Paper No. 13-3395

Arterial Traffic Signal Optimization: A Person-based Approach

Eleni Christofa, Ph.D.
(corresponding author)
Department of Civil and Environmental Engineering
University of Massachusetts
216 Marston Hall, 130 Natural Resources Road
Amherst, MA 01003
Phone: +1(413) 577-3016, Fax: +1(413) 545-9569
Email: christofa@ecs.umass.edu

Konstantinos Aboudolas, Ph.D.
Ecole Polytechnique Fédérale de Lausanne
School of Architecture, Civil and Environmental Engineering
Urban Transport Systems Laboratory
GC C2 390 (Bâtiment GC), Station 18
1015 Lausanne, Switzerland
Phone: +41-21-69-32964
Email: konstantinos.ampountolas@epfl.ch

Alexander Skabardonis, Ph.D.
Department of Civil and Environmental Engineering
Institute of Transportation Studies
109 McLaughlin Hall, University of California, Berkeley, CA 94720
Phone: +1(510) 642-9166, Fax: +1(510) 642-1246
Email: skabardonis@ce.berkeley.edu

For Presentation at the
Transportation Research Board
92nd Annual Meeting
Washington, D.C.
January 13–17, 2013

Word Count: 6,567
Figures/Tables: 5
Total: 7,817

Revised November 15, 2012

ABSTRACT

This paper presents a traffic responsive signal control system that optimizes signal settings based on minimization of person delay on arterials. The system's underlying mixed integer linear program minimizes person delay by explicitly accounting for the passenger occupancy of autos and transit vehicles. This way it can provide signal priority to transit vehicles in an efficient way even when they travel in conflicting directions. Furthermore, it recognizes the importance of schedule adherence for reliable transit operations and accounts for it by assigning an additional weighting factor on transit delays. This introduces another criterion for resolving the issue of assigning priority to conflicting transit routes. At the same time, the system maintains auto vehicle progression by introducing the appropriate delays for when interruptions of platoons occur. In addition to the fact that it utilizes readily available technologies to obtain the input for the optimization, the system's feasibility in real-world settings is enhanced by its low computation time. The proposed signal control system was tested on a segment of San Pablo Avenue arterial located in Berkeley, California. The findings have shown the system's capability to outperform static optimal signal settings and have demonstrated its success in reducing person delay for bus and in some cases even auto users.

INTRODUCTION

With the continuous growth of population and car ownership and the limited funds available, there is an imperative need to design and manage mutimodal transportation systems more efficiently while maximizing the use of existing infrastructure. With traffic signal control systems already widely deployed in urban street networks, one of the most cost-effective ways to improve efficiency and sustainability of urban transportation systems is to develop signal control strategies that enhance person mobility. This can be achieved with the development of signal control strategies that in addition to resolving conflicts between vehicles, give preferential treatment to high occupancy transit vehicles while accounting for the overall traffic conditions in the network.

Several advanced signal control systems have incorporated transit signal priority strategies in their algorithms in order to manage multimodal systems more efficiently. However, none of the systems is explicitly optimizing signal settings by minimizing person delay in a network. On the contrary, they usually minimize vehicle delays (*1, 2, 3*) and provide priority based on rules that are not directly included in the optimization process (*4*) or pre-select a subset of transit vehicles to apply their priority strategies (*5, 6*). As a result, existing systems lack an efficient way of assigning priority to transit vehicles when they are traveling on conflicting routes. Furthermore, they often ignore the importance of transit schedule adherence in providing priority which in some cases can cause further disruptions to the transit system. Recently, an adaptive signal control system was developed that utilizes information from mobile sources to optimize signal settings on arterial networks (*7*). However, the requirement for high penetration of probe vehicles and the long computation times constraint its applicability in the real world.

A person-based traffic responsive signal control system for isolated intersections was recently developed by the authors (*8, 9*). The system minimizes person delay by explicitly accounting for the passenger occupancy of autos and transit vehicles. This results to provision of signal priority to transit vehicles and introduces an efficient way for resolving the issue of priority assignment when transit vehicles travel in conflicting directions. The system uses information that can be obtained form currently deployable surveillance and communication technologies (i.e., detectors, Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) systems). This paper presents an extension of the existing traffic responsive signal control system to arterials, that accounts for vehicle progression and transit schedule adherence.

The paper is organized as follows: first, the optimization procedure and the underlying mathematical model that minimizes person delay are described. The next sections describe the test arterial and present the application of the signal control system. Finally, the paper discusses the findings of this work and outlines areas for future research.

MATHEMATICAL MODEL

The optimization of signal settings for an arterial is based on a pairwise optimization strategy introduced by Newell (*10, 11*). According to this strategy signal timings are optimized for a pair of consecutive intersections. Therefore, the mathematical program is formulated to minimize the total person delay at two consecutive intersections, r and $r+1$, for all vehicles that are present during the design cycle, T. The optimization process starts by determining the critical intersection of the subject arterial, which could be defined as the one with the highest intersection flow ratio (i.e., demand to saturation flow ratio for critical lane groups) or the one with the highest transit traffic. Starting with the critical intersection, progression is maintained for the heaviest direction of traffic on the arterial. Once the signal settings for the first two intersections, r and $r+1$, are optimized, the next pair, $r+1$ and $r+2$, will be optimized. For this optimization, the beginning of green for the coordinated phase (i.e., phase that serves the heaviest direction) at $r+1$ will be constrained by the optimization outcome of r and $r+1$. This constraint ensures that the beginning of the green for that phase will be held constant when optimizing the second pair of intersections. Assuming that

the yellow times are constant, this can be expressed as:

$$\sum_{i=1}^{c^{r+1}-1} g_{i,T}^{r+1}(2) = \sum_{i=1}^{c^{r+1}-1} g_{i,T}^{r+1}(1) \tag{1}$$

where c^{r+1} is the phase that serves the heaviest direction at intersection $r+1$, $g_{i,T}^{r+1}(1)$ are the optimal green times for phase i during cycle T at intersection $r+1$ obtained from the optimization of the first pair of intersections, and $g_{i,T}^{r+1}(2)$ are the corresponding green times obtained from the optimization of the second pair of intersections in which intersection $r+1$ belongs. The same pairwise optimization is repeated until all intersections in the subject arterial network are optimized. In case it is not clear which direction has the heaviest traffic, the same process can be repeated in the opposing traffic direction, and the signal settings that give the lowest total person delay can be chosen.

The mathematical program that minimizes total person delay at two consecutive intersections is formulated under the assumption of perfect information on traffic demand and passenger occupancies. It is also based on the assumption of deterministic vehicle arrivals (for delay estimation purposes), fixed phase sequence, and constant lane capacities. Vehicles arrive at each intersection in platoons when traveling on the arterial and on the cross-street links, since the subject arterial is considered to be part of a larger arterial signalized network. It is also assumed that there is negligible platoon dispersion. The cycle length is kept constant for the analysis period and it is common for all intersections along the arterial to maintain signal coordination. Finally, the model is formulated assuming that transit vehicles travel on mixed-use traffic lanes along with autos. However, the formulation of the mathematical model holds even when dedicated lanes for transit vehicles exist.

The generalized formulation of the mathematical program that minimizes person delay for two consecutive intersections and for a cycle T is as follows:

$$\min \quad \sum_{r=1}^{2} \left[\sum_{a=1}^{A_T^r} o_a d_a^r + \sum_{b=1}^{B_T^r} o_{b,T}^r (1+\delta_{b,T}^r) d_b^r \right] \tag{2a}$$

$$\text{s.t.} \quad d_{a,T}^r = d_a^r\left(g_{i,T}^r\right) \tag{2b}$$

$$d_{b,T}^r = d_b^r\left(g_{i,T}^r\right) \tag{2c}$$

$$g_{i\min}^r \le g_{i,T}^r \le g_{i\max}^r \tag{2d}$$

$$\sum_{i=1}^{I^r} g_{i,T}^r + L^r = C \tag{2e}$$

where:
r: intersection index
a: auto vehicle index
b: transit vehicle index
A_T^r: total number of autos present at intersection r during cycle T
B_T^r: total number of transit vehicles present at intersection r during cycle T
o_a: passenger occupancy of auto a $[\frac{\text{pax}}{\text{veh}}]$
$o_{b,T}^r$: passenger occupancy of transit vehicle b for cycle T at intersection r $[\frac{\text{pax}}{\text{veh}}]$
$d_{a,T}^r$: delay for auto a for cycle T at intersection r [sec]
$d_{b,T}^r$: delay for transit vehicle b for cycle T at intersection r [sec]
$\delta_{b,T}^r$: factor determining the weight for schedule delay of transit vehicle b for cycle T at intersection r
$d_a^r\left(g_{i,T}^r\right)$: function relating the delay for auto a to green times
$d_b^r\left(g_{i,T}^r\right)$: function relating the delay for transit vehicle b to green times

$g^r_{i,T}$: green time allocated to phase i in cycle T at intersection r [sec]
$g^r_{i\min}$: minimum green time for phase i at intersection r [sec]
$g^r_{i\max}$: maximum green time for phase i at intersection r [sec]
I^r: total number of phases in a cycle for intersection r
L^r: total lost time at intersection r [sec]
C: cycle length [sec].

The mathematical program is run once for every cycle for each pair of intersections and its objective function consists of the sum of the delay for all auto and transit passengers that are present at the intersection during that cycle T. Delays for autos, $d^r_{a,T}$, and transit vehicles, $d^r_{b,T}$, depend on the green times, $g^r_{i,T}$, which are the decision variables of the mathematical program. In fact, $d^r_{a,T}$ and $d^r_{b,T}$ also depend on the green times of the previous and next cycles, the yellow times, and the traffic demand. These are either pre-specified by the user or collected with the use of surveillance technologies. To simplify the notation, the delays for each platoon, lane group, and transit vehicle are included in the objective function as a variable and this variable is constrained to equal a function as shown in equations (2b) and (2c). The optimal green times also determine the beginning time of the coordinated phase (i.e., offset). As a result, offsets are effectively optimized through this process in those cases that the coordinated phase is not the first one in the cycle. Otherwise, offsets cannot be changed, because the cycle length remains constant. In those cases, in order to maintain progression in a selected direction, the offset between successive intersections is set equal to the average free flow travel time.

The delays of both autos and transit vehicles are weighted by their respective passenger occupancies in the objective function. The delays for transit vehicles are also weighted by a factor $(1+\delta^r_{b,T})$ in order to account for the schedule delay that a transit vehicle b has when arriving at intersection r during cycle T. This factor, $\delta^r_{b,T}$, which is user-specified, can be a linear function of the schedule delay of the transit vehicle or a binary variable indicating whether a transit vehicle is ahead or behind schedule based on a predetermined threshold. In either case, the delay for a transit vehicle that is behind schedule is weighted more than a transit vehicle that is arriving early or on time at the intersection.

Three constraints are introduced for the decision variables. The green times of each phase i and intersection r are constrained by their minimum and maximum green times (constraint (2d)). Minimum green times, $g^r_{i\min}$, are necessary to ensure safe vehicle and pedestrian crossings and guarantee that no phase is skipped. Maximum green times, $g^r_{i\max}$, are used to restrict the domain of solutions for the green times of the phases and reduce computation times. The phase green times are also constrained such that the sum of the green times for all phases at each intersection plus the total lost time adds up to the cycle length (constraint (2e)). The total lost time is assumed to be the summation of the yellow times per phase. After the first pair is optimized the additional constraint described in equation (1) is added for each subsequent pair of intersections to ensure that the optimal decision from the optimization of the previous pair is accounted for in the optimization of the next one.

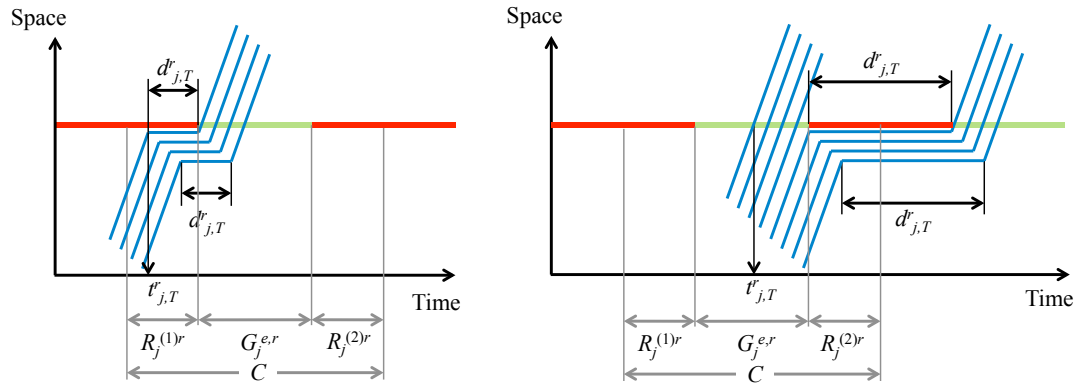
Auto Delay

For each pair of intersections, the auto delays that contribute to the objective function of the optimization consist of three terms: 1) the delay experienced by vehicles that travel in platoons on incoming links during cycle T, 2) the delay experienced by vehicles that travel on shared links during cycle T, and 3) the delay experienced by vehicles that did not get served during the previous cycle, which constitute the residual queues at the approaches of the two intersections during cycle T. This means that during cycle T a platoon could be experiencing delay while traveling on the incoming link (approaching the first intersection of the arterial it arrives at) and a portion of that platoon that continues in the subject network could be experiencing delay while traveling on the shared link (approaching the second intersection of the arterial it arrives at). These

two delay components ensure that the effect of disrupting progression is accounted for in both directions.

Since the optimization is conducted using a pairwise approach, the delays are calculated for each pair of intersections r and $r+1$ in order to optimize the signal settings for that pair. There is symmetry in the formulas for the delays of vehicles traveling in the direction of progression and those traveling in the opposing direction. Suppose that r is the first intersection of the pair being optimized, and the optimization sequence is r and then $r+1$, which is the second intersection of the pair. For any platoon, the first intersection at which it arrives is denoted by u, and the second intersection at which a portion of it arrives is denoted by v. This means that for a platoon traveling in the direction of progression $u = r$ and $v = r+1$, while for a platoon traveling in the opposing direction $u = r+1$ and $v = r$. The same holds for transit vehicles. This notation is used for the remainder of the paper for both auto and transit delays.

The auto delays are estimated based on the assumption that vehicles arrive and are served at both intersections at capacity since they travel in platoons with no dispersion. Consequently, assuming that kinematic wave theory (*12, 13*) holds, all vehicle trajectories are parallel at all times, as shown in Figure 1. This means that the first and last vehicle that get stopped in a platoon will experience the same delay. So, the collective delay for all vehicles can be easily estimated knowing only the arrival time of the first vehicle in a platoon at the back of its lane group's queue at intersection r, $t^r_{j,T}$, the size of that platoon, $P^r_{j,T}$, and the traffic conditions at the approach as expressed by the size of the residual queue of lane group j at the end of the previous cycle $T-1$, $N^r_{j,T-1}$.



(a) Stopping the head of the platoon (b) Stopping the tail of the platoon

FIGURE 1 Auto delay estimation for platoon arrivals.

The estimation of the auto delay components of the objective function is presented next. Note that for simplicity and reduced computation time, all equations are formulated assuming that there is only one platoon per cycle per lane group. However, the algorithm can be easily extended to include multiple platoons in a cycle for the same lane group as long as the arrival times and sizes of those platoons are known.

To further facilitate the notation of the mathematical program, the cycle time for each lane group can be split into three components which are functions of the green times for each phase. The first is the component of the red time from the start of the cycle to the beginning of the green for the subject lane group, $R_j^{(1)r}(g^r_{i,T})$, the second is the duration of the effective green time itself, $G_j^{e,r}(g^r_{i,T})$, and the third is the component of the red time from the end of the green until the end of the cycle, $R_j^{(2)r}(g^r_{i,T})$. These values are illustrated in Figure 1 and are calculated as follows:

$$R_j^{(1)r}(g_{i,T}^r) = \sum_{i=1}^{k_j^r-1} g_{i,T}^r + \sum_{i=1}^{k_j^r-1} y_i^r \tag{3a}$$

$$G_j^{e,r}(g_{i,T}^r) = \sum_{i=k_j^r}^{l_j^r} g_{i,T}^r + \sum_{i=k_j^r}^{l_j^r-1} y_i^r \tag{3b}$$

$$R_j^{(2)r}(g_{i,T}^r) = \sum_{i=l_j^r+1}^{I^r} g_{i,T}^r + \sum_{i=l_j^r}^{I^r} y_i^r \tag{3c}$$

where y_i^r is the yellow time interval after phase i at intersection r, $g_{i,T}^r$ is the green time for phase i in cycle T at intersection r, k_j^r is the first phase in a cycle that can serve lane group j at intersection r, and l_j^r is the last phase in a cycle that can serve lane group j at intersection r. Note that in order to simplify the illustration in Figure 1 the yellow time intervals have not been marked on the time axes, but are considered to be included at the end of each phase.

1) Auto Delay for Vehicles in Platoons on Incoming Links

The auto delay for vehicles in platoons that travel on incoming links to an intersection u consists of two components: 1) the delay caused by stopping the head of the platoon, $D_{j,T}^{(H)u}$, which includes any delay incurred before the vehicles start being served and 2) the delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$, which corresponds to the delay experienced by vehicles after the platoon starts being served or after the end of green phase that serves it (whichever comes first) and until the beginning of the green phase that serves it in the next cycle. Based on the arrival time of a platoon at the back of its lane group's queue, $t_{j,T}^u$, the size of the platoon, $P_{j,T}^u$, and the traffic conditions at the intersection, the delay for autos in that platoon can be estimated as in one of the following six cases:

- *Case P1: Arrival before residual queue served, entire platoon served in green*

  A platoon of size $P_{j,T}^u$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^u$ before the time that the corresponding residual queue of j from the previous cycle $T-1$, $N_{j,T-1}^u$, would have finished being served if there was enough green time available. There is enough available green time to serve the residual queue, and spare green time to serve all $P_{j,T}^u$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u} \tag{4a}$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u) s_j^u \tag{4b}$$

$$P_{j,T}^u \leq G_j^{e,u}(g_{i,T}^u) s_j^u - N_{j,T-1}^u \tag{4c}$$

  where s_j^u is the saturation flow for lane group j at intersection u and t_T^u is the beginning of cycle T for intersection u, which is determined as:

$$t_T^u = t_T + O_T^u. \tag{5}$$

  $t_T = (T-1)C$ is the beginning of cycle T at the critical intersection which is the first one to be optimized and O_T^u is the difference between the starting time of cycle T at intersection u and the

critical intersection. The number of vehicles in the residual queue, $N^u_{j,T-1}$, is calculated as:

$$N^u_{j,T-1} = \max\left\{P^u_{j,T-1} + N^u_{j,T-2} - G^{e,u}_j(g^u_{i,T-1})s^u_j, 0\right\}, \tag{6}$$

but it can also be obtained with the use of detectors upstream of the stop line.

All vehicles in the platoon experience delay caused by stopping the head of the platoon, $D^{(H)u}_{j,T}$:

$$D^{(H)u}_{j,T} = P^u_{j,T}\left(t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} - t^u_{j,T}\right) \tag{7}$$

but no delay caused by stopping the tail of the platoon, $D^{(T)u}_{j,T}$.

- *Case P2: Arrival before residual queue served, insufficient green to serve entire platoon*

  A platoon of size $P^u_{j,T}$ arrives at the back of its lane group's queue at time $t^u_{j,T}$ before the time that the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P^u_{j,T}$ vehicles. These conditions are summarized as:

$$t^u_{j,T} \leq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{8a}$$

$$N^u_{j,T-1} \leq G^{e,u}_j(g^u_{i,T})s^u_j \tag{8b}$$

$$P^u_{j,T} \geq G^{e,u}_j(g^u_{i,T})s^u_j - N^u_{j,T-1}. \tag{8c}$$

  All vehicles in the platoon experience delay caused by stopping the head of the platoon, $D^{(H)u}_{j,T}$, and a portion of the vehicles experience delay caused by stopping the tail of the platoon, $D^{(T)u}_{j,T}$:

$$D^{(H)u}_{j,T} = P^u_{j,T}\left(t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} - t^u_{j,T}\right) \tag{9a}$$

$$D^{(T)u}_{j,T} = \left(P^u_{j,T} - G^{e,u}_j(g^u_{i,T})s^u_j + N^u_{j,T-1}\right)\left(C - \frac{N^u_{j,T-1}}{s^u_j}\right). \tag{9b}$$

  The delay caused by stopping the tail of the platoon is equal to one cycle length minus the time it takes to serve the residual queue. This component is subtracted in order to avoid double counting since that delay component has already been captured by equation (9a). However, one can adjust this delay estimate to change the penalty imposed for stopping the tail of the platoon.

- *Case P3: Arrival before end of green, insufficient green to serve residual queue*

  A platoon of size $P^u_{j,T}$ arrives at the back of its lane group's queue at time $t^u_{j,T}$, before the end of the phase that can serve it, but there is not enough available green time to serve all $N^u_{j,T-1}$ vehicles in the

residual queue. These conditions are summarized as:

$$t^u_{j,T} \leq t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}) \tag{10a}$$

$$N^u_{j,T-1} \geq G^{e,u}_j(g^u_{i,T}) s^u_j. \tag{10b}$$

All vehicles in the platoon experience delay caused by stopping the head of the platoon, $D^{(H)u}_{j,T}$, and by stopping the tail of the platoon, $D^{(T)u}_{j,T}$:

$$D^{(H)u}_{j,T} = P^u_{j,T}\left(t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}) - t^u_{j,T}\right) \tag{11a}$$

$$D^{(T)u}_{j,T} = P^u_{j,T}\left(C - G^{e,u}_j(g^u_{i\,\text{next}})\right) \tag{11b}$$

where $g^u_{i\,\text{next}}$ is a pre-specified value for the green time of phase i for the next cycle $T+1$ at intersection u.

- *Case P4: Arrival after residual queue served, entire platoon served in green*

  A platoon of size $P^u_{j,T}$ arrives at the back of its lane group's queue at time $t^u_{j,T}$ after the time that the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served. There is enough available green time to serve the residual queue, and there is enough spare green time to serve all $P^u_{j,T}$ vehicles. These conditions are summarized as:

$$t^u_{j,T} \geq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{12a}$$

$$N^u_{j,T-1} \leq G^{e,u}_j(g^u_{i,T}) s^u_j \tag{12b}$$

$$P^u_{j,T} \leq \left(t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}) - t^u_{j,T}\right) s^u_j. \tag{12c}$$

  In this case, vehicles in the platoon do not experience any delay at intersection u. As a result, both the delay caused by stopping the head of that platoon, $D^{(H)u}_{j,T}$, and by stopping the tail of the platoon, $D^{(T)u}_{j,T}$, are zero.

- *Case P5: Arrival after residual queue served, insufficient green to serve entire platoon*

  A platoon of size $P^u_{j,T}$ arrives at the back of its lane group's queue at time $t^u_{j,T}$ after the time that the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P^u_{j,T}$ vehicles. These conditions are summarized as:

$$t^u_{j,T} \geq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{13a}$$

$$N^u_{j,T-1} \leq G^{e,u}_j(g^u_{i,T}) s^u_j \tag{13b}$$

$$P^u_{j,T} \geq \left(t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}) - t^u_{j,T}\right) s^u_j. \tag{13c}$$

A portion of vehicles in the platoon experience delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$:

$$D_{j,T}^{(T)u} = \left[P_{j,T}^{u} - \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u\right) s_j^u\right] \left(t_{T+1}^u - t_{j,T}^u + R_j^{(1)u}(g_{i\,\text{next}}^u)\right) \tag{14}$$

but no delay caused by stopping the head of the platoon, $D_{j,T}^{(H)u}$.

- *Case P6: Arrival after the green*

  A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$ after the end of the phase that can serve it. This case captures all arrivals not satisfying the conditions of cases P1 through P5, and it can also be expressed as:

$$t_{j,T}^u > t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u). \tag{15}$$

  All vehicles in the platoon experience delay caused by stopping the tail of the platoon, $D_{j,T}^{(T)u}$:

$$D_{j,T}^{(T)u} = P_{j,T}^u \left(t_{T+1}^u - t_{j,T}^u + R_j^{(1)u}(g_{i\,\text{next}}^u)\right) \tag{16}$$

  but no delay caused by stopping the head of the platoon, $D_{j,T}^{(H)u}$.

2) Auto Delay for Vehicles in Platoons on Shared Links

The auto delay for vehicles in platoons that travel on shared links on the main arterial from one intersection u to another v can be estimated with the use of the equations for the six cases described above. The only differences are the values for the platoon size that need to be adjusted based on the portion of the platoon continuing downstream and its arrival time at the second intersection, v, which is a function of its arrival and therefore service time at the first intersection u. In addition, some of the delay estimates for stopping the tail of the platoon are modified.

An estimate of the size of the platoon in lane group j at the second intersection, v, during cycle T, denoted by $\hat{P}_{j,T}^v$, is used in the optimization instead of the actual incoming platoon size. The estimate can be obtained with data from detectors located at the upstream end of the shared link between the two intersections and information on the percentage that is expected to join the subject lane group j. This information can be obtained by stop line detectors. The platoon size estimate is used instead of a direct calculation of platoon size from the signal settings and arrivals at the upstream intersection, because it reduces the number of bilinearities and trilinearities in the objective function, and as a result, it decreases the computation time of the optimization process. In addition, in cases that the number of lanes changes throughout the arterial, the saturation flow and delays at intersection v are normalized to represent saturation flow and delays on a per lane basis.

The arrival time for a platoon at the back of its lane group's queue at the second intersection, $t_{j,T}^v$, is estimated based on the arrival case for the first intersection as follows:

$$t_{j,T}^v = \begin{cases} t_T^u + R_j^{(1)u}(g_{i,T}^u) + tt_{j,u}^v & \text{for cases P1, P2, P3, and P6} \\ t_{j,T}^u + tt_{j,u}^v & \text{for cases P4, P5} \end{cases} \tag{17}$$

where $tt_{j,u}^v$ is the average free flow travel time to traverse the shared links between intersections u and v. For cases P1, P2, P3, and P6 the estimate of the platoon's arrival time at v is based on the assumption that vehicles from the incoming platoon join the vehicles in the residual queue at u and travel together as one

platoon. As a result, this new platoon is assumed to arrive at the downstream intersection $tt^v_{j,u}$ seconds after the beginning of green at u when the residual queue starts being served. For cases P4 and P5, the platoons are served by u as soon as they arrive. This implies that residual queues are short, and as a result, the majority of vehicles at the downstream intersection, v, can be assumed to be mainly vehicles that have just been served by the upstream intersection u. Therefore, it is assumed that the arrival time of the platoon at v depends only on the service time of the platoon at u.

Finally, the delay estimates for stopping the tail of the platoon for cases P5 and P6 (equations (14) and (16)) are adjusted as follows to avoid introducing non-linear terms in the objective function:

$$D^{(T)v}_{j,T} = \begin{cases} \left[\hat{P}^v_{j,T} - \left(t^v_T + R^{(1)v}_j(g^v_{i,T}) + G^{e,v}_j(g^v_{i,T}) - t^v_{j,T}\right) s^v_j\right] \left(C - G^{e,v}_j(g^v_{i\,\text{next}})\right) & \text{for case P5} \\ \hat{P}^v_{j,T}\left(C - G^{e,v}_j(g^v_{i\,\text{next}})\right) & \text{for case P6.} \end{cases} \tag{18}$$

3) Auto Delay for Vehicles in Residual Queues

The auto delay for vehicles in residual queues at both intersections of the pair r and $r+1$ are estimated based on the size of the residual queue and whether or not it can be entirely served during cycle T. Two cases arise which are described next along with the corresponding delay equations for an intersection r (but also hold for $r+1$):

- *Case R1: Residual queue served in green*

  The residual queue of a lane group j, $N^r_{j,T-1}$, can be entirely served during cycle T:

  $$N^r_{j,T-1} \leq G^{e,r}_j(g^r_{i,T}) s^r_j. \tag{19}$$

  So, the total delay experienced by all vehicles in the residual queue, $D^{(Q)r}_{j,T}$, is:

  $$D^{(Q)r}_{j,T} = N^r_{j,T-1} R^{(1)r}_j(g^r_{i,T}). \tag{20}$$

- *Case R2: Residual queue not entirely served in green*

  The residual queue of a lane group j, $N^r_{j,T-1}$, cannot be entirely served during cycle T:

  $$N^r_{j,T-1} \geq G^{e,r}_j(g^r_{i,T}) s^r_j. \tag{21}$$

  So, the total delay experienced by all vehicles in the residual queue, $D^{(Q)r}_{j,T}$, is:

  $$D^{(Q)r}_{j,T} = N^r_{j,T-1} R^{(1)r}_j(g^r_{i,T}) + \left(N^r_{j,T-1} - G^{e,r}_j(g^r_{i,T}) s^r_j\right) C \tag{22}$$

  because the vehicles that do not get served will have to wait for an extra cycle before they start being served.

The auto delay component of the objective function that minimizes person delay at two consecutive intersections is as follows:

$$\sum_{r=1}^{2} \sum_{a=1}^{A^r_T} o_a d^r_a = \sum_{u=1}^{2} \sum_{j \in J^u_{IN}} \bar{o}_a \left(D^{(H)u}_{j,T} + D^{(T)u}_{j,T}\right) + \sum_{v=1}^{2} \sum_{j \in J^v_{SH}} \bar{o}_a \left(D^{(H)v}_{j,T} + D^{(T)v}_{j,T}\right) + \sum_{r=1}^{2} \sum_{j=1}^{J^r} \bar{o}_a D^{(Q)r}_{j,T} \tag{23}$$

where $\bar{o}_a$ is the average auto passenger occupancy, J^u_{IN} is the set of lane groups for the incoming links at intersection u, J^v_{SH} is the set of lane groups for the shared links at intersection v, and J^r is the total number of lane groups at intersection r. The delay components can be estimated with the use of the corresponding case equation for each of the three delay types as presented above. For each platoon and lane group, each case introduces a binary decision variable and the related constraints. The summation of binary variables for a platoon or lane group for a specific type of delay (i.e., at intersection u, v or when in residual queue) is constrained to be one to ensure that only one of the delay equations will be added to the objective function for that delay type.

Transit Delay

In addition to auto delay, the objective function includes the delay for transit vehicles present at the two intersections during cycle T, which consists of two terms: 1) the delay transit vehicles experience when traveling on incoming links during cycle T, and 2) the delay they experience when traveling on shared links during cycle T. In addition, transit vehicles that do not get served during the cycle in which they arrive experience an extra component of delay equal to $R^{(1)r}_j(g^r_{i\,\text{next}})$, which is the time a transit vehicle would experience if it was the first one in the queue to be served in the next cycle. Since transit vehicles travel in mixed-use traffic lanes and perfect information about their arrivals is assumed, their delays can be estimated as in the cases of platoons, further assuming that a transit vehicle behaves similarly to a platoon of size one.

1) Transit Delay for Vehicles on Incoming Links

The transit delay for vehicles traveling on incoming links to an intersection u, depends on the actual arrival time at the back of its lane group's queue at that intersection, $t^u_{b,T}$, as well as whether the vehicle is served during cycle T or not, which also depends on traffic conditions on the subject approach. Note that the delay equations are formulated based on the assumption that a transit vehicle arrives at the back of the queue before or after the arrival of the platoon due to its dwell time at bus stops. This means that when arriving at the intersection, the bus observes only the residual queue in front of it. The four delay estimation cases for transit arrivals are presented next.

- *Case T1: Arrival before residual queue served, transit vehicle served in green*

  A transit vehicle arrives at the back of its lane group's queue at time $t^u_{b,T}$ before the time the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served and there is enough available green time to serve the residual queue in front of it. These conditions are summarized as:

$$t^u_{b,T} \leq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{24a}$$

$$N^u_{j,T-1} \leq G^{e,u}_j(g^u_{i,T}) s^u_j. \tag{24b}$$

  The transit vehicle is served during cycle T and its delay, $d^u_{b,T}$, can be expressed as:

$$d^u_{b,T} = t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} - t^u_{b,T}. \tag{25}$$

- *Case T2: Arrival before residual queue served, transit vehicle not served in green*

  A transit vehicle arrives at the back of its lane group's queue at time $t^u_{b,T}$ before the time the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served, but there is not enough available

green time to serve the residual queue in front of it. These conditions are summarized as:

$$t^u_{b,T} \leq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{26a}$$

$$N^u_{j,T-1} \geq G^{e,u}_j(g^u_{i,T}) s^u_j. \tag{26b}$$

The transit vehicle is not served during cycle T and its delay, $d^u_{b,T}$, can be expressed as:

$$d^u_{b,T} = t^u_{T+1} - t^u_{b,T} + R^{(1)u}_j(g^u_{i\,\text{next}}). \tag{27}$$

- *Case T3: Arrival after residual queue served and before the end of green*

  A transit vehicle arrives at the back of its lane group's queue at time $t^u_{b,T}$ after the time the corresponding residual queue, $N^u_{j,T-1}$, would have finished being served and before the end of the green time for the phase that can serve it. There is enough available green time to serve the residual queue in front of it. These conditions are summarized as follows:

$$t^u_{b,T} \geq t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} \tag{28a}$$

$$t^u_{b,T} \leq t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}) \tag{28b}$$

$$N^u_{j,T-1} \leq G^{e,u}_j(g^u_{i,T}) s^u_j. \tag{28c}$$

  The transit vehicle is served as soon as it arrives at the intersection and as a result its delay, $d^u_{b,T}$, is zero.

- *Case T4: Arrival after the end of green*

  A transit vehicle arrives at the back of its lane group's queue during cycle T at time $t^u_{b,T}$ after the end of the green time for the phase that can serve it, which is expressed as follows:

$$t^u_{b,T} > t^u_T + R^{(1)u}_j(g^u_{i,T}) + G^{e,u}_j(g^u_{i,T}). \tag{29}$$

  The transit vehicle is not served during cycle T, and its delay, $d^u_{b,T}$, can be expressed as:

$$d^u_{b,T} = t^u_{T+1} - t^u_{b,T} + R^{(1)u}_j(g^u_{i\,\text{next}}). \tag{30}$$

2) Transit Delay for Vehicles on Shared Links

As with autos, the delay for transit vehicles traveling on shared links on the main arterial from one intersection u to another, v, can be estimated with the use of the equations for the four cases presented above given that the transit arrival time at the second intersection, v, $t^v_{b,T}$, can be estimated. $t^v_{b,T}$ depends on the arrival case at the first intersection, u, and can be estimated as follows:

$$t^v_{b,T} = \begin{cases} t^u_T + R^{(1)u}_j(g^u_{i,T}) + \frac{N^u_{j,T-1}}{s^u_j} + tt^v_{b,u} & \text{for case T1} \\ t^u_{b,T} + tt^v_{b,u} & \text{for case T3} \end{cases} \tag{31}$$

where $tt^v_{b,u}$ is the expected travel time for the shared link between intersections u and v for a transit vehicle b, and it includes the lost time due to transit stops. For cases T2 and T4, the transit vehicle is not served

during cycle T at u. As a result, there is no delay at the downstream intersection, v, included in the objective function for cycle T.

The transit delay component of the objective function that minimizes person delay at two consecutive intersections is as follows:

$$\sum_{b=1}^{B_T^u} o_{b,T}^u (1+\delta_{b,T}^u) d_{b,T}^u + \sum_{b=1}^{B_T^v} o_{b,T}^v (1+\delta_{b,T}^v) d_{b,T}^v \quad (32)$$

where $d_{b,T}^u$ and $d_{b,T}^v$ can be estimated per one of the cases presented above. As for the auto delays, each case introduces a binary decision variable and the related constraints. The summation of binary variables for a transit vehicle at intersection u or v is constrained to add up to one to ensure that only one of the delay equations will be added to the objective function per transit vehicle for that intersection.

The objective function of the mathematical program has bilinearities and trilinearities caused by multiplication between continuous and binary decision variables, which introduces non-linearities in the objective function and the constraints. To avoid this problem, convex relaxations for bilinearities and trilinearities as described in (*14*) are used. After performing the convex relaxations, the mathematical program consists of an objective function that is linear in its continuous and binary variables and has linear constraints. Therefore, it is a mixed-integer linear program that can be solved very quickly with CPLEX (*15*). A detailed description of the complete mathematical program can be found in (*16*).

TEST SITE

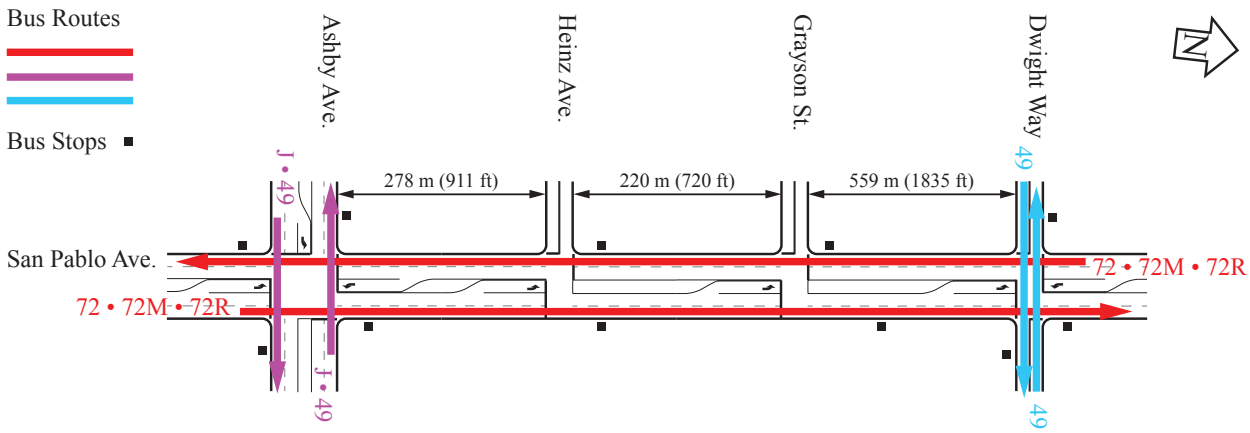
The performance of the person-based traffic responsive signal control system is tested at a four-intersection segment of a real-world signalized arterial. In particular, the test site consists of the intersections along San Pablo Avenue at Ashby Avenue, Heinz Avenue, Grayson Street, and Dwight Way located in Berkeley, California. This segment has been selected due to the variety in phasing schemes utilized on the four intersections and the existence of conflicting bus routes at two intersections.

The arterial segment's layout and bus routes are shown in Figure 2(a). Figure 2(b) presents the phasing and green times for all intersections during the evening peak hour (4–5pm). As indicated in the figure, the intersections consist of a variety of signal phasing schemes that cover all of the basic possible phasing schemes. All intersections operate under a common cycle length of 80 seconds and the demand used for the tests corresponds to the evening peak hour. During that time period, all four intersections operate in undersaturated traffic conditions with intersection flow ratios varying from 0.3 to 0.6. The intersection of Ashby and San Pablo Avenues is the critical one and the northbound is the heaviest traffic direction.

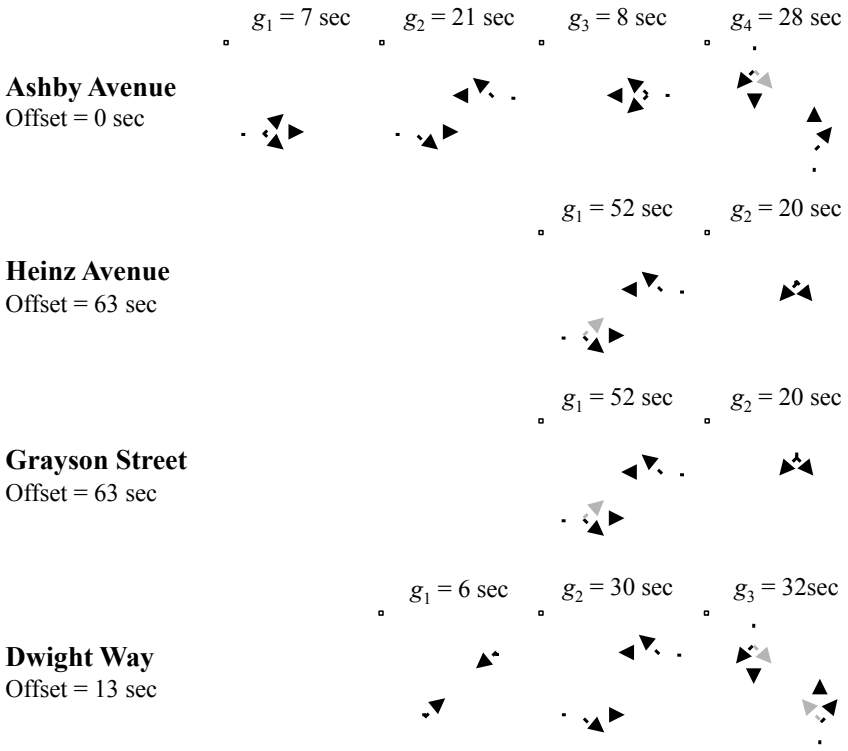
Five bus routes travel through the segment under consideration in mixed-use traffic lanes with headways that vary from 12 to 30 minutes on each route. This corresponds to an average of 24 buses per hour for the analysis period. The numbers next to the directional arrows in Figure 2(a) correspond to the different bus routes. Of the buses that travel in the corridor, 60% travel on San Pablo Avenue and 40% on the two cross streets: Ashby Avenue and Dwight Way. At these cross streets, buses travel in two conflicting directions. The location of the bus stops varies with some of them being located nearside and some others farside (Figure 2(a)). The bus schedule is available at the Alameda-Contra Costa Transit District's website (*17*).

APPLICATION

Data from the arterial segment of San Pablo Avenue are used to test the proposed signal control system. First, tests are performed for a few cycles assuming that perfect information exists on the platoon sizes and arrival times of platoons and transit vehicles at intersections (deterministic arrival tests). These give an idea of the maximum benefit that can be achieved by the proposed system. Next, tests are performed with



(a) Layout and bus routes (not to scale)



(b) Signal phasing and green times

FIGURE 2 San Pablo Avenue test site.

Emulation-In-the-Loop Simulation (EILS) to evaluate the system when perfect information is not available and predictions of inputs are based on measured quantities from the simulated network as it would be done in reality (stochastic arrival tests). A warm-up period equal to the common cycle length of all intersections is used. The computation times for both types of tests are on the order of 5-10 seconds which is sufficiently short time for real-world implementations.

The evening peak average flows are used as the input for auto demand and the bus arrivals are based on the actual schedule. The average auto occupancy, $\bar{o}_a$, is assumed to be 1.25 passengers per vehicle and the average bus occupancy, $\bar{o}_b$, is fixed to 40 passengers per vehicle for all buses. The green times for the

next cycle, $g^r_{i\,\text{next}}$, are set to be the same as the fixed optimal signal timings provided by TRANSYT-7F (*18*) for the specific traffic conditions under evaluation. In addition, the upper bounds for the green times of the phases, $g^r_{i\max}$, are set equal to $C - \sum_{i=1}^{I^r} y^r_i$ for each intersection r. Non-zero lower bounds for the green times of each phase, $g^r_{i\min}$, are also introduced for the green times for each intersection r to ensure that all phases are allocated some minimum green time. A total minimum green time of 7 seconds is assigned to each of the left-turn phases and 12 seconds to each of the through phases.

Deterministic Arrivals Tests

Deterministic arrival tests are performed for the four-intersection segment for five signal cycles and two optimization scenarios: 1) when only vehicle delay is minimized and 2) when total person delay for both auto and bus passengers is minimized. In addition, the results are compared to the ones obtained with the optimal signal settings provided by TRANSYT-7F. Each scenario is evaluated ten times and the average person delays of these ten replications for auto, bus, and all passengers are presented in Table 1. A comparison of the person-based with the vehicle-based optimization indicates that the proposed system can achieve a reduction in the total person delay of 2.7% by reducing the delay of the bus users by 9.7% and increasing auto user delay by only 1.5%. A comparison of the delays obtained from the vehicle-based optimization with the TRANSYT-7F optimal signal settings shows that the optimal signal settings of the vehicle-based optimization outperform those of TRANSYT-7F. Therefore, evaluation of the proposed signal control system for the simulation tests is performed by comparing person delays from the person-based optimization with the ones from the vehicle-based optimization.

TABLE 1 Person Delays on the Arterial Segment for $\bar{o}_b/\bar{o}_a = 40/1.25$ and Five Signal Cycles of Traffic Operations (deterministic arrivals)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 0: TRANSYT-7F (Fixed Settings)	5.91	2.70	8.61
Scenario 1: Vehicle-based Optimization	4.14	2.45	6.60
Scenario 2: Person-based Optimization	4.20	2.22	6.42
% Change in person delay between Scenarios 1 & 2	1.45%	–9.73%	–2.71%

Stochastic Arrival Tests–Simulation Experiments

EILS tests are performed with the AIMSUN microscopic simulation model (*19*) for one hour of operations and the average outcomes from the 10 replications of the two optimization scenarios are compared. The auto inter-arrival times on the incoming links are simulated to follow an exponential distribution. However, these vehicles stop at upstream fixed-time signalized intersections before they arrive at the intersections on the arterial. This ensures that vehicles arrive in platoons at the intersections being optimized. As a result, the auto demand for an approach is estimated based on measurements from detectors located at the upstream end of each incoming link under consideration. Exponential smoothing is used on the measured counts during the previous cycle in order to estimate the demand of the respective lane group for the next cycle. Predictions of auto arrival times at the intersections are based on an average free flow speed of 45 km/hr.

The timetable of the bus arrivals at the entry links of the network is fixed and based on headways obtained from the actual schedule. The arrival time of the buses at the intersections is predicted based on their location on a link at the end of the previous cycle. Information on the location of vehicles and bus stops

determine the estimated arrival time of a bus at the intersection. For simplicity, dwell times for all buses and stops are set to 30 seconds. For each bus that stops, an additional 6 seconds are added to its estimated travel time to reach the intersection in order to account for lost time due to acceleration and deceleration.

Table 2 shows the auto, bus, and total passenger delays for each arterial direction, the cross streets and the whole arterial segment. Although the southbound direction has lower auto traffic demand than the northbound, both directions have similar bus flows. Note also that both directions are served by the same phases at all intersections, except Ashby Avenue, and as a result, they experience similar green times. Therefore, the differences between the results of the person-based and vehicle-based optimization are similar for both directions. These results also show a reduction in delays for bus users by 5.6% and for auto users by 4.2% for cross streets with person-based compared to vehicle-based optimization. On cross streets, there are fewer buses and less auto traffic traveling compared to the main arterial. As a result, provision of priority to buses on cross streets leads to longer green time for them which also substantially benefits auto users. This outweighs the delay increase to cross-street auto traffic caused by priority provision to the high frequency transit vehicles traveling on the main arterial.

TABLE 2 Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$ and 1 Hour of Traffic Operations (simulation)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Main Arterial Northbound			
Scenario 1: Vehicle-based Optimization	48.43	29.18	77.61
Scenario 2: Person-based Optimization	48.18	28.59	76.77
% Change in person delay between Scenarios 1 & 2	–0.52%	–2.02%	–1.08%
Main Arterial Southbound			
Scenario 1: Vehicle-based Optimization	52.84	22.82	75.66
Scenario 2: Person-based Optimization	52.09	22.49	74.58
% Change in person delay between Scenarios 1 & 2	–1.42%	–1.45%	–1.43%
Cross Streets			
Scenario 1: Vehicle-based Optimization	36.34	7.29	43.63
Scenario 2: Person-based Optimization	34.80	6.88	41.68
% Change in person delay between Scenarios 1 & 2	–4.24%	–5.62%	–4.47%
System			
Scenario 1: Vehicle-based Optimization	137.61	59.29	196.90
Scenario 2: Person-based Optimization	135.07	57.96	193.03
% Change in person delay between Scenarios 1 & 2	–1.85%	–2.24%	–1.97%

On the system level, the proposed person-based signal control system achieves a reduction in total person delay of 1.8% for the arterial segment. This translates to a 2.2% reduction in bus passenger delay and a 1.9% reduction in auto user delay. Despite the expectations that person-based optimization would result in higher delays for auto users, the tests presented here show that it could lead to lower delays for all users compared to vehicle-based optimization. This can be attributed to three reasons:

1. Autos that are traveling in the same direction as transit vehicles benefit from the provision of priority, and as a result, their delays are reduced along with those of transit.

2. Since the proposed system is used to minimize total person delay for one cycle, at two intersections at

a time, it does not guarantee global optimality for the whole hour and arterial segment. This explains the fact that person-based optimization outperforms vehicle-based optimization for auto delays.

3. Performance of the two types of optimization through simulation is highly dependent on the accuracy of auto demand estimates and the auto and transit vehicle arrival predictions at the intersections. As a result, it is possible that person-based optimization does not operate as expected when inaccuracies exist in the arrival estimates because the mathematical program formulation does not account for such inaccuracies.

Tests are also performed under the assumption that all transit vehicles arrive late at the intersections in the network. Accounting for their schedule delay translates into weighting the delay of transit vehicles by a factor of 2 (i.e., $\delta^r_{b,T} = 1$). The results indicate that the benefit to transit users improves to a 2.9% reduction in their delay with person-based optimization when schedule delay is considered (Table 3). At the same time, autos benefit by the extra green provided to serve buses faster. The ratio of average passenger occupancy of buses over autos is also expected to affect the level of priority provided.

TABLE 3 Person Delays for $\bar{o}_b/\bar{o}_a = 40/1.25$, $\delta^r_{b,T} = 1$ and 1 Hour of Traffic Operations (simulation)

	Auto Passenger Delay (pax-hrs)	Bus Passenger Delay (pax-hrs)	Total Passenger Delay (pax-hrs)
Scenario 1: Vehicle-based Optimization	137.61	59.29	196.90
Scenario 2: Person-based Optimization	135.16	57.57	192.74
% Change in person delay between Scenarios 1 & 2	–1.78%	–2.90%	–2.11%

DISCUSSION

The paper has presented the formulation of an arterial-level person-based traffic responsive signal control system. The proposed pairwise optimization method accounts explicitly for the passenger occupancy of transit vehicles to provide priority to them by minimizing person delay at two consecutive intersections at a time. In addition to accounting for auto vehicle progression, by assigning the appropriate delays for interrupting the platoons, the system recognizes the importance of schedule adherence for reliable transit operations. Therefore, it introduces an additional weighting factor that reflects how early or late a transit vehicle is. The passenger occupancy and schedule adherence weighting factors facilitate priority assignment decisions for conflicting transit routes. The proposed system is flexible because the user can choose to weigh transit delay and delays caused by interrupting the head or tail of the platoon differently to reach different objectives. In addition, it relies on currently deployable technologies that can provide real-time information for estimating platoon sizes and arrival times at intersections (e.g., demand, travel times, and turning ratios from detectors with appropriate communication systems), transit vehicle arrival times (e.g., speed and location from AVL systems), and passenger occupancies (e.g., APC systems). In addition, its low computation time contributes to the feasibility and economic viability of its implementation in the real world.

Deterministic arrival tests have shown that the arterial traffic responsive signal control system outperforms static signal settings provided by TRANSYT-7F, even for auto delays. It has also been found that for stochastic arrival tests it can reduce delay for all travelers and bus users at an intersection by up to 5% for the selected test site and in most cases it can reduce auto user delay as well. This is due to the fact that autos traveling on links with buses also benefit by the provision of transit priority. In addition, since the proposed

system optimizes the signal settings at two intersections at a time, it does not guarantee global optimality for the whole arterial. Therefore, it is likely that after the first cycle, the two optimization scenarios result in different traffic conditions that could make the same simulated replication not be exactly comparable for the two scenarios. In addition, errors in the arrival estimates could lead to the person-based optimization not operating as expected, because the mathematical program formulation does not account for such inaccuracies. The tests have also shown that buses traveling on cross streets with low auto demand experience very high benefits when transit priority is provided. This is due to the much higher weight the cross street vehicle delays get when a bus is present and person-based optimization is used compared to when vehicle-based optimization is used. Finally, accounting for schedule delay provides additional benefit to transit users without negatively impacting auto users.

The proposed system is promising for reducing person delay for an arterial segment and providing priority to transit vehicles even when multiple transit routes run in conflicting directions at intersections. Ongoing and future work includes incorporating pedestrian delays in the system and testing it for a variety of bus to auto passenger occupancy ratios and schedule delay weighting factors. In addition, we plan to improve the arrival prediction algorithms and adjust the delay equations to account for inaccuracies in the input estimates. Future work will also relax the no platoon dispersion assumption by splitting the platoons for each approach into smaller ones and modifying the mathematical program formulation to appropriately capture delays for more than one platoons per lane group. Finally, we plan to expand the system to arterial signalized networks by using the proposed pairwise optimization method along multiple arterials.

REFERENCES

[1] Cornwell, P., J. Luk, and B. Negus, Tram priority in SCATS. *Traffic Engineering and Control*, Vol. 27, No. 11, 1986, pp. 561–565.

[2] Hunt, P., R. Bretherton, D. Robertson, and M. Royal, SCOOT on-line traffic signal optimisation technique. *Traffic Engineering and Control*, Vol. 23, 1982, pp. 190–192.

[3] Bretherton, D., G. Bowen, and K. Wood, Effective urban traffic management and control: SCOOT Version 4.4. *European Transport Conference*, 2002.

[4] 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. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1634, 1998, pp. 100–109.

[5] Mauro, V. and C. Di Taranto, UTOPIA. *Proceedings of the 6th IFAC-IFIP-IFORS Symposium on Control, Computers, and Communications in Transportation*, 1989, pp. 245–252.

[6] Henry, J. and J. Farges, P.T. priority and Prodyn. *Proceedings of the 1st World Congress on Application of Transport Telematics and Intelligent Vehicle-Highway Systems*, Vol. 6, 1994, pp. 3086–3093.

[7] He, Q., K. Head, and J. Ding, PAMSCOD: Platoon-based arterial multi-modal signal control with online data. *Transportation Research Part C: Emerging Technologies*, Vol. 20, No. 1, 2012, pp. 164–184.

[8] Christofa, E. and A. Skabardonis, Traffic signal optimization with application of transit signal priority to an isolated intersection. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2259, 2011, pp. 192–201.

[9] Christofa, E., I. Papamichail, and A. Skabardonis, Person-based traffic signal optimization for real-time applications. In *Procedings of the 91st Annual Meeting of the Tranportation Research Board*, 2012.

[10] Newell, G., Synchronization of traffic lights for high flow. *Quarterly of Applied Mathematics*, Vol. 21, No. 4, 1964, pp. 315–324.

[11] Newell, G., *Traffic signal synchronization for high flows on a two-way street*. Research report, Institute of Transportation and Traffic Engineering, University of California, 1967.

[12] Lighthill, M. and G. Whitham, On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317–345.

[13] Richards, P., Shock waves on the highway. *Operations Research*, Vol. 4, No. 1, 1956, pp. 42–51.

[14] Meyer, C. and C. Floudas, Trilinear monomials with mixed sign domains: Facets of the convex and concave envelopes. *Journal of Global Optimization*, Vol. 29, No. 2, 2004, pp. 125–155.

[15] IBM, *IBM ILOG CPLEX, Version 12.1: High Performance Mathematical Programming Engine*. http://www.ilog.com/products/cplex, 2011.

[16] Christofa, E., *Traffic Signal Optimization with Transit Priority: A Person-based Approach*. Ph.D. thesis, University of California, Berkeley, 2012.

[17] AC Transit, *Alameda-Contra Costa Transit District*. http://www.actransit.org, 2011.

[18] Wallace, C., K. Courage, M. Hadi, and A. Gan, TRANSYT-7F User's Guide. *Transportation Research Center, University of Florida, Gainesville, Florida*, 1998.

[19] Transport Simulation Systems, *Aimsun Users Manual v6.1*, 2010.