Transportation Research Part C 20 (2012) 164–184

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

PAMSCOD: Platoon-based arterial multi-modal signal control with online data

Qing He, K. Larry Head *, Jun Ding

University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Keywords:
Multi-modal traffic signal control
Platoon recognition
Mixed integer linear programming
Multi-modal dynamical progression
Vehicle-to-infrastructure communications

ABSTRACT

A unified platoon-based mathematical formulation called PAMSCOD is presented to perform arterial (network) traffic signal control while considering multiple travel modes in a vehicle-to-infrastructure communications environment. First, a headway-based platoon recognition algorithm is developed to identify pseudo-platoons given probe vehicles' online information. It is assumed that passenger vehicles constitute a significant majority of the vehicles in the network. This algorithm identifies existing queues and significant platoons approaching each intersection. Second, a mixed-integer linear program (MILP) is solved to determine future optimal signal plans based on the current traffic controller status, online platoon data and priority requests from special vehicles, such as transit buses. Deviating from the traditional common network cycle length, PAMSCOD aims to provide multi-modal dynamical progression (MDP) on the arterial based on the probe information. Microscopic simulation using VISSIM shows that PAMSCOD can easily handle two common traffic modes, transit buses and automobiles, and significantly reduce delays for both modes under both non-saturated and oversaturated traffic conditions as compared to traditional state-of-practice coordinated-actuated signal control with timings optimized by SYNCHRO.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Traffic signals are very important tools for urban traffic management. With correct installation and control strategy selection, they can improve both traffic throughput and safety for all road users. Traffic flow usually consists of more than one travel mode, including automobiles, transit buses, and emergency vehicles as well as commercial trucks, bicycles, and pedestrians. Different travel modes have their own specific characteristics, including travel speed, density, and priority level. Traditionally, traffic operations treat different travel modes separately (e.g., signal coordination for automobiles, signal preemption for emergency vehicles, and transit signal priority (TSP) for buses). However, treating each mode separately is likely to result in sub-optimal performance. In the environment of vehicle-to-infrastructure communications (Research and Innovative Technology Administration, 2011), enriched traffic data makes optimal multi-modal traffic control a real possibility. Systems of traffic signals, such as along an arterial, need to be coordinated based on actual traffic composition and the flow of vehicles. With the advent of vehicle-based positioning and communications, it is possible to know where vehicles are and to plan traffic signal control to best serve all vehicles in the entire network.

This paper presents a unified platoon-based formulation called Platoon-based Arterial Multi-modal Signal Control with Online Data (PAMSCOD) to concurrently optimize network traffic signal control for different travel modes given the

* Corresponding author. Tel.: +1 520 621 2264; fax: +1 520 621 6555.
E-mail address: larry@sie.arizona.edu (K.L. Head).

0968-090X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trc.2011.05.007

assumption that advanced communication systems are available between vehicles and traffic controllers. In this paper, two modes of traffic composition are considered: transit buses and passenger vehicles in a decision framework that can easily accommodate pedestrians and bicycles. First, when approaching an intersection, travelers are able to send a "green light" request to the traffic controller. The "green light" request includes travelers' travel mode, position, speed, and requested traffic signal phase (which they know from a digital map of the intersection that contains the relevant information). Single requests are categorized and clustered into platoons by priority level and phase. Finally, a mixed-integer linear program (MILP) is solved online for future optimal signal plans based on the real-time arterial platoon request data and traffic controller status.

One feature of PAMSCOD is network level dynamical signal coordination without the constraint of common cycle length and fixed offsets. State-of-the-practice coordinated-actuated control requires a common fixed cycle length and a predetermined set of offsets, which are generally developed only for automobiles under a certain range of traffic flow conditions.

In addition to the consideration of multi-modes, the large variance of traffic flow can degrade the performance of off-line optimized time-of-day plans (Yin, 2008). Robust traffic signal control is less sensitive to variances in traffic flow. To achieve both robust and multi-modal network traffic signal control, the concept of multi-modal dynamical progression (MDP) is proposed in this paper. Regardless of cycle length and offsets, multi-modal dynamical progression along an arterial is realized in the MILP formulation by adjusting phase durations to minimize the overall multi-modal delay both at the current intersection and downstream intersections.

Large MILP formulations for signal control problems are subject to the "curse of dimensionality" and are not easily solved in real-time (He et al., 2010). To address the issue of computational complexity, three measures are taken into account. First, the automobile requests are aggregated into platoons requesting specific phases to reduce the total number of integer variables (from each individual vehicle to platoons of vehicles), as the number of integer variables contributes significantly to the complexity of the MILP formulation. Second, the integer feasible region is enhanced by assuming a first-come, first-served discipline for the requests on the same approach. For example, suppose two platoons are requesting the same phase. The first platoon should always travel through the intersection before the second platoon. Third, multi-modal dynamical progression is only considered between adjacent intersections due to the large variations of real-time traffic data. It is assumed that the control problem can be resolved frequently to coordinate across time and space.

It is well known that traffic signal control is very difficult during oversaturated traffic conditions (Gazis, 1964). Oversaturated traffic situations are caused by traffic demands that exceed the available capacity and can produce queues that grow over time. Residual queues can overflow the storage capacity of urban streets and physically block intersections to cause de facto red (Abu-Lebdeh and Benekohal, 1997) when queue build-up leads to complete blockage of upstream signals and no traffic can discharge when the traffic signal is green. When real-time queue length and queue size are available from advanced vehicle-based communications, the link capacity between adjacent intersections is known. Therefore, PAMSCOD is able to control the discharge rate from upstream intersections to avoid de facto red.

This paper is organized as follows. Section 2 presents a fast algorithm to identify platoons. Section 3 develops a MILP to address online arterial (or network) platoon-based multi-modal traffic signal control. Both dynamical progression and queue management are discussed in this section. Section 4 compares the performance of PAMSCOD with different traffic signal control methods under different traffic volume and scenarios. Section 5 presents a summary of the model and performance comparison.

2. Platoon identification with probe vehicles

Many well-known traffic signal control algorithms have been based on platoon dynamics. Platoon dispersion and secondary flows were considered in the TRANSYT-7F model (Robertson, 1969; Wallace et al., 1998) which has evolved into the well known adaptive traffic signal control system SCOOT (Hunt, 1982). Dell'Olmo and Mirchandani (1995) proposed a real-time, platoon-based network level control algorithm called REALBAND which attempts to resolve future platoon conflicts. REALBAND generates platoon information based on loop detector data (Gaur and Mirchandani, 2001) and creates a decision tree based on the projected platoon flow to find the best root-to-leaf decision by minimizing the total platoon delay. REALBAND has been developed as one component of the RHODES adaptive traffic signal control system (Mirchandani and Head, 2001). Recently, a platoon-based traffic signal control algorithm was developed by Jiang et al. (2006) for intersections with major–minor traffic volume. However, their approach does not consider coordination as a primary consideration.

Traditional platoon recognition relies on pre-selected platoon detector locations. Platoons are identified by vehicle density (Gaur and Mirchandani, 2001), cumulative headways (Wasson, 1999; Chaudhary et al., 2006), or critical headways (Jiang et al., 2006). However, fixed location detectors have many limitations. First, the detector location needs to be adjusted based on traffic conditions. Distant detectors may be effected by platoon dispersion effects, whereas nearby detectors may not be able to identify platoons due to long queues. Second, the platoon information provided by fixed location detectors is untraceable. Platoons may grow, disperse, or even disappear when approaching an intersection. Third, it is costly to install and maintain advanced platoon detectors in each lane.

Traffic probe data (or floating car data) has become available with recent advances in communication technology including electronic transponders in Electronic Toll Collection systems (ETC), GPS-enabled mobile phones (Herrera and Bayen, 2010), and Connected Vehicles (Research and Innovative Technology Administration, 2011). Improving arterial traffic signal control performance utilizing the vast amount of new data is an emerging research problem.

In this paper, individual probe vehicles are clustered as pseudo-platoons and serve as input to the PAMSCOD algorithms. Pseudo-platoons only indicate clusters of equipped vehicles including moving and standing (queued) vehicles. A single pseudo-platoon includes only vehicles of the same mode and requesting the same traffic control phase. Buses are treated as individual vehicle pseudo-platoons due to their different movement characteristics.

It is assumed that the percentage of probe vehicles is known and that advance platoon detectors are not considered. Assuming the percentage of probe vehicles is reasonable if vehicle counts are available within the network to compare to the number of probe vehicles. The platoon recognition algorithm consists of the following steps:

1. *Classification*: Probe vehicles are classified as moving probes and queuing probes according to their speeds.
2. *Identification*: Queuing probes are considered to be in a single platoon while moving platoons are recognized by the headways between individual probe vehicles.
3. *Estimation*: Platoon parameters, such as the number of vehicles and time of arrival, are estimated by regression models or previous queuing information from upstream intersections.

There has been little research on the platoon identification problem using probe data. Smith et al. (2010) adopted classical statistical clustering algorithms, such as agglomerative hierarchical clustering algorithm, or *k*-means clustering. To correctly use these statistical clustering methods, one needs to determine the best number of cluster and how to delete outliers. Also it is difficult to interpret the clustering results from traffic engineers' perspective.

In this paper, probe headways are analyzed and applied to recognize significant platoons. Suppose h_1 indicates the headway with 100% penetration, and h_p is the headway with penetration rate p. Actually, $h_p = Kh_1$ holds by assuming vehicle headway as a constant, where K represents the number of vehicles since the last probe, as shown in Fig. 1. The probability that the Kth vehicle is the next equipped vehicle is modeled by a geometric distribution with parameter p. The expectation is

$$E(h_p) = \frac{h_1}{p} \tag{1a}$$

And the standard deviation is

$$Std(h_p) = \frac{h_1\sqrt{1-p}}{p} \tag{1b}$$

Therefore, platoons can be separated by comparing h_p with a critical headway $\overline{H_p}$, defined as

$$\overline{H_p} = \frac{h_1}{p} + \lambda\frac{h_1\sqrt{1-p}}{p} \tag{2}$$

where λ defines the number of standard deviations from the mean. Platoons are formed if the headway of two consecutive probes is less than $\overline{H_p}$. Fig. 2 illustrates platoon identification profiles under different penetration rates with $h_1 = 2$ and $\lambda = 3$, where the red solid and black dashed vertical lines represent equipped and unequipped vehicles, respectively, and the length of vertical lines indicates normalized vehicle speeds . Identified platoons are indicated by a function whose value is 1 if a platoon is present and 0 if no platoon is present as shown by the horizontal green line. Penetration rates influence how well the algorithm performs. The platoon span is extremely overestimated when $p = 0.2$ due to large $\overline{H_p}$. When $p = 0.4$, two platoons are captured. The higher the penetration rate p, the closer the identification matches to the real-world platoon. This variability will make a significant difference in determining traffic signal control.

After platoons are identified, platoon parameters need to be estimated for signal optimization. In terms of a standing platoon (existing queue), queue length is the only parameter to be determined. Comert and Cetin (2009) proposed a probability model to estimate the queue size based on probe data. They claimed that only the location information of the last probe vehicle in the queue is sufficient for the estimation. Neumann (2010) proposed a maximal likelihood estimation method to estimate traffic demand, density profile, and queue length. However, neither of these studies considered the length of a discharging queue, when vehicle speeds are actually not zero. For moving platoons, the platoon size includes both the

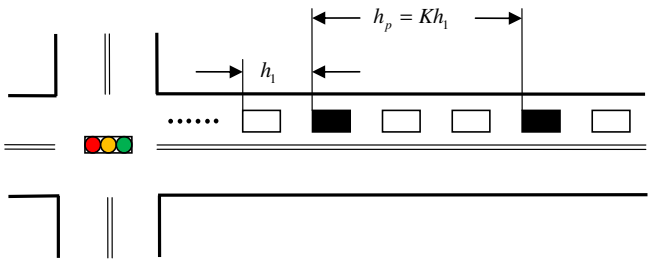

Fig. 1. Illustration of headways of probe vehicles.

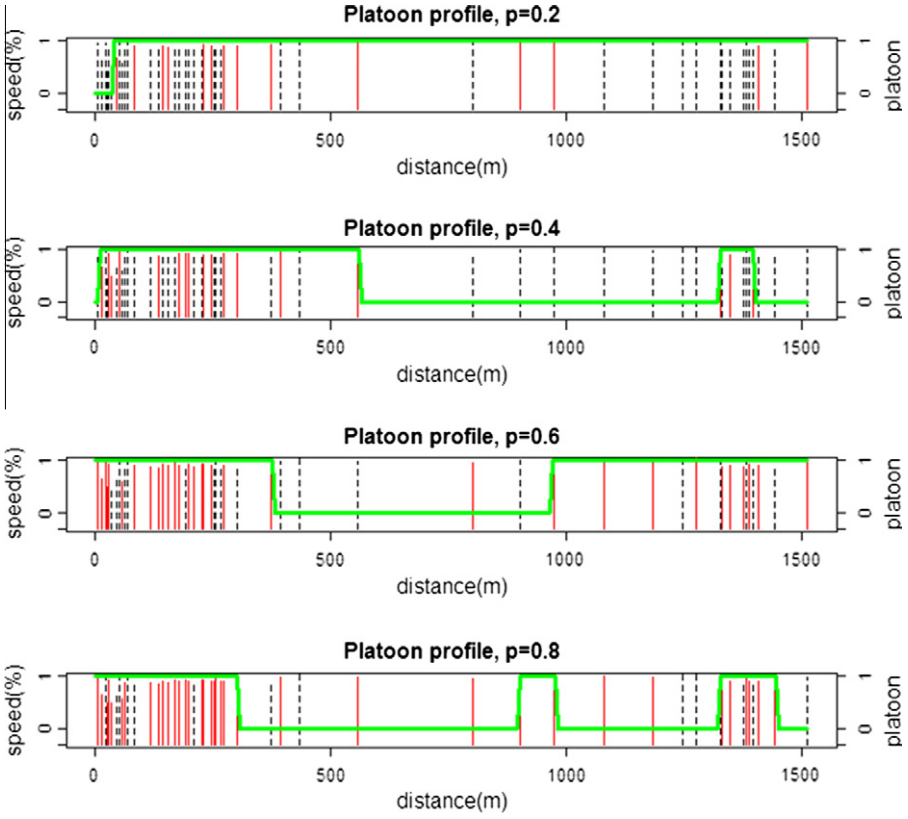

Fig. 2. Identification of platoon profiles under different penetration rates.

number of vehicles in the platoon (N_p) as well as the arrival time of the leading vehicle and the trailing vehicles ($\underline{T_a}$ and $\overline{T_a}$, respectively) are require for signal optimization.

In this paper, a linear regression model is used to estimate all of these important signal control parameters. The basic form of the regression model is depicted as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{3}$$

For a stopped platoon, queue length, N_q, is estimated as the dependent variable $Y = N_q$ with the first independent variable $X_1 = l_{\max} N_l k$ where $l_{\max}$ denotes the distance of the last probe vehicle in queue from stop bar, N_l is the number of lanes, k is the density of the queue in one lane which can be approximated by a linear relationship between density and speed, $k = k_{jam} - aV_p$; k_{jam} indicates the jam density in one lane, V_p represents the average speed of the probes, and a is a coefficient; The second variable $X_2 = n/p$; n is the total number of probes in the queue and p is the penetration rate. Empirical simulation results show that X_1 is a better estimator than X_2 in low penetration rates, while X_2 is a better estimator than X_1 in high penetration rates.

The size of a moving platoon N_p, can be similarly estimated. In this case, $X_1 = l_s N_l k$, where l_s denotes the distance span of probe vehicles in the moving platoon. $X_2 = n/p$, where n is the total number of probes in the moving platoon and p is the penetration rate.

Two additional variables are important in characterizing a moving platoon. These are the leading and training vehicle arrival times, denoted as $\underline{T_a}$ and $\overline{T_a}$, respectively. $\underline{T_a}$ and $\overline{T_a}$ are estimated using (3) again, with independent variables $X_1 = T_c$ and $X_2 = N_p/N_l$, where T_c denotes the median of arrival of all probe vehicles in the platoon. The estimation of arrival time starts from estimating the median, or "central", arrival time T_c of the platoon as the median of all individual probe arrival times. Suppose that the first half of platoon arrives at intersection before T_c, while the second half arrives after T_c. Hence the arrival time of leading vehicle and tail vehicle can be estimated by using the average headway in the platoon, which is interpreted as $2\beta_2$.

Queues and platoons were collected using the VISSIM microscopic simulation software. A total of 211,540 queues and 317,900 platoons were observed under different traffic scenarios and penetration rates. Results from statistical analysis are shown in Table 1. All the independent variables are statistically significant as shown by the very small p-values. The model fits the data very well as indicated by the high R-square value. The RMSE of each prediction are shown in Fig. 3. The critical value can be found when the penetration rate $p = 0.4$ RMSE ranges between 2 and 4 when $p > 0.4$. However, it

Table 1
Statistical results of platoon parameter estimation.

	β_0	t-value	p-value	β_1	t-value	p-value	β_2	t-value	p-value	R-square
N_q	0.1975	5.427	5.76e−08	0.2286	119.247	<2e−16	0.7098	294.717	<2e−16	0.9572
N_p	1.2496	23.6	<2e−16	0.2732	65.94	<2e−16	0.6036	115.47	<2e−16	0.8861
$\frac{T_a}{T_a}$	−0.5027	−9.225	<2e−16	0.9717	761.166	<2e−16	−1.0847	−173.084	<2e−16	0.9686
$\overline{T_a}$	1.1129	15.83	<2e−16	0.9480	575.65	<2e−16	1.23235	152.43	<2e−16	0.958

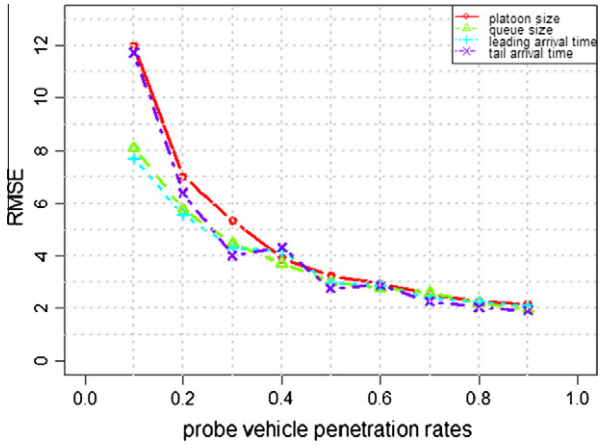

Fig. 3. RMSE of platoon estimation.

increases quickly when $p < 0.4$ One may note that the size of moving platoon has larger randomness and fluctuations than the size of queues. The prediction may yield large variations especially when the penetration rate is low. In such cases, one may consider vehicle queuing information at upstream intersections (Smith et al., 2010). If it is assumed that stop-bar queues will eventually evolve moving platoons at downstream links, the size of moving platoon could be roughly estimated by the queue position of each probe vehicle at upstream stop-bar.

3. Mixed integer linear program (MILP) in PAMSCOD

Several MILP formulations have been proposed to solve the network traffic signal control problem over the past few decades. The original formulation was MAXBAND (Little, 1966; Little et al., 1981). MULTIBAND, another well-known traffic signal control MILP, was developed in the early 1990s (Gartner, 1991; Stamatiadis and Gartner, 1996). Both MAXBAND and MULTIBAND aimed to maximize the "green wave" bandwidth by optimizing cycle length, splits and offsets. Neither of them was intended to address oversaturated conditions.

A number of papers have developed dynamic traffic signal control formulations based on the cell transmission model (CTM) (Daganzo, 1994; Daganzo, 1995). The CTM-based signal control formulation can address both unsaturated and oversaturated conditions by considering shockwaves and physical queues (Lo, 1999; Lo, 2001) formulated the network signal optimization problem as a mixed-integer linear programming problem using CTM and assuming that the cycle length is fixed. Lin and Wang (2004) developed a more computationally efficient version of the MILP for the signal optimization problem based on the CTM, but only two-phase signals were considered in their model. For the CTM-based formulations with the MILP approach, the problem size can grow very quickly with the size of the network and the time horizon. The "curse of dimensionality" makes it impossible to solve these formulations online directly, even when using commercially available packages such as CPLEX and LINDO.

In this paper, a platoon-based MILP that reduces the number of integer variables required and that accommodates both unsaturated and oversaturated conditions by considering shockwaves and physical queues is presented. The model notation is summarized in Appendix.

3.1. Unified precedence constraints

The signal controller model examined is based on the standard North American NEMA dual-ring, eight-phase controller. A four-legged intersection with eight movements is shown in Fig. 4a. Typically, each ring in the controller contains four phases, as depicted in Fig. 4b. A barrier exists across both rings between groups of conflicting movements such that all phases in one group must terminate before any phase in the other group starts. The dual-ring controller can be modeled by a precedence

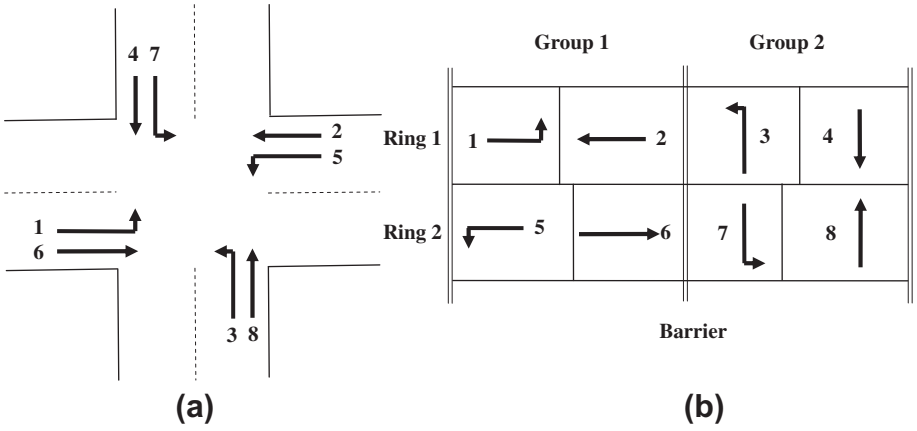

Fig. 4. (a) Intersection layout; (b) A dual-ring, eight-phase controller.

graph as depicted in Fig. 5 (Head et al., 2006). Arcs in the precedence graph represent the duration of phases, whereas nodes represent the phase transitions. Phase intervals can be easily visualized in the precedence graph. In this model, the precedence constraints are defined for each intersection in the system being optimized and then unified into one overall optimization problem.

Phase sequence and phase initialization are important considerations for online optimization. In formulating the MILP, the identification of the current phase to act as the starting phase in the MILP and accounting for the time that has already elapsed in the current phase define the phase time constraints (minimum green time) in the MILP formulation. In general, in an 8-phase controller, as shown in Fig. 5, the starting phase pair can only be 1 of 8 different cases: 1 and 5, 1 and 6, 2 and 5, 2 and 6, 3 and 7, 3 and 8, 4 and 7, or 4 and 8. The starting phases at each intersection can be different depending on the current status when the MILP problem is formulated. Similarly, different intersection may have different phase configurations (e.g., sequences) depending on the intersection geometry and operational considerations.

One important issue is how to handle different starting phases at different intersections in the unified precedence constraints formulation. To achieve this, it is assumed that a dual-ring structure is applicable to all of the intersections, and phase 1 and 5 are always considered as first phases in a phase configuration. Therefore, the phase precedence constraints are as follows:

$$t(i,1,1) = s_1(i) \quad \forall i \tag{4}$$
$$t(i,5,1) = s_2(i) \quad \forall i \tag{5}$$
$$t(i,p,1) = O_1(i) \quad \forall i, p \in \Delta_{s1}(i) \tag{6}$$
$$t(i,p,1) = O_2(i) \quad \forall i, p \in \Delta_{s2}(i) \tag{7}$$
$$t(i,2,k) = t(i,1,k) + v(i,1,k) + s(i,1,k) \quad \forall i,k \tag{8}$$
$$t(i,6,k) = t(i,5,k) + v(i,5,k) + s(i,5,k) \quad \forall i,k \tag{9}$$
$$t(i,3,k) = t(i,2,k) + v(i,2,k) + s(i,2,k) \quad \forall i,k \tag{10}$$
$$t(i,3,k) = t(i,6,k) + v(i,6,k) + s(i,6,k) \quad \forall i,k \tag{11}$$
$$t(i,7,k) = t(i,2,k) + v(i,2,k) + s(i,2,k) \quad \forall i,k \tag{12}$$
$$t(i,7,k) = t(i,6,k) + v(i,6,k) + s(i,6,k) \quad \forall i,k \tag{13}$$
$$t(i,4,k) = t(i,3,k) + v(i,3,k) + s(i,3,k) \quad \forall i,k \tag{14}$$

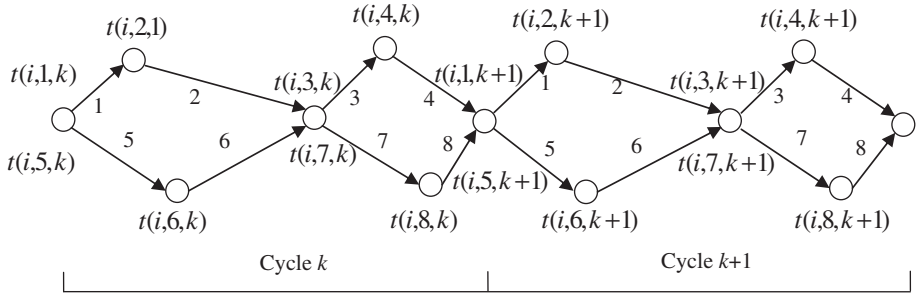

Fig. 5. Precedence graph representation of a dual-ring controller at intersection *i*.

$$t(i,8,k) = t(i,7,k) + v(i,7,k) + s(i,7,k) \quad \forall i,k \tag{15}$$

$$t(i,1,k+1) = t(i,4,k) + v(i,4,k) \quad \forall i,k \tag{16}$$

$$t(i,1,k+1) = t(i,8,k) + v(i,8,k) \quad \forall i,k \tag{17}$$

$$t(i,5,k+1) = t(i,4,k) + v(i,4,k) \quad \forall i,k \tag{18}$$

$$t(i,5,k+1) = t(i,8,k) + v(i,8,k) \quad \forall i,k \tag{19}$$

$$v(i,p,k) = g(i,p,k) + Y(i,p) + R(i,p) \quad \forall i, \quad p \notin \Delta_{non}(i) \cup \Delta_p(i), k \tag{20}$$

$$v(i,p,k) = 0 \quad \forall i, p \in \Delta_{non}(i) \cup \Delta_p(i), k \tag{21}$$

$$g(i,p,k) = 0 \quad \forall i, p \in \Delta_{non}(i) \cup \Delta_p(i), k \tag{22}$$

$$s(i,p,k) = 0 \quad \forall i,p,k \geqslant 2 \text{ or } \forall i, p \notin \Delta_{s1}(i) \cup \Delta_{s2}(i) \cup \Delta_p(i), k = 1 \tag{23}$$

$$\sum_p \sum_k s(i,p,k) \leqslant Z \quad \forall i,p,k \tag{24}$$

$$G_{\min}(i,p) \leqslant g(i,p,k) \leqslant G_{\min}(i,p) \quad \forall i, p \notin \Delta_{s1}(i) \cup \Delta_{s2}(i) \cup \Delta_p(i) \cup \Delta_{non}(i), k \tag{25}$$

$$g(i,p,k) \geqslant G_{\min}(i,p) - E(i,p) \quad \forall i, p \in \Delta_{s1}(i) \cup \Delta_{s2}(i), k \tag{26}$$

A static set Δ_{non} and three dynamic sets Δ_{s1},Δ_{s2} and Δ_p are introduced in this paper to address different phase configurations and starting phases. Δ_{non} represents those phases that are not used at a particular intersection. Δ_{s1} and Δ_{s2} denote the starting, or currently active, phases for ring 1 and ring 2, respectively. Δ_p is the set of past phases, or phases that have already completed service at the current time, in the first cycle, assuming that the cycle always starts from phase 1 and 5. Δ_{non} is fixed over the time horizon of consideration, whereas Δ_{s1}, Δ_{s2} and Δ_p are variable, depending on the starting phases and the elapsed time in the starting phases at the current time. Fig. 6 illustrates these sets in a standard NEMA eight-phase configuration with an example at one intersection i. The current timing phases are phases 4 and 7 with some elapsed time $E(i,4,1)$ and $E(i,7,1)$, respectively. Therefore, the minimal green time for phase 4 and 7 in the first cycle should be reduced by the elapsed times. Past phases in cycle 1 are phase 1, 2, 3, 5 and 6. Thus, the duration of phases in 1, 2, 3, 5 and 6 is set to 0 in cycle 1. Slack variables account for the green dwell time and maintain barrier constraints.

The unified precedence constraints have the following features:

1. Suppose phases 1 and 5 are the first phases in the dual-ring. The actual starting phases depend on the state of the controller at the time the MILP is formed and are denoted by sets Δ_{s1} and Δ_{s2}, constraint (4) and (5) in the above formulation.
2. If one or more of the starting phases is timing the yellow or red clearance, the issue is addressed in constraint (6) and (7) by introducing initial times $O_1(i)$ and $O_2(i)$, which are calculated by subtracting the elapsed yellow or all red times from the total clearance time $Y(i,p) + R(i,p)$.
3. The dual-ring structure is represented by the unified constraints (6)–(19).
4. Different starting phases are modeled by set $\Delta_p(i)$. If phase $p \in \Delta_p(i)$, then the phase duration of phase p should be 0, as depicted in constraint (21) and (22).

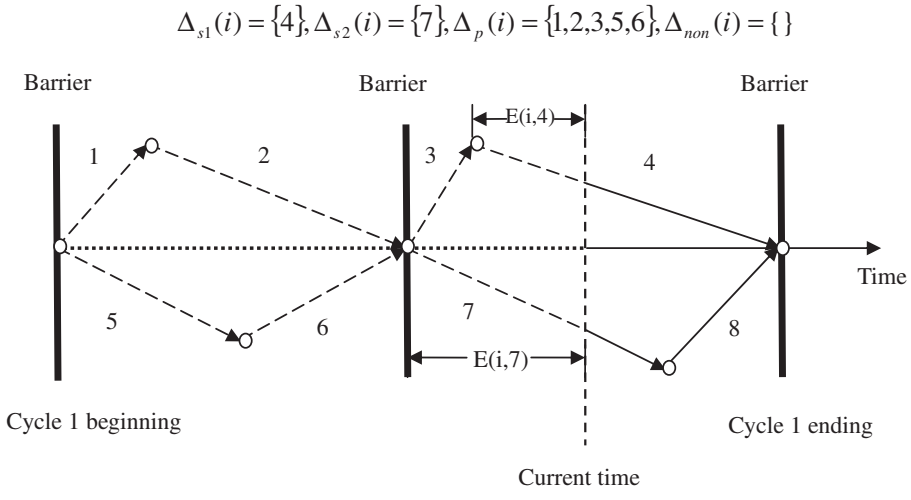

Fig. 6. An example illustrating the currently active phases at intersection i.

5. Slack variables are introduced only for green rest to fulfill the barrier constraints in the first cycle. This means the maximal green constraints could be violated if the first barrier constraints must be satisfied. Constraint (23) and (24) ensure that the slack variables are equal to zero in all of the phases after the starting phases. Green rest occurs when one ring reaches a barrier, or maximal green, but the other ring does not reach the barrier such that the ring will rest on the barrier to satisfy the barrier constraint, regardless of the maximal green time constraint.
6. The minimal green times of the starting phases can be less than the pre-defined minimum green time, $G_{\min}(i,p)$, as the starting phases have already been timing for $e(i,p)$.

3.2. Platoon delay categorization and evaluation

Delay is perhaps the most commonly reported performance index in the literature of traffic signal control. Within the PAMSCOD formulation, before delay can be assessed, it is necessary to determine which cycle is selected to serve platoon (m,i,p,j), thereby determining the correct phase starting time, $t(i,p,k)$, to calculate the delay of platoon (m,i,p,j).

3.2.1. Platoon serving cycle selection

Fig. 7 illustrates a platoon moving towards an intersection. This platoon can travel through the intersection either during cycle $k-1$ or during cycle k, assuming only two cycles are considered. If the platoon is served in cycle k, the platoon should pass through the intersection before the green ends, as described by constraint (27) and (28).

$$\underline{T_a}(m,i,p,j) + T_p(m,i,p,j) \leqslant t(i,p,k) + g(i,p,k) + M(1-\theta(m,i,p,j,k)) \quad \forall (m,i,p,j) \in \Gamma, k \tag{27}$$

$$\sum_k \theta(m,i,p,j,k) = 1 \quad \forall (m,i,p,j) \in \Gamma \tag{28}$$

where $\underline{T_a}(m,i,p,j)$ is the arrival time of the leading vehicle in the platoon, and $T_p(m,i,p,j)$ is the green time required to clear the platoon. Binary variables $\theta(m,i,p,j,k)$ are introduced to handle the cycle service selection problem. For each valid platoon $(m,i,p,j) \in \Gamma$, constraint (27) comes into effect if $\theta(m,i,p,j,k)$ equals one. Thus, the platoon (m,i,p,j) will be served in cycle k. Otherwise, platoon (m,i,p,j) will not be served in cycle k. Constraint (28) ensures there is only one cycle selected to serve platoon (m,i,p,j), as illustrated in Fig. 7.

3.2.2. Platoon split and link capacity

Constraint (27) only allows a platoon to be served completely during a cycle. However, there are some cases when platoons need to be split and served in two cycles or more. These cases include the following:

1. Platoon length may grow quickly during heavy demand or oversaturated traffic conditions. In this case, $T_p(m,i,p,j)$ could be greater than the maximal phase green times, $G_{\max}(i,p,j)$. In this case, the demand exceeds the service capacity.
2. Downstream dynamic link storage capacity $C_l(i_d,p)(i_d \in I_d(i,p))$ may be smaller than the size of the platoon $N_p(m,i,p,j)$. Hence, the storage capacity is not sufficient to store the entire platoon.
3. Two conflicting platoons are approaching the intersection on a main street and side street simultaneously. Suppose the size of the main street platoon is much larger than that of the side street platoon. Assuming the side street is currently being served (e.g., the side street phase is in the green state), the signal may decide to switch to serve the main street before completely serving the side street platoon to lower the total delay on the arterial. In this case, the side street platoons may be split to yield to main street platoons.

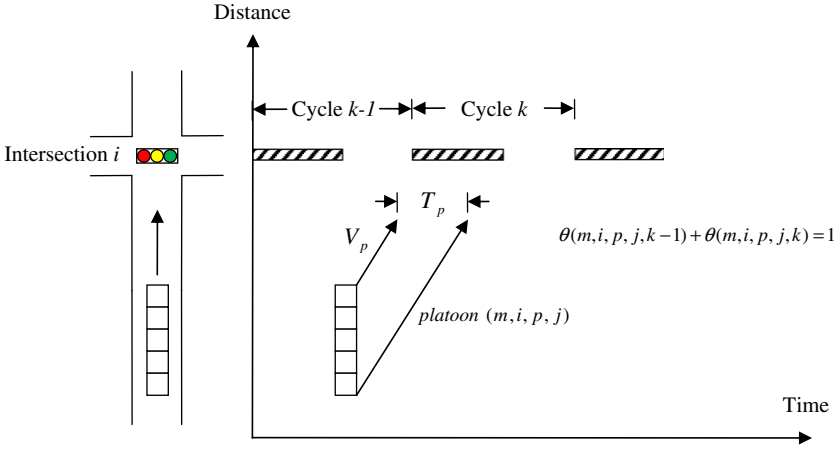

Fig. 7. Platoon serving cycle selection.

Platoon splitting in the first two cases is due to capacity constraints that cannot be violated. Platoon splitting in the third case is due to the performance benefit of platoon splitting, which increases arterial progression. The details of arterial progression will be discussed in the next section.

A positive variable $c_t(m,i,p,j) \in [0,1]$ is introduced to "cut" the platoons into two cycles, assuming that platoon can be served in no more than two cycles. Therefore, constraint (27) is rewritten as

$$\underline{T_a}(m,i,p,j) + T_p(m,i,p,j)^* c_t(m,i,p,j) \leqslant t(i,p,k) + g(i,p,k) + M(1-\theta(m,i,p,j,k)) \quad \forall (m,i,p,j) \in \Gamma, k \tag{29}$$

The first part of platoon (m,i,p,j) is served in cycle k, whereas the remaining is assigned to be served in cycle $k+1$. The remaining $N_p(m,i,p,j)^*(1-c_t(m,i,p,j))$ vehicles are delayed by at least one cycle. Assuming a platoon can be served in, at most, two cycles, a delay penalty, $d_{pen}(m,i,p,j)$, is added into the formulation to account for the delay as follows:

$$d_{pen}(m,i,p,j) = N_p(m,i,p,j)^*(1-c_t(m,i,p,j))^* C_r \quad \forall (m,i,p,j) \in \Gamma \tag{30}$$

where C_r is the reference cycle length, which represents an empirical estimation of cycle length, such as average over the historical cycles.

If one platoon is cut in the past cycle, the remaining vehicles in the platoon become the first queue at the stop bar. This means that the remaining vehicles in the platoon will be served in the next cycle without any queue delay. If the remaining platoon has to be further cut in the next cycle, the new delay penalty will be considered in the formulation by solving another MILP based on a rolling horizon optimization approach. Therefore, constraint (30) can be used to roughly estimate how much delay will be generated for platoon splits.

Green times are subject to the required time to clear the platoon in the selected cycle. Thus, additional lower bounds of $g(i,p,k)$ are added in the formulation as constraint (31):

$$g(i,p,k) \geqslant T_p(m,i,p,j)^* c_t(m,i,p,j) - M(1-\theta(m,i,p,j,k)) \quad \forall (m,i,p,j) \in \Gamma \tag{31}$$

Queue spillback occurs during oversaturated conditions, which can cause de facto red (Abu-Lebdeh and Benekohal, 1997). De facto red is an extreme case of queue build-up that is due to the complete blockage of upstream signals when no traffic can discharge during a green phase. To avoid de facto red and assign green times efficiently, the link storage capacity is utilized to set an upper limit on the split ratio in constraint (32):

$$c_t(m,i,p,j) \leqslant C_l(i_d,p)/N_p(m,i,p,j) \quad \forall (m,i,p,j) \in \Gamma \tag{32}$$

The experiments described in Section 4 used microscopic simulations to demonstrate the significance of constraint (32) during oversaturated conditions. The network throughput was increased by $10 \sim 20\%$, and overall delay was decreased by $5 \sim 10\%$ by considering platoon split and link capacity for the oversaturated arterial.

3.2.3. Delay evaluation

To evaluate the total delay of a platoon, delay is divided into two categories: signal delay and queue delay. These two types of delay can co-exist or exist separately for each platoon (Head, 1995). There are four combinations that must be considered. Platoons that travel through an intersection with free flow speed have no signal delay or queue delay. Platoons in the queue stopped at the stop bar by a red traffic light experience signal delay but no queue delay. Platoons that arrive at the back of the queue but eventually travel through the intersection during the green phase experience queue delay but no signal delay, as depicted in Fig. 8a. Platoons stopped by a queue delayed by the traffic signal will experience both queue delay and signal delay, as shown in Fig. 8b. Fig. 8a and b presents the case of two platoons requesting the same phase p. Platoon $(m,i,p,1)$ is a queue waiting at the stop bar, whereas platoon $(m,i,p,2)$ is a moving platoon approaching the intersection.

In Fig. 8a, t_1 is the time point when the leading vehicle of platoon 2 joins the back of platoon 1. $t_2$is the time point when the leading vehicle of platoon 2 starts to discharge $(t_2 - t_1)$ represents the leading vehicle queue delay $d_{q1}(m,i,p,j)$ in the platoon $(m,i,p,2)$. To simplify the calculation, the average queue delay can be approximated by the leading vehicle queue delay, shown as follows:

$$\begin{aligned} d_q(m,i,p,j)/N_p(m,i,p,j) \geqslant\ & t(i,p,k) + \sum_{m_1 \in M} \sum_{j_1=1}^{j-1} N_p(m_1,i,p,j_1)^* L_s/N_l(i,p)/V_s - (\underline{T_a}(m,i,p,j) \\ & - \sum_{m_1 \in M} \sum_{j_1=1}^{j-1} N_p(m_1,i,p,j_1)^* L_s/N_l(i,p)/V_p(m,i,p,j)) - M(1-\theta(m,i,p,1,k)) \quad \forall (m,i,p,j) \in \Gamma, k \end{aligned} \tag{33}$$

where $\sum_{m_1 \in M} \sum_{j_1=1}^{j-1} N_p(m_1,i,p,j_1)$ represents the total number of vehicles in front of platoon (m,i,p,j). Thus, $\sum_{m_1 \in M} \sum_{j_1=1}^{j-1} N_p(m_1,i,p,j_1)^*\ L_s/N_l(i,p)$ is the total queue length in front of platoon (m,i,p,j). t_2 can be found to be $t(i,p,k) + \sum_{m_1 \in M} \sum_{j_1=1}^{j-1}\ N_p(m_1,i,p,j_1)^* L_s/N_l(i,p)/V_s$ and t_1 is $\underline{T_a}(m,i,p,j) - \sum_{m_1 \in M}\ \sum_{j_1=1}^{j-1} N_p(m_1,i,p,j_1)^* L_s/N_l(i,p)/V_p(m,i,p,j)$. Also, the start of the phase p green time, $t(i,p,k)$, is a decision variable determined by the first queuing platoon decision variable $\theta(m,i,p,1,k)$.

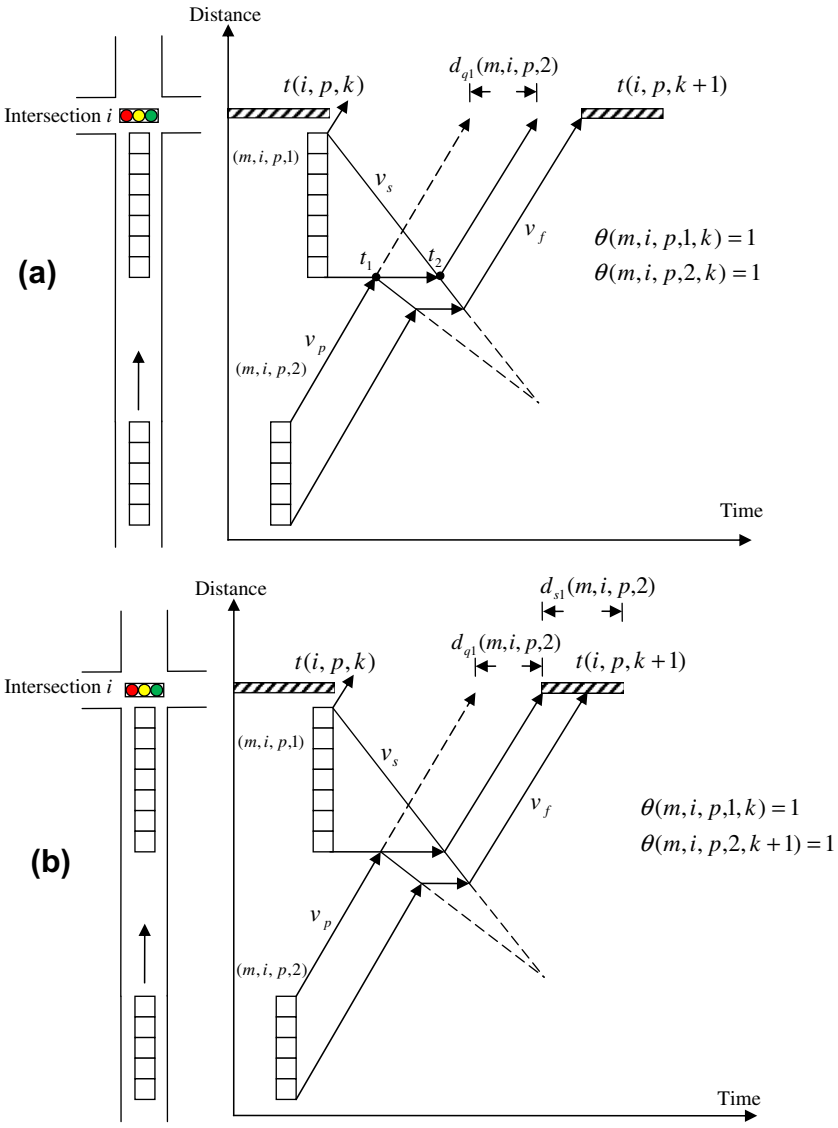

Fig. 8. (a) Queue delay when two platoons are served in the same cycle; (b) Queue delay and signal delay when two platoons are served in different cycles.

The queue delay can affect the arrival of a platoon; hence, constraint (29) holds only if the queue delay is equal to zero. Otherwise, queue delay should be included in constraint (29) to estimate the realized arrival time of the platoon at the stop bar, as described by constraint (34):

$$\begin{aligned}&\underline{T_a}(m,i,p,j)+d_q(m,i,p,j)/N_p(m,i,p,j)+T_p(m,i,p,j)c_t(m,i,p,j)\\&\quad\leqslant t(i,p,k)+g(i,p,k)+M(1-\theta(m,i,p,j,k))\quad\forall(m,i,p,j)\in\Gamma,k\end{aligned}\tag{34}$$

Signal delay is generated by stopping the moving platoon due to a red phase rather than the leading platoon. An example of signal delay is illustrated in Fig. 8b. The platoon $(m,i,p,2)$ is stopped by the discharging queue and then arrives during the red phase. The leading vehicle signal delay can be determined based on the starting time of phase p in cycle $k+1$, $t(i,p,k+1)$.

Delay can be calculated as the total area between the cumulative arrival curve and departure curve (Gazis, 1964). It is assumed that the headways between vehicles are the same in a platoon, so the platoon flow rates can be treated as constant arrival rates, as shown as Fig. 9.

In Fig. 9, the average arrival headway for a platoon is

$$H_p(m,i,p,j)=\frac{\overline{T_a}(m,i,p,j)-\underline{T_a}(m,i,p,j)}{N_p(m,i,p,j)}\quad\forall(m,i,p,j)\in\Gamma\tag{35}$$

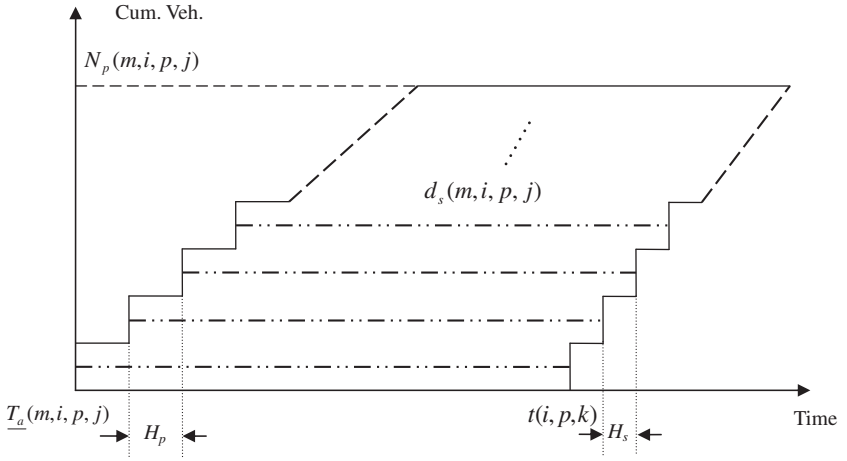

Fig. 9. Signal delay evaluation for the leading platoon.

where $\overline{T_a}(m,i,p,j) - \underline{T_a}(m,i,p,j)$ represents the arrival time difference between the leading vehicle and trailing vehicle in the platoon. If it is a queuing platoon, Eq. (35) holds as

$$\overline{T_a}(m,i,p,j) = \underline{T_a}(m,i,p,j) = 0 \tag{36}$$

which makes $H_p = 0$ for queuing platoons.

The headway for saturation departure can be calculated as

$$H_s(i,p) = \frac{H_{s1}}{N_l(i,p)} \quad \forall i,p \tag{37}$$

where H_s is the average headway of all the departure vehicles across all the output lanes, H_{s1} is saturation flow headway in a single lane (usually 2 s), and $N_l(i,p)$ is number of lanes being served by phase p at intersection i. The signal delay of the leading platoon is the summation of the vehicle delays measured between the arrival curve and departure curve illustrated in Fig. 9. In terms of non-leading platoons, the service time of platoon (m,i,p,j) will be delayed. It is necessary to account for the green time used to serve the preceding platoons, which is $\sum_{j_1=1}^{j-1}[\theta(m,i,p,j_1,k)^* T_p(m,i,p,j_1)]$, shown in constraint (38).

$$\begin{aligned} d_s(m,i,p,j) \geqslant \sum_{n=1}^{N_p(m,i,p,j)} \Bigg\{ & t(i,p,k) + \sum_{m_1 \in M} \sum_{j_1=1}^{j-1} [\theta(m_1,i,p,j_1,k)^* T_p(m_1,i,p,j_1)] + (n-1) H_s(m,i,p,j) \\ & - [\underline{T_a}(m,i,p,j)) + (n-1) H_p(i,p)] \Bigg\} - M(1 - \theta(m,i,p,j,k)) \quad \forall j,(m,i,p,j) \in \Gamma, k \end{aligned} \tag{38}$$

3.3. Multi-modal dynamical progression

The state of practice of traffic signal control strategies is coordinated-actuated signal control. Coordinated-actuated signals can offer additional flexibility compared with fixed-time traffic signals because of their ability to respond to cycle-by-cycle variation in traffic demand while still being able to provide progression for arterial movement. Traditionally, arterial coordination is managed through the determination of appropriate offsets, splits, and a common cycle length at each intersection. The "optimal" parameters of coordination signal plan can obtained by off-line signal optimization software, such as TRANSYT (Wallace et al., 1998), SYNCHRO (Trafficware, 2009) and PASSER (Chaudhary and Chu, 2003). However, the traffic flows optimized in signal optimization software are considered to be deterministic. In actuality, real-time traffic flow, even time-of-day traffic flow, can vary significantly (Yin, 2008). Previous on-line traffic-responsive (adaptive) signal control systems (Hunt, 1982; Luk, 1984) are able to adjust the coordination parameters (cycle length, offsets and splits) to fit the current detected traffic data, but the optimized plan assumes a constraint that every intersection must have the same cycle length, which may not be the best solution for every intersection on the arterial based on real-time traffic data. It is not necessary to maintain the concept of a common cycle length or an offset that is constant during each cycle, as it is assumed that multi-modal traffic data are available through vehicle-to-infrastructure communications.

A platoon-based dynamic coordination strategy is proposed as part of PAMSCOD to consider not only the current intersection but also the progression through downstream intersections given the path information of each vehicle. The concept is to consider platoon queue delay and signal delay in the current intersection as well as its potential queue delay and signal delay at the next downstream intersection. Another binary variable $\theta'(m,i,p,j,c)$ is introduced into the PAMSCOD formulation

to perform cycle selection at downstream intersection, $i_d \in I_d(i,p)$, for platoon (m,i,p,j) approaching the current intersection i for phase p. Again, it is assumed that only two cycles will be considered to serve a platoon. The platoon service cycle selection constraints are shown below:

$$t(i,p,k) + \sum_{m_1 \in M} \sum_{j_1=1}^{j-1} [\theta(m_1,i,p,j_1,k)^* T_p(m_1,i,p,j_1)] + F_d^* T_p(m,i,p,j)^* c_t(m,i,p,j) + T_t(i,i_d) + T_{st} + d_q'(m,i,p,j)/N_p(m,i,p,j)$$
$$\leqslant t(i_d,p,c) + g(i_d,p,c) + M(2 - \theta(m,i,p,j,k) - \theta'(m,i,p,j,c)) \quad \forall (m,i,p,j) \in \Gamma_e \cup \Gamma_w, k, c \tag{39}$$

$$\sum_c \theta'(m,i,p,j,c) = 1 \quad \forall (m,i,p,j) \in \Gamma_e \cup \Gamma_w \tag{40}$$

Constraint (39) is quite similar to constraint (27). One difference is that platoon dispersion is considered in (39) to ensure that enough phase green time is allocated at the downstream intersection to accommodate a platoon that is arriving at the upstream signal. The other difference is that (39) is taken into account when both θ and θ' are equal to 1, meaning that the departure and arrival occurs in cycle k and cycle c, at intersection i and i_d, respectively.

Queue delay and signal delay are evaluated based on a time–space diagram construction, as depicted in Fig. 10. Two additional constraints are included in the formulation to account for downstream queue delay and signal delay as follows:

$$d_q'(m,i,p,j)/N_p(m,i,p,j) \geqslant t(i_d,p,c) + \sum_{m_1 \in M} \sum_{j_1} N_p(m_1,i_d,p,j_1)^* L_s/N_l(i_d,p)/V_s - \{t(i,p,k) + [L(i,i_d)$$
$$- \sum_{m_1 \in M} \sum_{j_1} N_p(m_1,i_d,p,j_1)^* L_s/N_l(i_d,p)]/V_p(m,i,p,j)\} - M(2 - \theta(m,i_d,p,1,c) - \theta(m,i,p,j,k))$$
$$\forall (m,i,p,j) \in \Gamma_e \cup \Gamma_w, k, c \tag{41}$$

$$d_s'(m,i,p,j) \geqslant P_r(m,i,i_d)^* \sum_{n=1}^{N_p(m,i,p,j)} \Big\{ t(i_d,p,c) + \sum_{m_1 \in M} \sum_{j_1} [\theta(m_1,i_d,p,j_1,c)^* T_p(m_1,i_d,p,j_1)] + (n-1) H_s(m,i,p,j)$$
$$- [t(i,p,k) + T_t(m,i,i_d) + d_q'(m,i,p,j)/N_p(m,i,p,j) + T_{st} + (n-1) H_p(i,p)] \Big\}$$
$$- M(2 - \theta(m,i,p,j,k) - \theta'(m,i,p,j,c)) \quad \forall (m,i,p,j) \in \Gamma_e \cup \Gamma_w, k, c \tag{42}$$

In constraint (41), queue delay is assessed by determining the departure cycle k at intersection i and shockwave starting time from cycle c at downstream intersection i_d.

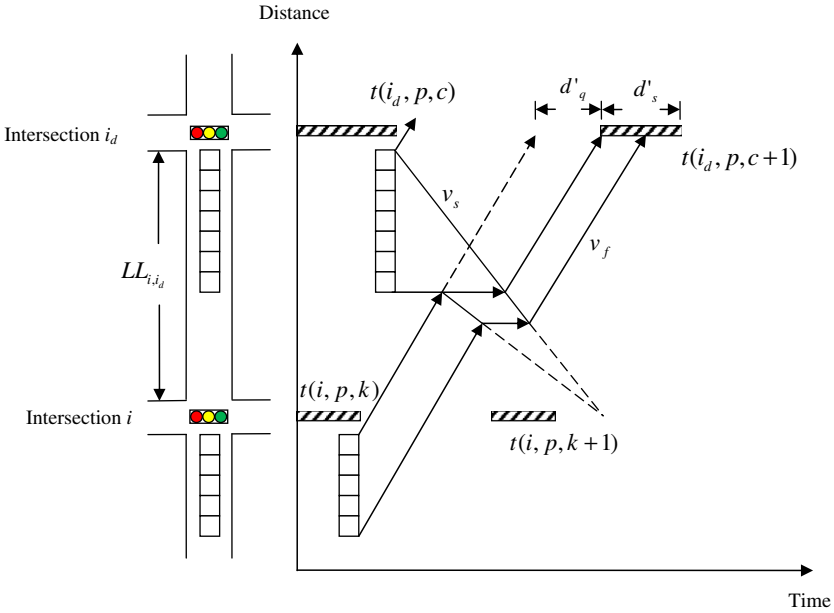

Fig. 10. Queue delay and signal delay at downstream intersection.

The platoon departure cycle is controlled by binary variable $\theta(m,i,p,j,k)$, whereas the shockwave starting time is controlled by binary variable $\theta(m,i_d,p,1,c)$, denoting that the first platoon to arrive at the downstream intersection is served in cycle c.

In constraint (42), signal delay is addressed by choosing the departure cycle k at intersection i and then the departure cycle c at downstream intersection i_d, controlled by binary variables $\theta(m,i,p,j,k)$ and $\theta'(m,i_d,p,j,c)$, respectively. $P_r(m,i,i_d) \in [0,1]$ represents the percentage of the vehicles remaining in the platoon after traveling through link (i,i_d).

3.4. Objective and summary of PAMSCOD

Usually, the objective of traffic signal control algorithms is to minimize a disutility function, such as travel time, delay or number of stops, or maximize a utility function, such as network throughput. The proposed formulation aims to serve all of the platoons in $|K|$ cycles using a rolling horizon approach. Hence, the throughput cannot be evaluated in the formulation. The number of stops is not computable due to the lack of an exact traffic flow model in the formulation. However, the total delay can be approximately assessed by the constraints derived in the previous sections.

The objective function will be to minimize the total weighted delay as well as the sum of slack variables representing the total green rest time, as described by Eq. (43). The delay weight factor $W(m,i,p,j)$ can be set to different values for each mode as well as each different platoon. Weight factors can depend on the priority level of the mode and can be adjusted for individual vehicles according to other real-time information, such as vehicle occupancy. It is assumed that emergency vehicles will receive a very high level of priority and can be considered in a complementary and separate formulation.

A summary of the formulation is depicted as follows:

$$\text{Objective}: \text{Minimize} \sum_{(m,i,p,j)\in\Gamma} W(m,i,p,j)^* \left[d_s(m,i,p,j) + d_q(m,i,p,j) + d_{pen}(m,i,p,j) + d'_s(m,i,p,j) + d'_q(m,i,p,j) \right] + \alpha \sum_{(i,p,k)} s(i,p,k) \tag{43}$$

Subject to

Precedence constraints: (4)–(26)
Selection constraints at current intersections: (28), (31), and (34)
Delay evaluation at current intersections: (30), (33), and (38)
Link capacity constraints: (32)
Selection constraints at downstream intersections: (39) and (40)
Delay evaluation at downstream intersections: (41) and (42)
Variables θ and θ' are binary decision variables, and all other variables are non-negative.

The maximal number of integer variables in PAMSCOD is equal to $2|\Gamma||K|$, where 2 indicates two sets of integer variables: θ and θ',$|\Gamma|$ the number of existing platoons, and $|K|$ the number of considered cycles. The solution times of this formulation vary depending on the size of the network and the traffic saturation condition. Rather, solution times are directly influenced by the total number of platoons, the size of platoons and number of cycles considered. Therefore, only a couple of the most significant platoons are considered in each phase. Moreover, a rolling horizon is adopted to solve MILP every 30 s for 4 cycles in the future.

The formulation was tested with GAMS/CPLEX 10.1 on a PC with a dual core 2.67-GHz processor and 3.5 Gb of memory. The test network is the same as that described in Section 4, shown in Fig. 11. For uncongested traffic conditions with no residual queue, the solution times are less than 1 s in most cases, as shown in Table 2. However, for congested and oversaturated conditions, the solution times extend from a few seconds up to a couple of minutes to achieve an optimality gap near 10%. The solution times have a significant impact on the implementation of an algorithm such as PAMSCOD in the field. Thus, the authors are actively developing simplified formulations and heuristic algorithms for future real-time applications.

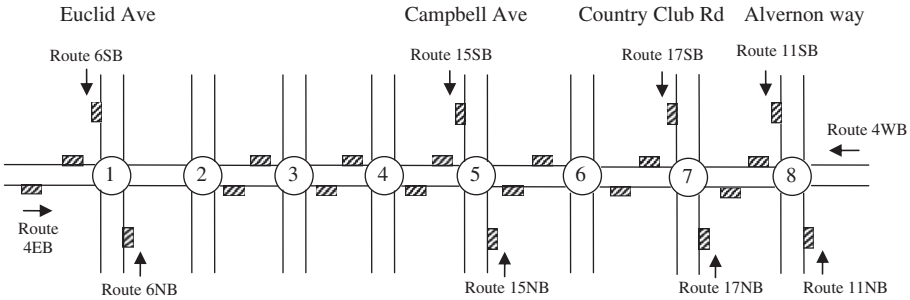

Fig. 11. Speedway arterial in Tucson, AZ.

Table 2
PAMSCOD solution times for an eight-intersection arterial.

Saturation rate	0.3	0.6	0.8	0.9	1	1.1	1.2
Average solution times (s)	0.264	0.461	1.104	7.136	78.589	85.131	91.146
Optimality gap (%)	10	10	10	10	10–20	10–20	10–20

4. Simulation experiments

Two importation travel modes are considered in the simulation experiments: automobiles and transit buses. Other motorized travel modes can be added in future with the same formulation. Bicycles and pedestrians may need to be treated differently within the control logic at each intersection.

The numerical experiments were conducted using VISSIM, a microscopic simulation tool used to model a traffic network. The ASC/3 SIL (software in the loop) feature allows a VISSIM user to use actual traffic signal controller logic, including actuated-coordinated and transit signal priority (TSP) (Econolite, 2009).

The entire evaluation platform contains VISSIM with Component Object Model (COM) and ASC virtual controller as the simulation environment and GAMS/CPLEX as the solver. Simulation in VISSIM can be easily controlled using COM, which can be created with a variety of programming languages, such as C++. First, COM runs a VISSIM model and continuously reads probe vehicle data including buses from VISSIM. Platoons are identified and estimated every 30 seconds. Once platoon information is updated, a MILP program will be formulated by COM and sent to GAMS/CPLEX. After retrieving the optimal signal plan from GAMS/CPLEX, COM implements the plan by sending phase control commands (phase hold and force-off) to ASC/3 SIL. Therefore, the entire system applies a network online platoon-based fixed time traffic signal control, which is updated every 30 s of simulated time.

The traffic network is based on an eight-intersection arterial, Speedway Blvd, from Euclid Ave. to Alvernon Way in Tucson, AZ. Ten bus routes were added to the model, as shown in Fig. 11, based on the actual bus routes operated by Sun Tran (2010). Two of the ten bus routes travel on Speedway, whereas the others travel on the side streets. All bus stops are located at the far-side of the intersections. Bus frequencies vary by $5 \sim 20$ min/bus. Every bus sends a priority request when it approaches an intersection, and each of them is treated as a single-vehicle platoon.

4.1. Illustration of solutions from PAMSCOD

Signal-timing optimization software such as TRANSYT-7F and SYNCHRO are commonly used as benchmarks for good arterial signal timing. In this paper, the performance of PAMSCOD was compared with SYNCHRO optimized signal timings under different levels of traffic demand. Seven different deterministic flows were designed as the basic flow levels of experiment. They corresponded to different levels of intersection saturation – 0.3, 0.6, 0.8, 0.9, 1.0, 1.1 and 1.2, which were estimated by the intersection capacity utilization in SYNCHRO (Husch and Albeck, 2003). The "optimal" solutions obtained from SYNCHRO can be considered as near-optimal solutions for deterministic flows, as the input flows of VISSIM are the same as the input flows of SYNCHRO. Four sets of stochastic volumes were generated by assuming a normal distribution using the deterministic flows as the mean and 20% of deterministic flow as variance. The random flow sets were selected to have similar total volumes.

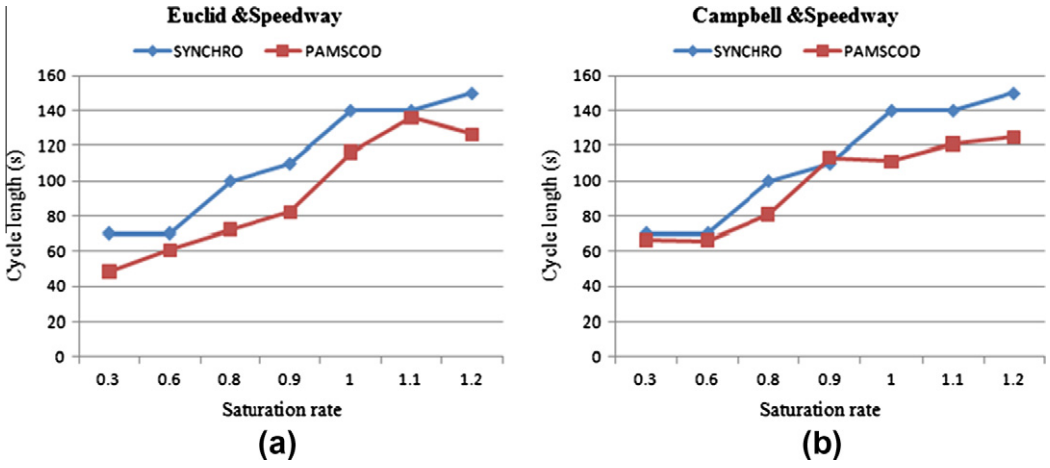

Fig. 12. Comparisons between optimized cycle length from SYNCHRO and average cycle length in PAMSCOD at two critical intersections: (a) Intersection at Euclid and Speedway and (b) Intersection at Campbell and Speedway.

First, the output cycle lengths of PAMSCOD were compared with SYNCHRO's optimal cycle length. Average cycle lengths with seven volume levels were recorded at the two most saturated intersections: Euclid and Speedway, and Campbell and Speedway. Fig. 12a and b plot the curves of cycle length with PAMSCOD and SYNCHRO at these two intersections, respectively. The results show that PAMSCOD's cycle lengths are slightly shorter than SYNCHRO's optimized cycle lengths, but PAMSCOD's cycle lengths follow the tendency that arterial cycle length is proportional to flow levels. Previous studies show

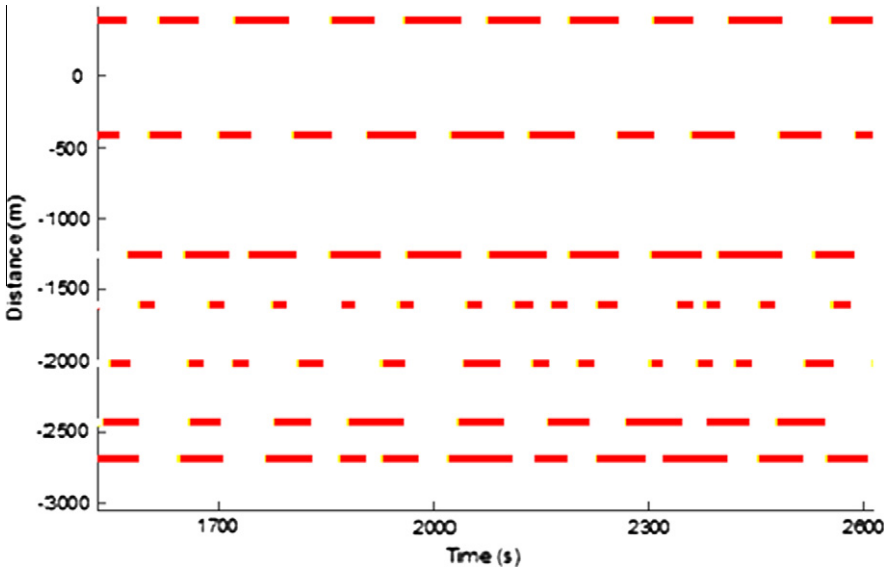

Fig. 13. Online optimized signal plan from PAMSCOD at a saturation rate of 0.9 (intersection 1 is in the bottom).

Table 3
Comparisons of the performance of four methods (Th. = throughput; Dv, Db = average delay for all vehicles and buses, respectively).

Set	SR[a]	ASC free			ASC coord			PAMSCOD			TSP coord		
		Th.	Dv	Db	Th.	Dv	Db	Th.	Dv	Db	Th.	Dv	Db
1	0.3	4600	11.18	10.66	4603	17.50	19.67	4601	14.82	12.57	4604	21.48	12.84
	0.6	9112	16.02	16.29	9123	18.48	22.47	9108	17.86	14.08	9121	25.04	15.53
	0.8	13,633	25.47	26.82	13,673	22.19	25.91	13,649	21.90	19.79	13,644	32.10	18.20
	0.9	17,997	42.63	44.73	17,294	29.91	33.00	18,283	29.79	25.34	16,351	59.31	40.31
	1	18,947	49.08	56.39	18,511	38.41	40.52	19,630	35.24	24.15	17,979	57.73	33.43
	1.1	19,385	57.73	56.65	18,401	45.25	46.99	19,626	42.11	32.83	17,135	58.05	32.61
	1.2	20,798	94.51	101.70	20,599	65.14	62.51	20,670	61.10	51.30	19,108	69.69	50.81
2	0.3	4613	10.81	10.05	4620	17.35	20.21	4617	15.97	12.49	4608	21.55	12.26
	0.6	9830	17.45	18.20	9839	18.82	20.67	9849	18.48	15.41	9787	27.53	15.23
	0.8	14,825	28.73	27.33	14,628	27.97	29.44	14,872	24.84	21.47	14,068	45.80	21.04
	0.9	16,724	37.25	40.98	16,026	37.17	41.40	16,881	27.82	24.16	15,138	48.42	25.21
	1	18,713	60.75	62.14	18,499	44.35	47.38	19,342	41.14	32.28	16,963	54.27	38.12
	1.1	19,840	80.94	83.10	19,319	51.78	58.21	20,098	52.62	41.90	17,957	59.98	38.52
	1.2	20,876	98.51	116.27	21,122	75.15	85.52	21,163	67.65	56.19	19,419	71.76	47.68
3	0.3	4690	11.29	10.01	4692	17.64	20.14	4696	16.20	13.60	4691	22.00	13.56
	0.6	8991	16.23	17.01	8990	17.43	21.89	9000	17.02	14.72	8972	24.36	14.23
	0.8	12,222	22.35	22.89	11,672	21.89	25.23	12,254	20.34	18.09	11,767	32.90	15.51
	0.9	18,152	44.46	41.71	17,168	38.10	33.37	18,818	34.03	27.44	16,048	52.24	27.20
	1	18,691	54.38	61.48	18,004	41.46	44.94	19,180	42.52	34.48	17,570	58.77	37.50
	1.1	19,022	69.14	73.90	19,702	51.04	54.22	20,859	54.16	43.46	18,160	59.34	41.57
	1.2	20,543	89.66	99.10	20,464	60.03	59.88	21,092	60.63	46.46	19,294	66.44	48.41
4	0.3	4744	11.34	10.22	4754	17.72	20.18	4752	15.80	13.93	4744	21.14	12.47
	0.6	9196	16.17	16.74	9194	18.35	21.55	9198	17.97	15.00	9184	23.46	12.98
	0.8	12,508	20.44	21.04	12,519	21.79	24.29	12,513	20.69	18.56	12,467	30.38	16.85
	0.9	18,579	49.35	52.20	17,836	41.34	44.97	19,006	40.41	30.37	16,550	50.86	30.84
	1	18,793	57.45	58.71	18,081	40.96	45.39	20,105	43.42	36.44	17,400	52.92	27.70
	1.1	20,311	67.81	74.55	19,593	49.30	51.10	20,637	49.85	35.38	18,488	59.27	29.24
	1.2	20,582	82.64	98.88	20,036	55.58	59.86	21,276	58.65	48.44	19,145	62.73	44.68

[a] Approximate saturation rate.

that cycle lengths in a small range may result in similar delays (Miller, 1963). Therefore, the output cycle lengths of PAMSCOD are reasonable.

The signal plans generated by PAMSCOD are visualized on the space–time diagram depicted in Fig. 13. The intersection at the bottom is intersection 1 and the intersection at the top is intersection 7. The results show a two-stage, two-way progression in the diagram. One two-way progression is from intersection 1 (Euclid and Speedway) to intersection 5 (Campbell and Speedway). The other progression two-way progression is from intersection 5 (Campbell and Speedway) to intersection 7(Country Club and Speedway). The results are intuitive as intersection 5 (Campbell and Speedway) is nearly saturated (saturation rate is 0.9) and has a larger volume on the northbound (2742 veh/hr) and southbound (2296 veh/hr) movements on Campbell as opposed to the eastbound (1489 veh/hr) and westbound (1262 veh/hr) movements on Speedway. PAMSCOD outputs smaller eastbound and westbound green times than northbound and westbound, which seems reasonable. The volume pattern of intersection 5 (Campbell and Speedway) is detected by PAMSCOD correctly, and good splits are achieved.

In addition to progression, a large network can be partitioned into a sub-network according to the time–space diagram from PAMSCOD. Network partitions can be used to reduce the total number of concurrently considered intersections, resulting in better MILP solution times. If the solutions in a sub-network show a good two-way progression in high-volume

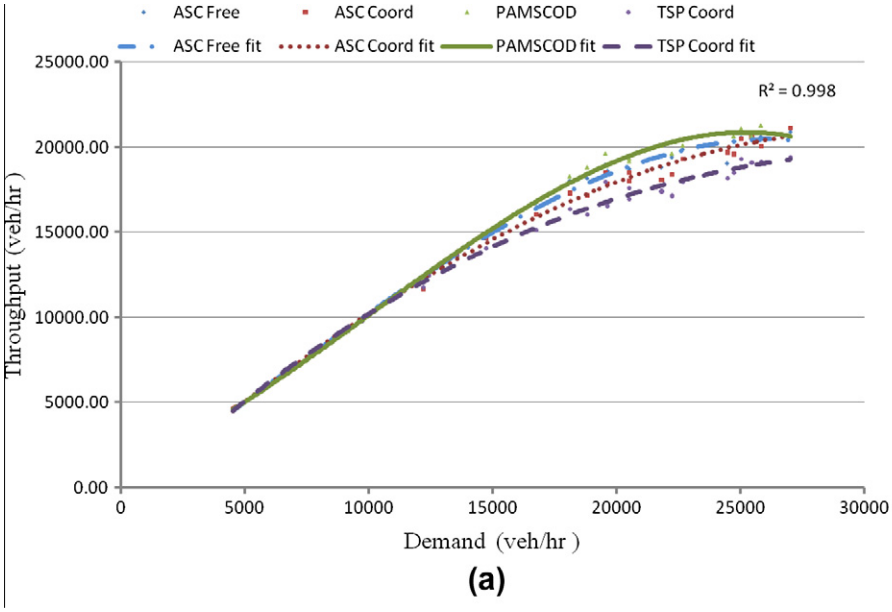

Fig. 14. (a) Network throughput. (b) Average vehicle delay. (c) Average bus delay.

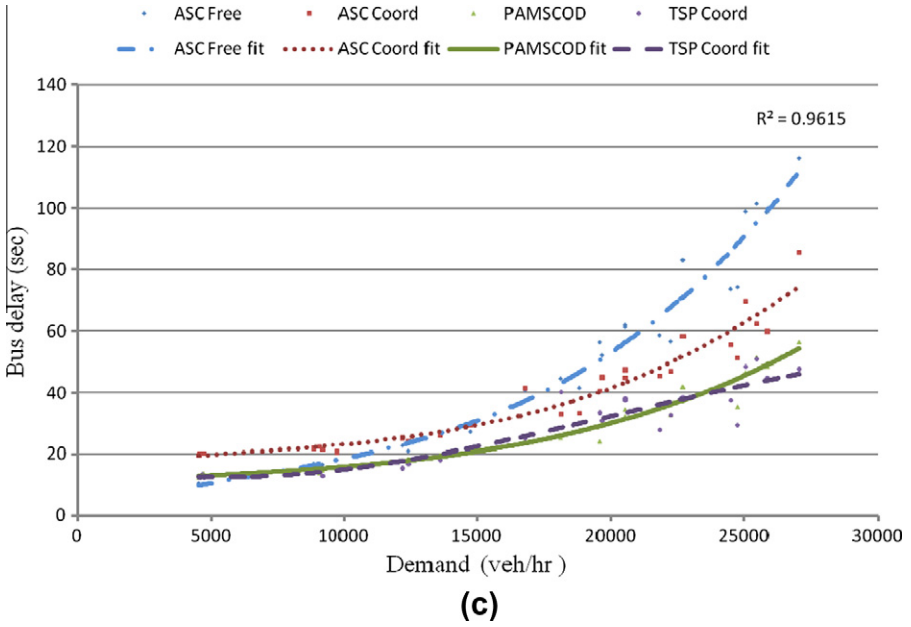

Fig. 14 (*continued*)

experiments, this sub-network can be treated as an isolated network. The details of network partition are not considered in this paper.

4.2. Experiment results

Three different traffic signal methods were compared with PAMSCOD under 100% penetration rates using the four stochastic volume sets. "ASC Free" treated each intersection independently and applied actuated control at each isolated intersection without bus priority. "ASC Coord" represented the coordinated–actuated non-priority signal control based on the optimal signal timing plans obtained from SYNCHRO 7.0. "TSP Coord" added transit signal priority (TSP) to "ASC coord" control. The detailed settings of TSP in the ASC/3 SIL are documented in Econolite (2009).

Table 3 presents the detailed results based on one-hour simulations in VISSIM after a 10-min warm-up period. The performances of the four different methods are compared based on network throughput, average total vehicle delay and stops for each phase at each intersection, average bus delay and stops per phase intersection. Polynomial regression lines are plotted in Fig. 14a–c for the four methods based on the data pairs of demand-throughput, demand-average vehicle delay and demand-average bus delay. Performance percentage changes of PAMSCOD are summarized in Table 4.

The following are some observations:

1. When the saturation rate is lower than 0.9, all of the vehicles are served by all four control methods. However, not all demanding vehicles can be served when the saturation rate is near or higher than 0.9. The trend line in Fig. 14a, as well as the data in Table 4, shows that PAMSCOD always has the highest throughput among all of the methods, especially when the traffic condition is oversaturated. When considering multi-modes (buss and automobiles) in coordination, the "TSP Coord" method has the lowest throughput among all of the methods for congested traffic conditions.

Table 4
Performance percentage changes of PAMSCOD compared with other methods.

	Methods	Low volume (%)	Medium volume (%)	High volume (%)	Oversaturated (%)
Throughput	ASC free	0.10	0.07	1.15	2.90
	ASC coord	−0.01	0.03	4.21	4.81
	TSP coord	0.11	0.24	8.21	11.33
Avg. vehicle delay	ASC Free	40.78	8.35	−16.54	−28.75
	ASC coord	−10.57	−2.40	−7.89	−1.22
	TSP coord	−27.13	−28.77	−36.92	−17.20
Avg. bus delay	ASC Free	28.59	−13.19	−30.10	−48.15
	ASC coord	−34.43	−31.48	−27.21	−26.48
	TSP coord	2.94	2.71	−0.55	3.22

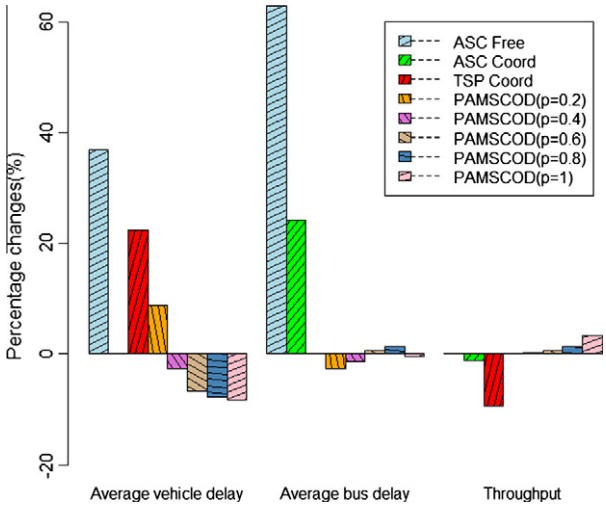

Fig. 15. Performance percentage changes of PAMSCOD under different penetration rates.

2. When the saturation rate is low, "ASC Free" always provides the minimal overall total vehicle delay, which is acknowledged by most researchers. However, "ASC Free" degrades and "ASC Coord" improves as the saturation rate increases. Although "ASC Coord" can be treated as a near optimal method, PAMSCOD outperforms "ASC Coord" in average total vehicle delay by about 8% under high volume scenarios and 1% under oversaturated scenarios, respectively, as shown in Table 4. Because "TSP Coord" holds absolute priority for buses, the average vehicle delay for all vehicles of PAMSCOD is around 20–30% less than the average total vehicle delay under the "TSP Coord" method.
3. With regard to average bus delay, the bus delay of "TSP Coord" can be treated as the lower bound, as it provides priority to any bus approaching intersection, regardless of other passenger vehicles' delay or throughput. PAMSCOD yields only 3% more bus delay than "TSP coord" but 10% more network throughput under saturated conditions. Compared with "ASC Coord", the average bus delay is reduced by about 25–30% under different scenarios, as shown in Table 4.

According to the above observations, PAMSCOD provides the lower bound of total vehicle delay, compared with "ASC Coord", which is optimized by SYNCHRO with deterministic flows. PAMSCOD can achieve performance very near the lower bound of bus delay, as compared with "TSP Coord".

Fig. 15 shows the performance percentage comparison of PAMSCOD under different penetration rates. Only under congested scenarios conditions are considered and simulated. "TSP Coord" represents the condition of coordinated traffic signals with Transit Signal Priority, which is chosen as the baseline method for bus delay and assumed to provide the best existing condition in terms of transit delay. "ASC Coord" is selected to be the baseline for delay of all vehicles. It is assumed that "ASC Coord", which represents coordinated traffic signals without Transit Signal Priority, provides the near-optimal average vehicle delay for all vehicles. "ASC Free" is considered as the baseline solutions when considering throughput, since it yields better throughput than either "ASC Coord" or "TSP Coord". When the penetration rate is at 20%, PAMSCOD reduces bus delay by 2.5% because all buses are equipped no matter what penetration is. Average vehicle delay is 8% higher compared with "ASC Coord". However, PAMSCOD reaches both near-optimal values when penetration rates are near 40% without any prior knowledge of traffic flow rate. So 40% penetration rate can be regarded as a critical value for future implementation of V2I communications based on this limited finding. The critical penetration also matches what was found in platoon identification.

5. Summary

A unified platoon-based formulation called PAMSCOD is presented in this paper to optimize arterial (network) traffic signals for multiple travel modes, given the assumption that advanced communication systems are available between vehicles and traffic controllers. One important feature of PAMSCOD is that the prior knowledge of network flow profile is not required. Also PAMSCOD does not assume the traditional common network cycle length. It aims to provide multi-modal dynamical progression on an arterial based on real-time platoon information.

Microscopic simulation shows that PAMSCOD can successfully coordinate traffic signals considering two traffic modes including buses and automobiles and significantly reduce vehicle delay for both modes. At about a 40% penetration rate, PAMSCOD can outperform state-of-practice signal control methods. Compared with coordinated-actuated traffic signal control optimized by SYNCHRO, PAMSCOD decreases the overall average vehicle delay by about 8%, while the average bus delay is reduced by about 25–30%. Compared with same SYNCHRO plans with TSP, the average overall vehicle delay in PAMSCOD is

reduced by about 20–30%, and average bus delay is increased by only 3%. However, the throughput is increased by more than 10% for congested cases, as compared with the TSP method.

Future research should mainly focus on how to reduce the solution times of PAMSCOD by considering network partition, reducing the complexity of the MILP or developing heuristic algorithms. In this paper, no prior knowledge of traffic flow is needed in the model. However, it would be interesting to incorporate short term traffic prediction with signal control. Also it is a potential research topic to explore signal control optimization with online incident data in addition to traffic data.

Appendix A

Notation definition of decision variables (lower case) and data (upper case)

Type	Symbol	Definition
Sets	$m, m_1 \in M$	The set of travel modes
	$i, n \in I$	The set of intersections
	$i_d \in I_d(i,p)$	The set of downstream intersections for current phase p at intersection i ($I_d(i,p) \subset I$)
	$p \in P$	The set of phases
	$p \in P_t$	The set of through traffic phases ($P_t \subset P$)
	$j, j_1 \in J$	The set of platoons
	$k, c \in K$	The set of cycles
	$p \in \Delta_{s1}(i)$	The starting phase in ring 1 at intersection i($\Delta_{s1}(i) \subset P$)
	$p \in \Delta_{s2}(i)$	The starting phase in ring 2 at intersection i($\Delta_{s2}(i) \subset P$)
	$(i,p) \in \Delta_{non}$	The set of unused phases (compared with a standard NEMA 8-phase controller) at intersection i ($\Delta_{non}(i) \subset P$)
	$(i,p,k) \in \Delta_p$	The set of past phases in the first cycle at intersection i($\Delta_p(i) \subset P$)
	$(m,i,p,j) \in \Gamma = \{(m,i,p,j) \mid Np(m,i,p,j) > 0\}$	The set of jth valid platoons in mode m, intersection i, and phase p
	$(m,i,p,j) \in \Gamma_e = \{(m,i,p,j) \mid i \leqslant \lvert I \rvert - 1, p = 2, Np(m,i,p,j) > 0\}$	The set of platoons on arterial Eastbound ($\Gamma_e \subset \Gamma$)
	$(m,i,p,j) \in \Gamma_w = \{(m,i,p,j) \mid i \geqslant 2, p = 6, Np(m,i,p,j) > 0\}$	The set of platoons on arterial Westbound ($\Gamma_w \subset \Gamma$)
Decision variables	$c_t(m,i,p,j)$	The cut ratio of serving platoon $(m,i,p,j) \in \Gamma$
	$d_{pen}(m,i,p,j)$	Delay penalty to cut a platoon to serve in two cycles for platoon $(m,i,p,j) \in \Gamma$
	$d_q(m,i,p,j)$	Queue delay at the current intersection for platoon $(m,i,p,j) \in \Gamma$
	$d_{q1}(m,i,p,j)$	Leading vehicle queue delay in current intersection for platoon $(m,i,p,j) \in \Gamma$
	$d_s(m,i,p,j)$	Signal delay in current intersection for platoon $(m,i,p,j) \in \Gamma$
	$d_{s1}(m,i,p,j)$	Leading vehicle signal delay in current intersection for platoon $(m,i,p,j) \in \Gamma$
	$d'_q(m,i,p,j)$	Queue delay in downstream intersection for platoon $(m,i,p,j) \in \Gamma_2 \cup \Gamma_6$
	$d'_s(m,i,p,j)$	Signal delay in downstream intersection for platoon $(m,i,p,j) \in \Gamma_2 \cup \Gamma_6$
	$e(i,p)$	Elapsed green time for phase p at intersection i
	$g(i,p,k)$	Green time for phase p during cycle k at intersection i
	$s(i,p,k)$	slack variables for phase p during cycle k at intersection i
	$s_1(i)$	slack variable for ring 1 at intersection i
	$s_2(i)$	slack variable for ring 2 at intersection i

Appendix A (*continued*)

Type	Symbol	Definition
	$t(i,p,k)$	Starting time of phase p during cycle k at intersection i
	$v(i,p,k)$	Phase duration time of phase p during cycle k at intersection i
	$\theta(m,i,p,j,k)$	0–1 binary variables. Whether to serve platoon (m,i,p,j) in cycle k at current intersection (if $\theta(m,i,p,j,k) = 1$, the platoon (m,i,p,j) is served in cycle k at current intersection; else, not served in cycle k)
	$\theta'(m,i,p,j,c)$	0–1 binary variables. Whether to serve platoon (m,i,p,j) in cycle c at downstream intersection (if $\theta'(m,i,p,j,c) = 1$, the platoon (m,i,p,j) is served in cycle c at downstream intersection; else, not served in cycle c)
Data	α	Weighting factor for the sum of slack variables
	$C_l(i,p)$	Link remaining storage capacity before solving the MILP
	C_r	Reference cycle length (s)
	$E(i,p)$	Elapsed green times for starting phase p at intersection i ($p \in \Delta_{s1}(i) \cup \Delta_{s2}(i)$) (s)
	$G_{\min}(i,p)$	Minimal green time for phase p at intersection i (s)
	$G_{\max}(i,p)$	Maximal green time for phase p during at intersection i(s)
	$F_d(i,n)$	Platoon dispersion factor on link (i,n)
	$H_p(m,i,p,j)$	Average headway of a platoon (m,i,p,j) (s)
	$H_s(i,p)$	Saturation flow headway for phase p at intersection i
	H_{s1}	Saturation flow headway for a single lane (s)
	$L(i,n)$	Link length between intersection i and n (m)
	L_s	Average vehicle spacing in queue (m)
	M	A large number
	$N_l(i,p)$	Number of lanes at intersection i served by phase p
	$N_p(m,i,p,j)$	Number of vehicles in platoon (m,i,p,j)
	$O_1(i)$	Initial time for ring 1 at intersection i(s)
	$O_2(i)$	Initial time for ring 2 at intersection i(s)
	$P_r(m,i,n)$	Percentage of a platoon arriving intersection n when traveling from intersection i ($P_r(m,i,n) \in [0,1]$)
	$R(i,p)$	Red clearance time for phase p at intersection i (s)
	$S_r(i,p)$	Saturation rate for phase p at intersection i (veh/h)
	$\underline{T_a}(m,i,p,j)$	Time arrival for leading vehicle in platoon (m,i,p,j) (s)
	$\overline{T_a}(m,i,p,j)$	Time arrival for tail vehicle in platoon (m,i,p,j) (s)
	$T_p(m,i,p,j)$	Green time needed to clear the platoon (m,i,p,j) (s)
	T_{st}	Start-up lost time (s)
	$T_t(m,i,n)$	Travel time between intersection i and n for mode m (s)
	$V_f(m,i,p)$	Free flow speed for model m in phase p at intersection i (m/s)
	$V_p(m,i,p,j)$	Average speed for platoon (m,i,p,j) (m/s)
	V_s	Shock wave speed (m/s)
	$W(m,i,p,j)$	Weight for platoon (m,i,p,j)
	$Y(i,p)$	Yellow clearance time for phase p at intersection i (s)
	Z	Allowable additional green rest time (s)

References

Abu-Lebdeh, G., Benekohal, R., 1997. Development of traffic control and queue management procedures for oversaturated arterials. Transportation Research Record: Journal of the Transportation Research Board 1603, 119–227.

Chaudhary, N., Abbas, M., Charara, H., 2006. Development and field testing of platoon identification and accommodation system. Transportation Research Record: Journal of the Transportation Research Board 1978, 141–148.

Chaudhary, N., Chu, C., 2003. New PASSER Program for Timing Signalized Arterials. Texas Transportation Institute.

Comert, G., Cetin, M., 2009. Queue length estimation from probe vehicle location and the impacts of sample size. European Journal of Operational Research 197 (1), 196–202.

Daganzo, C., 1995. The cell-transmission model, Part II: network traffic. Transportation Research Part B: Methodological 29B (2), 79–93.

Daganzo, C., 1994. The cell-transmission model: a simple dynamic representation of highway traffic. Transportation Research Part B: Methodological 28B (4), 269–287.

Dell'Olmo, P., Mirchandani, P., 1995. REALBAND: an approach for real-time coordination of traffic flows on networks. Transportation Research Record: Journal of the Transportation Research Board 1494, 106–116.

Econolite, 2009. Transit Signal Priority (TSP). User Guide for Advanced System Controller ASC/3.

Gartner, N. et al, 1991. A multi-band approach to arterial traffic signal optimization. Transportation Research Part B: Methodological 25 (1), 55–74.

Gaur, A., Mirchandani, P., 2001. Method for real-time recognition of vehicle platoons. Transportation Research Record: Journal of the Transportation Research Board 1748, 8–17.

Gazis, D., 1964. Optimum control of a system of oversaturated intersections. Operations Research 12 (6), 815–831.

He, Q. et al., 2010. Heuristic algorithms to solve 0–1 mixed integer LP formulations for traffic signal control problems. In: 2010 IEEE International Conference on Service Operations and Logistics, and Informatics. Qingdao, China, pp. 118–124.

Head, K., 1995. Event-based short-term traffic flow prediction model. Transportation Research Record 1324, 105–114.

Head, K., Gettman, D., Wei, Z., 2006. Decision model for priority control of traffic signals. Transportation Research Record: Journal of the Transportation Research Board 1978, 169–177.

Herrera, J.C., Bayen, A.M., 2010. Incorporation of Lagrangian measurements in freeway traffic state estimation. Transportation Research Part B: Methodological 44 (4), 460–481.

Hunt, P.B. et al, 1982. The SCOOT on-line traffic signal optimization technique. Traffic Engineering and Control 23 (4), 190–192.

Husch, D., Albeck, J., 2003. Trafficware Intersection Capacity Utilization.

Jiang, Y., Li, S., Shamo, D., 2006. A platoon-based traffic signal timing algorithm for major–minor intersection types. Transportation Research Part B: Methodological 40, 543–562.

Lin, W., Wang, C., 2004. An enhanced 0–1 mixed-integer linear program formulation for traffic signal control. IEEE Transactions on Intelligent Transportation Systems 5 (5), 238–245.

Little, J., 1966. The synchronization of traffic signals by mixed-integer linear programming. Operations Research 14 (4), 568–594.

Little, J., Kelson, M., Gartner, N., 1981. MAXBAND: a program for setting signals on arteries and triangular networks. Transportation Research Record (795), 40–46.

Lo, H., 2001. A cell-based traffic control formulation: strategies and benefits of dynamic timing plan. Transportation Science 35, 721–744.

Lo, H., 1999. A novel traffic signal control formulation. Transportation Research Part A 33 (6), 433–448.

Luk, J., 1984. Two traffic-responsive area traffic control methods: SCAT and SCOOT. Traffic Engineering and Control 25, 14–18.

Miller, A., 1963. Settings for fixed-cycle traffic signals. Operational Research 14 (4), 373–386.

Mirchandani, P., Head, K., 2001. A real-time traffic signal control system: architecture, algorithms, and analysis. Transportation Research Part C: Emerging Technologies 9 (6), 415–432.

Neumann, T., 2010. A cost-effective method for the detection of queue lengths at traffic lights. Traffic Data Collection and its Standardization. International Series in Operations Research & Management Science. Springer, New York, pp. 151–160. <http://dx.doi.org/10.1007/978-1-4419-6070-2_10>.

Research and Innovative Technology Administration, 2011. Connected Vehicle Research. <http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm> (accessed 31.03.11).

Robertson, D.I., 1969. TRANSYT method for area traffic control. Traffic Engineering and Control 11 (6), 276–281.

Smith, B.L., Venkatanarayana, R., Park, H., Goodall, N., Datesh, J., Skerrit, C., 2010. IntelliDriveSM Traffic Signal Control Algorithms. University of Virginia.

Stamatiadis, C., Gartner, N., 1996. MULTIBAND-96: a program for variable bandwidth progression optimization of multiarterial traffic networks. Transportation Research Record 1554, 9–17.

Sun Tran, 2010. Sun Tran – Tucson, Arizona. <http://www.suntran.com/> (accessed 10.02.11).

Trafficware, 2009. Synchro Studio 7.0 User's Guide.

Wallace, C. et al., 1998. TRANSYT-7F User's Guide.

Wasson, J. et al, 1999. Reconciled Platoon Accommodations at Traffic Signals. Purdue University, West Lafayette, IN.

Yin, Y., 2008. Robust optimal traffic signal timing. Transportation Research Part B: Methodological 42 (10), 911–924.