Traffic Signal Optimization with Application of Transit Signal Priority to an Isolated Intersection

Eleni Christofa and Alexander Skabardonis

Transit signal priority (TSP) is a control strategy that has been used extensively to improve transit operations in urban networks. However, several issues related to TSP deployment—including the effect of TSP on auto traffic and the provision of priority to transit vehicles traveling in conflicting directions at traffic signals—have not yet been addressed satisfactorily by existing control systems. This paper presents a real-time, traffic-responsive signal control system for signal priority on conflicting transit routes that also minimizes the negative effects on auto traffic. The proposed system determines the signal settings that minimize the total person delay in the network while assigning priority to the transit vehicles on the basis of their passenger occupancy. The system was tested through simulation at a complex signalized intersection located in Athens, Greece, that had heavy traffic demands and multiple bus lines traveling in conflicting directions. Results showed that the proposed system led to significant reductions in transit users' delay and the total person delay at the intersection.

There is an increasing need for designing efficient multimodal transportation systems to improve mobility in urban areas. To do so in a system optimal and equitable way, high-occupancy vehicles, mainly transit vehicles, should be treated differently from low-occupancy vehicles. One way to achieve that objective is by granting priority to transit vehicles at bottlenecks such as signalized intersections, which are responsible for a large portion of the vehicles' delay. But because of geometric or spatial restrictions it is not always feasible to provide priority to transit vehicles through improvements in facility design (e.g., bus lanes). As a result, there is a clear need to optimize signal control systems such that they balance their treatment for transit and auto users by minimizing total person delay in the system.

Transit signal priority (TSP) is an emerging operational strategy that facilitates efficient transit operations by providing priority to transit vehicles at signalized intersections. TSP strategies have been implemented in several urban areas in the United States and Europe. Many studies report significant reductions in transit vehicles' control delay and an overall improvement in their operations. However, these strategies are often disruptive for the auto traffic, leading to substantial increases in its delay. Most of the implementations are site specific. They are also restricted to provide priority only to transit vehicles traveling on nonconflicting routes, which leads to inequitable treatment among transit users.

Demand for balanced multimodal systems requires that signal control systems provide optimized signal settings for the auto traffic such that the negative effects of TSP on auto traffic delays are minimized. Combining traffic signal optimization with TSP strategies is the most cost-effective way to improve the level of service for transit operations and minimize the total person delay in signalized networks. This paper describes the development of a real-time, traffic-responsive signal control system with TSP for an isolated intersection. The system provides priority to transit vehicles even when they are traveling in conflicting directions, and at the same time it minimizes the negative effects on the auto traffic. This is part of a major effort to develop TSP strategies for a range of operating conditions on arterials and grid networks.

This paper is organized as follows. First, existing TSP strategies and signal control systems are briefly reviewed. The next sections present the methodology and the study site used for testing the proposed real-time traffic-responsive signal control system. The results from the simulation tests performed are presented next. The final section summarizes the study findings and outlines future research.

BACKGROUND

TSP Strategies

Existing transit signal priority strategies fall into two major categories: passive and active. Passive priority strategies are developed offline on the basis of historical data and do not require any detection system. They include mainly changes in the signal settings (green times, offsets, and cycle lengths). Examples include adjustment of offsets to account for slower bus speeds and midblock dwell times, addition of green time to the phases that serve transit vehicles so that the probability of a transit vehicle arriving at the intersection during the green interval is increased, and instances in which both of those changes are made (*1*). Reduction in the cycle length is another passive priority strategy commonly used because it increases the turnover of the phases and, as a result, decreases the delays for all vehicles.

Passive priority strategies are inexpensive to develop and easy to implement. However, their success depends on having traffic volumes with low variability. In addition, such strategies assume deterministic dwell times at the transit stops, which is not realistic for most transit operations.

Active priority strategies respond to traffic variations in real time and are therefore more effective than passive priority strategies. Information in real time about the transit vehicles' speed and location

Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Berkeley, 109 McLaughlin Hall, Berkeley, CA 94720. Corresponding author: E. Christofa, christofa@berkeley.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 192–201.
DOI: 10.3141/2259-18

obtained by sensing technologies is required for the design of such strategies. Active priority strategies include holding the green until the transit vehicle clears the intersection (phase extension) or advancing the start of the green for the phase serving a transit vehicle (phase advance). Other options include inserting a new phase that can serve the transit vehicle at the moment it arrives at the intersection (phase insertion) or rotating the phases so that transit vehicles are served as soon as possible (phase rotation).

Although active priority strategies can be more effective than passive priority strategies in improving transit operations, they require detection and communication systems that increase their cost without any guarantee of succeeding on a network-level basis. Active priority strategies often have detrimental effects on the nontransit traffic (especially the cross-street traffic), can cause confusion to motorists, and in many cases are responsible for loss of signal coordination and interruption in the progression of the vehicle platoons that can result from excessive delays (*1, 2*).

Signal Control Systems with TSP

A number of real-time signal control systems that incorporate active priority strategies exist in the literature. Such systems use detection of vehicular traffic at some point upstream or downstream or at points both upstream and downstream of an intersection to predict the traffic conditions and adjust the signal settings in real time. By using the available information, the signal settings are optimized on a decision horizon on the order of one cycle to a few minutes (traffic-responsive systems) or on a rolling horizon (adaptive systems).

Active priority strategies have been implemented with the split, cycle, and offset optimization technique (SCOOT), a cyclic model that optimizes phase splits and cycle lengths in real time on the basis of saturation-level constraints as well as offsets such that traffic progression improvements can be achieved (*3*). Priority is provided to transit vehicles through phase extension or advance conditional on schedule-based and headway-based criteria, only when traffic conditions are below user-defined levels of saturation (*4*). A similar TSP logic is followed by the Sydney Coordinated Adaptive Traffic System (SCATS) (*5*). Another traffic-responsive signal control system with TSP was recently developed by California Partners for Advanced Transit and Highways (*6*). The system provides priority based on a trade-off between bus delay savings and the effect on the rest of the traffic.

PRODYN (Programme Dynamique) incorporates transit priority by including cost elements for the transit vehicles in the objective function that is optimized over a rolling horizon. The cost elements are weighted on the basis of the priority level assigned a priori to each transit vehicle and its direction (*7*). The UTOPIA (Urban Traffic Optimization by Integrated Automation) system in Turin, Italy, is an adaptive signal control system that provides unconditional priority to selected bus routes by continuously optimizing the signal settings on a rolling horizon and simultaneously improving mobility for private vehicles (*8*). SPPORT (Signal Priority Procedure for Optimization in Real-Time) is another adaptive signal control system that incorporates rule-based transit priority while accounting for the effect of stopped transit vehicles on traffic operations (*9*). Finally, the centralized TSP system, an adaptive signal control system implemented in Los Angeles, California, provides priority to buses on the basis of their schedule lateness (*10*).

There are several issues that have not been successfully addressed by the systems described above. First, none of the existing systems systematically provides priority to transit vehicles traveling in conflicting directions. Existing work has dealt with this issue either by predetermining which route is given priority (*7*) or by constraining the implementation of the system on networks that include only transit vehicles traveling in nonconflicting directions (*5*). Moreover, transit priority is often provided unconditionally, without considering specific criteria, such as passenger occupancy and schedule delay (*6*). Such criteria could ensure improvement in the operations of transit vehicles while protecting cross streets from reaching oversaturated conditions. In addition, the existing systems do not account for the difference in the passenger occupancy of autos and transit vehicles, instead optimizing their systems on a per vehicle basis (*4, 5*). Finally, the provision of priority is often rule-based, and as a result it is not explicitly included in the optimization process.

METHODOLOGY

A real-time, traffic-responsive signal control system that minimizes the total person delay at the traffic signals is proposed. The goal of the formulation is to optimize the signal timings, such that conditional priority is granted for the transit vehicles on the basis of their passenger occupancy. Conditional priority is used as a way to assign priority when two or more transit vehicles are expected to arrive at the intersection at approximately the same time and compete for priority. In addition, the effect of TSP on the auto delays at the intersection is taken into account by including the total person delay in the objective function for all vehicles present at the intersection.

The formulation is based on the assumption of fixed cycle lengths with a fixed phase sequence. The vehicle arrivals and service times for all vehicles at the signalized intersection are assumed to be deterministic. The arrivals of the transit vehicles at the intersection are assumed to be known in real time. It is also assumed that transit vehicles travel on mixed traffic lanes. However, the formulation could be easily derived even when dedicated rights-of-way exist.

The mathematical program minimizes the total person delay at the intersection by changing the green times for each phase i, G_i, in the cycle under consideration (indexed by T), constrained by the minimum green times for each lane group j, $G_{j\,\min}$, and a fixed cycle length, C. A lane group is defined by the *Highway Capacity Manual* 2000 as one or more adjacent lanes at each intersection approach that can be served by the same phases (*11*). The mathematical program is run once for every cycle. The general form of the mathematical program that optimizes the signal settings for any design cycle T is as follows:

subject to

$$\min \sum_{a=1}^{A} o_a d_a + \sum_{b=1}^{B} o_b d_b$$

subject to

$$\sum_{i \in I_j} G_i \ge G_{j\,\min} \tag{1}$$

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

where

o_a = passenger occupancy of auto a (pax/veh),
o_b = passenger occupancy of transit vehicle b (pax/veh),
d_a = control delay for auto a (s),
d_b = control delay for transit vehicle b (s),
A = total number of autos that experience delay at intersection during cycles T and $T + 1$,
B = total number of transit vehicles present at the intersection during cycle T,

I_j = set of phases that can serve lane group j,
G_i = green time allocated to phase i (s),
$G_{j\,\min}$ = minimum green time allocated to lane group j (s),
P = number of phases in a cycle, and
C = cycle length (s).

The objective function consists of the summation of the person delay for the auto and transit vehicles passengers, which is explained in more detail in the following sections, along with the constraints of the mathematical program.

Auto Delay

The person delay for the auto passengers included in the objective function consists of the summation of two terms. The first is the person delay experienced by the autos during the design cycle T. The second is the estimated delay experienced during the next cycle, $T+1$. Both terms include the delays of the vehicles that remain in the residual queues in the case of oversaturated conditions (when demand for a lane group exceeds capacity or queues fail to clear). Delays for the passengers experienced during the next cycle $T+1$ need to be included to account for the effect that the design of the signal timings in the current cycle will have on the delays of the vehicles in the next cycle.

To be able to estimate the delay experienced by the autos in oversaturated conditions, the number of autos in the residual queues for each of the lane groups needs to be estimated. The number of autos in the residual queue of lane group j at the end of the last phase that can serve the respective lane group for a cycle T, $N_{j,T}$, is calculated as

$$N_{j,T} = N_{j,T-1} + q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1} + q_{j,T}\sum_{i=1}^{l_j} G_{i,T} - s_j \sum_{i\in I_j} G_{i,T} \tag{2}$$

where

$N_{j,T-1}$ = number of autos in residual queue of lane group j at end of last phase that can serve the respective lane group for previous cycle $T-1$,
$q_{j,T}$, $q_{j,T-1}$ = arrival rate for lane group j during cycle T and previous cycle $T-1$ (veh/h),
l_j = last phase in a cycle that can serve lane group j,
$G_{i,T}$, $G_{i,T-1}$ = green time allocated to phase i during cycle T and previous cycle $T-1$ (s), and
s_j = saturation flow for lane group j (veh/h).

Figure 1 shows an example of the cumulative number of vehicles present at an intersection for cycles $T-1$, T, and $T+1$ for a lane group j. According to Figure 1, lane group j can be served by Phases 4 and 5, so its green time will be $G_j = G_4 + G_5$. The shaded area beneath the solid lines represents the total delay experienced by the autos that belong to lane group j and are present at the intersection from the end of the green time for j in the previous cycle $T-1$ until the end of the green time for j in the design cycle T (denoted by $D_{j,T}$). The other shaded area beneath the dashed lines represents the estimate of the delay that the autos of lane group j will experience until the end of the green time for j in the next cycle $T+1$ (denoted by $\hat{D}_{j,T+1}$). Such queuing diagrams can be drawn for all lane groups to allow for the estimation of the delay for autos and transit vehicles under the assumption of a first-in-first-out queuing discipline. Next, the two cases for calculating auto vehicle delay for under- and oversaturated traffic conditions are described. The examples illustrate how to calculate the red times for a lane group in each of the cases.

1. The total delay for all autos for the design cycle T, D_T (as described above), is as follows:

$$D_T = \sum_{j=1}^{J} D_{j,T} \tag{3}$$

where J is the total number of lane groups.

If the autos of lane group j experience oversaturated conditions for cycle T (i.e., $N_{j,T} > 0$), their total delay is given by

$$\begin{aligned} D_{j,T} = {} & \frac{1}{2}\left(2N_{j,T-1} + q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1}\right)\sum_{i=l_j+1}^{P} G_{i,T-1} \\ & + \frac{1}{2}\left(2N_{j,T-1} + 2q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1} + q_{j,T}\sum_{i=1}^{k_j-1} G_{i,T}\right)\sum_{i=1}^{k_j-1} G_{i,T} \\ & + \left(N_{j,T-1} + q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1} + q_{j,T}\sum_{i=1}^{k_j-1} G_{i,T}\right)\sum_{i\in I_j} G_{i,T} \\ & + \frac{1}{2}\left(q_{j,T}\sum_{i\in I_j} G_{i,T} - s_j\sum_{i\in I_j} G_{i,T}\right)\sum_{i\in I_j} G_{i,T} \end{aligned} \tag{4}$$

where k_j is the first phase in a cycle that can serve lane group j.

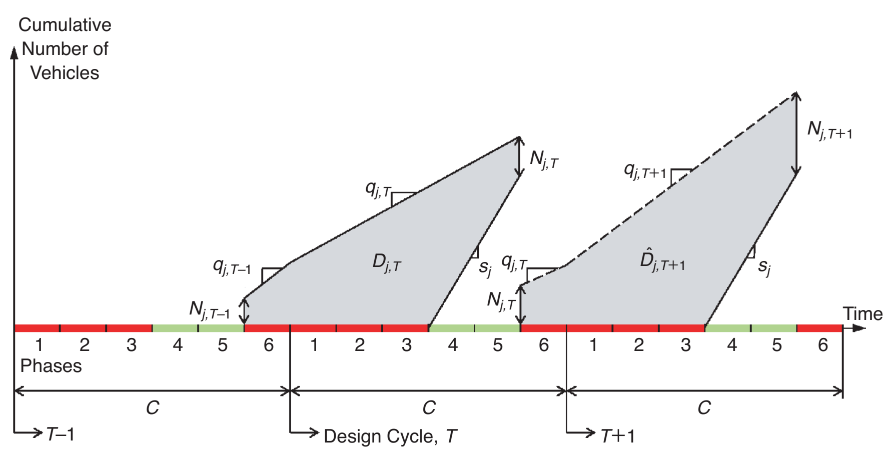

FIGURE 1 Calculation of auto delay for lane group *j*.

If the autos of a lane group j experience undersaturated conditions for cycle T (i.e., $N_{j,T} \le 0$), the third and fourth terms of the delay formula above change and the total delay for the autos of lane group j becomes

$$D_{j,T} = \frac{1}{2}\left(2N_{j,T-1} + q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1}\right)\sum_{i=l_j+1}^{P} G_{i,T-1} + \frac{1}{2}\left(2N_{j,T-1} + 2q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1} + q_{j,T}\sum_{i=1}^{k_j-1} G_{i,T}\right)\sum_{i=1}^{k_j-1} G_{i,T} + \frac{1}{2(s_j - q_{j,T})}\left(N_{j,T-1} + q_{j,T-1}\sum_{i=l_j+1}^{P} G_{i,T-1} + q_{j,T}\sum_{i=1}^{k_j-1} G_{i,T}\right)^2 \tag{5}$$

Example. Vehicles in lane group j, which can be served by Phases 4 and 5, experience red time equal to the summation of the green time of Phase 6 in the previous cycle $T - 1$ and the green times of Phases 1 through 3 in the design cycle T (Figure 1).

2. The estimated delay for all autos for cycle $T + 1$, $\hat{D}_{T+1}$, is derived from Equation 3 as follows:

$$\hat{D}_{T+1} = \sum_{j=1}^{J} \hat{D}_{j,T+1} \tag{6}$$

If the autos of lane group j experience oversaturated conditions for cycle T (i.e., $N_{j,T} > 0$), their total estimated delay for cycle $T + 1$ is given by

$$\hat{D}_{j,T+1} = \frac{1}{2}\left(2N_{j,T} + q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T}\right)\sum_{i=l_j+1}^{P} G_{i,T} + \frac{1}{2}\left(2N_{j,T} + 2q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T} + q_{j,T+1}\sum_{i=1}^{k_j-1} G_{i\,\text{next}}\right)\sum_{i=1}^{k_j-1} G_{i\,\text{next}} + \left(N_{j,T} + q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T} + q_{j,T+1}\sum_{i=1}^{k_j-1} G_{i\,\text{next}}\right)\sum_{i\in I_j} G_{i\,\text{next}} + \frac{1}{2}\left(q_{j,T+1}\sum_{i\in I_j} G_{i\,\text{next}} - s_j\sum_{i\in I_j} G_{i\,\text{next}}\right)\sum_{i\in I_j} G_{i\,\text{next}} \tag{7}$$

where

$G_{i\,\text{next}}$ = green time for phase i during the next cycle, $T + 1$, which is defined here as

$$G_{i\,\text{next}} = \begin{cases} G_{i\,\min} & \forall i < P \\ C - \sum_{i=1}^{P-1} G_{i\,\min} & \forall i = P \end{cases} \tag{8}$$

$G_{i\,\min}$ = minimum green time for phase i (fixed).

The delay estimate for the autos of lane group j for cycle $T + 1$ for undersaturated conditions (i.e., $N_{j,T} \le 0$) changes accordingly on the basis of Equation 5:

$$\hat{D}_{j,T+1} = \frac{1}{2}\left(2N_{j,T} + q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T}\right)\sum_{i=l_j+1}^{P} G_{i,T} + \frac{1}{2}\left(2N_{j,T} + 2q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T} + q_{j,T+1}\sum_{i=1}^{k_j-1} G_{i\,\text{next}}\right)\sum_{i=1}^{k_j-1} G_{i\,\text{next}} + \frac{1}{2(s_j - q_{j,T+1})}\left(N_{j,T} + q_{j,T}\sum_{i=l_j+1}^{P} G_{i,T} + q_{j,T+1}\sum_{i=1}^{k_j-1} G_{i\,\text{next}}\right)^2 \tag{9}$$

The estimate of the auto delay for the next cycle $T + 1$ is calculated here assuming that all lane groups experience similar traffic conditions. As a result, the minimum green times, $G_{i\,\min}$, are applied for all phases except the last one (it will be allocated green time equal to the residual of the cycle length), because that approach gives the minimum delay. However, the choice of the signal timings for the next cycle $T + 1$ is a user-specified factor that does not affect the structure of the formulas presented.

Example. Vehicles in lane group j, which can be served by Phases 4 and 5 in the next cycle $T + 1$, experience red time equal to the summation of the green time of Phase 6 in the design cycle T and the green times of Phases 1 through 3 in cycle $T + 1$ (Figure 1).

As a result, the first component of the objective function becomes

$$\sum_{a=1}^{A} o_a d_a = \bar{o}_a\left(D_T + \hat{D}_{T+1}\right) \tag{10}$$

An average occupancy per auto, $\bar{o}_a$, is used for simplicity because total vehicle delay is calculated in aggregate rather than accounting for each vehicle separately. However, the delays of each individual vehicle could be easily estimated with the use of queuing diagrams such as the one in Figure 1, given that the arrival time of each vehicle at the end of the queue is known.

Transit Delay

Person delay for the transit vehicles consists of the summation of two terms: (*a*) the person delay experienced by those present at the intersection during design cycle T before the end of their respective green times and (*b*) the estimated person delay for those that arrive before the end of the design cycle T but cannot be served in it. Under the assumption that information on the transit vehicles' location and arrival times is available only for the design cycle T, the transit vehicles that arrive during cycle $T + 1$ are not taken into account. The exclusion of such vehicles is not expected to affect the results significantly. Real-time information about the transit vehicles' location and passenger occupancies is required to determine their actual arrival times at the intersection and estimate their person delays so that priority is provided accordingly.

Because transit vehicles travel in mixed lanes with autos, the delay of the transit vehicle that arrives in its lane group's queue at some time t_b is the same as an auto vehicle that arrives at the same time at the back of that lane group's queue. The delay for a transit vehicle that belongs to a lane group j can be calculated by queuing diagrams as shown in Figure 2.

Estimation of the transit delay used in the optimization of each cycle T depends on the actual arrival time of the transit vehicles, t_b,

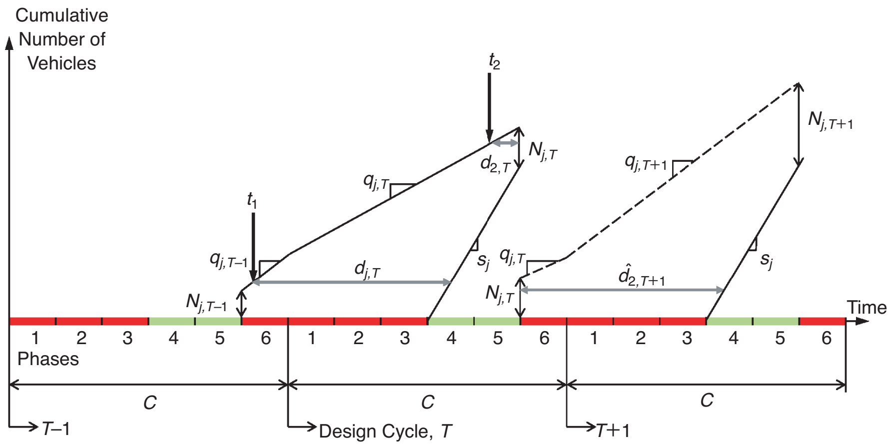

FIGURE 2 Calculation of transit delay for lane group *j*.

as well as whether or not the vehicle is served during cycle T. All cases, along with the respective formulas for estimating the transit delays, are summarized below:

1. If a transit vehicle b that belongs to lane group j has arrived at some time after the end of the last phase that can serve j in the previous cycle $T-1$ and before the end of the respective phases in the design cycle T (i.e., $(T-2)C+\sum_{i=1}^{l_j} G_{i,T-1} \le t_b \le (T-1)C+\sum_{i=1}^{l_j} G_{i,T}$), its delay for cycle T, $d_{b,T}$, is

$$d_{b,T} = \begin{cases} (T-1)C - t_b + \sum_{i=1}^{l_j} G_{i,T} & \text{if } n_{b,T} > 0 \\ (T-1)C - t_b + \sum_{i=1}^{k_j-1} G_{i,T} + \dfrac{n_{b,T-1}}{s_j} & \text{if } n_{b,T} \le 0 \end{cases} \tag{11}$$

where

t_b = arrival time of transit vehicle at the back of the queue;
$n_{b,T}$ = position in queue after the end of the last phase that serves the transit vehicle's lane group j in cycle T, which is calculated as

$$n_{b,T} = n_{b,T-1} - s_j \sum_{i \in I_j} G_{i,T} \tag{12}$$

$n_{b,T-1}$ = position in the queue before the end of the phases that serve the transit vehicle's lane group j in cycle T, which is calculated as

$$n_{b,T-1} = \begin{cases} N_{j,T-1} + q_{j,T-1}\left(t_b - (T-2)C - \sum_{i=1}^{l_j} G_{i,T-1}\right) & \text{if } t_b < (T-1)C \\ N_{j,T-1} + q_{j,T-1} \sum_{i=l_j+1}^{P} G_{i,T-1} \\ \qquad + q_{j,T}\left(t_b - (T-1)C\right) & \text{if } t_b \ge (T-1)C \end{cases} \tag{13}$$

A positive position (i.e., $n_{b,T} > 0$) means that the transit vehicle has not been served by the end of the last phase that can serve its lane group in cycle T and nonpositive means that it has been served.

2. If a transit vehicle b that belongs to lane group j has arrived at some time before the end of the green time that can serve j in the previous cycle $T-1$ (i.e., $t_b \le (T-2)C + \sum_{i=1}^{l_j} G_{i,T-1}$) and the transit vehicle is still present during cycle T, its delay for cycle T, $d_{b,T}$, is

$$d_{b,T} = \begin{cases} \sum_{i=l_j+1}^{P} G_{i,T-1} + \sum_{i=1}^{l_j} G_{i,T} & \text{if } n_{b,T} > 0 \\ \sum_{i=l_j+1}^{P} G_{i,T-1} + \sum_{i=1}^{k_j-1} G_{i,T} + \dfrac{n_{b,T-1}}{s_j} & \text{if } n_{b,T} \le 0 \end{cases} \tag{14}$$

where

$$n_{b,T-1} = n_{b,T-2} - s_j \sum_{i \in I_j} G_{i,T-1} \tag{15}$$

For the transit vehicles that are actually served during the design cycle T (i.e., $n_{b,T} \le 0$), it is possible that they are served the moment they arrive, in which case the delay formula above gives a negative number. For such cases the delay for the transit vehicle is 0. For the transit vehicles that arrive after the end of the green that can serve their respective lane groups, the delay experienced during the design cycle T is assumed to be 0 and the transit vehicle experiences only an estimated delay as shown next.

Example. If a bus in lane group j arrives during Phase 6 of cycle $T-1$, and it can be served during the design cycle T (e.g., $t_b = t_1$ in Figure 2), its delay, $d_{1,T}$, will be as indicated on the queuing diagram (Figure 2).

If a vehicle arrives before the end of the last phase that can serve its lane group but cannot be served, or it arrives after the end of that phase and before the end of the design cycle T, two occurrences are possible: (a) the phases that can serve it will be extended so that the transit vehicle can be served during the design cycle T or (b) the transit vehicle will be served during a subsequent cycle. For optimization purposes, the estimated delay that such a transit vehicle would experience if it cannot be served during cycle T is included in the calculation of the delay for the design of cycle T. The estimated delay of that transit vehicle in the next cycle is based on the assumption that the green times for the phases of the next cycle $T+1$, $G_{i\text{next}}$, are the same as the minimum green times for each phase, $G_{i\min}$, apart from the last phase, as expressed in Equation 8. The estimated delay of such a transit vehicle depends on whether or not it can be served during the next cycle $T+1$ (given the assumptions stated above) and on its arrival time, t_b. The estimation of the delay experienced during the next cycle $T+1$ follows one of the two cases below:

1. If a transit vehicle b that belongs to lane group j has arrived at some time before the end of the green time that can serve j in the design cycle T (i.e., $t_b \le (T-1)C + \sum_{i=1}^{l_j} G_{i,T}$) but cannot be served during that cycle, its estimated delay for cycle $T+1$, $\hat{d}_{b,T+1}$, is

$$\hat{d}_{b,T+1} = \begin{cases} \sum_{i=l_j+1}^{P} G_{i,T} + \sum_{i=1}^{l_j} G_{i\text{next}} & \text{if } n_{b,T+1} > 0 \\ \sum_{i=l_j+1}^{P} G_{i,T} + \sum_{i=1}^{k_j-1} G_{i\text{next}} + \dfrac{n_{b,T}}{s_j} & \text{if } n_{b,T+1} \le 0 \end{cases} \tag{16}$$

where

$$n_{b,T+1} = n_{b,T} - s_j \sum_{i \in I_j} G_{i,T+1} \tag{17}$$

2. If a transit vehicle b that belongs to lane group j has arrived at some time after the end of the green time that can serve j in the design cycle T (i.e., $(T-1)C + \sum_{i=1}^{l_j} G_{i,T} < t_b < TC$) and that phase is not extended to serve it, its estimated delay for cycle $T+1$, $\hat{d}_{b,T+1}$, is

$$\hat{d}_{b,T+1} = \begin{cases} TC + \sum_{i=1}^{l_j} G_{i\text{next}} - t_b & \text{if } n_{b,T+1} > 0 \\ TC + \sum_{i=1}^{k_j-1} G_{i\text{next}} + \dfrac{n_{b,T}}{s_j} - t_b & \text{if } n_{b,T+1} \le 0 \end{cases} \tag{18}$$

where

$$n_{b,T+1} = N_{j,T} + q_{j,T}\left(t_b - (T-1)C - \sum_{i=1}^{l_j} G_{i,T}\right) - s_j \sum_{i \in I_j} G_{i\text{next}} \tag{19}$$

Example. If a bus that belongs to lane group j arrives during Phase 5, but according to the base case signal timings it cannot be served during the design cycle T (e.g., $t_b = t_2$ in Figure 2), there are two options: either it can be served by Phase 5 of the design cycle T if it is possible to extend the green by a sufficient amount to serve the bus, or it will be served during a subsequent cycle. For the optimization process of cycle T, the delay for cycle T, $d_{2,T}$, and an estimate of the delay the bus would experience if it was to be served during the next cycle, $\hat{d}_{2,T+1}$, are taken into account.

Constraints

The first constraint refers to the minimum green times for each lane group. Minimum green times $G_{j\min}$ are necessary to ensure safe pedestrian crossings. In addition, they ensure that no phase is skipped and provide a safety buffer between the phases. Because each phase does not always coincide with serving one lane group, and a lane group can be served by more than one phase, there are many different combinations of minimum green times for the phases that can satisfy the minimum green time requirement. Thus, constraining the green times for each lane group makes it possible to reduce delays more than if minimum green times for the phases were fixed. The second constraint ensures that the green times for each phase, which will be the outcome of the optimization, add up to the cycle length, which is kept constant for every cycle.

Implementation Issues

The assumptions are not expected to affect the formulation of the mathematical program. Relaxing the assumption of deterministic delays will require a different formula for the estimation of delays. The assumptions of fixed phase sequence, design, and as a result, permission of only green extension and phase advance can be relaxed and lead to more flexibility in the design of signal settings, which will result in even lower total person delays.

The formulation and implementation of the proposed optimization process are based on real-time information about the traffic conditions, the arrivals of the transit vehicles, and their passenger occupancies, which need to be communicated to the signal controllers. Traffic conditions can be monitored in real time by inductive loop detectors placed upstream of the intersection at a distance from the stop line such that the arrivals are measured under free-flow conditions. The detectors can provide necessary information for the prediction of arrivals at the intersection.

Automated vehicle location technologies, such as Global Positioning Systems, can be used for tracking transit vehicles continuously and sending the information about their location to the control center. In the absence of automatic vehicle location technologies, other sensing systems such as radio control and special detectors that recognize transit vehicle signatures can be used for detecting and identifying the vehicles. Such detectors have been used by the Los Angeles Department of Transportation for its centralized TSP implementation (*10*).

Advanced technologies have also been developed recently to provide real-time information on passengers' boarding and alighting at the transit stops. Such technologies use mixed infrared sensors or cameras located near the doors that are able to determine the direction of passenger traffic and thus estimate with 95 to 98% accuracy how many people are exiting or entering a transit vehicle (e.g., INFODEV automatic passenger counting, ACOREL onboard counter, and Eurotech passenger counter). These systems can be connected to automatic vehicle location systems to provide real-time transmission of the data.

APPLICATION OF METHODOLOGY

The proposed traffic-responsive signal control system was applied to a real-world intersection: Katechaki and Mesogeion Avenues located in Athens, Greece. This test intersection was selected because of the high traffic volumes on all approaches and the existence of conflicting bus routes.

The intersection's layout is presented in Figure 3. As the figure shows, this is a complex intersection with through and turning traffic

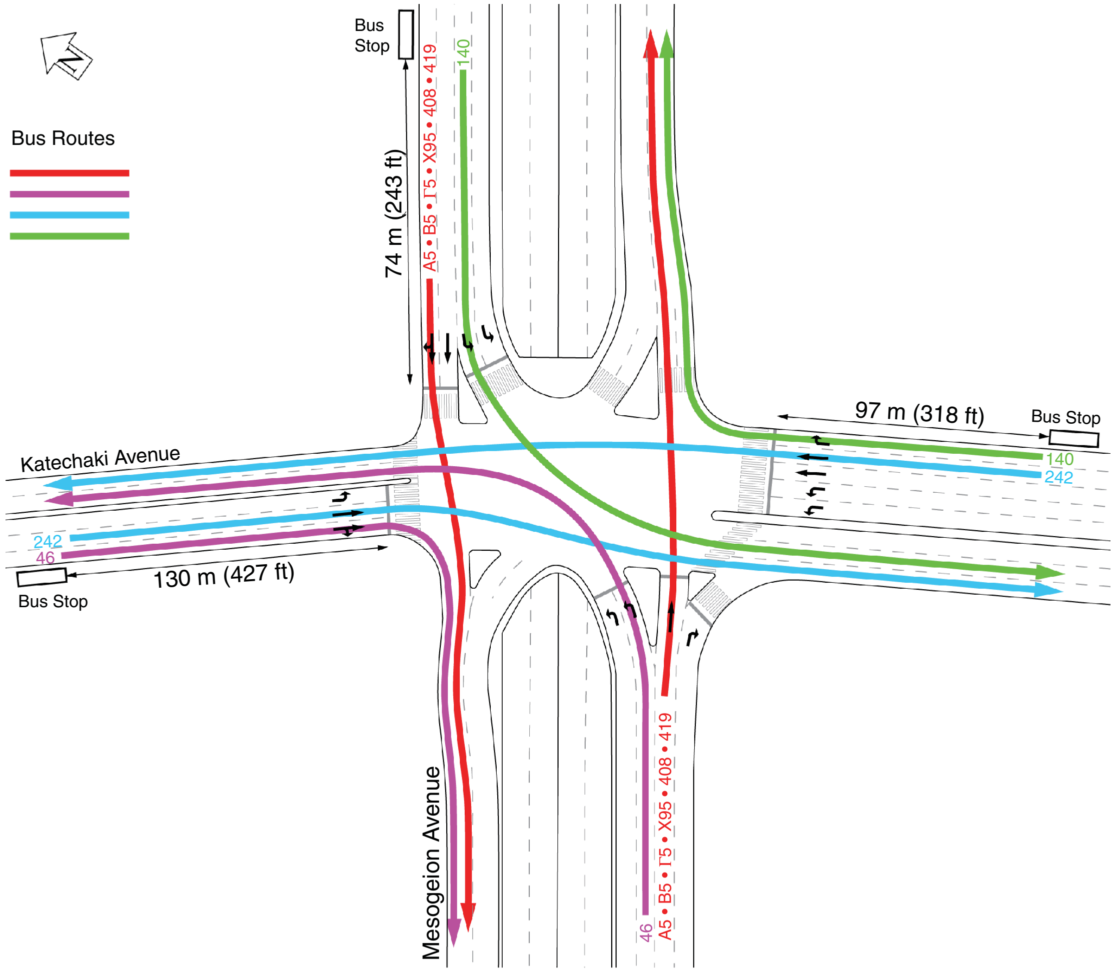

FIGURE 3 Test intersection configuration and bus routes.

in all directions (the main through movement is on Katechaki Avenue). Auto volumes are available from loop detectors placed 40 m (130 ft) upstream for each approach at a rate of once per second.

Nine bus routes travel through the intersection in mixed traffic lanes with headways that vary from 8 to 20 min during peak hours. The numbers next to the directional arrows in Figure 3 correspond to the different bus routes. The bus routes travel in four conflicting directions and their bus stops are located on the nearside (i.e., upstream of the intersection). The southwest approach bus stop is not shown here because of its longer distance from the stop line, which also diminishes its effect on the traffic operations of the intersection. However, the effect of all the bus stops on the operations of the intersection is ignored for the moment. Information about the bus schedule is available on the Athens Urban Transport Organisation website (*12*).

Traffic volumes during the morning peak hour (7 to 8 a.m.) were used as representative initial volumes. A 1-h flow profile was created by increasing the initial flows gradually by 2.5% per cycle up to 1.3 times the initial volumes and then decreasing them gradually by 5% per cycle until they reached values of 0.45 times the initial volumes. The minimum green times for each lane group were defined according to the initial volumes. So, as the volumes increase, traffic could soon operate in oversaturated conditions. The intersection signal is operated on a fixed six-phase cycle with a cycle length of 120 s. Figure 4 shows the lane groups (on the left, labeled 1 to 8r) and the phasing and splits for the intersection during peak hours.

Because no real-time information was available, the actual bus arrival times at the intersection were simulated on the basis of a shifted normal distribution around the buses' scheduled arrival times. The average auto occupancy $\bar{o}_a$ was assumed to be 1.25 passengers per vehicle, which is a reasonable value for autos at this specific intersection during the morning peak hours assumed for this analysis. For the transit vehicles, passenger arrivals at the bus stops are assumed to be deterministic and constant. As a result, the bus occupancy is a function of the time between the actual arrivals of two consecutive buses of the same route (i.e., the buses are assumed to operate as if they arrive empty at the bus stop just upstream of the intersection under consideration). The occupancy of each bus that arrives at the intersection is given by

$$o_b = \phi_r \left(t_{b,r} - t_{b-1,r} \right) \tag{20}$$

where

ϕ_r = passenger demand for bus route r (pax/h),
$t_{b,r}$ = actual arrival time of bus b of route r (h), and
$t_{b-1,r}$ = actual arrival time of bus $b - 1$ of route r (h).

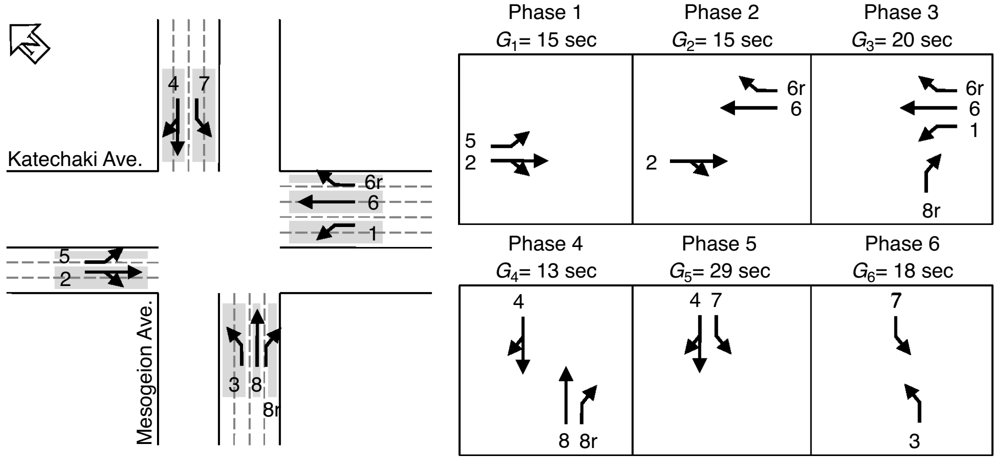

FIGURE 4 Lane groups (*left*), phasing, and splits (*right*) for the test intersection.

In this way, despite the fact that the schedule delay of the buses was not considered initially, it is implicitly included through the higher passenger occupancy expected of late buses. For the initial testing of the signal control optimization, an average bus occupancy of 40 passengers per vehicle was assumed.

RESULTS

Several scenarios with different average bus occupancies were evaluated through a 1-h simulation. For each scenario a warm-up period equal to one cycle length was used. In addition, each scenario was evaluated 10 times to account for the stochastic variation in bus arrivals at the intersection. A total of 90 test cases were performed.

Real data from the intersection of Katechaki and Mesogeion, described in the previous section, and the 1-h generated flow profile were used to test the outcomes of two optimization scenarios: (*a*) when only vehicle delay is minimized (i.e., vehicle-based optimization) and (*b*) when total person delay for both bus and auto passengers is minimized (i.e., person-based optimization). In addition, the base case signal timings were evaluated for each scenario. The corresponding person delays for auto passengers, bus passengers, and all passengers combined are shown in Table 1 for a 1-h simulation of the base case and the optimized signal timings. The table also contains a comparison of the person delays for the two optimized scenarios.

A comparison of the person delays obtained from the proposed optimization process (Scenario 2) with those obtained from minimizing vehicle delays (Scenario 1) indicates the magnitude of the improvement achieved even when traffic operates in oversaturated conditions. Total person delay for the intersection was reduced by 9.46%, and bus passenger delay was reduced by 35.45%. This indicates the improvement for transit operations achieved by using conditional priority. The 2.81% increase in the auto passenger delay is expected as a result of the autos' lower passenger occupancy, which weights them much less during the optimization process. The percentages translate to an increase in the auto delay on the order of 3 s per vehicle on average and a decrease in the bus delay on the order of 19 s per bus on average at the intersection. In general, the effects depend on the auto and transit demand (i.e., vehicle flows) as well as the transit characteristics such as passenger occupancies, headways, and the number of routes traveling through the intersection in combination with the degree of saturation for each of the lane groups.

SENSITIVITY OF RESULTS TO AVERAGE BUS OCCUPANCY

The sensitivity of the results to the buses' average occupancy was captured by simulating scenarios with average bus occupancies of 20, 30, and 40 passengers per vehicle. For each average bus occupancy level, different bus arrival scenarios were tested. The

TABLE 1 Person Delays for 1-h Simulation and Average Bus Occupancy of 40 Passengers per Vehicle

Scenario	Auto Delay (pax h)	Bus Delay (pax h)	Total Delay (pax h)
Base case (existing signal settings)	93.64	34.31	127.95
Vehicle-based optimization	90.07	42.53	132.60
Person-based optimization	92.60	27.46	120.06
Change of Scenario 2 compared with Scenario 1 (%)	2.81	–35.45	–9.46

NOTE: pax h = passenger hours.

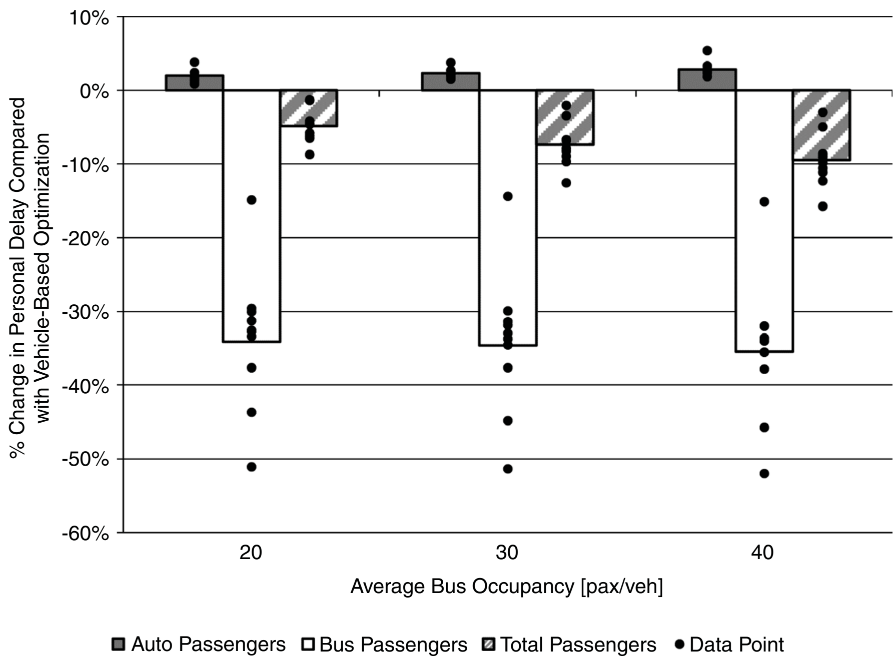

FIGURE 5 Change in person delay for different average bus occupancy scenarios.

average auto occupancy was kept constant and equal to 1.25 passengers per vehicle for all scenarios. Figure 5 shows the changes in the person delay of auto and bus passengers, as well as the total person delay, achieved by the person-based optimization compared with the vehicle-based optimization during the 1-h simulation period for the different average bus occupancy scenarios. Results indicate that although the higher occupancy of buses has a small effect on the reduction in delay for bus passengers, the higher occupancy leads to increased savings in the overall person delay at the intersection. This result is accomplished despite increased delay for auto users.

Figure 5 also shows the variation in the results of the 10 different simulation runs used to account for the stochasticity in bus arrivals, for each average bus occupancy scenario (shown as data points in the figure).

DISCUSSION OF RESULTS

A real-time, traffic-responsive signal control system with TSP has been developed and tested at an isolated intersection under varying traffic conditions. The optimization method explicitly accounts for the passenger occupancy of autos and transit vehicles to assign priority equitably, even for transit vehicles that travel in conflicting directions. At the same time, it prioritizes the approaches with longer residual queues to minimize person delay, thus minimizing the effect on auto traffic.

Results from the application of the optimization method on a real-world intersection show the effectiveness of the proposed traffic-responsive TSP system in reducing the overall person delay as well as the bus passengers' delay by providing priority to buses traveling in conflicting directions. The total person delay of all passengers was reduced by 9.5%, and the delay of bus passengers was reduced by 35.5% compared with the vehicle-based optimization results for the time-dependent volume profile that was generated. At the same time, the increase in the auto passengers' delay was on the order of only 2.8%. The optimization was shown to be effective in reducing total passenger delays at the isolated intersection for a variety of random bus arrivals and average bus occupancies. The tests showed that greater bus occupancy reduces the delay for bus passengers and for all travelers at the intersection and increases delay slightly for auto users.

The proposed real-time, traffic-responsive signal control system with TSP is generic and offers flexibility to weigh the relative merit of the passenger and vehicle delays. It allows for different trade-offs between auto and transit delays by adjusting the passenger weights in the objective function. Next steps in the study include testing additional TSP strategies (e.g., phase rotation) for a wide range of traffic and design characteristics. The test intersection is characterized by more complicated geometry and phasing (Figure 4) than at most intersections in the United States. Therefore, when implemented at simpler intersections, the proposed method is expected to yield similar or better results. Also planned is the evaluation of the proposed system on the basis of several performance measures (e.g., transit schedule adherence, energy consumption, and emissions). Next, by using the presented formulation for the isolated intersection as the stepping stone, the system will be extended to arterials and grid networks.

ACKNOWLEDGMENTS

This research was supported by the Volvo Center of Excellence on Future Urban Transport, University of California, Berkeley. The authors express their appreciation to Ioannis Papamichail for his valuable comments on this manuscript. The authors also thank Dimitris Sermpis for providing the data for the test intersection.

REFERENCES

1. Skabardonis, A. Control Strategies for Transit Priority. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1727,* TRB, National Research Council, Washington, D.C., 2000, pp. 20–26.
2. Chang, E., and A. Ziliaskopoulos. Data Challenges in Development of a Regional Assignment: Simulation Model to Evaluate Transit Signal Priority in Chicago. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1841,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 12–22.
3. Hunt, P., D. I. Robertson, R. D. Bretherton, and M. Royal. The SCOOT On-line Traffic Signal Optimisation Technique. *Traffic Engineering and Control,* Vol. 23, No. 4, 1982, pp. 190–192.
4. Bretherton, D., G. Bowen, and K. Wood. Effective Urban Traffic Management and Control: SCOOT Version 4.4. *Proc., European Transport Conference* (CD-ROM), Homerton College, Cambridge, United Kingdom, Association of European Transport, London, 2002.
5. Cornwell, P., J. Luk, and B. Negus. Tram Priority in SCATS Melbourne. *Traffic Engineering and Control,* Vol. 27, 1986, pp. 561–565.
6. Li, M. *Toward Deployment of Adaptive Transit Signal Priority Systems.* California Partners for Advanced Transit and Highways, Institute of Transportation Studies, University of California, Berkeley, 2008.
7. Henry, J., and J. Farges. P. T. Priority and PRODYN. *Towards an Intelligent Transport System: Proceedings, 1st World Congress on Application of Transport Telematics and Intelligent Vehicle-Highway Systems,* Vol. 6, Artech House, Norwood, Mass., 1995, pp. 3086–3093.
8. Mauro, V., and C. Di Taranto. UTOPIA. *Proc., 6th IFAC-IFIP-IFORS Symposium on Control, Computers, and Communications in Transportation*, Paris, 1989, pp. 245–252.
9. Conrad, M., F. Dion, and S. Yagar. Real-Time Traffic Signal Optimization with Transit Priority: Recent Advances in the Signal Priority Procedure for Optimization in Real-Time Model. In *Transportation Research Record 1634,* TRB, National Research Council, Washington, D.C., 1998, pp. 100–109.
10. Li, Y., P. Koonce, M. Li, K. Zhou, Y. Li, S. Beaird, W.-B. Zhang, L. Hegen, K. Hu, A. Skabardonis, and Z. S. Sun. FTA-CA-26-7049-2008.1. *Transit Signal Priority Research Tools.* FTA, U.S. Department of Transportation, 2008.
11. *Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 2000.
12. Athens Urban Transport Organisation. *Search Route.* www.oasa.gr. Accessed June 6, 2010.

The Traffic Signal Systems Committee peer-reviewed this paper.