Check for updates

Unified Framework for Dynamic Traffic Assignment and Signal Control with Cell Transmission Model

H. M. Abdul Aziz and Satish V. Ukkusuri

This research presents a unified framework for dynamic traffic assignment and signal control optimization. An objective function based on the system-optimal approach with an embedded traffic flow model (the cell transmission model) for dynamic traffic assignment was considered. The optimization model (a mixed-integer program) explicitly considered intersection delay and lost time from phase switches in addition to a traditional travel time objective. Two test networks were used to demonstrate the applicability of the proposed model. Results showed better performance of the models when they were compared with fixed-signal timing plans. The formulation of signal control design also accounted for the variation of cycle length, and results showed the variation of cycle lengths with different objective functions under different levels of congestion.

The *2010 Urban Mobility Report* (*1*) published by the Texas Transportation Institute (TTI) reveals that urban traffic congestion in the United States causes additional travel of 4.8 billion hours and consumption of an extra 3.9 billion gallons of fuel. All these factors combine into a total congestion cost of $115 billion in 2009. The conditions in 2009 were actually worse than those in 2008. The delay per commuter in 2009 increased to 34 h from the 33.7 h reported for 2008 (*1*). In addition, traffic congestion acts as a major contributor to roadside emissions that come from mass amounts of stop-and-go traffic during high levels of congestion. Clearly, traffic congestion has been a major problem in the United States over the past few decades. Although signal control is an inevitable component in assuring safety in the movements of vehicles, it is one of the major sources of traffic delays for road users. According to the Executive Summary of the *2007 National Traffic Signal Report Card* (*2*), signal timing plans without optimization cause about 5% to 10% of total travel delay, about 295 million vehicle hours just on major roadways. Furthermore, stops made at intersections are potential sources for emissions and undesired fuel consumption. So it is important to have optimal signal timing plans that (*a*) minimize intersection delay by ensuring maximum utilization of the intersection and (*b*) reduce emissions and fuel consumption.

However, the signal control design problem is a complex process that depends on route choice behavior, driving characteristics of individuals (e.g., smaller gap acceptance for left turning), and time-varying flow characteristics of the network. Signal optimization (determination of optimal signal timing plans) and traffic assignment (assignment of the demands to the paths in the form of flows or traffic volumes) show high interdependency in the context of the transportation planning process. Solutions for the traffic assignment problem are obtained under the assumption that the traffic signal settings are given. Similarly, the traffic signal settings are designed under the assumption of fixed flow distribution among the paths (as obtained from the traffic assignment). However, the route choice decisions made by road users depend on the state of the traffic network (e.g., link capacities, signal settings). Thus, signal settings influence the route choices made by road users, and route choice behavior affects the signal timing plans by changing the flow patterns in the network. Accordingly, researchers need to consider signal control design and the assignment problem in an integrated manner; failure to do so would lead to inconsistent results (*3–7*).

Recent progress in the area of intelligent transportation systems—including the advanced traffic management system, the advanced traveler information system, and the online routing information system—has shown the feasibility and applicability of dynamic traffic management tools to accommodate traffic dynamics in real time. One can apply these tools satisfactorily both to perform dynamic traffic assignment and to optimize networkwide signal settings in real time (*8, 9*) with the help of an integration framework of the signal control design and dynamic traffic assignment. This problem is commonly referred to as the "combined traffic assignment and control" problem. The assignment problem can take either of the two major approaches: user optimal or system optimal. The user-optimal approach implies that each road user attempts to minimize his or her generalized cost of travel (which can include travel time, emissions, fuel consumption, and other variables of interest) through rational decision making. In contrast, the system-optimal approach minimizes the generalized cost for all the users of the system as a whole. This research puts its focus on an integrated optimization framework of system-optimal dynamic traffic assignment and signal control design.

Two important parameters of signal timing plans are the cycle length and the lost time caused by phase switches within a cycle. Both of these are important for implementation of optimal signal settings. Phase transition lost time is a function of the number phase switches in the signal operation (*10*). The higher the number of the phase switch, the higher is the lost time. In addition, optimal cycle length is required for sound operation of signal settings (*11, 12*). In most fixed signal plans, cycle lengths are not allowed to vary (however, cycle lengths can vary after a specific period in cases in which different signal plans are used for different times of the day).

School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907. Corresponding author: S. V. Ukkusuri, sukkusur@purdue.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2311, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 73–84.
DOI: 10.3141/2311-07

Without flexibility in the cycle lengths, the delays are higher in cases in which the variation in demand pattern is reasonably higher. Therefore, it is important to account for the flexibility of cycle lengths in signal control settings.

This research aims to develop a mixed-integer program with the different objective of integrating the signal optimization and dynamic traffic assignment problems. This research makes the following contributions:

- Develops a joint framework of signal control and dynamic traffic assignment that accounts for route choice behavior that is system optimal,
- Considers explicitly the lost time caused by phase switches at the signalized intersection, and
- Increases the applicability of results by considering flexible cycle lengths.

The next section discusses some related works from previous literature and defines the signal control problem; a discussion of the proposed mathematical program and the results obtained from the proposed model follows. The paper concludes with a short summary of the contributions made by this optimization framework and of the future research direction of the current research.

BACKGROUND AND RELATED WORK

Previous researchers have addressed the problem of combined traffic assignment and control in different ways (varying both in solution techniques and framework). Allsop first made a comprehensive attempt to propose a mathematical model that combines traffic assignment and signal optimization problem (*13*). Previous researchers approached the optimization problem in two principal ways: global optimization and optimization through an iterative assignment procedure. The global optimization models are difficult to apply, as the solution algorithms in general are not sufficiently efficient. Iterative procedures mostly contain bilevel programming techniques, and exact solutions are difficult to find. Abdelfatah and Mahmassani formulated a simulation-based combined traffic assignment and signal optimization problem (*14*). However, the nonlinear nature of the formulation did not allow them to solve it analytically. Other works include Meneguzzer (*15*), Gartner and Al-Malik (*16*), Gartner and Stamatiadis (*17*), and Smith and Ghali (*18*). However, most of these initial models are in a static context that does not account for dynamic traffic interaction. Again, delays are estimated on the basis of steady-state queuing equations that have limitations while modeling near-saturation conditions. Emergence of the dynamic traffic assignment solves few of the limitations discussed in the context of deterministic static equilibrium.

Lo also attempted to model signal control by using the cell transmission model (CTM) (*19*). Lin and Wang proposed a mixed-integer formulation of traffic signal control (*4*). Beard and Ziliaskopoulos (*5*) formulated a joint model of dynamic traffic assignment and signal optimization for networks adopting the linear program framework proposed by Ziliaskopoulos (*20*). Karoonsoontawong developed a bilevel type of mathematical program to solve the integrated signal control and assignment problem (*3*). However, the problem setting was for only predefined timing plans (phase sequence and movements are input into the problem). Ukkusuri et al. incorporated demand uncertainty and formulated a robust optimization model (*6*). More recently, Zhang et al. applied the CTM model to incorporate signal control and to resolve the vehicle-holding problem by introducing mixed-integer transformation of piecewise linear conditional constraints (*21*). Thus, very few research works have attempted to integrate the dynamic traffic assignment and signal settings design problem. In most of the cases, analytical models were either too difficult to solve or did not capture the traffic dynamics. In addition, the multiple origin–destination pairs were not considered in most earlier works.

Besides these works, researchers have also focused on signal optimization with multiple objectives (*22, 23*). In addition, researchers looked at different objectives and attributes (e.g., impact of left-turn restriction, length of left-turn bays, minimum and maximum green-time settings, etc.) (*24, 25*). However, few research works focused on flexible cycle length and lost time caused by the phase change pattern. Within the context of joint traffic assignment and signal optimization, the authors did not find any significant work that considered both the lost time and a flexible cycle length.

Clearly, the existing literature lacks contributions related to the elements of signal planning discussed earlier. This research aims to develop the joint framework for a signal control and traffic assignment problem that considers system-optimal behavior at the level of a multiple origin–destination network. It develops a mixed-integer program with signal decision elements as binary variables that accounts for four objective functions.

DEFINITION OF PROBLEM

The assumption is made that demand for the given road traffic network is deterministic. The number of phases and movements for phases is predefined. In the problem, the decision variable that determines the activation of any particular phase at any simulation interval is binary (0–1). The output from the proposed model provides optimal signal timing plans that minimize intersection delay and total system travel time. "Signal timing plan" implies the green splits and sequence of the defined phases.

MATHEMATICAL FRAMEWORK

This section describes the joint formulation for signal optimization and dynamic traffic assignment. First, the embedded mesoscopic traffic flow model is described briefly. Next, the expressions to compute intersection delay and number of phase switches are explained in the context of the formulation. Finally, the mathematical program for optimization of signal settings with dynamic traffic assignment is presented.

Description of CTM

Daganzo (*26, 27*) proposed a CTM that is a discrete approximation of the Lighthill–Whitham–Richards hydrodynamic model (*28, 29*) under the assumption of a piecewise linear relationship between traffic flow and density for each segment on the road. CTM divides each link of the network into cells (however, cell size can be varied throughout the network). CTM can represent congestion and queue spillover effects on a road link and accordingly can serve as an ideal choice to capture traffic dynamics attributable to desirable properties such as link spillover and shockwave propagation (*5, 6, 20*). However, CTM represents a linear program that has few drawbacks such as the vehicle-holding problem; merging and diverging; and first-in, first-out violation for a multiple-destination network, details of which can be found elsewhere (*6, 20, 30, 31*).

The flow propagation conditions in the original CTM constitute a feasible region of a nonconvex set (*31*). Ziliaskopoulos (*20*) relaxed the formulation and proposed a linear programming formulation of a system-optimal objective with embedded CTM for a single destination with multiple origins, and Beard and Ziliaskopoulos (*5*) later formulated the problem of integrating signal control. Recently, Ukkusuri and Waller (*30*) and Ukkusuri et al. (*6*), respectively, developed a linear program for a network design problem and signal optimization that accommodated demand uncertainty. In addition, Ukkusuri et al. (*6*) considered multiple destinations in the formulation. In this paper, the flow propagations and demand satisfaction constraints follow the previous works of Ziliaskopoulos (*20*), Ukkusuri and Waller (*30*), and Ukkusuri et al. (*6*). Because the paper primarily focuses on the signal optimization aspect of the mathematical framework, the basic sets of constraints (flow propagation and demand satisfaction) are not explained in detail [detailed information on a system-optimal linear programming formulation that uses CTM appears elsewhere (*6, 19, 20*)].

Intersection Delay and Phase Switch Count Within CTM

Intersection Delay

Two primary variables in the CTM-based linear program formulation are the occupancy and flow in or out of the cell at each simulation interval. Table 1 is a list of all notations used in the equations for the formulation developed in this paper. In previous works (*4, 6, 19*), intersection delay in the context of linear program framework [see Ziliaskopoulos (*20*)] was expressed as

$$\sum_{i\in C_{\mathrm{IS}}}\sum_{t\in T}\tau\left(\sum_{\forall O,D}\left(x_i^{O,D,t}-\sum_{j\in\Gamma(i)}y_{i,j}^{O,D,t}\right)\right)\tag{1}$$

Equation 1 shows a cumulative measure of delay of all intersections in the network. In CTM, all vehicles move to the next cell if space is available to satisfy the flow constraints. Each vehicle that cannot move from the intersection cell to the next cell experiences a delay of one simulation interval.

Phase Switching

As described earlier in the section on problem definition, phase activation is formulated as a binary variable. Whenever a phase switch occurs, the binary variable changes value. At any point of simulation, two phase activation variables always change their respective values. If one (current) goes from 0 to 1, another (previous phase) should go from 1 to 0. Therefore, the absolute value of consecutive phase activation variables is calculated and the average is taken to measure the total number of switches made at any intersection. For any intersection, the number of phase switches can be expressed as

$$\frac{1}{2}\sum_{t\in T}\sum_{p\in P}\left|w_{\pi,p}^{t}-w_{\pi,p}^{t-1}\right| \qquad \forall\pi\in\Pi\tag{2}$$

Each time a phase is switched, some amount of time is lost because of the clearance interval (including both yellow and all red). To

TABLE 1 Notations and Symbols Used in Formulation

Notation	Description
T	Set of discrete time intervals {0, 1, 2, 3, . . .}
$\tilde{T}$	Subset of T; {1, 2, 3, 4, . . .}
C_R	Set of source cells
C_S	Set of sink cells
C_{IS}	Set of intersection cells (distinctly defined with three-digit notion in the figures)
C_{ISQ}	Set of cells immediate to the intersection cells (this is more applicable for links with high number of cells)
C	Set of cells except the sink cells ($C_O \cup C_M \cup C_D \cup C_{\mathrm{IS}} \cup C_R$)
E_S	Set of sink cell connectors
$\Gamma(i)$	Set of successor cells of cell i
$\Gamma^{-1}(i)$	Set of predecessor cells of cell i
P	Set of phases; here {1, 2, 3, 4}
Π	Set of intersections; {1, 2, . . . total number of intersections in the network}
τ	Discrete time interval (in seconds)
Q_i^t	The maximum number of vehicles that can flow into or flow out of cell i between the time intervals t and $t + 1$
N_j^t	Maximum number of vehicles that can be present inside a cell j at time t
$x_i^{O,D,t}$	Number of vehicles in cell i during the time interval t oriented from source cell O and destined to sink cell D
$y_{i,j}^{O,D,t}$	Number of vehicles moving from cell i to cell j during time interval t oriented from source cell O and destined to sink cell D
$d^{O,D,t}$	Demand from source cell O and destined to sink cell D at interval t
δ_j^t	Ratio of free flow speed to backward propagation speed at time step t for cell j
$w_{\pi,p}^t$	Decision variable for phase activation; 1, if phase $p \in P$ is active at intersection $\pi \in \Pi$ during simulation interval $t \in T$; 0 otherwise
$\sigma_{i,j}$	Set of phases that allows the movement from i to j; $i \in C_{\mathrm{IS}}$; $j \in \Gamma(i)$ the ordinary cells in the next road connected
$\kappa_{i,j}^t$	Function of w; 1, if movement from $i \in C_{\mathrm{IS}}$ to $j \in \Gamma(i)$ is allowed at interval $t \in T$, 0 otherwise
υ_i^π	Input variable; 1, if cell $i \in C_{\mathrm{IS}}$ belongs to the intersection $\pi \in \Pi$

reduce the amount of phase switching, a penalty term is introduced in the objective function to make the formulation more applicable in a real-world problem. The lost time for each phase switch is assumed to be 2.5 s (it can take any value chosen by the planner). Therefore, for all intersections in the network, the total lost time attributable to changing of phases can be expressed as

$$2.5 * \frac{1}{2}\sum_{\pi\in\Pi}\sum_{t\in T}\sum_{p\in P}\left|w_{\pi,p}^{t} - w_{\pi,p}^{t-1}\right| \tag{3}$$

This equation measures the total lost of time attributable to phase switching for all signalized intersections in the network. Equation 3 is converted into a piecewise linear formulation to ensure the linearity of the objective function.

Mathematical Formulation

This research considers four variants of the signal optimization objectives. All these are compared with the base case that is defined as the system-optimal dynamic traffic assignment with fixed signal settings. In the base case, the signal settings are input, and within the simulation, the flow of vehicles follows control by means of signal constraints (see Equation 17). First, the objective functions and the four variants of the signal optimization framework are explained. Afterward, the flow propagation, demand satisfaction, and signal control constraints are discussed.

Base Case. System Optimal Dynamic Traffic Assignment with Fixed Signal Settings

In this case, the objective function is solely the minimization of total system travel time (TSTT) of the network under consideration. In addition, the vehicles must follow the control at signalized intersections. The signal settings are predetermined on the basis of demand and act as an input to this formulation. Thus, signal control is obtained but without any kind of optimization. The objective function can be stated as

$$\underset{x_i^{O,D,t}}{\text{minimize}} \sum_{\forall O\in C_R, D\in C_S}\sum_{\forall i\in C}\sum_{t\in T}\tau x_i^{O,D,t} \tag{4}$$

Case 1. Signal Optimization with System Optimal Dynamic Traffic Assignment

In Case 1, the objective function remains the same as in Equation 4. However, signal settings are no longer input to the formulation. Rather, in this formulation, phase activation is a variable that determines when any particular phase should be switched on so that the total system travel time can be minimized.

Case 2. Signal Optimization with TSTT and Cumulative Intersection Delays

In Case 2, the objective function is a weighted combination of the TSTT and intersection delay. Similar to Case 1, the Case 2 signal settings are variable and determined in such a way that the combination of TSTT and aggregate intersection delays is minimized for the whole system. The objective function can be expressed as

$$\underset{x_i^{O,D,t},\, y_{i,j}^{O,D,t}}{\text{minimize}} \left\{ \begin{array}{l} \alpha \displaystyle\sum_{\forall O\in C_R, D\in C_S}\sum_{\forall i\in C}\sum_{t\in T}\tau x_i^{O,D,t} \\ + \beta \displaystyle\sum_{i\in C_{\text{IS}}\cup C_{\text{ISQ}}}\sum_{t\in T}\tau\left(\sum_{\forall O,D}\left(x_i^{O,D,t} - \sum_{j\in\Gamma(i)} y_{i,j}^{O,D,t}\right)\right) \end{array} \right\} \tag{5}$$

In this equation, α and β are the weights for the TSTT and intersection delays, respectively (e.g., 0.35 and 0.65 for the test networks, respectively).

Although intersection delay is a part of the total system travel time, it is considered separately in the objective function because the delay at an intersection must sometimes be minimized without consideration of travel time for the entire trip. One can think of the intersection delay as a local objective that maximizes the processing rate of the intersection in question.

Case 3. Signal Optimization with TSTT and Cumulative Lost Time

In Case 3, the objective function is a weighted combination of TSTT and cumulative lost time attributable to phase switches in signal operations. The objective function can be expressed as

$$\underset{x_i^{O,D,t},\, y_{i,j}^{O,D,t}}{\text{minimize}} \left\{ \alpha \sum_{\forall O\in C_R, D\in C_S}\sum_{\forall i\in C}\sum_{t\in T}\tau x_i^{O,D,t} + \gamma(2.5)\frac{1}{2}\sum_{\pi\in\Pi}\sum_{t\in T}\sum_{p\in P}\left|w_{\pi,p}^{t} - w_{\pi,p}^{t-1}\right| \right\} \tag{6}$$

Here, α and γ are the weights for the TSTT and cumulative lost time, respectively (0.35 and 0.65 for the test networks, respectively).

Case 4. Signal Optimization with TSTT, Cumulative Intersection Delays, and Cumulative Lost Time

The objective function now includes TSTT, intersection delays, and lost time as a weighted combination. It can be expressed as

$$\underset{x_i^{O,D,t},\, y_{i,j}^{O,D,t}}{\text{minimize}} \left\{ \begin{array}{l} \alpha \displaystyle\sum_{\forall O\in C_R, D\in C_S}\sum_{\forall i\in C}\sum_{t\in T}\tau x_i^{O,D,t} \\ + \beta \displaystyle\sum_{i\in C_{\text{IS}}\cup C_{\text{ISQ}}}\sum_{t\in T}\tau\left(\sum_{\forall O,D}\left(x_i^{O,D,t} - \sum_{j\in\Gamma(i)} y_{i,j}^{O,D,t}\right)\right) \\ + \gamma(2.5)\displaystyle\frac{1}{2}\sum_{\pi\in\Pi}\sum_{t\in T}\sum_{p\in P}\left|w_{\pi,p}^{t} - w_{\pi,p}^{t-1}\right| \end{array} \right\} \tag{7}$$

Here, α, β, and γ are the weights for TSTT, cumulative intersection delays, and lost time, respectively (0.2, 0.4, and 0.4, respectively, for the test networks).

Constraints for Traffic Flow Model

Initialization Constraints

Initialization constraints for the traffic flow model appear in Equations 8 through 11:

$$x_i^{O,D,t} \geq 0 \qquad \forall O\in C_R, D\in C_S, t\in T, i\in C \tag{8}$$

$$y_{i,j}^{O,D,t} \geq 0 \qquad \forall O\in C_R, D\in C_S, t\in T, i\in C, j\in C \tag{9}$$

$$x_i^{O,D,0} = 0 \qquad \forall O \in C_R, D \in C_S, t \in T, i \in C \tag{10}$$

$$y_{i,j}^{O,D,0} = 0 \qquad \forall O \in C_R, D \in C_S, i \in C \cup C_S, j \in C \cup C_S \tag{11}$$

Flow Conservation Constraints

Flow conservation constraints for the traffic flow model are shown in Equations 12 and 13:

$$x_i^{O,D,t} - x_i^{O,D,t-1} - \sum_{k \in \Gamma^{-1}(i)} y_{k,i}^{O,D,t-1} + \sum_{j \in \Gamma(i)} y_{i,j}^{O,D,t-1} = d^{O,D,t-1}$$
$$\forall i \in C_R, t \in \tilde{T}, O \in C_R, D \in C_S \tag{12}$$

$$x_i^{O,D,t} - x_i^{O,D,t-1} - \sum_{k \in \Gamma^{-1}(i)} y_{k,i}^{O,D,t-1} + \sum_{j \in \Gamma(i)} y_{i,j}^{O,D,t-1} = 0$$
$$\frac{\forall i \in C}{(C_R \cup C_S)}, t \in \tilde{T}, O \in C_R, D \in C_S \tag{13}$$

Demand Satisfaction Constraints

Demand satisfaction constraints for the traffic flow model are shown in Equation 14:

$$\sum_{t \in T} \sum_{i \in \Gamma^{-1}(\tilde{d})} y_{i,\tilde{d}}^{O,D,t-1} = \sum_{\forall t \in T} d^{O,D,t} \qquad \forall \tilde{d} \in C_S, t \in T, \forall O, D \tag{14}$$

Flow Propagation Constraints

Flow propagation constraints for the traffic flow model are shown in Equations 15 through 18:

$$\sum_{\forall j \in \Gamma(i)} y_{i,j}^{O,D,t} - x_i^{O,D,t} \le 0 \qquad O \in C_R, D \in C_S, t \in \tilde{T}, i \in C \tag{15}$$

$$\sum_{\forall O,D} \sum_{\forall j \in \Gamma(i)} y_{i,j}^{O,D,t} \le Q_i^t \qquad t \in \tilde{T}, i \in C \tag{16}$$

$$\sum_{\forall O,D} \sum_{\forall i \in \Gamma^{-1}(j)} y_{i,j}^{O,D,t} \le \kappa_{i,j}^t Q_j^t \qquad t \in \tilde{T}, \frac{j \in C}{C_R} \tag{17}$$

$$\sum_{\forall O,D} \sum_{\forall i \in \Gamma^{-1}(j)} y_{i,j}^{O,D,t} + \delta_j^t \sum_{\forall O,D} x_j^{O,D,t} \le \delta_j^t N_j^t \qquad t \in T, j \in C \tag{18}$$

Constraint Sets for Signal Settings

Constraint sets for the signal settings are shown in Equations 19 through 24:

$$\sum_{p \in P} w_{\pi,p}^t = 1 \qquad \forall \pi \in \Pi, \forall t \in T \tag{19}$$

$$\kappa_{i,j}^t = \sum_{p \in \sigma_{i,j}} \left(w_{\pi,p}^t * \upsilon_i^\pi\right) \qquad \sigma_{i,j} \subset P; \forall \pi \in \Pi; \forall i \in C_{\text{IS}}; j \in \Gamma(i) \tag{20}$$

$$\kappa_{i,j}^t \ge 0 \tag{21}$$

$$\sum_{\tilde{t}=t}^{t-\hat{r}} w_{\pi,p}^{\tilde{t}} \le g_{\max} \qquad \pi \in \Pi; p \in P, \tilde{t} = \{\hat{r}+1, \ldots, T\} \tag{22}$$

where $\hat{r}$ is an integer that denotes index of discrete time interval.

$$w_{\pi,p}^{(g_{\min})\hat{t}} = w_{\pi,p}^{(g_{\min})\hat{t}-1} = w_{\pi,p}^{(g_{\min})\hat{t}-2} = \cdots = w_{\pi,p}^{(g_{\min})\hat{t}-g_{\min}+1}$$
$$\forall \pi \in \Pi; \forall p \in P; \hat{t} = \left\{1, 2, 3, \ldots, \frac{T}{g_{\min}}\right\} \tag{23}$$

$$\sum_{\tilde{t}=t}^{t-9} w_{\pi,p}^{\tilde{t}} \ge 1 \qquad \pi \in \Pi; p \in P, \tilde{t} = \{9, \ldots, T\} \tag{24}$$

Constraint Set 19 implies that, for any intersection at any interval only one phase (of the four) will be active. Equation Set 15 constructs a mapping between the vehicle movement and the defined signal control phases. At any interval, if only the phases that allow the specific flow movement from $i \in C_{\text{IS}}$ to $j \in \Gamma(i)$ are considered and summed, the result is whether the flow is possible. The variable $\kappa_{i,j}^t$ need not necessarily be binary (however, settings yield only 0–1 value for this). Further, a nonnegativity constraint exists.

Next, Constraint Set 22 determines the maximum green ($g_{\max}$) allowed. Similarly, Constraint Set 23 determines the minimum green ($g_{\min}$). The choice of minimum and maximum green depends on the scope of the problem and the analyst's judgment. In addition, $g_{\min}$ and $g_{\max}$ are integer values taken from the set of discrete intervals. For example, if $g_{\max}$ is assigned a value of 10, the maximum green duration will be $10 * \tau$ seconds in total. The same interpretation can be used for the minimum green value.

Furthermore, Equation Set 24 imposes the maximum cycle length for the signal timing plan. This constraint set confirms the activation of each phase within nine successive intervals. If the interval is 10 s, then the maximum cycle length will be 10 ∗ 10 (100 s). Again, the minimum cycle length is 10 ∗ 4 (40 s) if the minimum green for each phase is set to 10 s.

NUMERICAL RESULTS WITH TEST NETWORKS

The proposed model was tested with an isolated intersection (Network 1) and a network containing three intersections (Network 2). Tables 2, 3, and 4 show different properties of the test networks. Tables 3 and 4 also show the signal phase descriptions and the settings for the pretimed signal controller. Furthermore, the proposed model was tested with different levels of demand to assess the performance of the model under different congestion levels. Figure 1 shows the general presentation of Network 2. Figures 2 and 3 show the cell representation of Networks 1 and 2, respectively.

To assess the performance of different models, a few metrics were determined for each of the networks. The metrics included total system travel time, cumulative intersection delays, and lost time attributable to phase switching. For each phase switch, the intersection experienced 2.5 s of lost time. Tables 5 and 6 show the results at different congestion levels from Networks 1 and 2, respectively. Table 7 shows the route choice results for the networks, and Table 8 shows sample output from the signal optimization models and cycle length flexibility. The description of the variants of the models (Cases 1 through 4) considered for analysis can be found in the previous section.

TABLE 2 Properties of Test Networks and Signal Settings

Attribute	Isolated Signalized Intersection	Network with Multiple Signalized Intersections
Total number of origins	4	7
Total number of destinations	4	5
Total number of origin–destination pairs	12	8
Saturation flow (vphpl)	2,160	2,160
Posted speed limit (mph)	40	40
Backward propagation speed (mph)	40	40
Minimum cell length (ft)	587	587
Typical length of a vehicle (ft)	27	27
Time step (s) τ	10	10
Jam density (maximum number of vehicles in a cell) N_j^t	22	22
Maximum flow allowed (vehicles per time step) Q_i^t	6	6

NOTE: vphpl = vehicles per hour per lane; mph = miles per hour.

Results for Isolated Intersection

An isolated signalized intersection was tested at two levels of demand: 1,800 and 900 vehicles per hour per lane (vphpl). Except for the lost time, the results show similar trends for the two levels of demand. The mixed-integer program was solved by using CPLEX version 12.1.0 solver on a server node consisting of 24 core processors and 48 GB of RAM. Objective functions for all the models are linear in nature.

TSTT for Cases 1 Through 4

At both congestion levels, Case 1 provided the minimum TSTT, and it was lower than that for the base case. This result was intuitive because, in the base case, the signal settings were not optimized and were merely input into the model. As the other cases had TSTT as a weighted component of the objective function, it was expected that Case 1 would provide the least travel time. Furthermore, higher TSTT was observed for Case 2, which included cumulative intersection delays as a major component of the objective function compared with Case 3, which included lost time in the objective function. Case 4 provided TSTT that lay within the TSTT values obtained in Cases 2 and 3. Because a weighted combination of TSTT, intersection delays, and lost time was considered, Case 4 rendered a balance for these objectives.

TABLE 3 Phases in Signal Settings

Phase	Description
1	East–west and west–east approaches (right turn and straight through)
2	Dual left turn (east to north and west to south)
3	North–south and south–north approaches (right turn and straight through)
4	Dual left turn (north to west and south to east)

TABLE 4 Parameters in Pretimed Cycle

Demand	Cycle Length	Sequence
Low flow (900 vphpl)	60	1 + 1 + 2 + 3 + 3 + 4
Medium flow (1,350 vphpl)	80	1 + 1 + 1 + 2 + 3 + 3 + 3 + 4
High flow (1,800 vphpl)	90	1 + 1 + 1 + 1 + 2 + 3 + 3 + 3 + 4

When lost time was added to TSTT, the actual systemwide travel time was obtained. For both congestion levels, Case 3 provided the minimum systemwide travel time when lost time is considered. Case 4 provided the next-best solution.

Intersection Delay

As expected, Case 2 had the minimum intersection delay for all levels of congestion. The next-best results were found in Case 4, which considered a weighted combination of TSTT, intersection delay, and lost time.

Lost Time

Lost time depends on the number of switches made. The base case with pretimed signal control has a deterministic nature and does not respond to real-time traffic demand; the number of switches made is fixed. Results showed that the higher-flow base case had the lowest lost time (however, not the lowest TSTT, even when lost time was considered), and Case 3 provided the next-best solution. But, for the low-congestion level, Cases 3 and 4 provided the minimum lost time. For high demand, the switches were made more frequently as a response to the demand, and accordingly the number of switches was higher.

Results for Network with Multiple Intersections (Network 2)

Network 2 was tested at three levels of congestion: 1,800, 1,350, and 900 vphpl.

Results for TSTT

Results showed higher TSTT for higher demand. Case 1 provided the lowest TSTT for both levels of congestion. For the highest demand, Case 3 provided the next-best solution. However, for medium- and low-demand levels, Case 4 provided the next-minimum TSTT. In addition, Case 2 provided the highest travel time among cases with those optimized signal settings.

TSTT values including lost time showed a slightly different trend. Case 3 provided the lowest value for high and low demands. For the medium level of congestion, Case 4 gave the minimum value.

Intersection Delay

As expected, Case 2 provided the lowest value of intersection delay for all levels of demand. Case 4 provided the next-best solution.

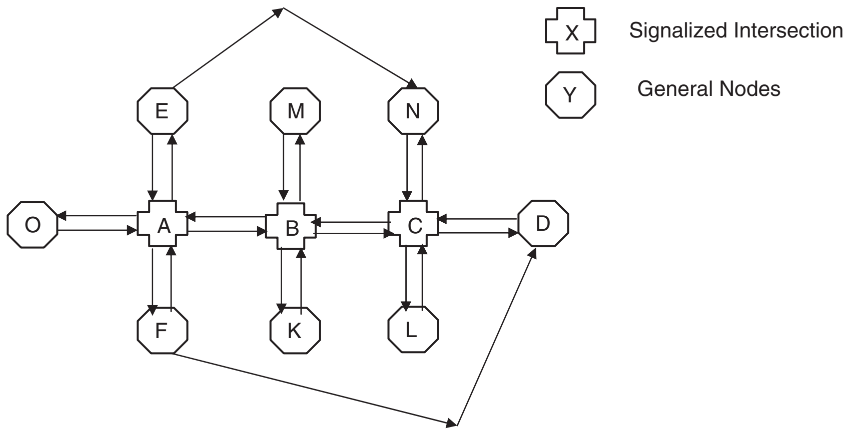

FIGURE 1 Network with multiple signalized intersections (Network 2).

Lost Time

Case 3 provided the lowest lost time for medium and low levels of congestion. For the high level of flow, the base case had the minimum number of switches, which were observed for the isolated intersection (Network 1) as well.

Route Choice Behavior

Because the formulation is based on a system-optimal dynamic traffic assignment, it is important to assess the effect of signal optimization on route choice behavior. Figure 1 shows a specific origin–destination pair (Node O to Node D) and the existence of three paths to reach Node D from Node O. The base case is a system-optimal formulation without signal optimization, and Case 1 has the same formulation with signal optimization. Cases 3 and 4 are not system optimal with respect to travel time in the true sense. For these cases, the objective function contains components other than systemwide travel time. However, one can define Cases 2 to 4 as system-optimal formulations with modified definitions of cost that contain components other than travel time only.

Table 7 shows the variation in route choice behavior for two levels of congestion. The results are based on the origin–destination pairs and paths shown in Figure 1 (Path 1, O–A–B–C–D; Path 2, O–A–E–N–C–D; and Path 3, O–A–F–D). Results from Table 7 indicate different flow distribution for the base case and Case 1. Path 1 includes three successive intersections, and without signal optimization, delay is higher for this path. As a result relatively low flow for Path 1 is observed. However, when signals are optimized (Case 1), the flow switches from Path 3 to Path 1. This switch indicates travel time improvement in Path 1 because of signal optimization. The trend is the same for both high (1,800 vphpl) and medium (1,350 vphpl) levels of congestion. However, the route change pattern is more prominent in the medium level of congestion than in the high level.

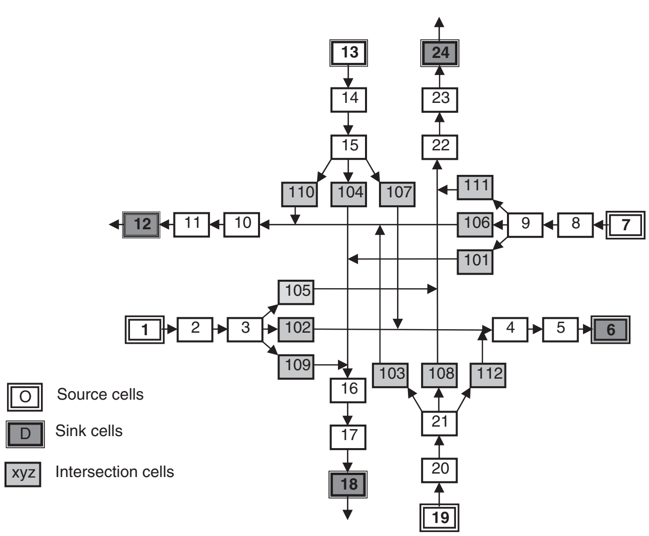

FIGURE 2 Cell representation of isolated signalized intersection (Network 1).

Variation in Cycle Length

In the formulation, variation of cycle length in the signal settings is considered. The minimum cycle length is 40 s, and the maximum is 100 s. Figures 4 and 5 show the variations of cycle lengths at two levels of demand: 1,800 vphpl and 1,350 vphpl, respectively. Cases 3 and 4 show a higher frequency of long cycle lengths (90 and 100 s). Medium cycle lengths (70 and 80 s) are found in all cases considered for analysis. One can see similar trends for both levels of congestion.

CONCLUSION AND FUTURE WORK

This study developed an analytical mathematical programming–based framework for integrated signal control and dynamic traffic assignment. One of the major contributions of this research is the framework that considers the dynamic traffic assignment and signal control optimization in a unified manner. The formulation accounts for intersection delays, lost time caused by phase switches, and the like, as a weighted combination in the objective function of the optimization model. In addition, flexibility in cycle length allows the model to account for variation in demand and traffic states. Furthermore, the proposed model provides optimal phase durations, sequence, and cycle length. The authors expect that the model can serve as a useful tool for devising signal-timing plans.

TSTT and total system time with lost time (TSTL) are two important measures used for comparison with fixed signal timing

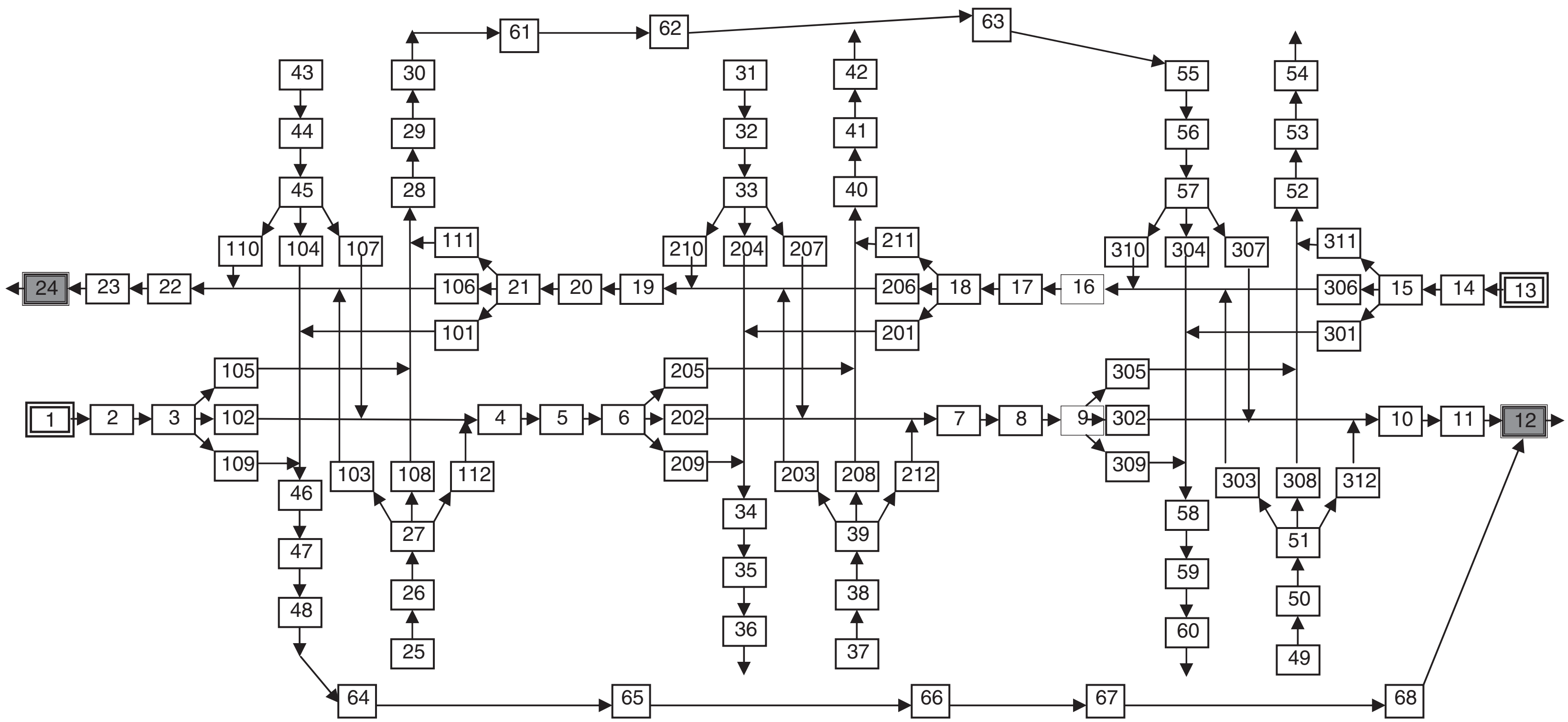

FIGURE 3 Cell representation of network with multiple signalized intersections (Network 2).

TABLE 5 Results Obtained for Isolated Signalized Intersection (Network 1)

Metric	Case Base	1	2	3	4
Results for High Demand (1,800 vphpl)					
Total system travel time (s)	53,360	44,860	45,020	44,876	44,890
Total intersection delay (s)	12,660	3,230	1,700	4,760	1,780
Total number of switches	39	68	72	41	43
Travel time including lost time (s)	53,458	45,030	45,200	44,979	44,998
Results for Light Demand (900 vphpl)					
Total system travel time (s)	10,180	9,380	9,700	9,392	9,400
Total intersection delay (s)	1,610	1,120	80	930	200
Total number of switches	59	72	73	31	31
Travel time including lost time (s)	10,328	9,560	9,883	9,470	9,478

TABLE 6 Results Obtained from Network with Multiple Signalized Intersections (Network 2)

Metric	Case Base	1	2	3	4
Results for High Demand (1,800 vphpl)					
Total system travel time (s)	79,960	63,680	64,620	63,690	64,020
Total intersection delay (s)	10,220	2,620	900	3,040	1,040
Total number of switches at Intersection 1	39	78	70	49	59
Total number of switches at Intersection 2	39	72	71	39	48
Total number of switches at Intersection 3	39	70	64	48	55
Travel time including lost time (s)	80,253	64,230	65,133	64,030	64,425
Results for Medium Demand (1,350 vphpl)					
Total system travel time (s)	62,000	50,550	50,990	50,800	50,600
Total intersection delay (s)	6,870	1,780	420	1,860	500
Total number of switches at Intersection 1	50	74	71	48	51
Total number of switches at Intersection 2	50	69	65	40	45
Total number of switches at Intersection 3	50	72	75	46	51
Travel time including lost time (s)	62,375	51,087	51,518	51,135	50,968
Results for Light Demand (900 vphpl)					
Total system travel time (s)	19,420	17,460	17,490	17,465	17,470
Total intersection delay (s)	2,020	380	120	460	250
Total number of switches at Intersection 1	59	64	69	41	40
Total number of switches at Intersection 2	59	58	64	36	38
Total number of switches at Intersection 3	59	65	75	42	41
Travel time including lost time (s)	19,863	17,928	17,788	17,763	17,768

TABLE 7 Route Choice Behavior Pattern Resulting from Signal Optimization

Variable	System Optimal Without Signal Optimization (Base Case) Path 1	Path 2	Path 3	System Optimal With Signal Optimization (Case 1) Path 1	Path 2	Path 3
High Demand (1,800 vphpl)						
Number of vehicles	22	14	44	32	12	36
Percentage of total demand	27.5	17.5	55	40	15	45
Medium Demand (1,350 vphpl)						
Number of vehicles	3	10	53	19	10	37
Percentage of total demand	4.55	15.15	80.30	28.79	15.11	56.1

NOTE: Path 1 = straight through; Path 2 = left turn; Path 3 = right turn.

TABLE 8 Sample Output (Phase Activation Values) from Signal Optimization Model

	Phase			
Time Step	1	2	3	4
15	1	0	0	0
16	0	1	0	0
17	0	1	0	0
18	0	1	0	0
19	1	0	0	0
20	0	1	0	0
21	0	1	0	0
22	0	1	0	0
23	1	0	0	0

in this paper. Results show that Case 3 (which considers a weighted combination of total system travel time and lost time caused by phase switching) gives the minimum TSTL for the network. Although Case 1 gives the minimal TSTT, it has a higher lost time. Furthermore, Case 2 minimizes the weighted combination of total travel time and intersection delay without consideration of lost time caused by phase switching. One drawback of Case 3 is that some of the intersections will experience higher delay than others in the network. As a response, Case 4 was introduced to consider a weighted combination of total system travel time, intersection delay, and lost time caused by phase transitions. Case 4 provides slightly higher TSTL than Case 3, but the intersection delay is much lower. Therefore, one can conclude that, if TSTL is considered, Case 3 is the best option. However, if fairness at local intersections (delays at intersections cannot be too high) is considered, Case 4 can be seen as a better choice. These conclusions apply to results for all demand levels. In addition, the difference in TSTL between Case 3 and Case 4 becomes smaller at a low demand level.

The phase duration and sequence results (Table 8) show the quasi-adaptive nature of the signal settings design. The optimization framework presented here is not in real time in the context of traditional real-time systems like SCOOT (*32, 33*), RHODES (*34*), CRONOS (*35*), and the like. However, the objective function accounts for the dynamic variations in demand within the signal optimization setting. In addition, consideration of lost time caused by phase switches and cycle length variation makes the model more appropriate from the perspective of traffic engineers.

The traffic flow model has well-known drawbacks attributable to the linear nature of the constraint set (approximating the minimum operator as a set of linear equations). This causes the well-known vehicle-holding problem (*20, 30, 31*). Because the scope of the paper is primarily signal timing optimization, the authors do not attempt to resolve this in this research. Again, the formulations do not consider diverging and merging, and lane changing phenomenon within the traffic flow model explicitly. In addition, only the system-optimal type of route choice behavior that is more appropriate from the system operator's perspective is considered. One might also focus on user-optimal type of behavior to see results that are appropriate from the user's perspective.

The mixed-integer program proposed in this research is computationally expensive because of the large number of integer variables. One of the future goals is to reduce the number of integer variables to allow the formulation to be applicable for medium-sized traffic

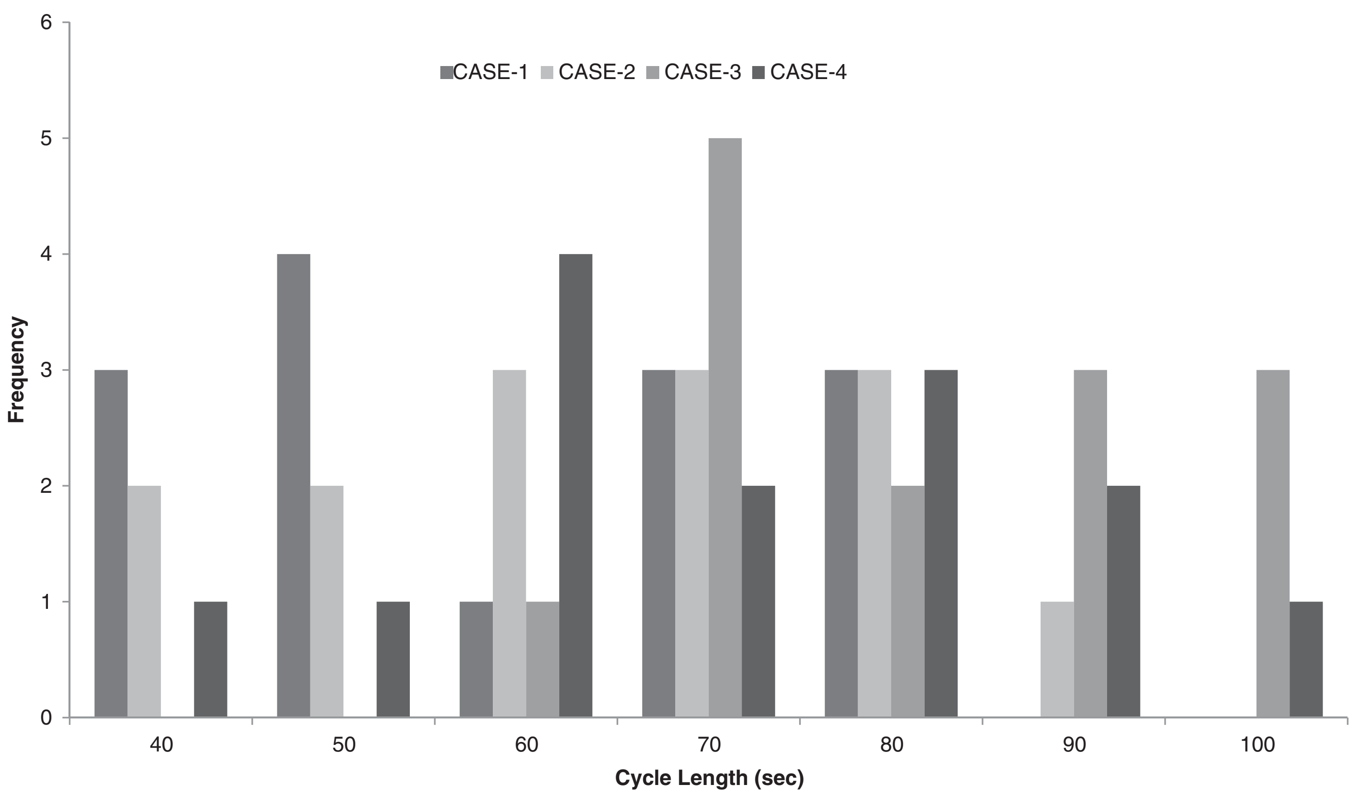

FIGURE 4 Cycle length variation for demand of 1,800 vphpl (Network 2).

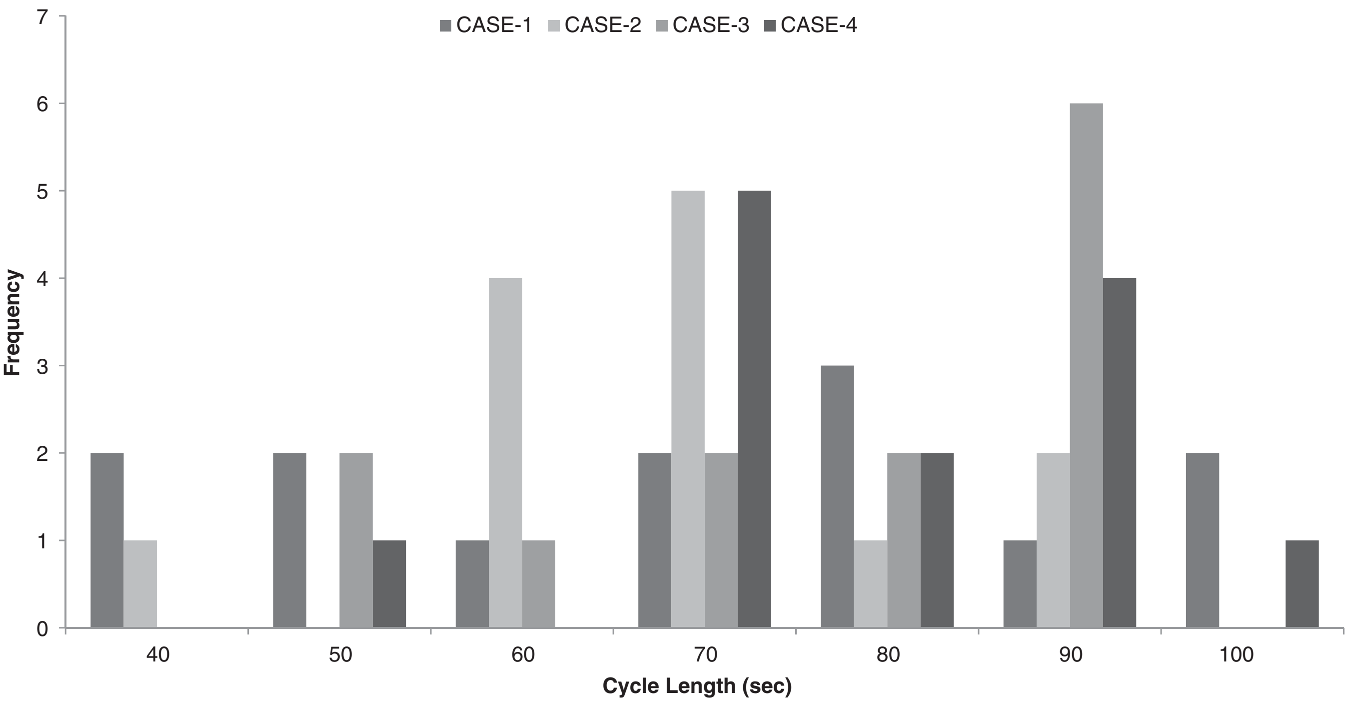

FIGURE 5 Cycle length variation for demand of 1,350 vphpl (Network 2).

networks. In addition, the weighted combination approach to solve a multiobjective problem in many cases would produce only sub-optimal solutions. A more rigorous approach can be determination of Pareto-optimal solutions by means of an exact approach or heuristics like evolutionary algorithms.

REFERENCES

1. Schrank, D., T. Lomax, and S. Turner. *TTI's 2010 Urban Mobility Report.* Texas Transportation Institute, Texas A&M University System, College Station, 2010.
2. *2007 National Traffic Signal Report Card: Executive Summary.* National Transportation Operations Coalition, 2007. http://www.ite.org/reportcard.
3. Karoonsoontawong, A., and S. Waller. Integrated Network Capacity Expansion and Traffic Signal Optimization Problem: Robust Bi-Level Dynamic Formulation. *Networks and Spatial Economics,* Vol. 10, No. 4, 2010, pp. 525–550.
4. Lin, W.-H., and C. Wang. An Enhanced 0–1 Mixed-Integer LP Formulation for Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 5, No. 4, 2004, pp. 238–245.
5. Beard, C., and A. Ziliaskopoulos. System Optimal Signal Optimization Formulation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1978,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 102–112.
6. Ukkusuri, S. V., G. Ramadurai, and G. Patil. A Robust Transportation Signal Control Problem Accounting for Traffic Dynamics. *Computers & Operations Research,* Vol. 37, No. 5, 2010, pp. 869–879.
7. Yang, H., and S. Yagar. Traffic Assignment and Signal Control in Saturated Road Networks. *Transportation Research Part A: Policy and Practice,* Vol. 29, No. 2, 2005, pp. 125–139.
8. Peeta, S., and A. Ziliaskopoulos. Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics,* Vol. 1, No. 3, 2001, pp. 233–265.
9. Papageorgiou, M. Dynamic Modeling, Assignment, and Route Guidance in Traffic Networks. *Transportation Research Part B: Methodological,* Vol. 24, No. 6, 1990, pp. 471–495.
10. Furth, P. G., B. Cesme, and T. H. J. Muller. Lost Time and Cycle Length for Actuated Traffic Signal. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2128,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 152–160.
11. Lan, C. J. New Optimal Cycle Length Formulation for Pretimed Signals at Isolated Intersections. *Journal of Transportation Engineering,* Vol. 130, No. 5, 2004, pp. 637–647.
12. Lee, J., and B. M. Williams. Development and Evaluation of a Constrained Optimization Model for Traffic Signal Plan Transition. *Transportation Research Part C: Emerging Technologies,* Vol. 20, No. 1, 2012, pp. 185–198.
13. Allsop, R. E. Some Possibilities for Using Traffic Control to Influence Trip Distribution and Route Choice. *Proc., 6th International Symposium on Transportation and Traffic Theory* (D. J. Buckley, ed.), Elsevier, New York, 1974, pp. 345–373.
14. Abdelfatah, A. S., and H. S. Mahmassani. System Optimal Time-Dependent Path Assignment and Signal Timing in Traffic Network. In *Transportation Research Record 1645,* TRB, National Research Council, Washington, D.C., 1998, pp. 185–193.
15. Meneguzzer, C. Review of Models Combining Traffic Assignment and Signal Control. *Journal of Transportation Engineering,* Vol. 123, No. 2, 1997, pp. 148–155.
16. Gartner, N. H., and M. Al-Malik. Combined Model for Signal Control and Route Choice in Urban Traffic Networks. In *Transportation Research Record 1554,* TRB, National Research Council, Washington, D.C., 1996, pp. 27–35.
17. Gartner, N. H., and C. Stamatiadis. Integration of Dynamic Traffic Assignment with Real-Time Traffic Adaptive Control System. In *Transportation Research Record 1644,* TRB, National Research Council, Washington, D.C., 1998, pp. 150–156.
18. Smith, M. J., and M. Ghali. The Dynamics of Traffic Assignment and Traffic Control: A Theoretical Study. *Transportation Research Part B: Methodological,* Vol. 24, No. 6, 1990, pp. 409–422.
19. Lo, H. K. A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans. *Transportation Science,* Vol. 35, No. 2, 2001, pp. 148–164.
20. Ziliaskopoulos, A. K. A Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem. *Transportation Science,* Vol. 34, No. 1, 2000, pp. 37–49.
21. Zhang, L., Y. Yin, and Y. Lou. Robust Signal Timing for Arterials Under Day-to-Day Demand Variations. In *Transportation Research Record:*

Journal of the Transportation Research Board, No. 2192, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 156–166.
22. Abdelgawad, H., B. Abdulhai, and M. Wahba. Multiobjective Optimization for Multimodal Evacuation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2196,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 21–33.
23. Kesur, K. B. Generating More Equitable Traffic Signal Timing Plans. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2192,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 108–115.
24. Kikuchi, S., and N. Kronprasert. Determining Lengths of Left-Turn Lanes at Signalized Intersections Under Different Left-Turn Signal Schemes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2195,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 70–81.
25. Li, P., M. M. Abbas, R. Pasupathy, and L. Head. Simulation-Based Optimization of Maximum Green Setting Under Retrospective Approximation Framework. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2192,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1–10.
26. Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B: Methodological,* Vol. 28, No. 4, 1994, pp. 269–287.
27. Daganzo, C. F. The Cell Transmission Model, Part II: Network Traffic. *Transportation Research Part B: Methodological,* Vol. 29, No. 2, 1995, pp. 79–93.
28. Lighthill, M. J., and G. B. 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.
29. Richards, P. I. Shockwaves on the Highway. *Operations Research,* Vol. 4, No. 1, 1956, pp. 42–51.
30. Ukkusuri, S. V., and S. T. Waller. Linear Programming Models for the User and System Optimal Dynamic Network Design Problem: Formulations, Comparisons and Extensions. *Networks and Spatial Economics,* Vol. 8, No. 4, 2008, pp. 383–406.
31. Zheng, H., and Y.-C. Chiu. A Network Flow Algorithm for the Cell-Based Single-Destination System Optimal Dynamic Traffic Assignment Problem. *Transportation Science,* Vol. 45, No. 1, 2011, pp. 121–137.
32. Hunt, P. B., D. I. Robertson, R. D. Bretherton, and M. C. Royle. The SCOOT On-Line Traffic Signal Optimization Technique. *Traffic Engineering and Control,* Vol. 23, No. 4, 1982, pp. 190–192.
33. Robertson, D. I., and R. D. Bretherton. Optimizing Networks of Traffic Signals in Real Time—The SCOOT Method. *IEEE Transactions on Vehicular Technology,* Vol. 40, No. 1, 1991, pp. 11–15.
34. Mirchandani, P., and L. Head. A Real-Time Traffic Signal Control System: Architecture, Algorithms, and Analysis. *Transportation Research Part C: Emerging Technologies,* Vol. 9, No. 6, 2001, pp. 415–432.
35. Boillot, F., S. Midenet, and J.-C. Pierrelée. The Real-Time Urban Traffic Control System CRONOS: Algorithm and Experiments. *Transportation Research Part C: Emerging Technologies,* Vol. 14, No. 1, 2006, pp. 18–38.

The Traffic Signal Systems Committee peer-reviewed this paper.