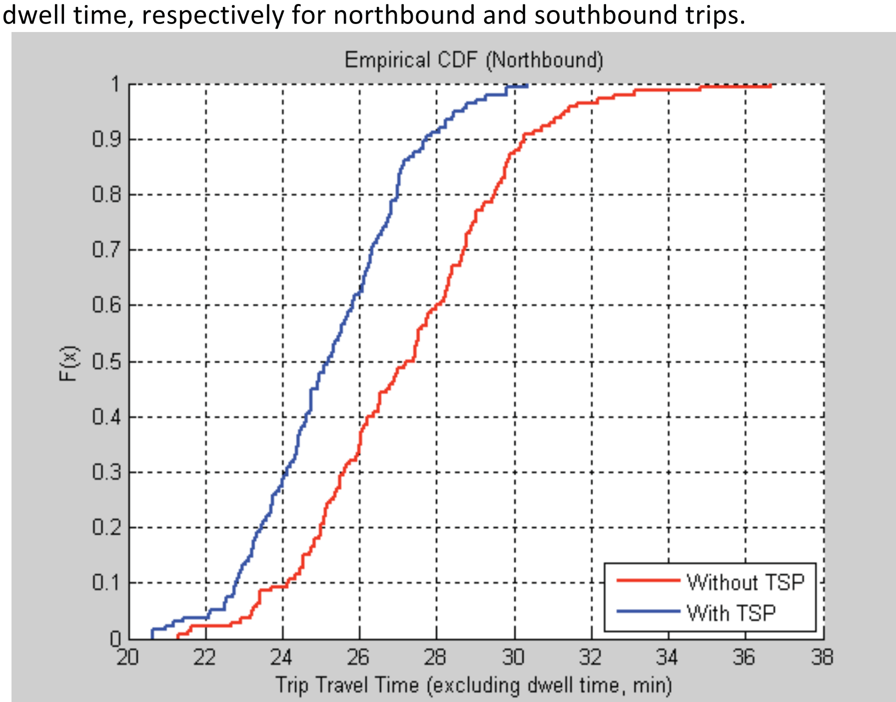


Figure 4.14 Comparison of Northbound Travel Time (excluding dwell time)

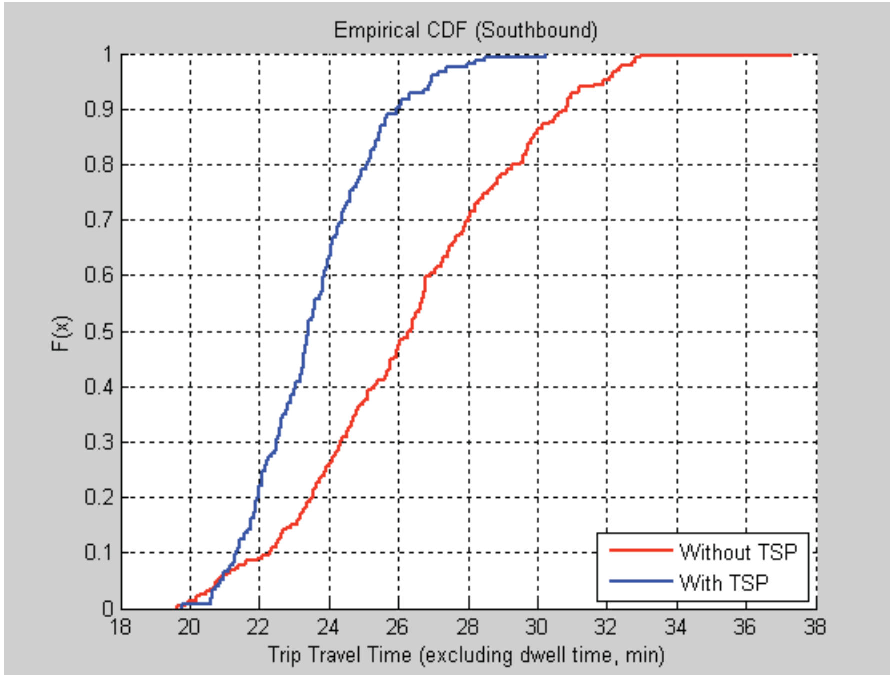


Figure 4.15 Comparison of Southbound Travel Time (excluding dwell time)

As summarized in Table 4.10, TSP operation reduced bus trip travel time (excluding dwell time) by 7.3% (or 2.0 minutes) for northbound direction and 10.4% (or 2.7 minutes) for southbound direction. The small p-values of the t-test indicate the changes are statistically significant. Section 4.6 of this report presents the estimated time savings on routes 390 and 391, respectively.

Table 4.10 Travel Time Comparison (Excluding Dwell Time)

Direction	Without TSP (minutes)	With TSP (minutes)	Change		p-value of t-test
			Value	%	
Northbound	27.2	25.2	-2.0	-7.3%	1.16E-11
Southbound	26.3	23.6	-2.7	-10.4%	2.44E-19

4.2.4 Average Trip Running Speed

Figure 4.16 and Figure 4.17 show CDF plots for average bus running speed on route. For northbound trips, bus average running speed increased from 18.3 mph without TSP to 19.2 mph with TSP on, while for southbound trips, the average running speed increased from 19.3 mph to 20.4 mph. The speed increments (4.9% on northbound, 5.6% on southbound) are statistically significant, as shown in Table 4.1Table 4.11.

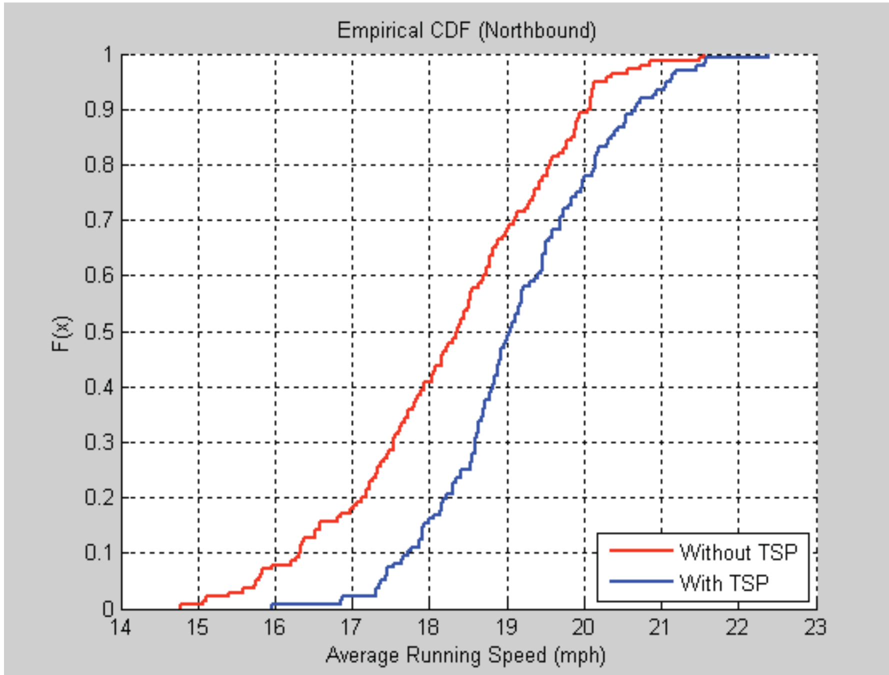


Figure 4-16 Comparison of Northbound Average Running Speed

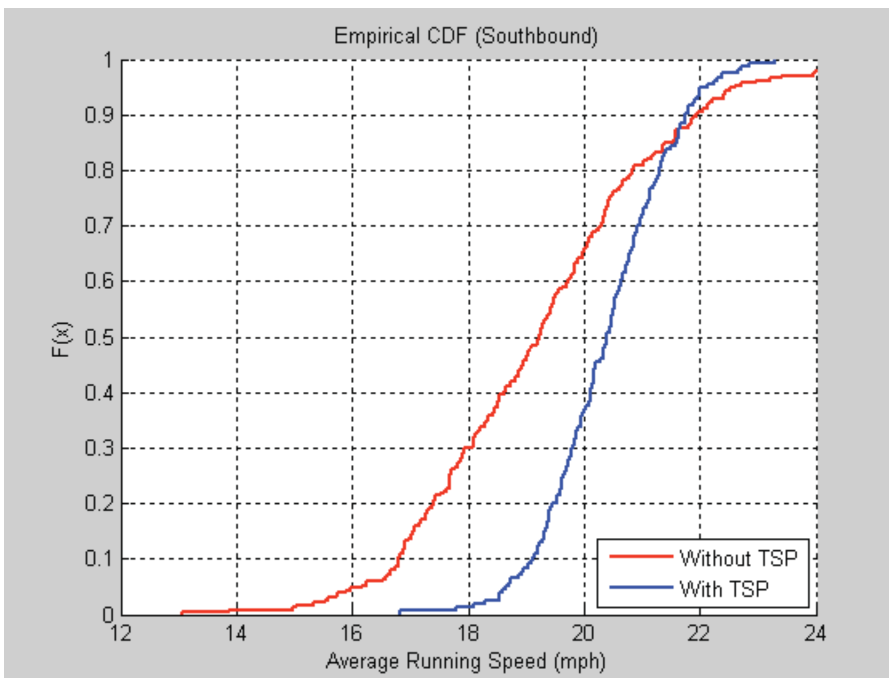


Figure 4.17 Comparison of Northbound Average Running Speed

Table 4.11 Average Running Speed Comparison

Direction	Without TSP (minutes)	With TSP (minutes)	Change		p-value of t-test
			Value	%	
Northbound	18.3	19.2	0.9	4.9%	2.22E-08
Southbound	19.3	20.4	1.1	5.6%	1.71E-07

4.2.5 Potential Saving on Route Travel Time

Bus trip travel time (TT) consists of running time (RT), dwell time (DT) at bus-stops and waiting time (WT) at intersections, i.e.,

$$\begin{aligned} \mathrm{TT} &= \mathrm{RT} + \mathrm{DT} + \mathrm{WT} \\ &= \mathrm{TT} \cdot (P_{\mathrm{RT}} + P_{\mathrm{DT}} + P_{\mathrm{WT}}) \end{aligned} \qquad \textbf{(4-1)}$$

where P_{RT}, P_{DT} and P_{WT} are the percentages of running time, dwell time and waiting time in the total travel time. Assuming TSP reduced the waiting time by r_{WT} percent and increased the average running speed by r_{V} percent, the percentage of reduction in total travel time is then given by

$$r_{\mathrm{TT}} = -\frac{r_{\mathrm{V}}}{1 + r_{\mathrm{V}}} \cdot P_{\mathrm{RT}} - r_{\mathrm{WT}} \cdot P_{\mathrm{WT}} \qquad \textbf{(4-2)}$$

with the same amount of dwell time for TSP on and off causes.

With TSP off, the average total trip travel time on study section was 33.3 minutes and 31.2 minutes on northbound and southbound, respectively. For northbound trips, dwell time accounted for 18.1% of trip travel time and intersection waiting time accounted for 14.7% of trip travel time; while for southbound trips, dwell time accounted for 15.7% and intersection waiting time accounted for 13.2% of total trip travel time. The reduction in waiting time was 18.5% and 32.0%, respectively for northbound and southbound trips. The increment in average running time was 4.9% for northbound and 5.6% for southbound. Applying Eq. 4-2, the potential reduction in travel time along the study section is 5.8% (or 1.9 minutes) for northbound and 8.0% (or 2.5 minutes) for southbound, as shown in Table 4.12.

Table 4.12 Estimated Travel Time Saving

	Travel Time (TSP off, min)	Percentages in Travel Time			TSP Benefits		Travel Time Saving	
		Waiting Time	Dwell Time	Running Time	WT Reduction	RT Reduction	Minutes	%
Northbound (study section)	33.3	14.7%	18.1%	67.2%	18.5%	4.6%	1.9	5.8%
Southbound (study section)	31.2	13.2%	15.7%	71.1%	32.0%	5.3%	2.5	8.0%
390 NB	127.4	14.7%	19.1%	66.3%	18.5%	4.6%	7.4	5.8%
390 SB	129.5	15.9%	19.0%	65.2%	32.0%	5.3%	11.1	8.5%
391 NB	114.9	14.8%	19.3%	65.8%	18.5%	4.6%	6.7	5.8%
391 SB	107.5	15.0%	17.3%	67.7%	32.0%	5.3%	9.0	8.4%

The same method was also applied for the whole routes of 390 (between Palo Alto Caltrain station and Daly City BART station) and 391 (between Redwood City Caltrain station and Evergreen) to estimate the potential travel time saving due to TSP. The travel time, percentages of waiting time, dwell time and running time in travel time were calculated from archived transit trip data. The percentage of waiting time reduction and the percentage of running speed increment obtained on the study section were applied for the whole route. Potentially, TSP could reduce northbound 390/391 travel time by approximately 7 minutes and southbound 390/391 by approximately 10 minutes, as shown in Table 4.1Table 2.

$$r_{\mathrm{TT}} = -\frac{r_{\mathrm{V}}}{1 + r_{\mathrm{V}}} \cdot P_{\mathrm{RT}} - r_{\mathrm{WT}} \cdot P_{\mathrm{WT}} \qquad \text{(4-2-3)}$$

In summary, field operational tests have demonstrated the effectiveness of ATSP. The findings of TSP benefits are summarized as follows:

- Bus waiting time at intersections was reduced by 18.5% for northbound trips and 32.0% for southbound trips;
- Number of stops for red was reduced by 9.5% (northbound) and 16.8% (southbound);
- Average waiting time per stop for red was reduced by 10.5% (northbound) and 18.5% (northbound);
- Average bus running speed was increased by 4.9% (northbound) and 5.6% (southbound);
- All of above changes are statistically significant at the 5 percent significant level.

This evaluation study also demonstrated that TSP benefits might be greater for headway-based bus service. For schedule-based bus service, re-designing the bus schedule to take advantage of TSP would achieve more significant benefits.

4.3 ATSP Impacts on other Traffic

Essentially, TSP is to "borrow" green times from other phases/approaches in order to reduce buses' delays and numbers of stops at signalized intersections. Accordingly, the major concern for TSP systems is its impact on other traffic. Therefore in the before-and-after comparison, the green and red durations for all approaches together with the intersection delays are the MOEs to evaluate ATSP impacts on other traffic.

4.3.1 Signal Phase Durations under Semi-Actuated Signal Control

A big segment of traffic signal control systems in the U.S. are coordinated and semi-actuated. The advantage of actuated signal control systems is that the signal indications are actuated by detections at inductive loop detectors or other sensors. Thus the signal timings are adaptive to the traffic arrival patterns. Figure 4.18, Figure 4.19 and Figure 4.20 illustrate the distributions of green durations at 31st Avenue @ El Camino Real, San Mateo, CA for the morning peak, the mid-day off-peak and the afternoon peak, respectively. Phases 2 and 6, which are coordinated phases along the El Camino Real corridor, are named as major phases. While phase 4 which is along the side-street 31st Avenue together with all left-turn phases are named as minor phases. Phases 7 and 8 are not permitted for this intersection. 31st Avenue @ El Camino Real is a typical intersection along the El Camino Real corridor which has moderate traffic demands on both major and minor phases.

As can be seen, the overall green assignments among phases are similar for peak-hour and off-peak. However, the distributions of individual phases, particularly the minor phases, are different in the three time periods. For example, the medians (at the notch place) for phase 4 in the three time periods are in the middle of 10~20 seconds, slightly over 20 seconds and over 30 seconds, and the box areas between 25th and 75th percentiles are quite different. It indicates that the combinations of traffic arrival patterns on major and minor phases during these times of day (TOD) are different. Therefore, the impact of TSP on traffic should be divided into different TOD segments.

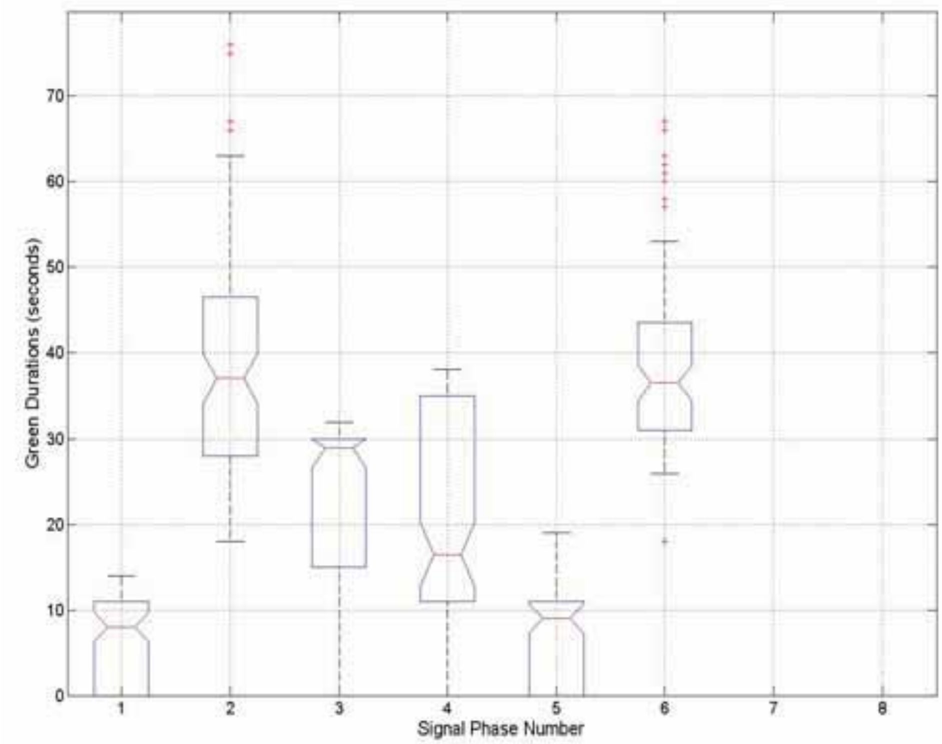


Figure 4.18 Box-plot of Green Durations for 31st Ave. @ El Camino Real, San Mateo, CA (Morning peak, 7AM~10AM)

Within each of the three defined TOD periods, green durations follow a normal-like distribution. The probability density curves for most signal phases are close to the bell shape. Tilton Ave. @ El Camino Real is another typical intersection in our testing site with moderate traffic demands. It is one of the major intesections in city of Burlingame, CA. Figure 4.21 shows the bell shaped histogram of the distribution for phase 8 durations at Tilton Ave. Figure 4.22 further plots the sample data versus the standar normal quantiles. The QQ-plot illustrates the similarity of the phase green distribution and standard normal distribution. It is believed that the traffic pattern within each of the three TOD periods is stable and thus the phase green distribution is normal-like.

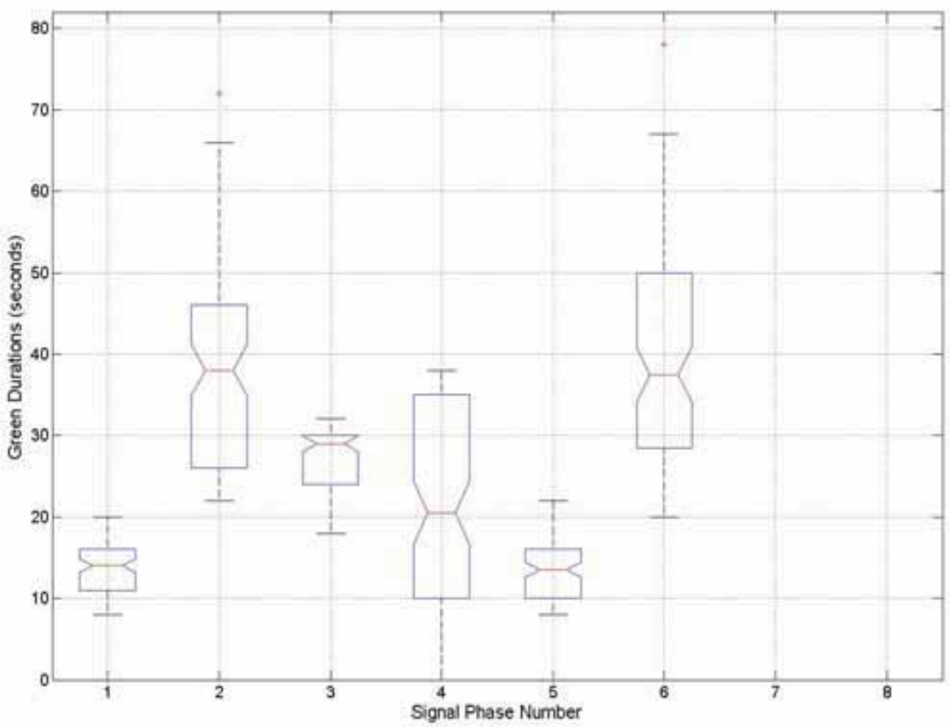


Figure 5.19 Box-plot of Green Durations for 31st Ave. @ El Camino Real, San Mateo, CA (Off-peak, 11AM~2PM)

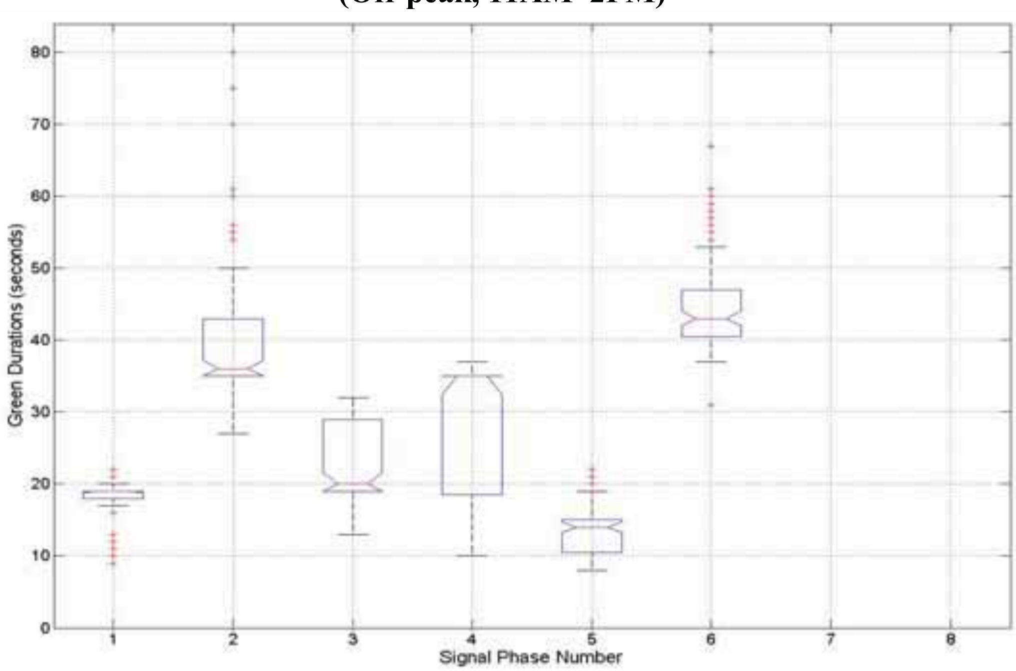


Figure 4.20 Box-plot of Green Durations for 31st Ave. @ El Camino Real, San Mateo, CA (Afternoon peak, 4PM~7PM)

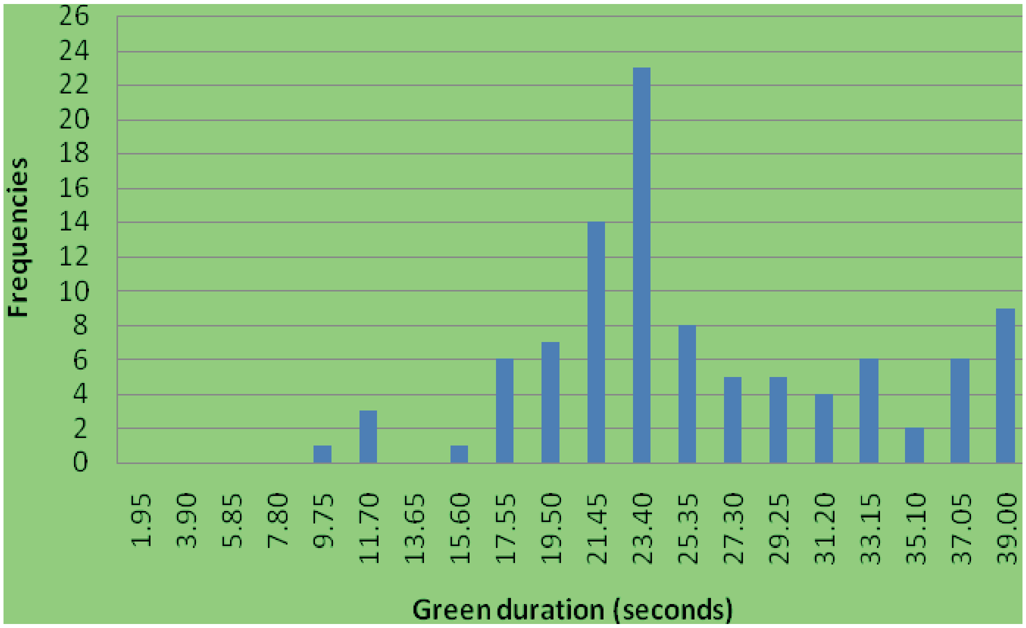


Figure 4-21 Histogram of Green Durations for Phase 8 at Tilton Ave. @ El Camino Real, Burlingame, CA (Morning peak, 7AM~10AM)

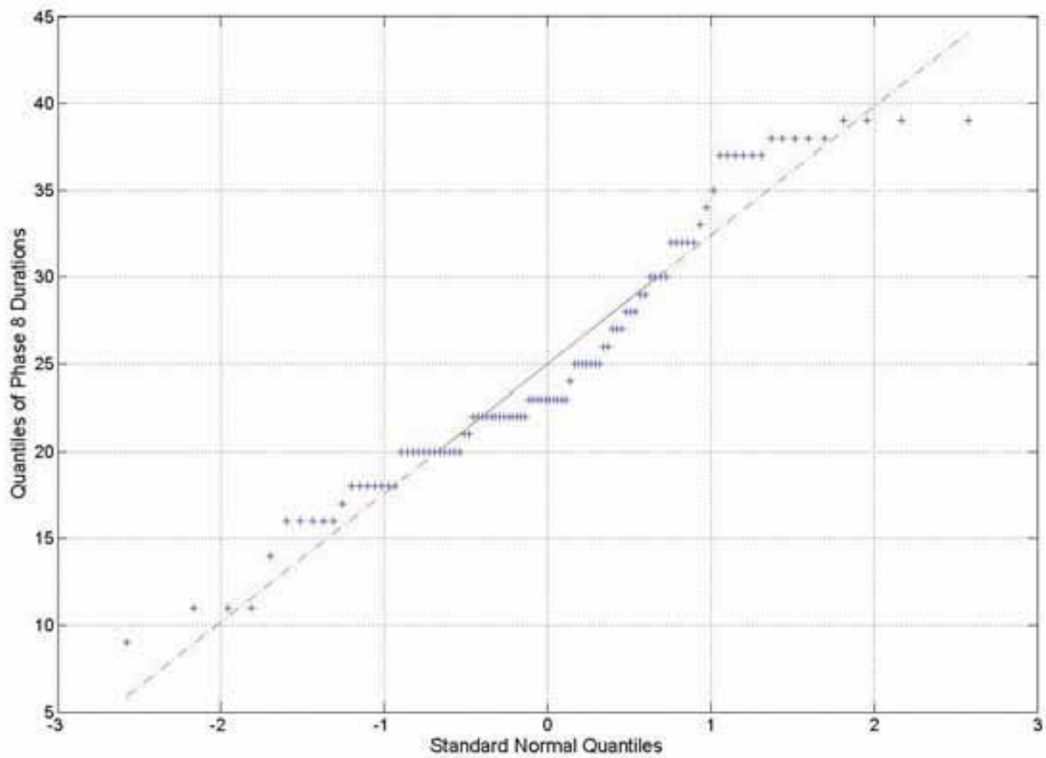


Figure 4.22 QQ-plot of Green Durations for Phase 8
Verus Standard Normal Quantiles
(Tilton Ave. @ El Camino Real, Burlingame, CA, morning peak, 7AM~10AM)

4.3.2 Field Testing Data Set

The data set applied to the traffic impact analysis is from July 15th, 2008 to September 19th, 2008. There are a total of 42 weekdays with valid TSP data; however, this does not necessarily mean the ATSP system and the data collection system were under successful operation for the whole day. Due to various system issues, e.g. the instabilities of the operating system and wireless communications, the ATSP system or the data collection system were not properly operating at some times during these weekdays. As a result, some bus trips on those weekdays did not experience the prioritized treatment at all the testing intersections. Therefore, the aforementioned benefits analysis on bus trip data only applies to a subset of the complete data set and only when the system was successfully operating for the whole day.

From the signal coordination perspective, there are three segments in the testing site. They are the Burlingame, downtown San Mateo, and southern San Mateo segments. The Burlingame segment is from Ray/Rosedale Avenue to Peninsula/Park Avenue. The downtown San Mateo segment is from Bellevue Avenue to 5th Avenue. The southern San Mateo segment is from 9th Avenue to 42nd Avenue.

Table 4.13 summarizes the executed priority treatments for the three segments. The total number of executed early green (EG) treatments is about the same for the Burlingame and downtown San Mateo segments. However, the number of executed green extension (GE) treatments is interestingly quite different. In the southern San Mateo segment, there are many more executed EG treatments. The ratios of EG/GE are similar for the downtown San Mateo and southern San Mateo segments. If the buses' arrivals are random, the arrival time would be uniformly distributed on the local signal controller clock. Buses will request GE only when its arrivals are in the GE window, which is occurs after the local cycle ends, e.g. 10 seconds or 10% in this study. The ideal ratio of executed EG and GE treatments, $\gamma_{\frac{EG}{GE}}$, can be derived by equation 5-4. Given the typical green duration ranges between $20\%C$ and $70\%C$. Then the reasonable range for $\gamma_{\frac{EG}{GE}}$ is 2 to 7. As shown in Table 5-13, $\gamma_{\frac{EG}{GE}}$ for the Burlingame segment is much smaller than the ideal lower bound. It can be concluded that buses' arrivals at the intersections in the Burlingame segment is not random. Due to the bus stop locations and/or the signal coordination, buses have a much higher chance of arriving during the GE time window than in the EG time window.

$$\gamma_{\frac{EG}{GE}} = \frac{C - G - T_{GE}^{Max}}{T_{GE}^{Max}} \qquad \textbf{(4-4)}$$

where: $\gamma_{\frac{EG}{GE}}$ is the ratio of executed EG and GE treatments;
C is the signal cycle length;
G is the green duration;
T_{GE}^{Max} is the maximum GE window.

Table 4.13 also illustrates the distribution of executed priority treatment at different times of day. Buses have obtained similar number of priority treatments in the morning and afternoon peak periods . While in the mid-day off-peak period from 11AM to 2PM, buses have experienced about double the number of priority treatments than in the peak periods. Considering the bus service frequency is the same, the difference is due to the combinations of signal coordination, passenger demand, and traffic demand.

Table 4.13 Summary of Executed Priority Treatments

	# of signals	*Priority Types*			*Time of Day*		
		Early Green	Green Extension	$\gamma_{\frac{EG}{GE}}$	Morning (7~10AM)	Mid-day (11~2PM)	Afternoon (4~7PM)
Burlingame, CA	14	571	836	0.68	247	367	210
Downtown San Mateo, CA	9	562	244	2.30	119	216	140
Southern San Mateo, CA	12	1151	542	2.12	185	466	244
Total	35	2284	1622	1.41	551	1049	594

4.3.3 ATSP Impacts on Green Durations and Traffic Delays

Among the 33 intersections with TSP capabilities, seven intersections have more than 200 executed TSP treatments during the 42 weekdays. They are 31st Avenue, 42nd Avenue, 9th Avenue, 12th Avenue, 17th Avenue, and 28th Avenue in San Mateo, CA, and Tilton Avenue in Burlingame, CA. Because most of the priority treatments were executed during the mid-day off-peak period, the following data analysis will focus on this period of time.

Figure 4.23 shows the comparison of minor phase green durations in the "before" scenario without TSP and the "after" scenario with TSP. Among the seven intersections, the maximum timing change is at 17th Ave. where the average minor phase green duration has been truncated by 3.1 seconds.

Figure 4.23 Comparison of Minor Phase Green Durations (during mid-day off-peak 11AM~2PM)

Table 4.14 further illustrates the changes of green durations for all phases at 17th Ave. As aforementioned, the green duration without the interference of TSP is assumed to follow a normal-like distribution with mean and standard deviation as shown in Table 4.14 below. Thus the statistical tests are conducted by calculating the z-values. With a significance level of 0.05, all the timing changes after executing TSP are statistically insignificant.

Table 4.14 Statistical *z*-test on Changes of Green Durations
(at 17th Ave. during mid-day off-peak 11AM~2PM)

Green Duration (seconds)	*Without TSP*		*With TSP*	*Change (seconds)*	*z*	*Significant**
	Mean	Standard deviation				
Phase 1	21.27	7.19	15.41	-5.86	-0.82	No
Phase 2	47.49	8.79	47.54	0.05	0.01	No
Phase 4	31.42	7.24	28.71	-2.71	-0.37	No
Phase 5	11.79	5.91	10.77	-1.02	-0.17	No
Phase 6	57.18	7.56	51.38	-5.80	-0.77	No
Phase 8	31.65	8.63	28.75	-2.90	-0.34	No

Note: * under the significant level of 0.05.

If the assumption of green durations to follow a normal-like distribution is not valid, we can conduct *t*-tests to evaluate whether the green durations in the "before" and "after" scenarios follow the same underlying distribution, which possesses the same mean. In the statistical test, the null hypothesis is that the average green durations in the "before" and "after" scenarios are the same. Here we apply the intersection with the second largest phase truncation, which is Tilton Ave. with 2.5 seconds reduction, as the example for the *t*-test. As shown in Table 4.15, the changes of green durations for all phases are statistically insignificant. Both of the two statistical tests conclude that the ATSP impacts on green durations are statistically insignificant.

Table 4.15 Statistical *t*-test on Changes of Green Durations
(at Tilton Ave. during mid-day off-peak 11AM~2PM)

Green Duration (seconds)	*Without TSP*		*With TSP*		*Change (seconds)*	*T*	*Significant**
	Mean	Standard deviation	Mean	Standard deviation			
Phase 2	77.79	8.03	78.67	58.10	0.88	0.18	No
Phase 6	77.79	8.03	78.67	58.10	0.88	0.18	No
Phase 8	22.22	7.35	19.76	12.95	-2.46	-1.86	No

Note: * under the significant level of 0.05.

Traffic intersection delay is an important MOE when evaluating traffic impacts. As described in a previous research report, the intersection delay for the three priority cycles can be calculated by equation 4-5. Table 4.16 demonstrates the ATSP impacts on the intersection delays at 17th Ave during the mid-day off-peak period. Because the existing actuated signal control is not optimal, the proposed ATSP system with timing optimization algorithm can further improve the existing traffic signal control. The average delay for all the three TSP control cycles is even smaller than the average delay under the existing timing. It is noted that the TSP delay calculation is based on the historical traffic demand information which might not fully represent the delay in a real situation.

$$d_T = \sum_{i=1}^{8}[\frac{\mu_i}{2}\rho_i(r_{1i}+r_{2i})^2 - r_{2i}\mu_i \min(g_{1i},\rho_i r_{1i}) + \frac{\mu_i}{2}\rho_i r_{0i}^2] \quad \textbf{(4-5)}$$

Table 4.16 ATSP Impacts on Intersection Delays (at 17th Ave. during mid-day off-peak 11AM~2PM)

		1	*2*	*4*	*5*	*6*	*8*
Traffic demand (#/hour)		210	1789	163	210	2406	81
Average red	TSP cycle 0	81.89	50.58	69.91	95.15	43.50	70.02
	TSP cycle 1	73.08	38.05	67.04	67.97	36.89	67.04
	TSP cycle 2	49.96	43.20	62.65	61.11	36.87	63.58
Vehicle delay without TSP	Total (secs)	223.14	944.20	116.78	301.28	1128.05	56.14
	Average (secs/veh)	42.50	21.11	28.66	57.39	18.75	27.72
Vehicle delay for three TSP control cycles	Total (secs)	483.91	2167.41	317.94	579.33	2749.70	153.88
	Average (secs/veh)	30.72	16.15	26.01	36.78	15.24	25.33
Change on Average delay	(sec/veh)	-11.78	-4.96	-2.65	-20.60	-3.52	-2.39
	%	-27.71%	-23.48%	-9.24%	-35.90%	-18.75%	-8.63%

The data analysis for all the intersections show that the ATSP impacts on green durations and vehicular intersection delays are statistically insignificant for most of the intersections at most of the TOD periods. The proposed ATSP system which provided significant benefits on buses' trips did not incur significant side impacts on other traffic.

5. Conclusion and Next Steps

The work reported in this report presents PATH's continuous research Adaptive Transit Signal Priority System and efforts for moving toward deployment. It summarizes the development of the key ATSP elements developed early in the project and describes the field operational tests.

Summary of development and improvement of the ATSP system

The large-scale implementation raises two major issues in system design. One is on the system architecture design, particularly the communications system. The other issue is on the system integration and the efficiency of system software design to handle the large-scale system. In this study, both of these issues have been addressed.

The ATSP system architecture has been finalized based on NTCIP 1211. Accordingly, the interface with the central AVL/ACS system at the transit fleet management center has been developed. The prior study investigated the feasibility of using existing ACS as a communications backbone for facilitating TSP request. Further investigation on the feasibility of employing limited AVL/ACS channel for El Camino Real application is conducted using an emulation platform developed under this project and actual field data. The emulation results show that ACS can indeed accommodate the ATSP communications needs.

In order to enable ATSP to work for larger size of transit fleets and number of intersections, the ATSP functional elements, such as bus arrival time prediction (ATP), PRG and PRS, were modularized into layered structure, with the upper layer dealing with the coordination and data exchanging and the lower layer handling individual buses and individual intersections. For the data analysis, we also developed a comprehensive transit and traffic database under MySQL. Software tools have been also developed to pre-process and clean the data, save the data into database, monitor bus movements and ATSP operations in real-time, and to calculate the measures of effectiveness (MOEs).

The elements of Priority Request Generator (PRG) and Priority Request Server (PRS) have been located to Traffic Management Center (TMC). It aims to effectively utilize the limited communications channels of the wildly deployed Automatic Vehicle Location (AVL) system and Advance Communications System (ACS) that have been wildly deployed by transit.

The Field Testing Results

Field testing was conducted at the central part of Routes 390 and 391, from 42nd Ave in the City of San Mateo to Rosedale Ave in Burlingame, a 6.6 miles long arterial stretch and covers 35 signalized intersections. Bus Routes 390 and 391 share the 27 northbound and 28 southbound bus-stops along the section, with 3 bus-stops as time-points on both directions. The existing traffic control systems are operated under the Caltrans C-8 Traffic Signal Control software together with Model 170E Control System hardware. The systems provide coordinated semi-actuated traffic control with multiple times-of-day plans. The control software was upgraded with the enhancement to support the ATSP operation.

After, The results from a six-week “before” and a six-week “after” field operational tests showed very positive results. Filed operational tests have demonstrated the effectiveness of ATSP. The findings of TSP benefits are summarized as follows:

- Bus waiting time at intersection was reduced by 18.5% for northbound trips and 32.0% for southbound trips;
- Number of stops for red was reduced by 9.5% (northbound) and 16.8% (southbound);
- Average waiting time per stop for red was reduced by 10.5% (northbound) and 18.5% (northbound);
- Average bus running speed was increased by 4.9% (northbound) and 5.6% (southbound);

This evaluation study also demonstrated that TSP benefits might be greater realized for headway-based bus service. For schedule-based bus service, re-design bus schedule to take advantage of TSP would achieve more significant benefits.

The major concern for TSP systems is its impacts on other traffic. The improved ATSP system optimized signal timings for two consecutive cycles based on traffic arrivals and historical signal timings. The FOT results show no statistically significant impacts on traffic intersection delays.

This project successfully addressed a number of issues that remained for eventual deployment of the ATSP system, including the system architecture for large-scale implementations, and the interface with the existing ACS as a communications backbone for facilitating TSP request. The FOT results along the 35-signal arterial demonstrate that bus service has been significantly improved while the impacts on traffic intersection delays are not statistically significant.

Additionally, the simulation tools developed under this project can also be used to evaluate the magnitude of benefits generated by ATSP.

Next Steps

We are working with Samtrans and Caltrans to expand the field testing site to include more TSP capable intersections. "Before" and "after" data collection will be conducted. The evaluation will include a systematic, comprehensive, and objective assessment of 1) technical/system performance, 2) traveler response, 3) emission reduction for buses due to the time saving obtained from TSP and estimation of emission reduction of automobiles due to mode shift because of the introduction of TSP and 4) institutional issues.

After the successful completion of the FOT, one of the immediate next steps is to assist transit agencies, which have already deployed AVL/ACS and have desires of implementing integrating ATSP with their AVL/ACS system, to assess possibilities of deployment of ATSP. The interface with a typical AVL/ACS system will be developed, so the central ATSP elements of PRG and PRS can directly communicate with the transit ACS system.

Future research can be conducted in a number of for further improvement of the deployability of ATSP, including:

- The ATSP system will build upon GPS-based bus AVL systems. GPS signals can be blocked by tall buildings, trees or overpasses, which may degrade the TSP performance at certain geographic locations. We plan to develop means to maintain the ATSP system functional when GPS signals are not reliable.
- We will explore possibility of applying Dedicated Short Range Communications System (DSRC). DSRC is a low-cost wide bandwidth communications infrastructure that is currently developed under the National and California Vehicle-to-Infrastructure Integration (VII) program and is expected to be massively deployed for automobile applications in the near term. DSRC-based ATSP system will be an alternative to the GPS-based system for no-GPS bus systems.
- The predicted bus arrival time can be used for other transit applications, such as prediction of bus arrival time at stations/stops for dynamic passenger information (DPI) and for real-time integrated transit and traffic information. Integrating ATSP with DPI and other features will support integrated application of ITS technologies.

References

AASHTO, ITE, and NEMA (2004). National Transportation Communications for ITS Protocol (NTCIP), Object Definitions for Signal Control and Prioritization

Baker, R. J., J. Collura, J. J. Dale, L. Head, B. Hemily, M. Ivanovic, J. T. Jarzab, D. McCormick, J. Obenberger, L. Smith and G. R. Stoppenhagen (2004). An Overview of Transit Signal Priority, Final version. Washington DC, ITS America.

Bar-Shalom, Yaakov, X. Rong Li, and Thiagalingam Kirubarajan, *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, 2001.

Bretherton, R.D., Current Develop in SCOOT: Version 3, Transportation Research Record 1554, National Research Council, Washington, D.C.

Dailey, D. J. and L. Li, "An Algorithm to Estimate Vehicle Speed Using Un-Calibrated Cameras", *79th Annual Meeting of the Transportation Research Board,* Washington, D. C., 9-13 January 2000.

Dailey, D. J., S. D. Maclean, F. W. Cathey, and S. R. Wall, "Transit Vehicle Arrival Prediction: An Algorithm and a Large Scale Implementation", TRB, 2001.

Duerr, P.A., Dynamic Right-of-Way for Transit Vehicles: An integrated Modeling Approach for Optimizing Signal Control on Mixed Traffic Arterials, Transportation Research Board 79th Annual Meeting, 2000.

Kalaputapu, Ravi and Michael J. Demetsky, "Modeling Schedule Deviations of Buses Using Automatic Vehicle Location Data and Artificial Neural Networks", *Transportation Research Record* 1497, pp. 44-52, 1995.

Yue Li, Peter Koonce, Meng Li, Kun Zhou, Yuwei Li, Scott Beaird, Wei-Bin Zhang, Larry Hegen, Kang Hu, Alex Skabardonis, and Z. Sonja Sun, Transit Signal Priority Research Tools, report to Federal Transit Administration, FTA-CA-26-7049-2008.1

Lin, Wei-Hua and Jian Zeng, "An Experimental Study of Real-time Bus Arrival Time Prediction with GPS Data", *78th Annual Meeting of the Transportation Research Board*, Washington, D. C., 10-14 January, 1999.

Lin, W.H. Quantifying Delay Reduction to Buses with Signal Priority Treatment in Mixed Mode Operation. *Transportation Research Record 1811,*Transportation Research Board, National Research Council, Washington, D.C. 2002

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.

Mishalani, Rabi, Sungjoon Lee, and Mark R. McCord, "Evaluation of Real-Time Bus Arrival Information Systems", *79th Annual Meeting of the Transportation Research Board,* Washington, D. C., 9-13 January 2000.

Montgomery, Douglas C., Elizabeth A. Peck, and Geoffrey Vining, *Introduction to Linear Regression Analysis*, third edition, Wiley Series in Probability and Statistics, 2001.

Newell, G.F. The Rolling Horizon Scheme of Traffic Signal Control. *Transportation Research*, Vol.32A, 39-44, 1998.

Schrank, D. and T. Lomax (2007). The 2007 Urban Mobility Report. Texas Transportation Institute, September.

Smith H. R. and B. Hemily (2005). Transit Signal Priority (TSP): A Planning and Implementation Handbook. Washington DC, ITS America, May.

TCRP Synthesis 73 (2008). AVL Systems for Bus Transit: A Synthesis of Transit Practice, Transportation Research Board, Washington DC.

Wall, Z. and D. J. Dailey, "An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location Data", *78th Annual Meeting of the Transportation Research Board*, Washington, D. C., 10-14 January, 1999.

Zhang W.-B., K. Zhou, M. Li, Y. Yin, H. Liu, C.-W. Tan, M. Miller, S. Park, A. Skabardonis, S. Sun, K. Leung, and J. Lau (2004). Development of Adaptive Transit Signal Priority Systems. California PATH Research Report, RTA65A0026, Vol. A.

Zhang, W-B, Y. Yin, K. Zhou, M. Li, S. Johnston, S. Sun, K. leung, J. Lau, and L. Cabugao (2006). Field Evaluation of VTA Rapid 522 Corridor Transit Signal Priority System. California PATH Research Report, TO4406.

Zhang W.-B., M. Li, K. Zhou, Y. Yin, and C.-W. Tan (2007). Towards Deployment of Adaptive Transit Signal Priority Systems. California PATH Research Report, TO5404.

Appendix A: Prediction Based on Historical Data and Adaptive Time-Until-Arrival Prediction Model

The bus arrival time predictor is the central element of the PRG. A major problem with existing TSP concepts and algorithms is that they often employ the same strategy to provide priority (e.g. extending the main-street green or returning early to the main-street green) regardless of when in the signal cycle the bus will arrive at the intersection. An "intelligent" or "efficient" priority scheme should involve collection of "better" information (e.g. bus location) and execution of "better" controls. In particular, the signal timing control algorithm depends critically on the accuracy of the predicted arrival time of a transit bus at the next traffic light. The predicted travel time can then be communicated to a priority request system that adaptively controls the signal timing with the transit priority service so that the transit bus can operate more efficiently, especially in the event when it is behind schedule. Our objective was to design and implement a reliable, efficient and accurate real-time travel time prediction algorithm for adaptive traffic signal control. A well-designed adaptive signal control is vital to implementing an efficient BRT, which will also help determine bus schedules that are optimal with respect to both the operational constraints and waiting times at bus stops.

In order to accurately predict the arrival time of a transit bus at a signalized intersection, real-time bus location information needs to be made available. A wide range of Advanced Vehicle Location (AVL) systems have been developed by providing vehicle position information and communicating the information to a traffic management center. For public transport, AVL applications include fleet management (real-time vehicle scheduling and control), real-time passenger information at bus stops, and signal priority. The key components of AVL systems are location, communications and central processing. Typically in an AVL system, bus location information is provided by a Global Positioning System (GPS), a network of satellites that transmits high-frequency radio signals containing time and distance data that can be picked up by a receiver, allowing the user to pinpoint their precise location anywhere around the globe. The GPS provides the absolute position of a vehicle and its corresponding coordinated universal time (UTC). Several researchers (Kalaputapu, Lin, Mishalani, Dailey, Wall, Z., Dailey) have used AVL data to develop models specifically for bus travel time prediction. Their motivation for developing these models was mostly to provide information to transit riders on expected bus arrival times with virtually no sensitivity of such models to operational control strategies, such as signal priority. These models have simple independent variables such as historical link travel times, upstream schedule deviations, and headway distributions, in addition to the current location of the bus.

The prediction of bus travel times between signalized intersections and bus stops is very challenging, since the travel times depend on a number of factors. These factors include (stochastic) traffic flow uncertainties along the route, the queue length in front of a traffic light, route length, ridership variation at bus stops (hence the uncertainties in dwell times), weather condition, time of day, statistical fluctuations in historical data (with large standard deviations), and GPS data errors. Furthermore, for signal priority application, the predicted arrival time at a traffic light must be within a *required strict level of tolerance* (e.g. within a 5-second error bound) after a priority call is executed. The work in (Dailey) presents a Kalman filter based algorithm for predicting the arrival time of a transit vehicle with a data structure that involves a combination of historical and AVL data. The prediction algorithm predicts arrival times up to an hour. It uses historical data to determine the vehicle speed over different links of the route. It does not reflect real-time speed information and will not meet the accuracy requirements for signal priority control. For the same accuracy requirement reason, the algorithms presented in

(Kalaputapu) are also not suitable for signal priority control. In their study, four algorithms are considered for predicting the arrival of the next bus in rural areas, where headways between buses are large and congestion is normally not a factor. So for their applications, none of the algorithms needs to take real-time traffic flow conditions into consideration. According to the authors, this is the first attempt in designing a bus arrival time prediction algorithm that can be incorporated with a signal priority control algorithm.

In principle, long-term and short-term arrival time predictions are fundamentally quite different. Long-term prediction typically only requires a minute-level accuracy requirement, since it involves a longer time and more uncertainties before the bus arrives at the next signalized intersection. Short-term prediction, on the other hand, requires second-level accuracy. So for short-term prediction, an accurate prediction model is needed and the prediction methodology relies more on real-time traffic data and downstream traffic conditions. A long-term prediction model will rely more on historical traffic data. In our approach to designing an algorithm for predicting the arrival time of a bus at the next traffic light, or the *time-until-arrival* prediction, we developed an algorithm that puts weights on a historical model and an *adaptive* prediction model using real-time data. The weights are adjusted adaptively according to statistical error variances generated by the observed historical data and real-time data.

A.1 Prediction using historical AVL data

Most of the prediction models discussed in the literature have included bus dwell time along any link in addition to link travel time (Dailey, Kalaputapu, Lin, Wall). In our approach, link travel time and bus dwell time are modeled separately but in a consistent single modeling framework. In this section, we construct a historical model for the historical AVL data that we have collected so far. As mentioned earlier, such a historical model predicts an "average" travel time between two nodes (bus stops or signalized intersections). The bus trajectories are separated into drive and stop sections, and we construct a stochastic model for each of them. A ***drive section*** is defined as a *continuous* section of a bus travel timeline when the bus moves at *non-zero speed*. Typically, this is the time when the bus travels along a link connecting two neighboring nodes. Since a bus may stop because of an incident, in general, this is a time section when the bus starts from zero speed and stops again. A ***stop section*** of a bus travel timeline refers to the time when the bus stops at *zero speed* at a bus stop or in front of a traffic light. For a drive section, we have constructed several stochastic functions for travel distance, travel time and average speed, and concluded that the simple first-order average speed model performs reasonably well, especially for short and medium range travel time predictions. In what follows, we first construct historical models for travel time as a function of traveled distance, waiting times at traffic lights and dwell times at bus stops. We then present an algorithm to predict the time-until-arrival using the historical AVL data.

A.1.1 A Linear model for historical AVL data

The central problem in constructing a prediction model is to predict the ***time-until-arrival*** or *time-to-go,* t_g. This is defined as the time that it will take from the current bus location to the next node, which is either a bus stop or a signalized intersection. The assumption is that a bus stops at a bus stop or an intersection *with probability one*. So this is the time until a bus reaches the next stop section of its travel timeline. It is possible that the bus stops due to traffic congestion or other incidents. In that case, the prediction algorithm restarts when the bus moves again, using the GPS coordinate of the location where the bus last stopped as its starting point. If the bus stops at a bus stop, the algorithm also restarts when the bus moves again. For a signal priority control application, the objective is to predict the time t_g until the bus reaches the next intersection.

A typical space-time diagram is shown in Figure A.1, where a drive section with travel time T_D seconds covers a distance of length *D* meters. Note that the pair (*D*, T_D) is for one drive section, with the understanding that the bus starts from zero speed at the beginning of this drive section and stops at the end of the section. So T_D is the travel time from "zero speed to zero speed". The variable t_{dw} is the dwell time at a bus stop or the waiting time in front of a traffic light, depending on the type of the node. So t_{dw} represents a stop section of the bus travel timeline. If the bus has already travelled a distance *d* from the previous node, the time to travel the remaining distance, *D* – *d*, until it reaches the next node is the time-until-arrival or time-to-go, t_g. Our historical model predicts this t_g based solely on the historical AVL data. Note that the distance *d* is the straight-line distance between the current bus location and the location of the start node.

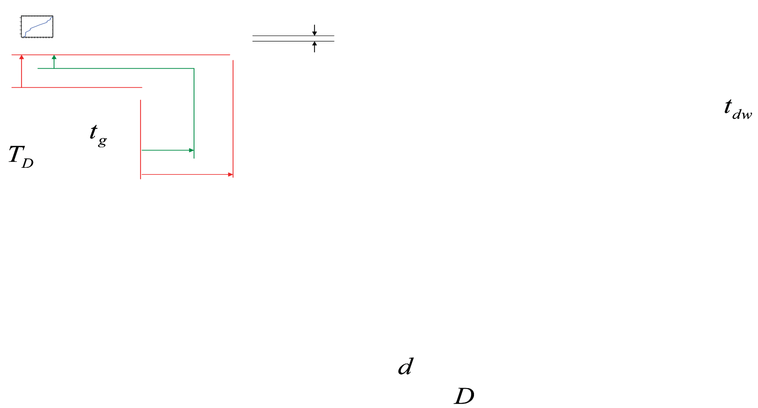


Figure A.1 A typical space-time diagram for a transit bus travelling on the test route

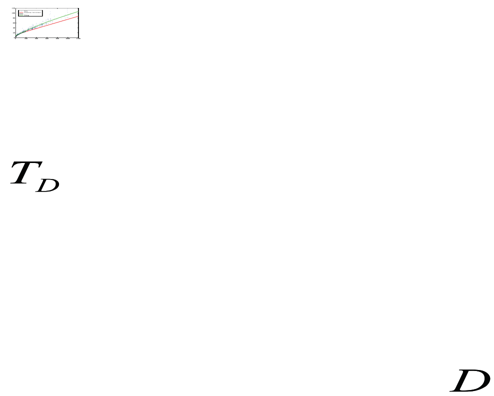


Figure A.2 Drive section lengths and corresponding travel times for 7-9 am, and a linear regression model for their relationship

The first issue we investigate is whether there is a relationship between *D* and T_D. That is, if the bus starts moving (from zero speed) now and stops at the next node after traveling *D* meters, what is a "good" estimate of the travel time T_D using historical data? Is there an expression for T_D as a function of *D* that "best fits" our data sets? We assume that the driving conditions (e.g. speed limit) are *homogeneous* for the entire test route. (If this does not hold, we can divide the route into homogeneous sections and develop a historical model for each section.) So the relationship $D \mapsto T_D$, if it exists, holds for all the links along the test route. A program was written to extract all the drive sections in our data sets, and then obtain the length *D* of each of these drive sections and their corresponding travel times T_D. The set of observation data (*D*, T_D) of all the drive sections for 7-9 am weekdays is extracted from our data sets and plotted in Figure A.2. We should think of *D* and T_D as two random variables, so the observations or *sample points* in Figure A.2 form a joint distribution for the pair (*D*, T_D). This distribution changes as more data for this specific TOD become available. Regression analysis is a statistical technique for investigating and modeling the relationship between variables. The goal of regression analysis is to determine the values of parameters for a function that cause the function to best fit a set of observed data. In *linear regression* (see Montgomery), the function is a linear (straight-line) equation that best predicts T_D from D. We have tried several regression techniques using the simulation tools available in MATLAB and noted that a simple first-order *linear* average speed model is a good fit for the historical data available for the different TODs. This linear relationship between *D* and T_D is of the form:

$$\hat{T}_D = \alpha \hat{D} + \beta \tag{A-1}$$

Here the "hat" indicates that it is a "best fit" estimate using linear regression. This is the green solid line in Figure A.2. Very often, the variables *D* and T_D are referred to as the *regressor* (or predictor) and *response* variables, respectively. The parameters α and β are chosen such that the *regression error*

$\left\| T_D - \hat{T}_D \right\|^2$ over the region of the regressor variable *D* in the observed data, where (*D*, T_D) is an observation, is minimized.

It is obvious that the sample points do not fall exactly on a straight line, so the regression equation (A-1) is only an approximation of the true functional relationship between the variables *D* and T_D. For a given drive section length *D*, the travel time T_D has a certain conditional distribution that depends on *D*. This distribution can be modeled as a *statistical error* so that for each *D*, we can model the sample points (*D*, T_D) as a linear regression with a distance-dependent error. That is:

$$T_D = \alpha D + \beta + w_{TD} \qquad \text{(A-2)}$$

The term β is a bias and w_{TD} is an error modeled as a *Gaussian* distribution with zero mean and variance $\sigma_{TD}^2(D)$ that depends on the length *D*. For simplicity, we also use the notation $w_{TD} \sim N(0, \sigma_{TD}^2(D))$. This distance-dependent variance is determined by the distribution of the observed data at $D = \hat{D}$, and it varies as observations at different times of day are considered and when more data become available.

To gain additional insight into the linear regression model, consider a drive section length *D* and recall that a drive section is typically a time interval when a bus travels along a link connecting two neighboring nodes. The accuracy of the length *D* is affected by the uncertainty in the location of a bus stop or signalized intersection. This uncertainty is due to the uncertainty in the GPS data. We assume the GPS measurement error has a Gaussian distribution with zero mean and variance σ_d^2, where σ_d = 15 meters is specified by the manufacturer. So each drive section length *D* has an uncertainty that can be modeled approximately as $N(0, \sigma_d^2)$. Then the *conditional expectation* of the travel time for a given length $D = \hat{D}$ is:

$$E(T_D \mid D = \hat{D}) = \alpha \hat{D} + \beta \qquad \text{(A-3)}$$

If we assume the uncertainty in *D* and the error process w_{TD} are independent, then the variance of T_D for the given $D = \hat{D}$ is:

$$\sigma_T^2(\hat{D}) = Var(T_D \mid D = \hat{D}) \approx \alpha^2 \sigma_d^2 + \sigma_{TD}^2(\hat{D}) \qquad \text{(A-4)}$$

From (A.3) the linear regression model, $\hat{T}_D = \alpha \hat{D} + \beta$, is a line of mean values; that is, the coordinate on the regression straight line (A.1) at any value of *D* is simply the expected value of T_D for that given *D*. The slope α can be interpreted as the change in the conditional expectation of T_D for a unit change in *D*. The variance of T_D is determined by the GPS data accuracy and the variance of the error component of the model (A-2). If the GPS measurement error is smaller, σ_d will also be smaller, thus implying a more accurate estimate of T_D. The linear relationship (A.1) generated by the "Linear-in-the-Parameters Regression" module in MATLAB is the green solid line shown in Figure A.2. The program also calculates the distance-dependent variance $\sigma_{TD}^2(D)$. For example, for northbound traffic between 7-9 am, this linear relationship is given by:

North-bound, 7-9am: $\hat{T}_D = 0.109\hat{D} + 8.0177$

North-bound, all day: $\hat{T}_D = 0.1072\hat{D} + 8.1614$

Compared to (A.1), a second-order regression model has the form $\hat{T}_D = \alpha_2 \hat{D}^2 + \alpha_1 \hat{D} + \beta$. For our historical AVL data, α_2 is typically of the order 10^{-4}. Also the regression error for the second-order model is larger than that for the first-order model. So the simple first-order model performs reasonably well for our observed data.

A.1.2 Prediction of Time-Until-Arrival using historical data

Next we develop a model for predicting the time-until-arrival using the historical data. The space-time diagram in Figure A.1 reveals approximately the "average" bus traveling speed between nodes. The historical AVL data are therefore only good for predicting an average travel time that a bus will take until it arrives at the next traffic light. The empirical data also suggest that in each drive section of a bus travel timeline, the bus average speed can be modeled as a constant with uncertainty. This will be our historical model for predicting the time-until-arrival. Let $D = \hat{D}$ be the length of a drive section, the predicted travel time for this section is $\hat{T}_D$, where $(\hat{D}, \hat{T}_D)$ satisfies the linear relationship (A.1). The average speed along this section is modeled as a constant plus a Gaussian error term:

$$v = \frac{\hat{D}}{\hat{T}_D} + w_v \quad , \quad w_v \sim N(0, \sigma_v^2) \qquad \text{(A-5)}$$

The error term w_v represents the variations in the constant average speed $\frac{\hat{D}}{\hat{T}_D}$. If we model the variations as first-order perturbations, then by taking derivatives, we can approximate variations in $v = \frac{\hat{D}}{\hat{T}_D}$ as:

$$\delta v = \frac{1}{\hat{T}_D}\delta D - \frac{\hat{D}}{\hat{T}_D^2}\delta T_D \quad \text{, at operating point } (D, T_D) = (\hat{D}, \hat{T}_D) \qquad \text{(A-6)}$$

As mentioned above, the variation in the drive section length, δD, is due to uncertainty in the GPS location which has a Gaussian distribution $N(0, \sigma_d^2)$. The error in the travel time, δT_D, has a distribution with variance σ_T^2 as given in (A-4). The "quality" of an estimate is usually measured by its variance. If we assume the uncertainty in *D* and the travel time T_D are independent, then by using (A-6) the variance of v for the given $D = \hat{D}$ is:

$$\sigma_v^2 = Var(v \mid D = \hat{D}) \approx \left(\frac{1}{\hat{T}_D}\right)^2 \sigma_d^2 + \left(\frac{\hat{D}}{\hat{T}_D^2}\right)^2 \sigma_T^2 \qquad \text{(A-7)}$$

where σ_d = 15m and σ_T^2 is given by (A-4). The *conditional expectation* of the average speed given that $D = \hat{D}$ is:

$$\hat{v}(\hat{D}) = E(v \mid D = \hat{D}) = \frac{\hat{D}}{\hat{T}_D} \tag{A-8}$$

A historical model for predicting the time-until-arrival (TUA) can now be constructed. We will denote it by t_{gH} to distinguish it from another TUA predicted by a real-time adaptive model that will be developed in the next section. Referring to Figure A.1, if the drive section length is $D = \hat{D}$ and the bus has already traveled a distance *d* from the beginning of this section, the time to travel the remaining distance, $\hat{D} - d$, until it reaches the end of the section (our interest is when the end node is an intersection) is the time-until-arrival or time-to-go. Since the bus speed along a drive section is modeled as a constant average speed with uncertainty, the TUA is thus modeled as the time to complete the remaining distance at the constant average speed $\hat{v}$ plus an error term. That is:

$$t_{gH}(d) = \frac{\hat{D} - d}{\hat{v}} + w_{tgH} \quad , \quad w_{tgH} \sim N(0, \sigma_{tgH}^2(d)) \tag{A-9}$$

By using the same first-order perturbation analysis for obtaining the variance σ_v^2 in (A-7), we can approximate variations in t_{gH} as:

$$\delta t_{gH} = \frac{1}{\hat{v}}\delta D - \frac{1}{\hat{v}}\delta d - \frac{\hat{D} - d}{\hat{v}^2}\delta v \tag{A-10}$$

The variance of t_{gH} for the given $D = \hat{D}$ is thus given by:

$$\sigma_{tgH}^2(d) = Var(t_{gH} \mid D = \hat{D}) \approx \frac{2}{\hat{v}^2}\sigma_d^2 + \left(\frac{\hat{D} - d}{\hat{v}^2}\right)^2 \sigma_v^2 \tag{A-11}$$

where σ_d = 15m and σ_v^2 is given by (A-7). The TUA predicted by the historical model is the *conditional expectation* of t_{gH} given that $D = \hat{D}$. This is given by:

$$\hat{t}_{gH}(d) = E(t_{gH} \mid D = \hat{D}) = \frac{\hat{D} - d}{\hat{v}} = \hat{T}_D\left(1 - \frac{d}{\hat{D}}\right) \tag{A-12}$$

This is expected since we use a *constant* average speed model. The most important issue in the development of the historical model is to calculate the variations in the observed data. These are expressed in the variances σ_T^2, σ_v^2 and σ_{tgH}^2 obtained in (A-4), (A-7) and (A-11), respectively. The variance σ^2_{tgH} will be used in a weighted combination of the prediction algorithm to be presented in Section A.2 of this document.

A historical dwell time model is developed considering the dwell time distribution at each bus stop. To model the dwell time t_{dw}, the dwell times at all the northbound and southbound bus stops are extracted from the observed data. For each of the bus stops, we calculate the mean $\bar{t}_{dw}$ and the variance σ^2_{tdw} of the observed dwell times. These two parameters are regularly updated as more AVL data are made available. As suggested by the *law of large numbers*, the accuracies of the mean and variance will improve and converge as more data become available.

A.2 An Adaptive Time-Until-Arrival Prediction Model

The historical model predicts the TUA in one calculation by using a *constant* speed given in (A-8) for the entire drive section. In this model, the traveled distance *d* increases linearly with time, and is clearly insufficient to capture any real-time traffic flow and speed variations. In this section, we develop another prediction model to complement the historical TUA prediction model. This is an *adaptive* model that uses real-time AVL data to adaptively estimate the *downstream* bus average speed. To correct for the GPS location error due to unknown GPS latency and CDPD modem transmission delay and data packet losses, a Kalman filter that uses bus wheel speed is implemented to obtain a more accurate estimate of the current bus location, and thus, the traveled distance from the start node. Suppose the bus has traveled a distance of $d(t)$ from the start node in *t* seconds. The average speed is approximately $d(t)/t$, subject to statistical measurement errors. The approach we propose is to use an average speed model to predict the TUA. That is, if it takes an average speed $a(t)$ to cover a distance of $d(t)$ in t sec., then the time to cover the remaining distance, predicted at the current time t, is $\hat{D} - d(t)/a(t)$. In contrast to using a fixed average speed in the historical model, the average speed $a(t)$ is adaptively tuned to the speed variations downstream as the bus travels. In this adaptive average speed model, we express the traveled distance as:

$$d(t) = a(t)t + b(t) + w_d(t) = H(t)x(t) + w_d(t) \quad \text{(A-13)}$$

Here $a(t)$ is the average speed at time t, $b(t)$ is the distance residual, $H(t) = [t \;\; 1]$, and $x(t) = [a(t) \;\; b(t)]^T$ is the state. The traveled distance is calculated using the current (real-time) bus GPS location, which has an error standard deviation of 15 m. So we assume $w_d(t)$ is an independent identically distributed (i.i.d.) measurement noise process with zero mean and variance σ_d^2, where σ_d = 15 m. The observation $d(t)$ is updated every Δt = 1 second, so $d(k)$ is the kth observation at time $t = k\Delta t$. For real-time implementation, the discrete-time adaptive model is:

$$d(k) = a(k)k\Delta t + b(k) + w_d(k) = H(k)x(k) + w_d(k)\;,\;\; k = 1,2,3,\cdots \quad \text{(A-14)}$$

We formulate it as a *linear least squares* (LS) estimation problem, so it is desirable to estimate the state $x(k)$ from the observation $d(k)$ such that the quadratic error defined below is minimized.

$$\text{LS estimator } \hat{x}(k) = ArgMin\sum_{j=1}^{k} | d(j) - H(j)x |^2 \quad \text{(A-15)}$$

The LS estimator that minimizes (A-15) is obtained by setting its gradient with respect to x to zero and has a *recursive* form (Bar-Shalom) given by:

$$\hat{x}(k) = \hat{x}(k-1) + W(k)[d(k) - H(k)\hat{x}(k-1)] \tag{A-16}$$

In standard LS estimation, the recursive update estimate $\hat{x}(k)$ equals the previous estimate plus a correction term. The correction term consists of a gain $W(k)$ multiplied by the residual, which is the difference between the observation $d(k)$ and the predicted value of this observation from the previous *k-1* measurements. The filter gain $W(k)$ is:

$$W(k) = P(k-1)H(k)^T S(k)^{-1} \tag{A-17}$$

Here $S(k)$ is the covariance of the residual and $P(k)$ is the covariance of the LS estimator $\hat{x}(k)$. They can be recursively expressed as:

$$S(k) = H(k)P(k-1)H(k)^T + \sigma_d^2 \tag{A-18a}$$

$$P(k) = P(k-1) - W(k)S(k)W(k)^T \tag{A-18b}$$

Next we develop a model for predicting the TUA using the adaptive model. Since the optimal state is $\hat{x}(k)$ = [$\hat{a}(k)$ $\hat{b}(k)$] and $\hat{a}(k)$ is the average speed at time $k\Delta t$, we model the TUA as the time to complete the remaining distance at the average speed $\hat{a}(k)$ plus an error term. That is:

$$t_{gA}(d(k)) = \frac{\hat{D} - d(k)}{\hat{a}(k)} + w_{tgA} \quad , \quad w_{tgA} \sim N(0, \sigma_{tgA}^2) \tag{A-19}$$

By using a similar first-order perturbation technique to obtain σ_{tgH}^2 for the historical model, we can approximate variations in t_{gA} as:

$$\delta t_{gA} = \frac{1}{\hat{a}}\delta D - \frac{1}{\hat{a}}\delta d - \left(\frac{\hat{D} - d}{\hat{a}^2}\right)\delta a \tag{A-20}$$

The variance of t_{gA} for the given $D = \hat{D}$ is thus given by:

$$\sigma_{tgA}^2(d(k)) = Var(t_{gA} \mid D = \hat{D}) \approx \frac{2}{\hat{a}(k)^2}\sigma_d^2 + \left(\frac{\hat{D} - d(k)}{\hat{a}(k)^2}\right)^2 P(1,1) \tag{A-21}$$

where σ_d = 15 m. The variance of $\hat{a}(k)$ is $P(1,1)$, the (1,1)-entry of the covariance matrix $P(k)$. The TUA predicted by this adaptive model is the conditional expectation of t_{gA} given that $D = \hat{D}$. From (A-19), it is:

$$\hat{t}_{gA}(d(k)) = E(t_{gA}(d(k)) \mid D = \hat{D}) = \frac{\hat{D} - d(k)}{\hat{a}(k)} \tag{A-22}$$

Considering t_{gA} as an observation of t_{gH}, which is assumed to be a Gaussian random variable with the mean and variance given by (A-11) and (A-12), respectively, i.e.,

$$t_{gA} = t_{gH} + w_t, \quad w_t \sim (0, \sigma_{tgA}) \tag{A-23}$$

the maximum a *posteriori* (MAP) estimate of (A-23) is given by

$$\hat{t}_g = w\hat{t}_{gH} + (1-w)\hat{t}_{gA} \tag{A-24}$$

with

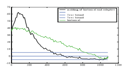
$$w_H = \frac{\sigma_{tgA}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \tag{A-25}$$

The corresponding error covariance of (A-24) is given by

$$\sigma_{tg}^2 = \frac{\sigma_{tgA}^2 \sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \tag{A-26}$$

A.3 Simulation Results

The performance of the time-until-arrival (TUA) prediction algorithm is examined by means of simulation. Simulations are compared with the *actual* TUA calculated from the empirical data. Figure A.3 shows a typical long-distance run between two nodes for northbound traffic during 7-9 am. The section length is *D* = 1,083.1 meters and the *actual* travel time is T_D = 89 seconds. The difference between the predicted TUA and the actual TUA, $\Delta\hat{t}_g = \hat{t}_g - t_g$, is the solid curve. The difference between the TUA predicted by the historical model and the actual TUA, $\Delta\hat{t}_{gH} = \hat{t}_{gH} - t_g$, is the dashed curve. We note that $\Delta\hat{t}_{gH}$ is within a bound of ± 5 seconds when the bus is about 200m from the location of the end node. In terms of travel time, this is about 25 seconds from arriving at the end node. The respective cutoffs for $\Delta\hat{t}_g$ are about 650m and 50 seconds. The predicted TUA thus converges to the actual TUA much faster than that predicted by using the historical model alone. So in the weighted combination of the historical and adaptive models, the adaptive model improves the convergence rate by including real-time AVL and speed information in the prediction. We also note that the prediction algorithm tends to be less reliable initially, with a significantly large error $\Delta\hat{t}_g$ ~ 60 seconds near the beginning of the drive section. This is because the bus starts with zero speed, so initially $\hat{a}(t)$ is small and $\hat{t}_{gA}$ tends to be large. In general, a drawback of the least-squares algorithm is that the initial phase of convergence is not monotonic, so the parameter update is initially "not good". During the initial phase the estimator may be able to keep the error small but the model is not reliable as a predictor. This initial inaccuracy is compensated by the historical TUA in our prediction algorithm. So in a way the historical TUA provides a fairly good initial estimate of the average flow condition. Figure A.4 shows a medium-distance run between two nodes for the same TOD, with *D* = 512.67 m and actual T_D = 47 seconds. Again the convergence is faster with the inclusion of the adaptive model.



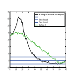
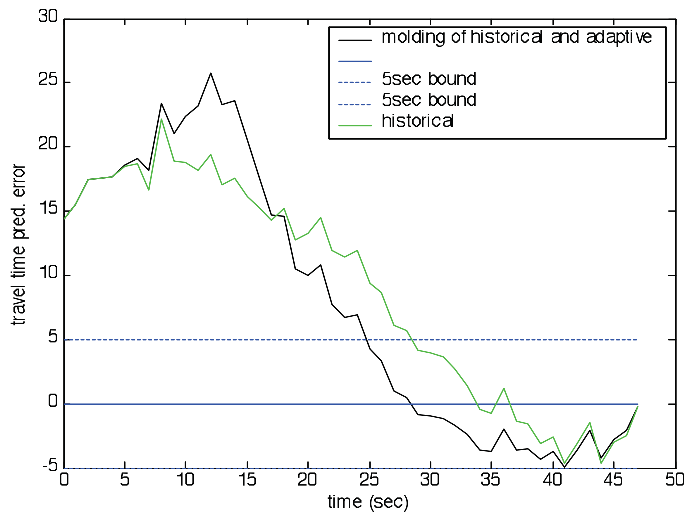


Figure A.3 Simulations comparing the prediction errors using the weighed model and historical model for a long-distance run

Next we obtain some statistics of the prediction error $\Delta\hat{t}_g$ for two adjacent intersections with no bus stop between them. We consider the Barneson and Hobart Avenues (cross street NSI08 and NSI09 in Table 4.1) in the northbound direction. The distance between these two nodes is 257 m. The red solid line in Figure A.5 shows the mean prediction error for the corresponding time-until-arrival. As the TUA gets smaller and the bus is closer to the end node, the error converges to zero. The plot also shows the standard deviation on both sides of the mean. Figure A.6 is a 3-D plot of the prediction error histogram. Note that there is a large peak around the coordinate (TUA = 0 sec, $\Delta\hat{t}_g$ = 0 sec), indicating that the solution almost surely converges and the prediction algorithm works well.



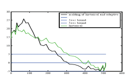


Figure A.4 Simulations comparing the prediction errors using the weighed model and historical model for a medium-distance run

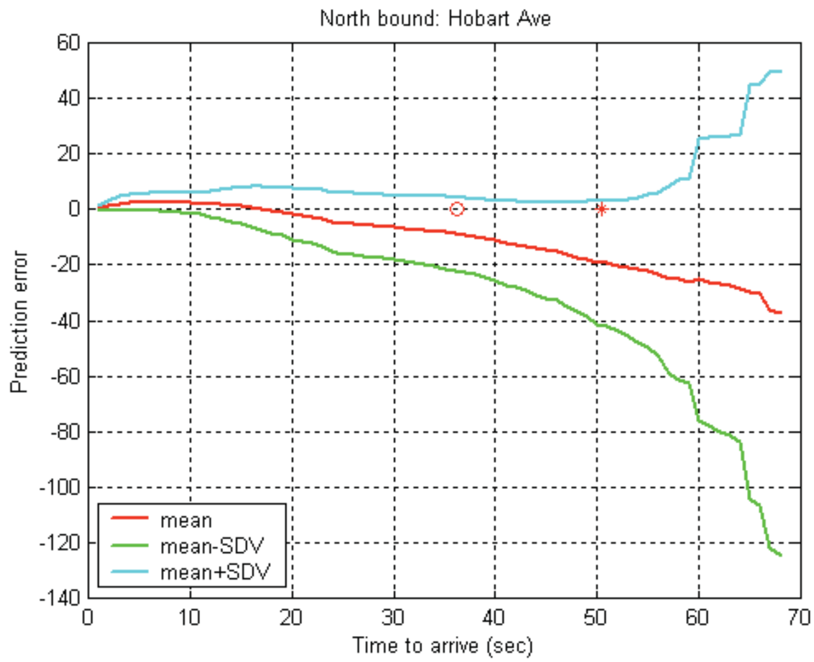


Figure A.5 Mean and standard deviations of the prediction error for transit vehicles approaching the Hobart Avenue in the north-bound direction

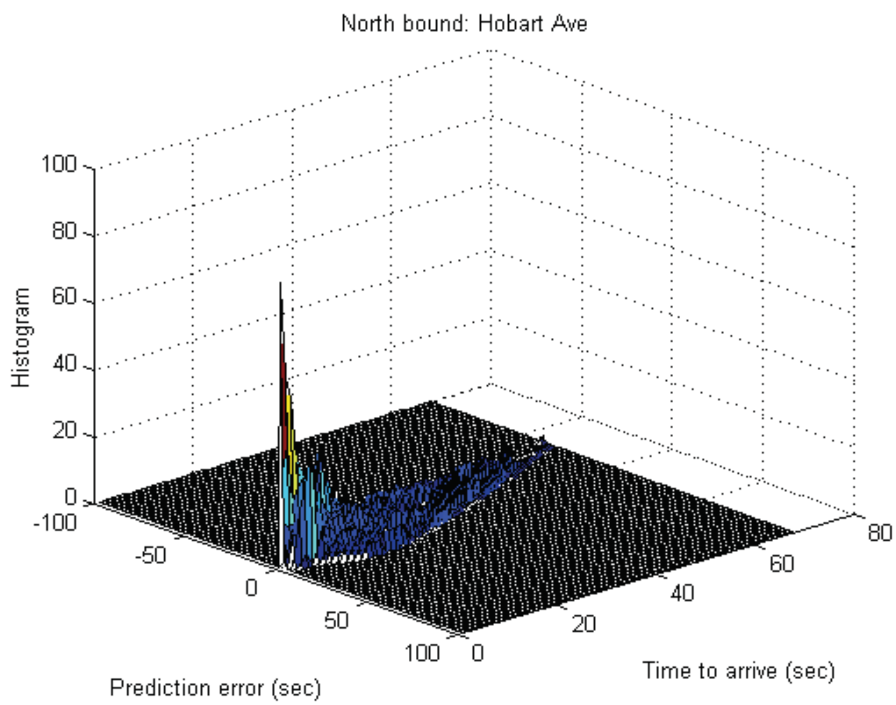
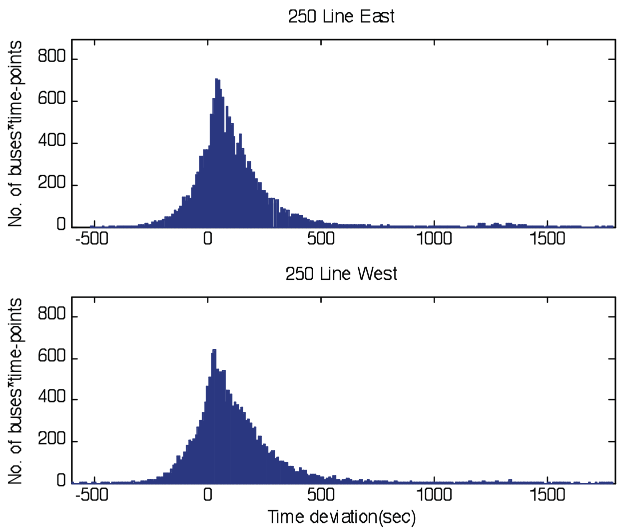
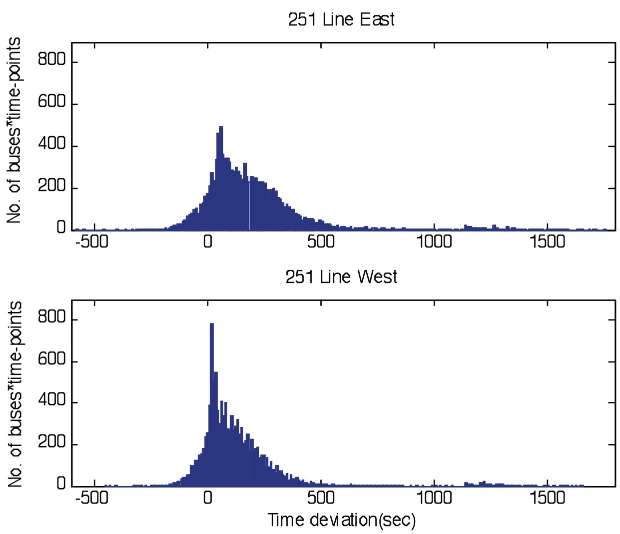
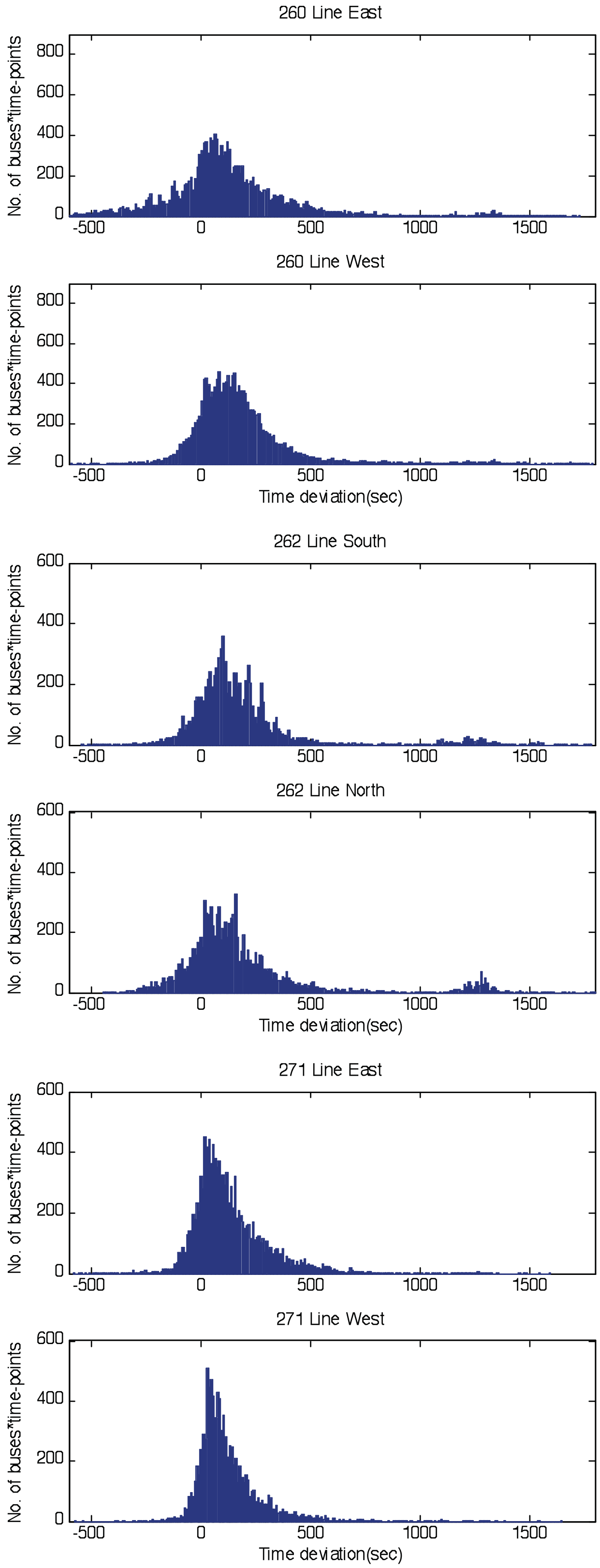


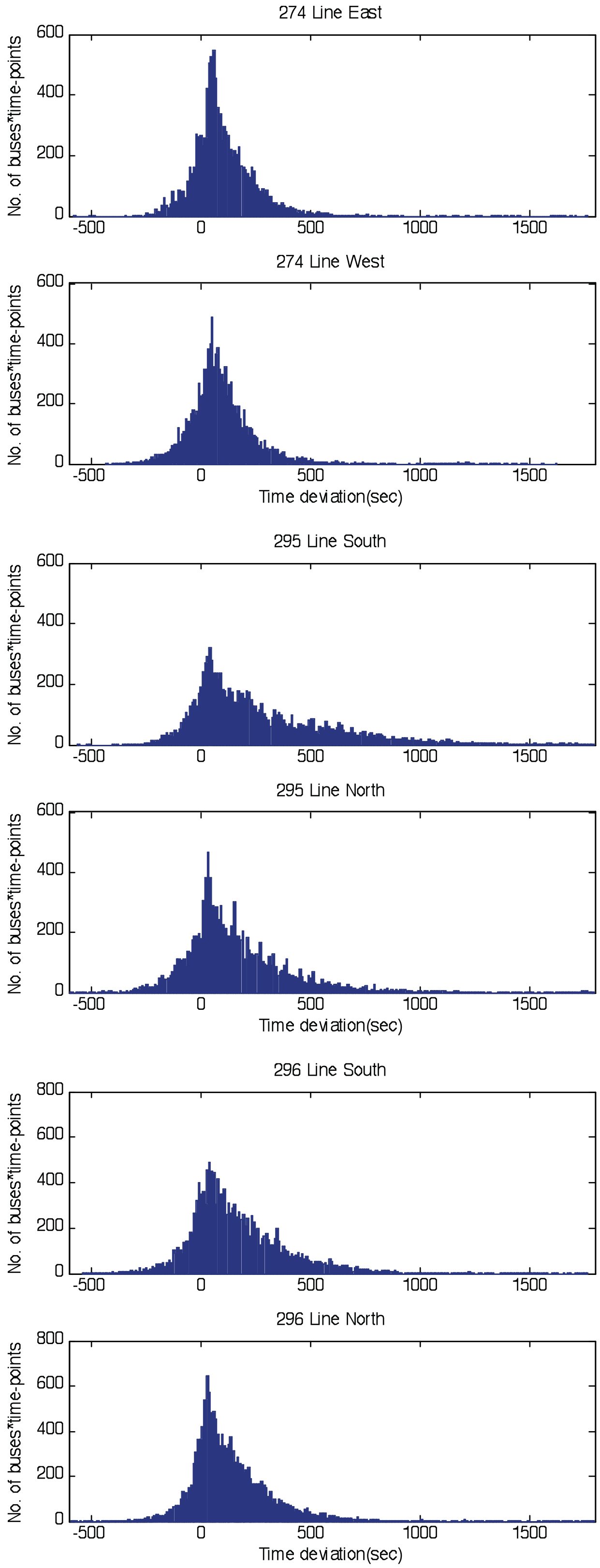
Figure A.6 Prediction error histogram for transit vehicles approaching the Hobart Avenue in the north-bound direction

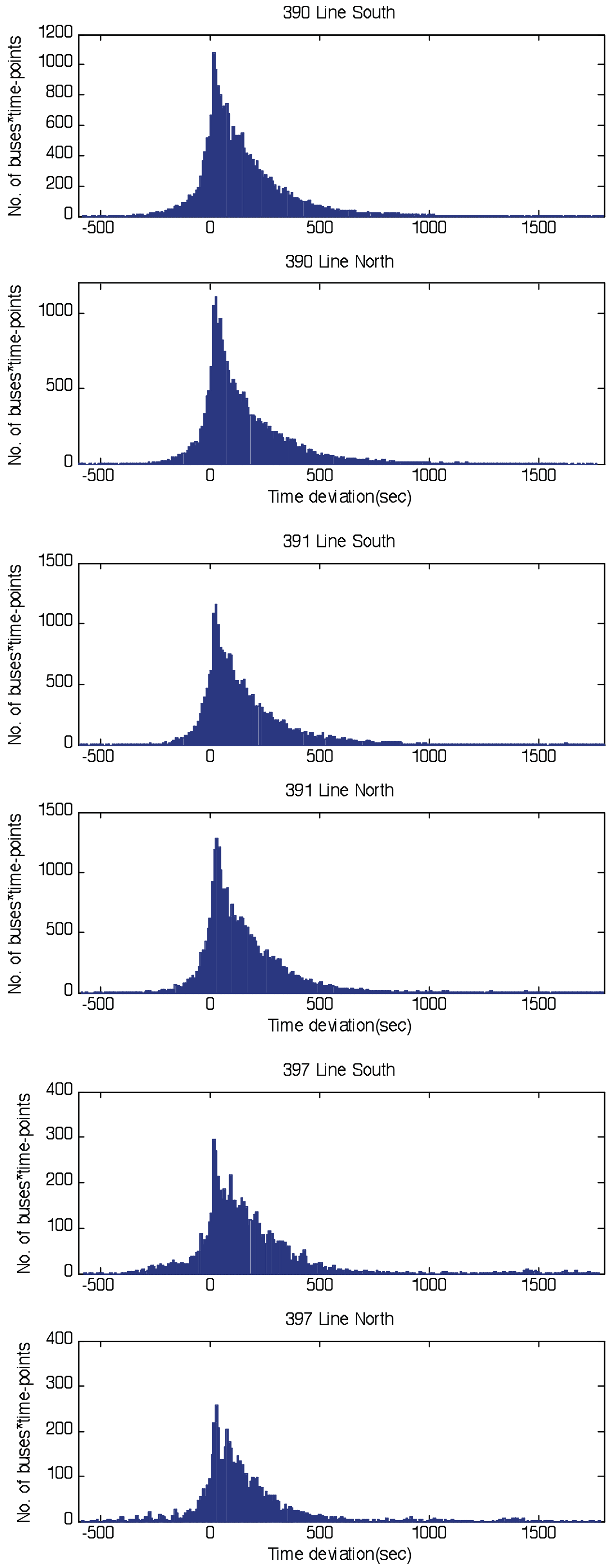
Appendix B: Arrival Time Deviation Distribution of 15 Lines

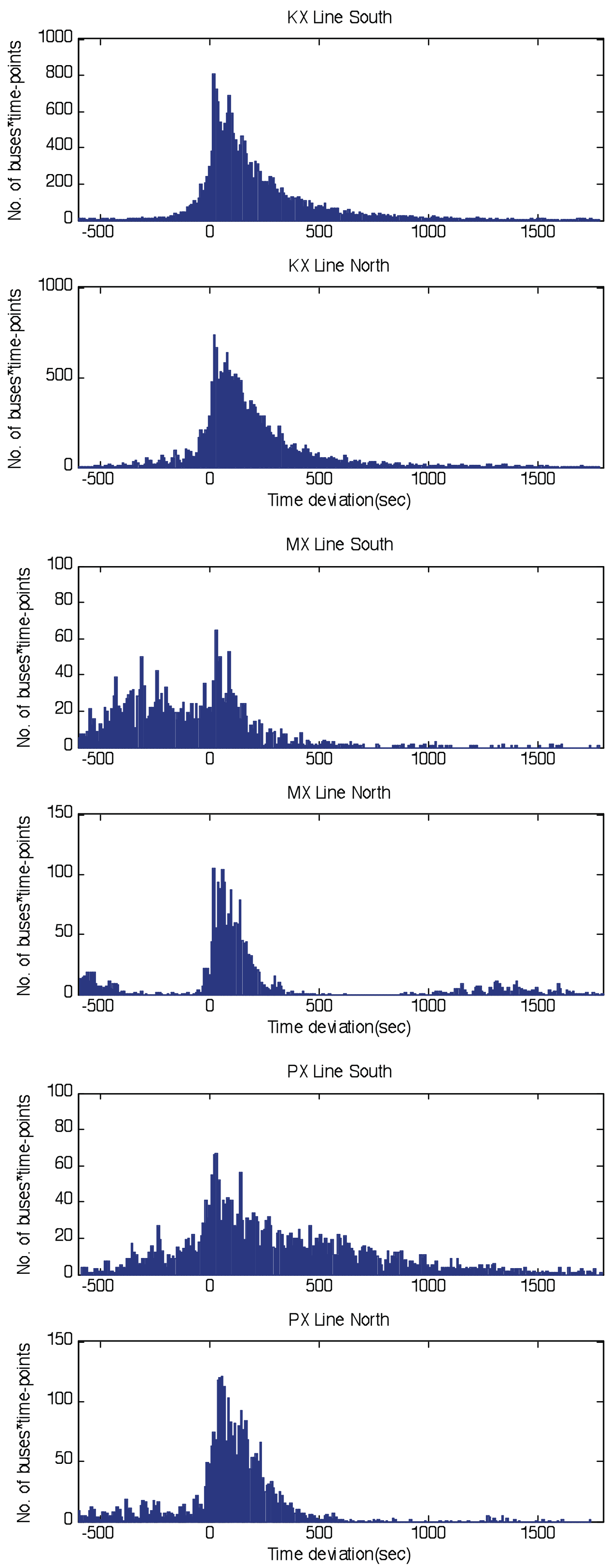


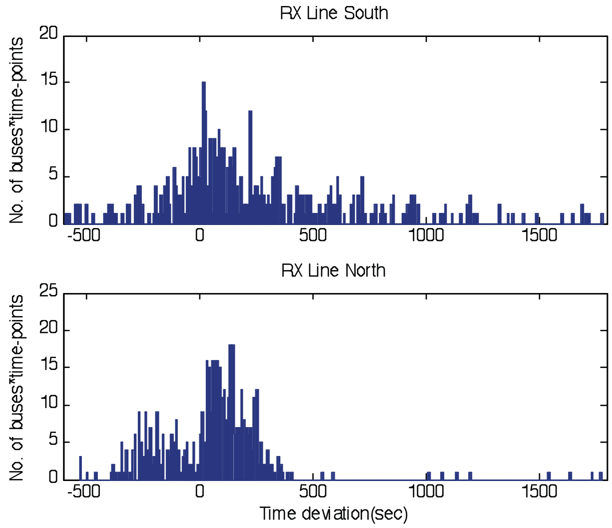



260 Line East
260 Line West
262 Line South
262 Line North
271 Line East
271 Line West
No. of buses*time-points
Time deviation(sec)
-500
0
500
1000
1500
0
200
400
600
800


274 Line East
274 Line West
295 Line South
295 Line North
296 Line South
296 Line North
No. of buses*time-points
Time deviation(sec)
-500
0
500
1000
1500
0
200
400
600
800


390 Line South
390 Line North
391 Line South
391 Line North
397 Line South
397 Line North
No. of buses*time-points
Time deviation(sec)


KX Line South
KX Line North
MX Line South
MX Line North
PX Line South
PX Line North
No. of buses*time-points
Time deviation(sec)
-500
0
500
1000
1500


RX Line South
No. of buses*time-points
20
15
10
5
0
-500
0
500
1000
1500
RX Line North
No. of buses*time-points
25
20
15
10
5
0
-500
0
500
1000
1500
Time deviation(sec)