# An operational strategy for Advanced Vehicle Location system based transit signal priority 

Hongchao Liu ${ }^{1}$, A.M. ASCE, Wei-Hua Lin ${ }^{2}$, and Chin-woo Tan ${ }^{3}$


#### Abstract

Future deployments of Transit Signal Priority (TSP) in the United States depend largely on improving TSP strategies to better accommodate transit vehicles while at the same time minimizing the negative impacts on the vehicles of the non-prioritized approaches. Advanced Vehicle Location (AVL) technology holds great potential in this regard. This paper develops a methodology that incorporates the predicted bus arrival time information into an AVL based TSP system to improve its performance. It is demonstrated analytically that the time to trigger the traffic signal for priority operation, especially Early Green, is of particular importance to both transit and passenger vehicles. A theoretical model is developed to identify the optimal time to place a priority call based on the predicted bus arrival time information. A simulation analysis is conducted to verify the theoretical approach and further identify the optimal calltime points for general cases. The research is focused on operation of TSP under moderately congested and congested traffic conditions wherein the concern about the adverse impact of TSP exists. It shows that in general starting the priority operation when the bus is about 20 to 30 seconds away from the intersection produces good results for both bus and general traffic. The findings of the research can be easily integrated into an AVL based TSP system and may potentially enhance the performance of such a system.


[^0]CE Database subject headings: Traffic signals; Simulation; Traffic models; Public transportation; Bus.

## Introduction

Transit Signal Priority (TSP) has long been regarded as an effective way of improving transit service and reducing operational cost. However, concerns regarding the possible negative impacts of TSP on otherwise well controlled traffics have been a major barrier for its implementation in North America. Future deployment of transit signal priority depends largely on the development of TSP strategies that can better accommodate transit vehicles while at the same time minimizing negative effects on general traffic.

Current TSP systems rely on either Advanced Vehicle Location (AVL) technologies or Selective Vehicle Detection systems (SVD) to sense the presence of transit vehicles. AVL is closely tied with the GPS technology, whereas SVD systems communicate with roadside equipment or inductive loops via onbus units. In contrast to SVD based systems in which the priority operation is launched upon the detection of an approaching bus at the single-point location where the detector is located, AVL based systems have the advantage of placing a priority call at anytime while a bus is approaching the intersection.

How real-time bus movement information can be used in developing more advanced transit signal priority strategies is of great interest to traffic professionals and of great importance for future deployment of TSP systems in the United States. Given the fact that most of the advanced priority strategies (e.g. transit phase, skipped phase, etc.) have not been widely accepted in the United States, it is of particular significance to enhance the basic ones, such as the Early Green and Extended Green, and make them an integral part of closed-loop traffic signal control systems.

The control logics of Early Green and Extended Green are identical in the sense that they both would shorten opposing phases, most commonly the green phases for the minor approaches, to favor transit vehicles on the prioritized phase. Starting the Early Green too early and/or operating the Extended Green too long will over shorten the opposing phases, which may increase additional delay to non-transit vehicles. Late start or "short notice" of Early Green, on the other hand, may result in insufficient time to serve transit vehicles within the current cycle. For instance, actuating an Early Green request when bus is

10 seconds away from the intersection can grant a maximum of exactly the same amount of time, i.e., 10 seconds to the bus.

A bus equipped with an AVL device can continuously provide its location information, along with a priority request, to the traffic management center or location where the TSP control algorithm is located. The TSP control algorithm will subsequently decide whether to grant or deny the request. If the request is accepted, the time to actuate the traffic signal will be issued. The purpose of the priority service is normally for enhancing bus schedule adherence, regulating their inter-arrival times, and/or reducing bus travel times. The strategy as to when to launch the priority operation has not been addressed in the previous studies.

The purpose of this research is to investigate whether and how the predicted bus arrival information produced by AVL data can be incorporated into a TSP system for finding the optimal time to initiate the priority operation, especially the early green operation. The study was conducted by both theoretical analysis and simulation verification and focused on the cases in which no near-side bus stops exist. The theoretical approach demonstrated the possible existence of best call-time points and a theoretical model was developed to identify the optimal call/actuation time for a priority request. The model was then implemented in a simulation environment for validation and experimental analysis.

## Literature Review

Ever since the emergence of the concept of transit signal priority, researchers and traffic engineers have been seeking best solutions to prioritize transit vehicles while at the same time minimizing their impacts on other vehicles (for example, Furth and Muller 2000, Ngan et al. 2004). The majority of the existing transit signal priority systems were built on the basis of either fixed-time or closed-loop traffic signal systems associated with the Early Green and Extended Green signal timing strategy.

The "2Es" strategy, as termed by the authors, facilitates the passage of transit vehicles by either shortening the preceding non-prioritized phases or extending the requested phases in favor of transit
vehicles. Evaluation studies, mostly conducted on a simulation platform, have focused primarily on analyzing TSP benefits to transit vehicles as well as possible adverse impacts to other vehicles (for example, Dion et al. 2004, Rakha and Zhang 2004). Literature addressing improved methodologies that can be used to enhance the fundamental "2Es" strategy in the field is very limited (Larry Head, 2002).

The performance of TSP relies on, to a considerable extent, the detection technologies of transit vehicles. For instance, in a point detection based system, the transit vehicle is equipped with a transponder, usually installed on the underside of the vehicle, which sends a unique code to the inductive loop that identifies the vehicle. The inductive loop in return sends an impulse to the signal controller to initiate the priority operation. Since the priority operation is launched at single-point locations where bus detectors are located, the detector location plays a pivotal role. In a previous study (Liu et al. 2004), a statistical approach was introduced to demonstrate the importance of bus detector locations.

Several vehicle detection technologies hold the potential to initiate the priority operation in a flexible manner other than from fixed locations. The Opticom ${ }^{\text {TM }}$ system by 3 M is probably the most widely deployed system in the nation for providing preemption/priority to emergency vehicles/buses. The system works by an emitter in the vehicle that sends an optical flashing signal at a certain rate and at an exact duration when activated. The signal is then detected at the traffic signal to launch the preemption/priority operation. The range of the system is programmable but typically set at approximately 400 meters for fast moving emergency vehicles and 90 to 150 meters for TSP-equipped vehicles (ITSA 2003). The range appears to be arbitrary and no literature was found addressing the effectiveness of the setting.

Advanced Vehicle Location technology uses GPS to track vehicle movement. However, there is a general agreement that this advantage has not been used to its full potential (Furth et al. 2003). Given the fact that the priority operation can be initiated at anytime for a centralized TSP system combined with GPS equipped buses, the issue of concern is then mainly on the selection of the most appropriate time point to actuate the traffic signal in response to a priority call for best performance.

## The Time Strategy for Priority Request Actuation

When a check-in call is received in an AVL based system, Extended Green is normally started automatically at the end of the requested phase if a check-out call has not been received. Operation of Early Green can be initiated at any time upon the receipt of a check-in call sent from the bus as long as the minimum green requirement is satisfied. The focus of this study is placed on the call-time strategy for Early Green. An AVL equipped bus can pick any time to actuate traffic signal while it is approaching the intersection. It is intuitive that actuating an early green operation 5 seconds ahead of the bus arrival time at an intersection may have different effects to buses and other vehicles in affected approaches than actuating the priority operation 20,30 , or 40 seconds ahead.

## Analytical illustration of possible existence of the optimal call-time points

The optimal time to trigger the signal is a product of three factors: bus travel time, the signal timing, as well as the queuing status of the subject intersection. The travel time of a bus to a specific location such as a bus stop or a signalized intersection is a function of the distance-to-go to the destination. If there is no near-side bus stop, the time it takes for the bus to traverse the intersection is then determined mainly by the signal timing plan and consequently, the state of queued vehicles on the associated approach. The relationship between the bus travel time and the queue clearance time can be illustrated in graphic terms by considering two curves, one describing the predicted bus arrival time at the intersection and the other describing the queue clearance time of the associated approach.

As is evident, bus travel time (in terms of time-to-go) is a decreasing function of distance in a timedistance plane and the length of queue (in terms of number of vehicles) is a stepwise function of time in a vehicle count-time plane. For illustrative purpose, a specific diagram needs to be constructed in which the two curves can be presented in a same domain. Fig. 1 is a product of such a diagram. The x-axis is the local signal time and the y-axis is the time, both in seconds.

The local signal time is a recursive time series starting from a referencing point in the cycle (normally at or close to the "yield point" in a traffic actuated signal). Because current signal priority systems seldom start the priority operation one cycle earlier than the time bus arrives at the intersection, we may start counting the local signal cycle at the onset of the red interval of the to-be-prioritized phase, for example, phase $i$ in this case. The number of queued vehicles on the associated approach, or the associated time needed to clear the queue, is thereby approximately an increasing function up to the time point when the shockwave from the stopline reaches the tail of the queue and a decreasing function thereafter during the green phase. This is presented by the curve $t_{q}^{i}$ in the figure (note that for illustrative purpose, the breaking point is set at the signal switching point). For the same reason, the travel time information of a bus one cycle ahead of its arrival time (to the intersection) is less important in decision making. If we start counting down the bus arrival time at the same time, i.e., the onset of the red interval of phase $i$, the curve of bus travel time (to the intersection) will yield a declining function on the signal cycle-time plane, which is presented by $t_{->s}$ in the figure.

In general, if the predicted bus arrival time (the time at which time-to-go $=0$ ) falls into the red interval of the phase $i$, i.e., $R_{i}$, a priority service may apply. Note that early green may also apply when the predicted bus arrival time falls into the early part of the green interval so that queue can be cleared further in advance before bus arrives. This situation was not considered in this study due to the complexity of determining the threshold of the "early part" of the green interval as well as its potential impacts to general traffic. Assume that by the time the phase of interest was terminated by a normal signal operation (e.g. gap-out, maximum green), a bus is $T$ seconds away from the intersection ( $T<C$, where $C$ is the duration of a cycle). Within $T$, the bus shall provide the location information along with the priority request continuously to the TSP control center (for a centralized system). A predefined control algorithm, in return, will make decisions upon whether to deny or grant the request as well as the time to actuate the traffic signal. The actuation time $t_{\text {call, }}$, is therefore a specific time point on the $t_{->s}$ curve as shown in the figure.

When a priority call is placed on phase $i$ at time $t_{\text {call }}$, the requested phase will be started after a certain period of time $G_{p}$, which is the summation of all shortened green intervals of the preceding phases. For example, if a call for early green is placed when a bus is 50 seconds-to-go to the intersection, i.e., $t_{\text {call }}=50$ seconds, the signal will go through the preceding phases quickly under the predefined principle (e.g. minimum green) which determines the duration of $G_{p}$. If $G_{p}$ is, say, 30 seconds in this case, the requested phase will start 30 seconds later, i.e., when the bus is 20 seconds away from the intersection. If the call is placed too early, for instance, at point 1 of the $t_{->s}$ curve, by the time phase $i$ is started, we will have $t_{q}^{i} \ll t_{\rightarrow>s}$. In such a case, queue will be cleared way before the time when bus arrives at the intersection, which implies that the time wasted may be as large as $t_{->s}-t_{q}^{i}$.

Note that early start of the prioritized phase is often made possible by "stealing" green time from opposing phases. The more green time stolen by the priority operation the less is left in a signal cycle for other phases. In this case, the non-prioritized phases could be over shortened by a total of $t_{->s}-t_{q}^{i}$ seconds. Although this might be beneficial to buses in practice, there is however, overcompensation that may induce additional delay to vehicles on the affected phases, especially when the demand is not negligible on the non-prioritized approaches.

If, on the other hand, the priority call is placed too late, say, at point $2, t_{q}^{i}$ will then be much larger than $t_{\rightarrow>s}$. This is another undesirable situation in which the bus has to wait in the queue for $t_{q}^{i}-t_{->s}$ seconds before it can clear the intersection. Clearly, there exists in theory such a time point (ideal $t_{\text {start }}$ in the figure) on the $t_{->s}$ curve; when a priority call is placed at the close proximity of that point, the benefit to buses will be enhanced while the negative impact to non-transit vehicles on the affected approaches is minimized.

## Identification of the optimal call-time points

With the relationship between the time of priority request actuation and the time of queue clearance in the requested phase being established, attention can now be directed towards identifying the optimal call/actuation time. For convenience in illustration, another diagram is constructed by projecting some of the critical time points on the $t_{\rightarrow s}$ and $t_{q}^{i}$ curves onto the signal cycle pie as shown in Figure 2.

Although the origin of the diagram in Figure 2 may coincide with the yield point/local zero of the signal cycle when the requested phase $i$, is the coordinated phase, it is not necessary to set local zero as the origin for our purposes. $t_{\text {call }}^{c}$ and $t_{\text {start }}^{c}$ represent the time points within a local cycle when the green interval of the prioritized phase will start (as a result of signal priority) and when the priority request actuation is placed, respectively. $t_{\rightarrow s}^{0}$ is another important time point which indicates the bus arrival time within a local signal cycle. As demonstrated in the previous section, the relationship of the desired start time of the prioritized phase, the bus arrival time, and the time of queue clearance would lead to:

$$
\begin{equation*}
t_{\text {start }}^{c}=t_{\rightarrow s}^{0}-t_{q}^{i} \tag{1}
\end{equation*}
$$

The number of vehicles accumulated within phase $i$ during the red interval by time $t$ is a function of traffic demand and the start time of the green phase:

$$
\begin{equation*}
q_{i}(t)=q_{i}(0)+\int_{0}^{t} \lambda_{i}(t) d t \quad\left(0 \leq t \leq t_{\text {start }}^{c}\right) \tag{2}
\end{equation*}
$$

where $\lambda_{i}(t)=$ traffic arrival rate in phase $i ; q_{i}(0)=$ the residual queue at the beginning of phase $i$ carried over from the previous phase; and $t=$ current time in a signal cycle. Consequently, if we assume there is no start-up lost time, the time needed to clear the queue at time $t$ is determined by the saturation flow of phase $i$ and the number of vehicles in the queue, which is:

$$
\begin{equation*}
t_{q}^{i}(t)=\frac{q_{i}(t)}{S_{i}} \tag{3}
\end{equation*}
$$

where $S_{i}$ is the saturation flow rate for phase $i$.

The time when the requested phase will be in operation is subject to many factors, of which the time at which the signal is triggered and the Early Green policy is determinant. Many existing transit signal priority systems terminate the non-prioritized minor phases at the end of the minimum green interval if there are no pedestrian calls. Some attempts have been made to optimize green time allocation among signal phases during priority operations (for example, Wadjas et al. 2003, Liu et al. 2003). However, the results have not been assessed by field experiments in part because most of the advanced control algorithms require additional vehicle detection devices on the side streets. As a result, although it is practically feasible to terminate an actuated phase during the Passage or Extension intervals, this strategy is only applicable to signal preemption at the moment. For illustrative purpose, we assume that the policy of "minimum green" applies. The relationship between $t_{\text {call }}^{c}$ and $t_{\text {start }}^{c}$ then yields:

$$
\begin{equation*}
t_{\text {start }}^{c}=t_{\text {call }}^{c}+G_{p}+t_{e l}+C L_{l}=t_{\text {call }}^{c}+\sum_{j}\left(\operatorname{MinG}_{j}+C L_{j}\right)+\gamma \operatorname{MinG}_{l}+C L_{l} \tag{4}
\end{equation*}
$$

where $t_{\text {call }}^{c}=$ time of priority request actuation/call as reflected on the local signal cycle; $l=$ the phase in which a priority call is placed $(l \neq i)$; $t_{e l}=$ the elapsed time of phase $l ; j=$ the phases in between the requested phase and the phase $l(j \neq i)$; MinG $=$ minimum green interval, which is given by $\max ($ Minimum green $\mid$ Walk + Flashing Don't Walk); $C L=$ clearance interval; and $\gamma=$ dummy variable which is 0 if a call is placed after the minimum interval of phase $l$, otherwise, $\gamma \in(0,1)$. Substituting equation (4) into (1) gives the optimal time for priority call/actuation:

$$
\begin{equation*}
t_{\text {call }}^{c}=t_{\rightarrow s}^{0}-\left[\sum_{j}\left(\operatorname{MinG}_{j}+C L_{j}\right)+\gamma \operatorname{MinG}_{l}+C L_{l}\right]-t_{q}^{i}(t) \tag{5}
\end{equation*}
$$

## Simulation Analysis

The analysis presented thus far has provided important insight into the nature of the problem. However, because both the predicted bus arrival time $t_{\rightarrow s}^{0}$ and the queue clearance time $t_{q}^{i}(t)$ are full of noise in the
real life, extension of the similar analytical approach to a real-life TSP system with actuated signals and time varying traffic demand is likely to be much more complex. Traffic simulation is advantageous in conducting "what if" studies and comparison analysis. In the following, we use simulation to continue our investigation. The purpose of the simulation study is to validate the theoretical model and identify the optimal call-time points for general cases.

The simulation analysis consists of four major processes which are (1) developing and coding a bus arrival time prediction model of reliable accuracy into a TSP-capable simulation model, (2) proof of concept, (3) representing the theoretical model in simulation, and (4) applying the theoretical model to identify the optimal call-time points, if any, and testing its generality. The prototype of the simulation model used in this study was developed earlier by the authors in support of a study for developing advanced transit signal priority strategies (Liu et al. 2004), a project sponsored by the California Department of Transportation in cooperation with the San Mateo Transit District in San Francisco Bay Area. Detailed description of the simulation model is beyond the scope of this study. Focuses are placed on using the simulation model to address the call-time problem.

A short-term bus travel time prediction model with a reasonable accuracy is prerequisite for this study for attaining $t_{\rightarrow s}$. In the following, we briefly introduce the bus travel time prediction model used in our simulation.

## Bus Arrival Time Prediction

A variety of models have been developed to predict bus travel time by using historical and real-time AVL data (see for example, Dailey and Maclean 2001, Chien and Kuchipudi 2003). The accuracy of these models is reportedly reasonable as long as the GPS data is reliable (Lin and Zeng 1999). However, none of existing models takes real-time traffic flow conditions into consideration or meets the accuracy requirements that are needed for signal priority control. In this study, we used historical and real-time GPS data and wheel speed data to predict bus arrival time to the intersection. In addition to bus movement
data, the simulation records also the time when a bus is released into the network, the time when a bus stops at a signalized intersection, the time when a bus stops at a bus stop, the time when a bus leaves a bus stop; the time when a bus passes the stop line of an intersection, and the time when a bus ends its trip.

In the model, the travel distance $D$ and travel time $T$ of a bus is assumed to follow a linear relationship:

$$
\begin{equation*}
T=\alpha D+\beta \tag{6}
\end{equation*}
$$

where $\frac{1}{\alpha}$ represents average bus travel speed. Bus speed data available from each simulation run can be used later on as historical bus data for estimation of $\alpha$. When the distance to the next node (intersection or bus stop) is $d_{n}$, the time-to-go-next-node estimated by historical data, $t_{h, \rightarrow>n}$, is given by:

$$
\begin{equation*}
t_{h, \rightarrow n}=\alpha d_{n} \tag{7}
\end{equation*}
$$

A recursive least-square (LS) method is then used online to estimate the average bus travel speed. Assume that a bus has traveled distance $d$ from the last zero speed motion within time $t$, the following linear relationship is used:

$$
d=a t+b=\left[\begin{array}{ll}
t & 1
\end{array}\right]\left[\begin{array}{l}
a  \tag{8}\\
b
\end{array}\right] \stackrel{\operatorname{def}}{=} H(t) X
$$

where $a$ is the average travel speed and $H(t)=\left[\begin{array}{ll}t & 1\end{array}\right]$. The recursive least-square method can be expressed by:

$$
\begin{align*}
& \hat{X}_{k+1}=\hat{X}_{k}+W_{k+1}\left(d_{k+1}-H_{k+1} \hat{X}_{k}\right) \\
& S_{k+1}=H_{k+1} P_{k} H_{k+1}^{T}+\sigma_{d}^{2} \\
& W_{k+1}=P_{k} H_{k+1}^{T} S_{k+1}^{-1}  \tag{9}\\
& P_{k+1}=P_{k}-W_{k+1} S_{k+1} W_{k+1}^{T}
\end{align*}
$$

where $\sigma_{d}^{2}, S$, and $P$ are the simulated noise covariance, residual covariance, $L S$ is the estimator covariance, respectively, $k$ is measurement step, and $W$ is the filter gain. When the distance to the next node is $d_{n}$, the time-to-go-next-node from adaptive model, $t_{r,->n}$, is given by:

$$
\begin{equation*}
t_{r, \rightarrow n}=\frac{d_{n}}{a} \tag{10}
\end{equation*}
$$

The predicted time-to-go-next-node $t_{->n}$ is given by:

$$
\begin{equation*}
t_{->n}=\frac{\sigma_{r}^{2}}{\sigma_{r}^{2}+\sigma_{h}^{2}} t_{h, \rightarrow n}+\frac{\sigma_{h}^{2}}{\sigma_{r}^{2}+\sigma_{h}^{2}} t_{r, \rightarrow n} \tag{11}
\end{equation*}
$$

where $\sigma_{r}^{2}$ and $\sigma_{h}^{2}$ are estimation error covariance of historical and adaptive model, respectively.
If the next node is an intersection, time-to-go-next-node is equal to time-to-go-next-signal, i.e., $t_{->s}=t_{->n}$. If the next node is a bus stop, assuming that there are $m$ bus stops $\left(n_{1}, n_{2}, \ldots, n_{m}\right)$ before the next signalized intersection, time-to-go-next-signal is obtained by:

$$
\begin{equation*}
t_{->s}=t_{->n_{1}}+t_{s, n_{1}}+\sum_{i=1}^{m-1}\left(t_{h, n_{i}->n_{i+1}}+t_{s, n_{i+1}}\right), \tag{12}
\end{equation*}
$$

where $t_{s, n_{i}}$ is the average bus dwell time at stop $n_{i}, t_{h, n_{i} \rightarrow n_{i+1}}$ is the historical travel time from stop $n_{i}$ to stop $n_{i+1}$.

## Verification of the existence of optimal call-time points

Next, a proof-of-concept study was conducted by using the simulation model. The primary aim of the proof-of-concept study was to verify the existence of the "optimal" call-time points by simulation. Recall that the speculation of the "optimal" call-time strategy came out of our investigation to Fig. 1 in the analytical study, the true existence of such call-time points must be validated before any further investigation. The analysis was conducted on two intersections of the study corridor by manually setting bus call-times as if we knew nothing about the "best" call-time. We used real traffic and signal timing data in the test and the buses were arranged to make priority calls from various time points at a 5 seconds interval while approaching to the intersection. The relationship between the call-times and the intersection delay of buses and general traffic was analyzed to see whether there exist such call-time points that meet
our request, i.e., bring significant reduction in bus delay while at same time affect the general traffic at minimum.

The test site comprises of twelve signalized intersections from 2nd Ave. to 28th Ave. to the south of El Camino Real corridor, the California State Highway 82 between southern San Jose and the southern outskirts of San Francisco. Figure 3 shows the schematic representation of the arterial including the locations of bus stops, spacing between intersections, and signal phasing of selected intersections. As the TSP systems working with nearside bus stops requires explicit estimation of bus dwell time (e.g. Kim and Rilett, 2005), most of the bus stops were far-sided in this study and no priority service was operational at intersections with nearside bus stops. In addition, signal priority was applied to main street phases 2 and 6 for every other cycle (the bottom of Fig. 3 shows NEMA phases), the consecutive cycle following the TSP operation was set as the transitional cycle. Traffic signal control along the corridor is an actuated system coordinated on the El Camino Real on phases 2 and 6, the background signal cycles are 90, 110, and 100 seconds during morning peak, midday, and afternoon peak, respectively. The typical phase sequence is the lead-lag design on phases 1 and 5 that serve the south/north bound left-turn movement on El Camino Real.

The test site selected for this research has many common features, as listed below, for a signalized arterial to which TSP usually applies.

- The functional classification of the highway is a typical urban major arterial
- The signal timing plan is representative for an urban arterial (combined semi-actuated and actuated control with lead-lag and/or lag-lead settings)
- The $\mathrm{v} / \mathrm{c}$ ratio ranges from 0.4 to 0.94 during the study period
- More than one bus routes on the arterial

The spacing of the intersections varies from 90 to 610 meters along the study segment of the arterial. For closely spaced intersections, the priority actuation time was set equal to the bus departure time from either the upstream intersection or the far side bus stop. The link lengths between $17^{\text {th }}$ Ave., $20^{\text {th }}$ Ave., and $25^{\text {th }}$

Ave. are 610 meters and 580 meters respectively which are ideal for this study. Therefore, these three intersections were selected for detailed tests. Table 1 gives the signal timing, traffic demand, and degree of saturation of the intersections.

As shown in the table, the differences between minimum green and normal green are the maximum amount of green time that can be used in the operation of early green. For example, the minor phases 1,5 , 4, 8 of the 17th Ave have the flexibility of $10,10,16$, and 14 seconds, respectively, for signal priority. This would allow, at maximum, 50 seconds of early green for the prioritized phase. For this case, because phases 4,8 are overlapped, the maximum is 20 to 22 seconds if the priority process can be operated early enough.

In the simulation, the normal signal control without bus priority was set as the baseline condition. The rest of the scenarios were designed to investigate different actuation/call time strategies varying from 5 to 40 seconds ahead of bus arrival to the intersection at the onset of the red interval of the prioritized phase with a step increment of 5 seconds. In total, nine scenarios were simulated:
(1) No priority
(2) $t_{\text {call }}=5 \mathrm{sec}$
(3) $t_{\text {call }}=10 \mathrm{sec}$
(4) $t_{\text {call }}=15 \mathrm{sec}$
(5) $t_{\text {call }}=20 \mathrm{sec}$
(6) $t_{\text {call }}=25 \mathrm{sec}$
(7) $t_{\text {call }}=30 \mathrm{sec}$
(8) $t_{\text {call }}=35 \mathrm{sec}$
(9) $t_{\text {call }}=40 \mathrm{sec}$

Since the priority operation was set to once every other cycle, buses were released to the network at a relatively high frequency of each every five minutes for over two hours from both directions to capture enough samples. Excluding the warm-up time period, simulation data was recorded for two continuous hours and five simulation runs were executed for each scenario to mitigate the random effect. As a result,
an average of 230 events (i.e., a bus passes the intersection) were recorded at each of the three intersections under each scenario. On the intersection of $25^{\text {th }}$ Ave., for example, 84 Early Green and 67 Extended Green requests on phase 2 and 6 were captured, out of which 66 and 49 were responded, respectively.

In terms of traffic analysis at signalized intersections, intersection delay is arguably the single most important aspect of signal/signal priority performance. In this study, bus intersection delay and traffic delay on the minor phases were selected for measurement of effectiveness. The eight scenarios with bus signal priority were grouped into late call (scenarios 2 and 3), mid call (scenarios 4, 5, and 6), and early call (scenarios 7, 8, and 9) for detailed analysis.

As shown in Figure 4, bus delay drops significantly as the call/actuation time changes from 5 seconds to 25 seconds ahead of the bus arrival time at the intersection and keeps rather stable thereafter. The late call strategy brings to bus an average of only 8.5 percent reduction in intersection delay, which is considerably limited as compared to the 54.3 and 73.3 percent brought by the mid and early calls respectively. It also can be seen that there is no noticeable benefits for buses when the priority calls were placed earlier than 25 seconds of bus arrival, which confirmed the conclusion from the theoretical analysis that provision of excessive early green would not bring significant benefits to buses.

Also noticeable from the figure is the slight increase in vehicle delay in minor phases (phases $1,4,5$, 7, and 8) with the changes of bus call-time strategies. The average delay to vehicles of minor phases increased an average of 15 percent in total of the eight prioritized scenarios, within which, about 4 percent was brought by late calls, 14.6 percent by mid calls, and 22.7 percent by early calls. The focal scenario 6 which gave maximum reduction in bus delay brought about 17.2 percent increase which is slightly above the average.

The v/c ratio of the minor phases varies from 0.44 to 0.77 at $25^{\text {th }}$ Ave., further investigation was conducted on the intersection of $17^{\text {th }}$ Ave., where the degree of saturation of minor phases was relatively higher ranging from 0.65 to 0.92 but still remains undersaturated during the study period. As depicted in Figure 5, the same trend in bus intersection delay can be observed as the call-time changes from 5
seconds to 40 seconds ahead of bus arrival. However, the vehicle delay of minor phases turns out to be more sensitive to call-time changes than that on the $25^{\text {th }}$ Ave. Average vehicle delay increased 40.4 percent in total, of which, late calls, mid calls, and early calls account for 6.9 percent, 20.1 percent, and 83.1 percent respectively. The early call strategy stands out significantly in increasing vehicle delay on minor phases, in the worst case, scenario 9, the delay was doubled. The observed negative impact to vehicles in minor phases, however, did not seem to bring noticeable benefit to mainline buses as shown in the figure.

The relationship between bus call-time strategy and the effect on minor phase vehicles can also be observed from the statistics of average queue lengths. Figure 6 shows the average queue length in number of vehicles on each minor phase of $17^{\text {th }}$ Ave under the test scenarios. The v/c ratio of phase $1,5,4$, and 8 are $0.88,0.65,0.92$, and 0.74 respectively. It can be seen that although queues increased on all four phases, they are more sensitive to the bus call-time changes especially for the phases with higher v/c ratio. For instance, the average queue length in phase 4 increased only 2 vehicles from the case with no priority to scenario 6 which is to place priority call 25 seconds ahead of the bus arrival time. The queue length increased abruptly to 9 vehicles when the call was placed 15 seconds earlier, i.e., 40 seconds ahead of the bus arrival time at the intersection.

## Verification and proof of generality of the theoretical model

Now that the existence of such time points was verified, our analysis was then extended to implement and apply the theoretical model in simulation to identify the location of these time points for generalized conditions. Since the manual test has also identified the approximate range of the optimal time points, the number of simulation runs needed for the proof of generality analysis was significantly reduced. As the major issue of concern is the variation in the length and clearance time of the queue on the prioritized approach, changes in one or two of the parameters that are determinant to the queuing condition would be sufficient.

In order to represent the theoretical model in simulation, the queue clearance time at time $t$, i.e., $t_{q}^{i}(t)$, needs to be modeled. For this purpose, the simulation needs to record the times when a vehicle enters a queue and the time it passes the stop line as well as its number and location in queue. The corresponding queue clearance time is approximated by the time difference between the two time stamps. The statistics of the clearance time corresponding to a specific queue length (in terms of the number and/or location of a vehicle in queue) can be obtained and used for representing $t_{q}^{i}(t)$. In simulation, priority calls were made based on the queue clearance time of the prioritized approach and the predicted bus arrival as defined by the theoretical model. A number of simulation tests were made on the test site, of which that contains enough early green calls (about 10 to 12) at each intersection were recorded. Fig. 7 shows the results of five simulation tests obtained from six intersections of the south bound of the arterial. Each intersection has different traffic demands on the major and minor approaches. Note that the concern about the adverse impact of TSP exists mostly with regard to the operation in moderate congested and congested traffic conditions. Traffic demands were selected from peak hour demands with an average $\mathrm{v} / \mathrm{c}$ ratio of 0.82 . The optimal call-time points were found falling into the range of 18.6 (17th Ave.) to 25.4 seconds (25th Ave.). The simulation result of 25th Ave. is very close to that from the manual test ( 25 to 30 seconds), while the result from the 17th Ave. (with high v/c ratio on side streets) is much conservative ( 18.6 seconds versus 20 to 25 seconds from manual test).

Signal cycle length is another pivotal factor to queue. We have also analyzed the condition of an intersection with multiple cycle lengths. The intersection of $17^{\text {th }}$ Ave. was isolated from the network and given varied cycles from 60 to 150 seconds with a step increase of 10 seconds. Ten samples were obtained for each cycle length. The v/c ratios were set to about 0.90 for major phases and 0.80 for minor phases to represent a moderate congested condition and to correspond to the assumption of a fixed signal timing plan made for the analytical analysis. Figure 8 depicts the simulation result.

As shown in the figure, the optimal call-time increases slightly with the increase in cycle length. Since the length of the queue is an increasing function of the signal cycle, the result agrees well with the
one obtained from the theoretical analysis, i.e., the queue clearance time of the prioritized phase is strongly correlated with the choice of the time for priority actuation. Furthermore, observations of the result revealed that in 8 out of 70 samples, or 11.4 percent of the time, the call/actuation time was less than 20 seconds (ahead of bus arrival to the intersection), 10 or 14.3 percent, were larger than 30 seconds, and about 74.3 percent of the optimal call/actuation time fell between 20 and 30 seconds. The result further confirmed the generality test and is of practical significance in determining the appropriate detection range of zone or area based bus detection devices under various background traffic signal cycles.

## Conclusion

This paper develops a methodology that incorporates the predicted bus arrival time information into an AVL based TSP system to improve its performance. It examines the best time to initiate the signal priority operation for early green based on the predicted bus arrival time at the intersection. In the paper, the existence of the optimal time for placing a signal priority call/actuating the priority operation was first demonstrated by a theoretical investigation and then by a detailed simulation analysis. The simulation analysis was conducted using a real-life arterial in the San Francisco Bay Area. The same arterial was used in a previous study sponsored by the California Department of Transportation (Caltrans) aimed at developing advanced traffic signal priority strategies in cooperation with the San Mateo Transit District (SamTrans). It concludes that, if applicable, starting the priority operation when the bus is about 20 to 30 seconds away from the intersection may produce good results for both buses and general traffic. The findings of this research may help traffic professionals set an appropriate range of their bus detection systems for best bus signal priority performance.

Being the first attempt of such an approach, the research has focused on operation of TSP under moderately congested and congested traffic conditions wherein the concern about the adverse impact of TSP exists. We have selected a typical urban major arterial for testing considering that a TSP system usually applies to urban arterials. Although the traffic and signal timing conditions of the selected site are
pretty typical, the conclusion may still not be universal to other arterials due to the complexity and diversity of the real situation. Future research will be focused on developing statistic models for quantitatively assessing the generality of the findings.

## Acknowledgement

The authors would like to thank the anonymous reviewers for their detailed comments and suggestions to improve the quality of the paper. The authors are grateful to Kai Leung, senior traffic engineer of the Headquarter of California Department of Transportation for his support and advices through the course of the research.

## Notation

The following notations are used in the paper:
$t_{\text {call }}=$ a specific time on $t_{->s}$ at which the bus check-in call is placed;
$t_{\text {call }}^{c}=$ time to place check-in call as reflected on the local signal cycle;
$t_{h,->n}=$ the time-to-go to the next node estimated by historical data;
$t_{r, \rightarrow>n}=$ the time-to-go to the next node estimated by adaptive model;
$t_{h, n_{i}>n_{i+1}}=$ the historical travel time from stop $n_{i}$ to stop $n_{i+1}$.
$t_{->s}=$ time-to-go to the next signalized intersection in seconds;
$t_{\rightarrow s}^{0}=$ predicted bus arrival time as reflected on the local signal cycle in second;
$t_{\text {start }}=$ a specific time on $t_{->s}$ at which the priority operation is started;
$t_{\mathrm{s}, n_{i}}=$ the average bus dwell time at stop $n_{i}$;
$t_{\text {start }}^{c}=$ desired start time of the requested phase as reflected on the local signal cycle in second;
$\alpha=$ reverse of average bus travel speed;
$\sigma_{d}^{2}=$ simulated noise covariance;
$\sigma_{h}^{2}=$ estimation error covariance of adaptive model;
$\sigma_{r}^{2}=$ estimation error covariance of historical model.

## References

Chien, Steven I. and Kuchipudi, C. M. (2003). "Dynamic travel time prediction with real-time and historic data." J. Transp. Eng., 129(6), 608-617.

Dailey, D.J. and Maclean, S.D. (2001). "Transit vehicle arrival prediction: algorithm and large-scale implementation." Transportation Research Record 1771, Transportation Research Board, Washington D.C., 46-51.

Dion, F., Rakha, H., and Zhang, Y. (2004). "Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial." J. Transp. Eng., 130(3), 294-303.

Furth, P.G., and Th.H.J.Muller. (2000). "Conditional bus priority at signalized intersections: Better service with less traffic disruption." Transportation Research Record 1731, Transportation Research Board, Washington D.C., 22-30.

Furth, P.G., Hemily, B.J, Th.H.J. Muller, and Strathman J.G. (2003). "Uses of archived AVL-APC data to improve transit performance and management: review and potential." Transit Cooperative Research Program Web document 23 (Project H-28): Contractor's final report.

ITS America. (2003). "An overview of transit signal priority." Advanced Traffic Management System and Advance Public Transportation System Committees, Washington, D.C.
K. Larry Head et al. (2002). "Improved Traffic Signal Priority for Transit." TCRP project A-16 Interim report. http://www4.nas.edu/trb/crp.nsf/All+Projects/TCRP+A-16A, (Dec. 12, 2006).

Kim, W. and L.R. Rilett. (2005). "New bus signal priority system for network with near-side bus stops." Proc., 84th TRB Conf, Transportation Research Board, Washington D.C.

Lin, W.H. and Zeng, J. (1999). "An Experimental Study on Real Time Next Bus Arrival Time Prediction with GPS Data." Transportation Research Record 1666, Transportation Research Board, Washington D.C., 101-109.

Liu, H., A. Skabardonis, Zhang, W.B., and Li, M. (2004). "Optimal detector location for transit signal priority." Transportation Research Record 1867, Transportation Research Board, Washington D.C., 144150.

Liu, H., et al. (2004). "Development and application of a simulation tool for transit signal priority." Proc., 83rd TRB Conf, Transportation Research Board, Washington D.C.

Liu, H., A. Skabardonis, and Zhang, W.B. (2003). "A dynamic model for adaptive bus signal priority." Proc., 82 th TRB Conf, Transportation Research Board, Washington D.C.

Ngan, V., T. Sayed, andA. Abdelfatah. (2004). "Impact of Various Parameters on Transit Signal Priority Effectiveness." J. Public Transportation, 7(3), 71-93.

Rakha, H. and Zhang, Y. (2004). "Sensitivity Analysis of Transit Signal Priority Impacts on Operation of a Signalized Intersection." J. Transp. Eng., 130(6), 796-804.

Wadjas, Y. and P. G. Furth. (2003). "Transit signal priority along an arterial using advanced detection." Transportation Research Record 1856, Transportation Research Board, Washington D.C., 220-230.

## LIST OF FIGURES

Fig. 1. Graphical illustration of optimal priority call-time for early green
Fig.2. The parameters as reflected on local signal cycle
Fig.3. The studied arterial
Fig.4. Delay of buses, minor phase vehicles as a result of various call-time strategies (25th Ave.)
Fig.5. Delay of buses, minor phase vehicles as a result of various call-time strategies (17th Ave.)
Fig.6. Queue lengths of minor phases (17th Ave.)
Fig.7. Optimal call-times for the test intersections (southbound)
Fig.8. Call-time strategy versus variable cycle length (17th Ave.)

## LIST OF TABLES

Table 1. Traffic \& signal timing condition at the selected intersections


Fig. 1. Graphical illustration of optimal priority call-time for early green


Fig.2. The parameters as reflected on local signal cycle


Fig.3. The studied arterial


Fig.4. Delay of buses, minor phase vehicles as a result of various call-time strategies (25th Ave.)


Fig.5. Delay of buses, minor phase vehicles as a result of various call-time strategies (17th Ave.)


Fig.6. Queue lengths of minor phases (17th Ave.)


Fig.7. Optimal call-times for the test intersections (southbound)
(Original figures of Fig.7. is provided on page 29-31)


Fig.8. Call-time strategy versus variable cycle length (17th Ave.)


Element 1 of Fig. 7


Element 2 of Fig. 7


Element 3 of Fig. 7


Element 4 of Fig. 7


Element 5 of Fig. 7


Element 6 of Fig. 7

Table 1. Traffic \& signal timing condition at the selected intersections

| Intersection | Signal phase | Phase sequence | Minimum green(sec) | Green (sec) | $\begin{gathered} \hline \text { Lost time } \\ \text { (sec) } \end{gathered}$ | Cycle (sec) | Demand (veh/hr.) | $\qquad$ | Number of lanes | $\begin{gathered} \hline \text { Bus speed } \\ (\mathrm{km} / \mathrm{hr}) \\ \hline \end{gathered}$ | Degree of saturation |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 17th Ave. | 1(lead) | 2 | 8 | 18 | 9 | 90 | 285 | 1620 | 1 | 48 | 0.88 |
|  | 2 | 3 | 10 | 39 |  |  | 1882 | 1800 | 3 |  | 0.8 |
|  | 4 | 1 | 8 | 24 |  |  | 440 | 1800 | 1 |  | 0.92 |
|  | 5(lag) | 3 | 8 | 18 |  |  | 212 | 1620 | 1 |  | 0.65 |
|  | 6 | 2 | 10 | 39 |  |  | 1720 | 1800 | 3 |  | 0.74 |
|  | 8 | 1 | 8 | 22 |  |  | 324 | 1800 | 1 |  | 0.74 |
| 20th Ave. | 1(lead) | 3 | 6 | 16 | 12 | 90 | 188 | 1620 | 1 | 48 | 0.65 |
|  | 2 | 4 | 15 | 32 |  |  | 1768 | 1800 | 3 |  | 0.92 |
|  | 3 | 2 | 8 | 16 |  |  | 245 | 1620 | 1 |  | 0.85 |
|  | 4 | 1 | 6 | 14 |  |  | 210 | 1800 | 1 |  | 0.75 |
|  | 5(lag) | 4 | 6 | 16 |  |  | 114 | 1620 | 1 |  | 0.4 |
|  | 6 | 3 | 15 | 32 |  |  | 1682 | 1800 | 3 |  | 0.88 |
| 25th Ave. | 1(lead) | 3 | 6 | 14 | 12 | 90 | 141 | 1620 | 1 | 48 | 0.56 |
|  | 2 | 4 | 10 | 30 |  |  | 1662 | 1800 | 3 |  | 0.92 |
|  | 4 | 1 | 8 | 18 |  |  | 242 | 1800 | 1 |  | 0.67 |
|  | 5(lag) | 4 | 8 | 14 |  |  | 108 | 1620 | 1 |  | 0.43 |
|  | 6 | 3 | 10 | 30 |  |  | 1690 | 1800 | 3 |  | 0.94 |
|  | 7 | 2 | 8 | 16 |  |  | 221 | 1620 | 1 |  | 0.77 |
|  | 8 | 1 | 8 | 18 |  |  | 245 | 1800 | 1 |  | 0.68 |


[^0]:    ${ }^{1}$ Assistant professor of Department of Civil Engineering, Texas Tech University, Box 41023, Lubbock, TX 79409. Tel: (806)742-2801 ext.229, Fax: (806)742-4168, E-mail: hongchao.liu@ttu.edu
    ${ }^{2}$ Associate professor of Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721. E-mail: weilin@ sie.arizona.edu
    ${ }^{3}$ Assistant research engineer of Institute of Transportation Studies, University of California at Berkeley, Berkeley, CA 94720. E-mail: tan@eecs.berkeley.edu

