An Enhanced 0–1 Mixed-Integer LP Formulation for Traffic Signal Control

Wei-Hua Lin and Chenghong Wang

***Abstract*—An enhanced 0–1 mixed-integer linear programming formulation based on the cell-transmission model is proposed for the traffic signal optimization problem. This formulation has several features that are currently unavailable in other existing models developed with a similar approach, including the components for handling the number of stops, fixed or dynamic cycle length and splits, and lost time. The problem of unintended vehicle holding, which is common in analytical models, is explicitly treated. The formulation can be utilized in developing strategies for adaptive traffic-control systems. It can also be used as a benchmark for examining the convergence behavior of heuristic algorithms based on the genetic algorithm, fuzzy logic, neural networks, or other approaches that are commonly used in this field. The discussion of extending the proposed model to capture traffic signal preemption in the presence of emergency vehicles is given. In terms of computational efficiency, the proposed formulation has the least number of binary integers as compared with other existing formulations that were developed with the same approach.**

***Index Terms*—Mathematical programming, signalized intersections, traffic control.**

I. Introduction

FOR the past three decades or so, signal-control strategies have evolved from ones that are developed based entirely on historical information, often referred to as the fixed timing plan, to the third-generation control strategies in which the control system is fully responsive, i.e., cycle length, offset, and split are all determined online based on real-time information. With the advances in communication systems, optimization algorithms for adaptive control have been incorporated into various traffic signal-control systems, such as SCOOT, SCATS, OPAC, and PRODYN.

The performance of a traffic signal-control system depends on the optimization procedures embedded in the system. Optimization procedures are conventionally either based on either heuristic or exact models. Heuristic optimization models, such as those developed with feedback control, genetic algorithms, fuzzy logic, or neural network approaches, if applied properly, can provide solutions that are good approximations to the optimal solution. It is sometimes difficult to assess if a heuristic procedure indeed achieves global optimization, even though the solution exhibits certain asymptotic behavior of convergence. On the other hand, an exact model is often computationally demanding, making it difficult to implement for real-time control. Research in this area, however, is equally important, since closed-form solutions would provide us with insights into the problem and reveal properties that would help us to develop effective control strategies. The closed-form solution can also be used as a benchmark to examine the performance of heuristic procedures that require less computational effort. Unfortunately, many existing exact models for traffic signal control are very limited in scope. Although attempts have been made to formulate the signal-optimization problem as a mathematical programming problem, formulations developed with this approach do not usually cover all the features that are important for signal control. Improta and Cantarella [1] formulated the problem as a binary mixed-integer linear program. Lo [2] also formulated the problem as a binary mixed-integer linear program recently based on the cell-transmission model. Both formulations, however, assume that the cycle length is fixed. Ramanathan *et al.* [3] formulated the problem of optimizing the sequence of phase over a long-time horizon and the signal is under no constraints except for minimum and maximum green durations. The formulation, however, is nonlinear and nonconvex. The solution approach is yet to be developed. Gartner and Stamatiadis [4] developed an optimal arterial-based progression scheme of the signal-control problem using mixed-integer linear programming. The underlying model, however, is static in nature.

This paper provides an enhancement to the existing 0–1 mixed-integer linear programming formulation for the traffic signal optimization problem. In particular, we consider models that handle physical queues. The proposed enhancement covers several elements that are currently unavailable in other formulations that were developed with a similar approach yet important for the traffic signal-control problem. The proposed model explicitly handles the total number of stops. It also allows variable lengths for cycles and splits, constrained by minimum and maximum green durations. The lost time is modeled by a penalty function corresponding to every switch of the traffic signal. In terms of computational efficiency, the proposed formulation has the least number of the total binary integers compared with other formulations. The number of 0–1 integer variables is equal to the product of the total number of intersections and the total time steps for the entire optimization duration. This formulation can be readily solved using standard mathematical programming optimizers such as CPLEX, LINDO, etc.

The remainder of this paper is organized as follows. In Section II, we provide an overview of the cell-transmission

Manuscript received November 28, 2003; revised July 15, 2004 and July 29, 2004. The Associate Editor for this paper was S. Tang.

W.-H. Lin is with Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85721-0020 USA (e-mail: weilin@sie.arizona.edu).

C. Wang is with the National Natural Science Foundation of China, Beijing 100085, China.

Digital Object Identifier 10.1109/TITS.2004.838217

model, which is the theoretical basis for our model formulation. Section III presents a complete description of all the modeling components. This is followed by an example presented in Section IV that shows the performance of the proposed 0–1 mixed-integer LP model. Section V discusses the extension of this model to capture the process of signal preemption in the presence of emergency vehicles. A numerical example showing the movement of an emergency vehicle and the traffic-recovery process is given. The conclusion and discussion for future study are given in the final section.

II. Cell-Transmission Model

The mixed-integer linear programming formulation proposed in this paper is based on the cell-transmission model [5], [6]. The cell-transmission model is a discrete version of the hydrodynamic theory of traffic flow or the Lighthill–Whitham–Richards (LWR) model [7], [8]. The central component of the theory is the classic continuity equation for flow conservation that defines the relationship between flow (q) and density (k) over time and space in the form

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{1a}$$

supplemented by an assumption that traffic flow at location x at time t is a function of traffic density only, i.e.,

$$q = Q(k, x, t). \tag{1b}$$

For homogeneous roads and a time-invariant flow-density relationship, this equation can be further simplified as $q = Q(k)$. The assumed flow-density relationship can generally be calibrated with three parameters: free-flow speed; capacity; and jam density. These parameters are functions of the roadway configuration. The point on the flow-density curve that corresponds to the capacity flow and optimal density separates the curve into two regimes. The left regime represents uncongested traffic and the right regime represents congested traffic. Equations (1a) and (1b) imply that a slight change in flow would propagate along a kinematic wave, which can be solved mathematically by a characteristic line. Abrupt changes in flow or density could arise in two situations. In the situation in which an obstruction is suddenly removed from the roadway, such as in the case when a traffic light turns green from red, it is assumed in the model that the lead vehicle waiting at the intersection accelerates to the maximum allowed speed instantly with a wave velocity equal to the free-flow speed. At the same time, a wave propagates back-passing through the waiting vehicles. The densities between the two waves are fully defined by the functional form of the flow density relationship. This treatment eliminates the possible multivalued solutions for the process of queuing dissipation. In the situation in which a sudden drop in speed occurs in the traffic stream, such as the case when vehicles join a queue, there is a discontinuity in flow and density. The velocity of the discontinuity that separates two traffic regions can be represented by the shockwave equation in the form

$$u = \frac{q_u - q_d}{k_u - k_d}, \tag{1c}$$

where (q_u, k_u) and (q_d, k_d) represent the traffic states upstream and downstream of the interface, respectively. Equation (1a)–(1c) can be solved exactly. However, the solutions are often tedious, especially when the flow–density relationship is parabolic, a functional form that is widely assumed in practice.

The cell-transmission model is a discrete version of the LWR model. The network representation in the cell-transmission model differs from the traditional link–node representation: cells actually carry physical length, whereas links serve only the purpose for connectivity. In the cell-transmission model, the roadway is partitioned into discrete segments or "cells" labeled $1, 2, 3, \ldots, i$ and the time is partitioned into discrete steps labeled $1, 2, 3, \ldots, t$. The partition is done in such a way that it takes a single time step $(\Delta t = 1)$ to traverse a cell at free-flow travel speed. When the functional form of the relationship between flow and density is assumed to be piecewise linear, represented by two-wave speeds as shown in Fig. 1, (1a)–(1c) can be substituted by a set of difference equations as

$$z_i(t) = \min\{n_i(t), Q_i(t), Q_{i+1}(t), \alpha(N_{i+1}(t) - n_{i+1}(t))\} \tag{2a}$$

$$y_{i+1}(t) = z_i(t) \tag{2b}$$

$$n_i(t+1) = n_i(t) - z_i(t) + y_i(t) \tag{2c}$$

where $z_i(t)$ is the number of vehicles leaving cell i at time $[t, t+1), y_{i+1}(t)$, the number of vehicles entering cell $i+1$ at time $[t, t+1)$ and $n_i(t)$ the number of vehicles inside cell i at time $[t, t+1)$. Q_i is the maximum number of vehicles that are allowed to leave cell i, an entity equivalent to capacity. N_i is the maximum number of vehicles that can reside within cell i, an entity equivalent to jam density. The geometry of a cell is characterized by its capacity and density. Equation (2a) determines the outflow for cell i and (2b) ensures flow conservation such that the inflow of a cell is equal to the outflow of its upstream cell. Equation (2c) is a state function for flow conservation. This model was proven to be convergent to the continuum model when the discretized time and space elements approach zero [9]. In the network model, boundary conditions were derived for merges and diverges, consistent with traffic rules and the first-in–first-out (FIFO) sequence in the merges and diverges (see [6]). In recent years, the cell-transmission model has been applied to various transportation network problems [10], [11].

III. Model Formulation

This section describes in detail the model formulation for the signal-optimization problem. A list of notation is given. In our formulation, all the exogenous variables are represented by symbols in lower case and the endogenous variables and sets are represented by symbols in upper case.

Sets:

- O set of origin cells.
- D set of destination cells.
- E set of ordinary cells.
- I set of intersection cells.

Exogenous variables:

- $n_{i,t}$ number of vehicles in cell i at time t.
- $y_{i,t}$ number of vehicles leaving cell i at time $[t, t+1)$.

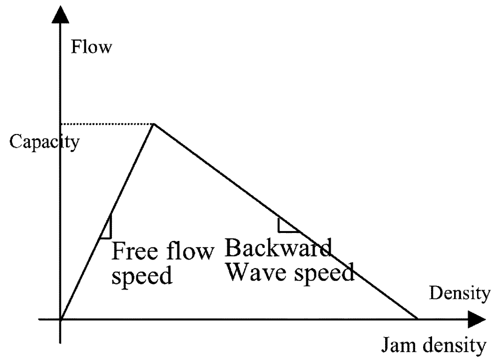


Fig. 1. Assumed two-wave speed flow-density relationship.

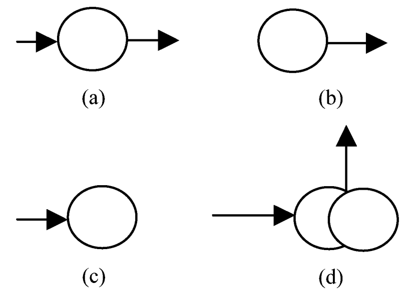


Fig. 2. Cell types used in the model formulation. (a) Type 1 cell: ordinary cell. (b) Type 2 cell: origin cell. (c) Type 3 cell: destination cell. (d) Type 4 cell: intersection cell.

$q_{i,t}$	actual capacity for intersection cell i in time interval $[t, t+1)$.
$w_t^{(ij)}$	0–1 variables for switching traffic light at intersection $(i,j) \in I$ in time interval $[t, t+1)$.

Endogenous variables:

$D_{i,t}$	number of vehicles entering the origin in $[t, t+1)$.
Q	capacity, maximum number of vehicles that can enter or exit cell i in $[t, t+1)$.
N:	jam density, maximum number of vehicles that can reside in cell i in $[t, t+1)$.
C	cycle length in time steps.
W	wave coefficient (a dimensionless unit representing the ratio of wave speed over free-flow speed).
GMIN	minimum green (in time steps).
GMAX	maximum green (in time steps).
α	coefficient for delay and the number of stops.
β	coefficient for treating vehicle holdings.
λ	penalty term for signal switching.

For the purpose of formulation, cells are classified into four types: ordinary; origin; destination; and intersection, as shown in Fig. 2(a)–(d). The definition of each cell type is given as follows.

Ordinary cell	cell that has a single downstream cell with a single upstream cell;
Intersection cells	pair of cells that represents an intersection;
Origin cell	ordinary cell connecting to a trip generator upstream;
Destination cell	cell that has no downstream cell.

A. Objective Functions

1) Minimization of Total Delay: One of the objectives often considered in the signal-optimization problem is the minimization of total delay. We will show that a linear objective function can be formulated directly from the definition of the system optimum condition. If the network is cleared at the end of the optimization period, then the total vehicle delay can be expressed exactly in the form

$$\sum_{i \in D} \sum_t t y_{i,t} - \sum_{i \in O} \sum_t t D_{i,t} - T.$$

The first two terms represent the total travel time for all vehicles in the unit of vehicle time, which is the area between the cumulative vehicle arrival curves at the origin and the destination for all O-Ds. The third term is the free-flow travel time for all the vehicles and is a constant. Note that if there is no elasticity in demand, then the second term is also a constant. The objective function can be simplified as

$$\min \sum_{i \in D} \sum_t t y_{i,t}. \tag{O1}$$

This linear objective function is exact and is independent of the underlying traffic-flow model adopted in representing traffic dynamics in the system.

2) Minimization of Total Number of Stops: The definition of the total number of stops used in practice has many different versions. We define the number of stops in a manner that is consistent with that defined in the kinematic wave model. A stop occurs only when a vehicle joins a stationary queue, i.e., a queue with jam density. By assuming that the flow–density relationship is of triangular shape, the queuing region immediate to the intersection can only be in two states, either in the state corresponding to zero flow and jam density (when traffic light turns red) or in the state corresponding to full capacity and optimal density (when traffic light turns green). Note that in the cell-transmission model it takes a single time step for a vehicle to traverse a cell under the free-flow condition. If there is no congestion, the outflow of a cell (denoted by i) at a time step should be equal to the outflow of its upstream cell (denoted by $i+1$) in the previous time step, i.e., $y_{i,t} = y_{i-1,t-1}$. When a cell is fully congested due to an obstruction downstream, we also have $y_{i,t} = y_{i-1,t-1} = 0$ in a steady state. When a queue forms in a cell, the outflow of the cell should be less than the outflow of its upstream cell, i.e., $y_{i,t} \le y_{i-1,t-1}$. The opposite is true in the queue-dissipation process $y_{i,t} \ge y_{i-1,t-1}$. The summation of the absolute value of the difference between $y_{i,t}$ and $y_{i-1,t-1}$ over a time period for a given region is proportional to the number of stops incurred during that time period in that region. One can further show that the number of stops can be approximated by $(1/2)\sum_i \sum_t |y_{i,t} - y_{i-1,t-1}|$. This approximation is compared with the exact solution obtained by the LWR model. The result is given in Fig. 3. In this figure, the x axis represents the traffic intensity and the y axis represents the number of stops. The number of stops increases as the ratio increases. As displayed in Fig. 3, the number of stops calculated with the approximation matches very well with that from the LWR model.

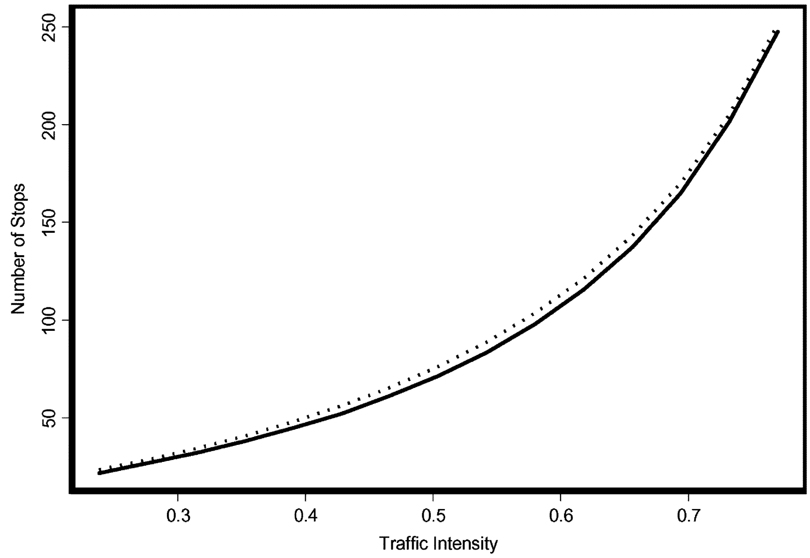


Fig. 3. Comparison of the number of stops calculated from the proposed model (dashed line) and the number of stops calculated from the LWR model (solid line).

The minimization of the total number of stops, i.e., $\min(1/2)\sum_i \sum_t |y_{i,t} - y_{i-1,t-1}|$, can be transformed into a linear objective function as

$$\min \frac{1}{2}\sum_i \sum_t \left(d_{i,t}^{(1)} + d_{i,t}^{(2)}\right)$$
$$\text{subject to} \quad y_{i,t} - y_{i-1,t-1} = d_{i,t}^{(1)} - d_{i,t}^{(2)}.$$

For the objective function that considers both the total number of stops and total delay, we can modify the objective function (O1) by introducing a weight $\alpha, 0 \le \alpha \le 1$, as

$$\min \alpha \sum_{i\in D}\sum_t ty_{i,t} + (1-\alpha)\frac{1}{2}\sum_i \sum_t \left(d_{i,t}^{(1)} + d_{i,t}^{(2)}\right) \quad \text{(O2)}$$

subject to the constraints discussed previously.

B. Constraints for Ordinary Cells

The state of a cell at any time step is determined by its inflow, outflow, and its state at the previous time step. An equivalent mathematical programming formulation to (2a) for this piecewise linear model at time t is

$$y_{i,t} \le n_{i,t}, \quad \forall t$$
$$y_{i,t} \le W(N - n_{i+1,t}), \quad \forall t$$
$$y_{i,t} \le Q, \quad \forall t.$$

These constraints along with the nonnegativity constraint for $y_{i,t}$, however, would yield a multivalued feasible solution set, including all the values between 0 and $\min\{n_{i,t}, Q, W(N - n_{i+1,t})\}$. The multivalued solution set is the major source of the unintended vehicle holding problem, i.e., a vehicle is held at a cell even though there is capacity available downstream for the vehicle to advance. In order to eliminate the unintended vehicle holding problem, one should only choose the upper bound of this multivalued solution set, ensuring that vehicles would advance to a downstream cell whenever possible. This condition is preserved when $\sum_{i\notin D}\sum_t ty_{i,t}$ is minimized for all t, by adding this term into (O2). The resulting objective function becomes

$$\min \alpha \sum_{i\in D}\sum_t ty_{i,t} + \frac{1}{2}(1-\alpha)\sum_i \sum_t \left(d_{i,t}^{(1)} + d_{i,t}^{(2)}\right) + \beta \sum_{i\notin D}\sum_t ty_{i,t}. \quad \text{(O3)}$$

Note that the value of β should be sufficiently small. Otherwise, it may interfere with the first two terms in the objective function. For ordinary cells, the inflow of a cell is equal to the outflow of its upstream cell. Suppose that cells are numbered in the ascending order in the direction of vehicle flow. The number of vehicles in a cell at a time interval can then be determined by the constraint

$$n_{i,t+1} = n_{i,t} + y_{i-1,t} - y_{i,t}, \quad \forall t.$$

C. Constraints for Origin Cells

Since origin cells are defined as ordinary cells connecting to a set of endogenously defined inflow representing the demand entering the system, the set of constraints for origin cells is exactly the same as that for ordinary cells, except for the inflow. Suppose that the arrivals at time t for origin i is $D_{i,t}$. The following constraints are used for origin cell i:

$$y_{i,t} \le n_{i,t}, \quad \forall t$$
$$y_{i,t} \le W(N - n_{i+1,t}), \quad \forall t$$
$$y_{i,t} \le Q, \quad \forall t$$
$$n_{i,t+1} = n_{i,t} + D_{i,t} - y_{i,t}, \quad \forall t.$$

D. Constraints for Destination Cells

The destination cells are assumed to be connected with a destination zone with infinite space, i.e., $N = \infty$. It is also assumed that flows entering the destination zone are not subject to capacity constraints, i.e., $Q = \infty$. With this simplification, the exit flow of the destination cells is simply the number of vehicles in the cell. The constraints for destination cells are reduced to

$$
\begin{aligned}
y_{i,t} &= n_{i,t}, \quad \forall t \\
n_{i,t+1} &= n_{i,t} + y_{i-1,t} - y_{i,t}, \quad \forall t.
\end{aligned}
$$

E. Constraints for Intersection Cells

Although the formulation discussed later is for one-way streets, it can be readily extended to four approaches with slight modifications. For one-way streets, an intersection can be characterized by a pair of cells (i,j). The two cells overlap in space, representing northbound (southbound) and eastbound (westbound) approaches at an intersection. The constraints can be expressed as

$$
\begin{aligned}
y_{i,t} &\le n_{i,t}, \quad \forall t \\
y_{i,t} &\le W(N - n_{i+1,t}), \quad \forall t \\
y_{i,t} &\le q_{i,t}, \quad \forall t \\
n_{i,t+1} &= n_{i,t} + y_{i-1,t} - y_{i,t}, \quad \forall t \\
y_{j,t} &\le n_{j,t}, \quad \forall t \\
y_{j,t} &\le W(N - n_{j+1,t}), \quad \forall t \\
y_{j,t} &\le q_{j,t}, \quad \forall t \\
n_{j,t+1} &= n_{j,t} + y_{j-1,t} - y_{j,t}, \quad \forall t \\
q_{i,t} &= w_t^{(ij)} Q, \quad \forall t \\
q_{j,t} &= \left(1 - w_t^{(ij)}\right) Q, \quad \forall t.
\end{aligned}
$$

Clearly, intersection cells have the same characteristics as the ordinary cells. Thus, the first eight constraints are for cells i and j, respectively, similar to the ones defined for ordinary cells. The only difference is that instead of using Q for capacity, we use $q_{i,t}$ and $q_{j,t}$ here. $q_{i,t}$ and $q_{j,t}$ are defined by the last two constraints, which are capacity restrictions associated with traffic lights, given by $y_{i,t} \le q_{i,t}$ and $y_{j,t} \le q_{j,t}$. $q_{i,t}$ or $q_{j,t}$ must satisfy the condition

$$
q_{i,t} = \begin{cases} 0, & \text{if approach } i \text{ is red at time } t \\ Q, & \text{if approach } i \text{ is green at time } t \end{cases}
$$

where $w_t^{(ij)}$ are 0–1 integer variables used as an indicator to determine if the signal for a specific approach should be red or green at time t, i.e.,

$$
w_t^{(ij)} = \begin{cases} 0, & \text{if approach } i \text{ is red at time } t \\ 1, & \text{if approach } i \text{ is green at time } t. \end{cases}
$$

The above formulation can be readily extended to capture the congestion effect at an intersection resulting from queues propagating from an immediate downstream intersection. The following constraints should be imposed to ensure that vehicles entering an intersection are reduced once the intersection is blocked by a static queue from another approach:

$$
\begin{aligned}
y_{j,t} &\le W(N - n_{i+1,t}), \quad \forall t \\
y_{i,t} &\le W(N - n_{j+1,t}), \quad \forall t
\end{aligned}
$$

where i and j are a pair of cells representing an intersection. In the case where $n_{i+1,t} = N$, meaning that the queue in approach i blocks the intersection, we have $y_{j,t} = 0$, i.e., vehicles from approach j are prohibited from entering the intersection.

1) Constraints to Maintain the Minimum Green: We use a moving time window with fixed width $\mathrm{GMIN} + 1$ to check if the minimum green duration is always preserved except for the offset, the first few time steps in the sequence. For the time steps covered by the moving time window, it is required that traffic light can be switched at most once. For example, suppose that we use 0 to indicate the red traffic light at a particular time step and 1 for the green traffic light. If $\mathrm{GMIN} = 4$, then the width of the moving time window is 5. The signal timing sequence 000 111 100 000 011 111$\underline{10\,001}$111 is not valid. The subsequence 10 001 violates the minimum red requirement. This formulation automatically chooses the best offset at the very beginning of the sequence. Mathematically, the constraint can be formulated as

$$
\sum_{\tau=t}^{t+\mathrm{GMIN}} |q_{i,\tau} - q_{i,\tau-1}| \le Q, \quad \forall t, (i,j) \in I
$$

which can be transformed into a set of linear constraints by using dummy variable $v_t^{(ij)}$

$$
\begin{aligned}
q_{i,t} - q_{i,t-1} &\le v_t^{(ij)}, \quad \forall t, (i,j) \in I \\
-q_{i,t} + q_{i,t-1} &\le v_t^{(ij)}, \quad \forall t, (i,j) \in I \\
\sum_{\tau=t}^{t+\mathrm{GMIN}} v_\tau^{(ij)} &\le Q \quad t = 0, 1, \ldots, T - \mathrm{GMIN}, \quad \forall (i,j) \in I.
\end{aligned}
$$

Dummy variable $v_t^{(ij)}$ represents the capacity difference between two neighboring time steps, which can be utilized to model the cost incurred due to lost time. Note that each switch of the traffic light is associated with a lost time. One can introduce a penalty function in the objective function such that a cost is imposed whenever a phase change occurs. The resulting objective function can be modified as

$$
\min \alpha \sum_{i \in D} \sum_t t y_{i,t} + \frac{1}{2}(1-\alpha) \sum_i \sum_t \left(d_{i,t}^{(1)} + d_{i,t}^{(2)}\right) + \beta \sum_{i \notin D} \sum_t t y_{i,t} + \lambda \sum_t v_t^{ij} \quad \text{(O4)}
$$

where λ is the penalty term.

2) Constraints to Maintain the Maximum Green: The moving time window concept can also be used to ensure that the green light provided to any approach would not exceed the maximum green duration allowed with the constraints

$$
\sum_{\tau=t}^{t+\mathrm{GMAX}} q_{i,\tau} \le Q\,\mathrm{GMAX}
$$

and

$$
\begin{aligned}
&\sum_{\tau=t}^{t+\mathrm{GMAX}} q_{j,\tau} \le Q\,\mathrm{GMAX}, \\
&t = 0, 1, \ldots, T - \mathrm{GMAX}, \quad \forall (i,j) \in I.
\end{aligned}
$$

These constraints can keep the cycle of a signal-timing plan to be dynamic. The optimal sequence of phases and the best offset will be determined automatically based on the level of traffic volumes entering the intersection. If the cycle length is required to be fixed, then additional constraints should be introduced, which is discussed later.

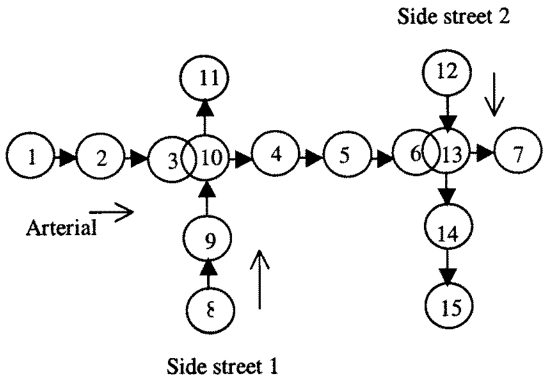


Fig. 4. Example network: an arterial with two intersections.

TABLE I
CELL OCCUPANCY FOR THE ILLUSTRATIVE EXAMPLE

Time	Arterial					Side Street 1			Side Street 2		
0	000	-	000	>	0	000	>	0	000	-	0
1	400	-	000	>	0	100	>	0	400	-	0
2	440	>	000	>	0	110	-	0	440	-	0
3	444	>	000	-	0	111	-	0	444	>	0
4	444	-	400	-	0	112	>	0	444	>	4
5	447	>	040	-	0	111	-	2	444	>	4
6	446	>	504	>	0	112	-	0	444	-	4
7	444	-	550	-	4	113	>	0	447	>	0
8	448	-	055	-	0	111	>	3	446	>	5
9	44B	>	00A	>	0	111	-	1	444	-	5
10	449	>	505	>	5	112	-	0	448	-	0
11	448	-	550	-	5	113	>	0	44B	>	0
12	44B	-	055	-	0	111	>	3	449	>	5
13	44F	-	00A	-	0	111	>	1	448	>	5
14	46G	>	00A	>	0	111	-	1	447	-	5
15	49B	>	505	-	5	112	-	0	44A	>	0
16	48B	-	555	-	5	113	>	0	448	>	5
17	47G	>	05A	-	0	111	-	3	447	>	5
18	4AB	>	50F	>	0	112	-	0	446	-	5
19	48B	-	55A	>	5	113	>	0	449	-	0
20	47G	-	05A	-	5	111	>	3	44D	>	0
21	4AG	-	00F	>	0	111	>	1	44B	-	5
22	4EG	>	00A	>	0	111	-	1	44E	-	0
23	5GB	>	505	>	0	112	-	0	46G	-	0
24	5BB	-	550	-	5	013	>	0	09G	>	0
25	0BG	-	055	-	5	001	>	3	09B	>	5
26	0BG	>	00A	-	0	000	-	1	04B	>	5
27	0BB	>	50A	>	0	000	-	0	00A	-	0
28	06B	-	555	>	5	000	>	0	00A	-	0
29	01G	>	055	>	5	000	-	0	00A	-	0
30	01B	>	505	-	5	000	-	0	00A	>	0
31	007	>	555	>	0	000	-	0	005	-	5
32	002	-	555	>	5	000	>	0	005	-	0
33	002	>	055	>	5	000	-	0	005	-	0
34	000	>	205	-	5	000	-	0	005	>	0
35	000	-	025	>	0	000	>	0	000	-	5
36	000	-	002	>	5	000	>	0	000	-	0
37	000	-	000	>	2	000	>	0	000	-	0
38	000	>	000	-	0	000	-	0	000	>	0
39	000	>	000	-	0	000	-	0	000	>	0
40	000	>	000	-	0	000	-	0	000	>	0

“ >” denotes the green light and “-“ denote the red light.
(A=10, B=11, C=12, D=13, E=14, F=15, G=16) .

3) Constraints to Ensure a Fixed Cycle Length: For a signal-timing plan with fixed cycle length, the following constraint should be added to the basic constraint set:

$$q_{i,t} - q_{i,t+C} = 0, \quad q_{j,t} - q_{j,t+C} = 0, \quad \forall t, (i,j) \in I.$$

These two constraints will also yield the best offset at the beginning of the period of optimization while preserving the fixed cycle length.

F. Constraints to Ensure Flow Conservation at the Network Level

It is required in our formulation that all arrivals will exit the network by the end of the optimization period. This can be ensured by the following constraint, imposed on the cells for origins and destinations:

$$\sum_i \sum_t y_{i,t} = \sum_j \sum_t D_{j,t}, \quad \forall t, \quad i \in D, j \in O.$$

IV. ILLUSTRATIVE EXAMPLE

A numerical example is presented here to illustrate the performance of the model. In this example, we consider a one-way arterial with two intersections. The roadway segment for the eastbound approach is partitioned into sections represented by seven cells, labeled $i = 1, 2, \ldots, 7$. The two side streets are partitioned into four cells each, $i = 8, 9, 10, 11$ for side street one and $i = 12, 13, 14, 15$ for side street two. The cell representation of the roadway geometry is given in Fig. 4.

In Fig. 4, cells 1, 8, and 12 are origin cells; cells 7, 11, and 15 are destination cells; and cell pairs (3,10) and (6,13) represent the two intersections. The rest of the cells are ordinary. Time is partitioned into discrete time steps $t = 1, 2, \ldots, 40$. Each time step is equivalent to 10 s. The free-flow speed is 30 mi/h. The capacity of a cell is five vehicles per time step and the jam density is 20 vehicles per cell. The minimum green time required is 10 s and the maximum green allowed is 30 s.

The arrival rates for the arterial, side street 1 and side street 2 are 4, 1, 4 vehicles per time step, respectively. In the example, the cycle length and splits are all decision variables. The mixed-integer linear programming formulation was programmed using CPLEX optimizer. The result is displayed in Table I.

Given the demand level, the downstream intersection is critical. Side street 1 is a minor street. It is interesting to see that this street has a longer overall green duration than the arterial. The queue for the arterial is held at the upstream intersection rather than the downstream intersection. This is reasonable, since the combined demand from the two side streets is higher than the arterial demand. The optimal strategy is to give more greens to the side street so as to maximize the system throughput instead of maintaining a maximum green band for the arterial. Other scenarios have been tested with the model. In terms of computation time, there is significantly variation with different scenarios. Overall, the scenario with fixed cycle time requires less computation time than the scenario with dynamic cycle length and split.

V. Extension to Modeling Signal Preemption for Emergency Vehicles

With minor modifications, the model discussed previously can be conveniently utilized to evaluate control strategies at signalized intersections for various other scenarios. In this section, we discuss the situation of signal preemption in the presence of emergency vehicles.

In practice, microscopic traffic-simulation models have often been used to evaluate the performance of signal timing plans under various traffic conditions. In the microscopic traffic-flow models, vehicles are treated as single entities. Vehicle dynamics, such as speed, acceleration and deceleration, and lateral movements, are captured in detail by car-following theory and gap-acceptance functions. When an emergency vehicle is present in the traffic stream, it is usually given the right-of-way according to the traffic rule. To accurately model how other vehicles would react to the presence of an emergency vehicle under free-flow and congested traffic conditions is a nontrivial task, involving the behavior of individual drivers that we know little about. To our knowledge, there is no theory available describing this particular behavior. The requirement for details in microscopic traffic-flow models makes it very challenging to microscopically model the traffic stream associated with emergency vehicles.

With a macroscopic traffic-flow model, the behavior of traffic in the presence of emergency vehicles can be handled implicitly. In essence, the passage of an emergency vehicle corresponds to a temporary reduction in roadway capacity when other vehicles yield to the emergency vehicle. Based on this observation, we made two assumptions for the movement of the emergency vehicle: 1) the emergency vehicle travels at a speed independent of the prevailing traffic stream and 2) the emergency vehicle behaves like a moving bottleneck that incurs a capacity reduction along its passage of travel. The first assumption can be relaxed by reducing the speed of the emergency vehicle only at locations where the traffic volume is extremely high. In reality, such temporary slow-down may occur from time to time. However, one would expect that the drivers of emergency vehicles would constantly adjust their speeds if they were delayed at a certain location of the network. It is reasonable to assume that, on average, the speed of an emergency vehicle is the free-flow speed. For the second assumption, since the actual level of capacity reduction corresponding to the passage of the emergency vehicle is not known, we use an indicator described later that can be calibrated through repeated experiments when more data become available. The indicator can also be utilized to determine if capacity reduction resulting from emergency vehicles can become a dominant factor to the overall delay.

Suppose that, at time step t, an emergency vehicle is traveling across cell i. As discussed previously, the presence of emergency vehicles would temporarily reduce the capacity. This is captured by the following constraint to the outflow of cell i at time t:

$$z_i(t) \leq \delta(t) Q_i(t)$$

where $\delta(t) \in [0,1]$. $\delta(t)$ can be viewed as an indictor that captures the degree of responsiveness of the traffic stream in the presence of an emergency vehicle. A very responsive traffic stream could be modeled by a full capacity reduction in which we choose $\delta = 0$, indicating that vehicles close to the emergency vehicle come to a full stop when they yield to the emergency vehicle. Partial responsiveness, meaning that some drivers are cooperative whereas others are not, is associated with a partial capacity reduction in which the value of δ is chosen to be less than 1, depending on the level of responsiveness. Note that the underlying assumption for this formulation is that the influence area of the emergency vehicle is a single cell at each time step. This assumption can be relaxed as well by considering a larger influence area, covering more cells downstream and/or upstream if one wants to model the scenario more realistically.

TABLE II
Cell Occupancy for the Illustrative Example That Captures Traffic-Signal Preemption for Emergency Vehicles

Time	Arterial					Side Street 1			Side Street 2		
0	000	>	000	-	0	000	-	0	000	>	0
1	400	>	000	-	0	100	-	0	400	>	0
2	440	-	000	>	0	110	>	0	440	-	0
3	444	-	000	>	0	111	>	0	444	-	0
4	447	>	000	-	0	112	>	1	447	>	4
5	446	>	500	-	0	112	-	0	446	>	5
6	444	-	550	-	0	113	>	0	444	>	5
7	448	-	055	>	0	111	>	3	444	-	4
8	44B	-	005	>	5	111	>	1	447	-	0
9	44E	>	000	-	5	111	-	1	44A	>	0
10	46B	>	500	-	0	112	-	0	449	>	5
11	44B	>	550	-	0	113	-	0	448	>	5
12	**4**4A	-	555	>	0	114	>	0	446	-	5
13	7**0**C	-	055	-	5	111	>	4	449	>	0
14	65**C**	>	00A	-	0	112	-	0	448	>	5
15	48G	>	**0**0A	-	0	112	-	0	447	>	5
16	4BB	>	5**0**A	>	0	113	-	0	445	-	5
17	4AB	-	55**5**	>	5	114	>	0	448	-	0
18	48G	-	05A	>	**0**	111	>	4	44C	-	0
19	4CG	>	00A	-	5	111	-	1	44F	>	0
20	4FB	>	50A	-	0	112	-	0	47B	>	5
21	7BB	-	55A	>	0	113	>	0	45B	-	5
22	5BG	-	05B	>	5	111	>	3	44G	-	0
23	4GG	>	00A	>	5	111	-	1	47G	-	0
24	4GB	>	505	-	5	012	-	0	0AG	>	0
25	4BB	-	555	>	0	003	>	0	0AB	-	5
26	09G	-	055	>	5	000	>	3	06G	-	0
27	09G	>	005	>	5	000	-	0	06G	-	0
28	09B	>	500	-	5	000	-	0	06G	>	0
29	05B	>	550	-	0	000	-	0	06B	>	5
30	00A	-	555	>	0	000	>	0	01B	-	5
31	00A	>	055	>	5	000	-	0	00B	-	0
32	006	>	505	-	5	000	-	0	00B	>	0
33	001	>	555	>	0	000	>	0	007	-	5
34	000	-	155	>	5	000	>	0	007	-	0
35	000	>	015	>	5	000	-	0	007	-	0
36	000	>	001	-	5	000	-	0	007	>	0
37	000	>	001	-	0	000	-	0	002	>	5
38	000	-	001	>	0	000	>	0	000	-	2
39	000	-	000	>	1	000	>	0	000	-	0
40	000	-	000	>	0	000	>	0	000	-	0

" >" denotes the green light and "-" denote the red light. (A=10, B=11, C=12, D=13, E=14, F=15, G=16) .

The result of an illustrative example for signal preemption in the presence of emergency vehicles is given in Table II. The demand level and system constraints used in this example are identical to the ones used in the previous example. An emergency vehicle enters the network at time step 12. The passage of the emergency vehicle through the network corresponds to the cells with bold numbers. Here, we assume that $\delta(t) = 0$, i.e., vehicles sharing the same cell with an emergency vehicle would respond with a full stop. For example, as shown in the table, vehicles in cell 6 at time step 17 are all held in the cell, even though traffic light downstream is green. An important issue related to this scenario is how to optimize the traffic-recovering process after the initiation of a signal preemption call. This is, however, beyond the scope of this paper. The model formulated in this paper has the potential to be used to obtain insights into developing alternative optimum control strategies.

VI. Conclusion and Future Research

This paper proposes an enhanced 0–1 mixed-integer linear programming formulation for the signal-optimization problem based on the cell-transmission model. This model is capable of capturing physical queues, fixed and variable cycle length, and the number of stops, while preserving the minimum and maximum green durations. A penalty term for the phase change is used in the objective function to capture the cost associated with the lost time. Computationally, the proposed formulation has the least number of integer variables compared with other formulations developed with the same modeling approach. The total number of 0–1 integers is equal to the total number of intersections multiplied by the total number of time steps for the entire optimization period.

The model has been used for deriving optimal signal sequence when signal preemption is provided to emergency vehicles, which were modeled as a moving obstruction. The preliminary results are encouraging. The proposed model also has the potential to be extended to model other ITS components, such as a bus signal priority system. Future research may include developing efficient solution procedures to take advantage of the special characteristics of the constraint structure in the formulation so as to reduce the computation time.

References

[1] G. Improta and G. E. Cantarella, "Control system design for an individual signalized junction," *Transport. Res. B.*, vol. 18, no. 2, pp. 147–167, 1984.

[2] H. K. Lo, "A novel traffic signal control formulation," *Transport. Res. A.*, vol. 33, no. 6, pp. 433–448, 2000.

[3] B. Ramanathan, M. G. McNally, and R. Jayakrishnan, "Formulation of modern signal control operations as a nonlinear mixed integer program," in *Proc. IEEE Intelligent Transportation Systems Council (ITSC'95)*, Kobe, Japan, 1995, pp. 165–171.

[4] N. H. Gartner and C. Stamatiadis, "Arterial-based control of traffic flow in urban grid networks," *Math. Comp. Modeling*, vol. 35, pp. 657–671, 2002.

[5] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transport. Res. B.*, vol. 28, no. 4, pp. 269–287, 1994.

[6] ——, "The cell transmission model: Network traffic simulation," *Transport. Res. B.*, vol. 29, pp. 79–93, 1995.

[7] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," in *Proc. Roy. Soc. A*, vol. 229, 1955, pp. 317–345.

[8] P. I. Richards, "Shockwaves on the highway," *Oper. Res.*, vol. 4, pp. 42–51, 1955.

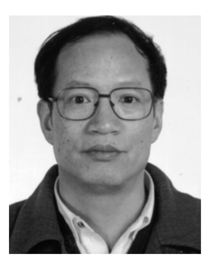
[9] C. F. Daganzo, "A finite difference approximation of the kinematic wave model of traffic flow," *Trans. Res. B.*, vol. 29, no. 4, pp. 261–276, 1995.

[10] H. K. Lo, "A dynamic traffic assignment formulation that encapsulates the cell-transmission model," presented at the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, 1999.

[11] T. Ziliaskopoulos, "A linear programming model for the single destination system optimum dynamic traffic assignment program," *Transport. Sci.*, vol. 34, no. 1, 2000.

Wei-Hua Lin received the B.S. degree from Brigham Young University, Provo, UT, the M.S. degree from Rensselaer Polytechnic Institute, Troy, NY, and the Ph.D. degree from the University of California, Berkeley.

He was a Postdoctoral Researcher in the Partners for Advanced Transit and Highways (PATH) program, University of California, Berkeley. Before joining the faculty of the University of Arizona, Tempe, he was an Assistant Professor in the Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg. He currently is an Assistant Professor in the Department of Systems and Industrial Engineering, the University of Arizona. His research interests include optimization in the intelligent transportation systems, logistics, and transportation network analysis.



Chenghong Wang received the B.S. degree from Heibei University of Science and Technology, Tianjin, China, in 1982, the M.S. degree from China University of Mining and Technology, Beijing, in 1988, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1997.

From 1997 to 1999, he was a Postdoctoral Fellow at the Institute of Mathematics and Systems Sciences, Chinese Academy of Sciences. Since 1999, he has been with Division III, Information Sciences Department, National Natural Science Foundation of China, Beijing, where he is an Associate Professor and Director. His main research interests include control theory, reliability theory, and operational research.