PERGAMON

Transportation Research Part A 33 (1999) 433–448

TRANSPORTATION RESEARCH PART A

www.elsevier.com/locate/tra

A novel traffic signal control formulation

Hong K. Lo*

Hong Kong University of Science and Technology, Department of Civil Engineering, Clear Water Bay, Hong Kong

Received 19 August 1997; received in revised form 8 July 1998; accepted 15 July 1998

Abstract

A novel traffic signal control formulation is developed through a mixed integer programming technique. The formulation considers dynamic traffic, uses dynamic traffic demand as input, and takes advantage of a convergent numerical approximation to the hydrodynamic model of traffic flow. As inherent from the underlying hydrodynamic model, this formulation covers the whole range of the fundamental relationships between speed, flow, and density. Kinematic waves of the stop-and-go traffic associated with traffic signals are also captured. Because of this property, one does not need to tune or switch the model for the different traffic conditions. It "automatically" adjusts to the different traffic conditions. We applied the model to three demand scenarios in a simple network. The results seemed promising. This model produced timing plans that are consistent with models that work for unsaturated conditions. In gridlock conditions, it produced a timing plan that was better than conventional queue management practices.

1. Introduction

Traffic signal is an essential element to manage the transportation network. Nevertheless, it is widely accepted that the benefits of traffic control signal systems are not being fully realized (US GAO, 1994). Along with the movement of Intelligent Transportation Systems (ITS) in the United States, traffic signal control remains one of the most heavily funded research and development items (Santiago, 1993). The research on traffic signal control is by no means complete.

Given its importance, a number of traffic signal control models have been developed in the past. Despite substantial differences among them, they can be roughly classified by two approaches. The first approach is developed mainly for unsaturated traffic. A basic assumption of this approach is that traffic flows more or less at the design speed or in the uncongested regime of a

* Tel.: +852-2358-8742; Fax: +852-2358-1534; E-mail: hong.k.lo@ust.hk

PII: S0965-8564(98)00049-4

speed-flow relationship. Furthermore, it is assumed that traffic is in steady-state; hence optimizing a set of fixed or repeatable timing plans is sufficient. Different timing plans are selected for different times of the day or when traffic becomes markedly different. Another basic assumption is that queuing is represented by the results of classical queuing theory, in terms of average vehicle waiting time, average service rate, etc. The fundamental principle and theory of this approach, as well as the possible extensions, can be found in Newell (1989). In the cases in which queues are modeled explicitly, a point-queue concept is used (e.g. TRANSYT): i.e. a queue has no physical dimensions. Hence, a link may be modeled to hold more vehicles than its length. Most existing models belong to this category. In a low to moderate degree of congestion, this approach serves reasonably well. To cite some examples: TRANSYT (Courage and Wallace, 1991) remains a popular software to determine fixed timing plans, MAXBAND (Little et al., 1981) and PASSER (TTI, 1991) focused on progression bandwidth optimization; Gartner et al. (1991) extended this concept to a multiband approach; Gartner (1985) developed a demand responsive traffic signal control approach, etc. Shepherd (1992) or Wood (1993) provided a good summary of these models.

The second approach is developed mainly for congested or over-saturated traffic conditions where queues persist and cannot be cleared totally. The dynamics of queue formation and dissipation becomes critical. Therefore, a dynamic formulation is necessary. Over the past few decades, most of the procedures proposed in this approach were ad hoc in nature. Their main objective was to provide rules-of-thumb based on experience or simple analysis to help practising engineers (example, Pignataro et al., 1978; Rathi, 1988; Quinn, 1992). The first two formulations were reported in Gazis (1964); D'ans and Gazis (1976), and Michalopoulos and Stephanopoulos (1977). Actually, the two were similar in many ways, with the latter adding the queue length constraint. Both formulations simplified the problem substantially, however, by making these two assumptions. Firstly, signal control was modeled as a continuous function with an average flow rate of $Q(t) = sg(t)/C$, where $Q(t), s, C, g(t)$ are, respectively, the maximum outflow at time t, saturation flow, cycle time, and green time. In reality, traffic flow at a signalized intersection follows a step function: $Q(t) = s$ when the time t is in a green phase; $Q(t) = 0$ when t is in a red phase. Smoothing $Q(t)$ would remove the stop-and-go traffic generated naturally by a signal. Secondly, and more importantly, both formulations modeled a queue by a simple first order condition:

$$\frac{dv}{dt} = \begin{cases} b(t) - Q(t) & \text{if } v > 0 \\ \max\{0, b(t) - Q(t)\} & \text{if } v = 0 \end{cases}$$

where $v, b(t)$ are, respectively, the number of vehicles on a link and its inflow. In this simplification, important traffic flow characteristics, such as kinematics waves and the fundamental speed-flow-density relationships, are ignored. More recently, a number of dynamic formulations have been proposed (examples, Eddelbuttel and Cremer, 1994; Abu-Lebdeh and Benekohal, 1997; Wey and Jayakrishnan, 1997). They introduced improvements by capturing parameters of queue dynamics, such as queue length, position of the end of queue, etc. However, the derivation of these parameters was based on assumptions that may not be appropriate for all traffic conditions; for example, "queues always exist on each link". None is intended to capture the full range of traffic characteristics.

The objective of this paper is to develop a dynamic traffic signal control formulation that is applicable for all traffic conditions. It aims to unify the two different approaches discussed earlier, thus removing the need to switch from one approach to another as traffic is gradually building up or dissipating. This is achieved by taking advantage of a recent traffic flow model—the Cell-Transmission Model (CTM) (Daganzo, 1994, 1995). CTM provides a convergent approximation to the Lighthill and Whitham (1955) and Richards (1956) (LWR) model which is arguably the most recognized traffic flow model. CTM covers the full range of the fundamental flow–density–speed relationships and has been validated by field data (Lin and Daganzo, 1994; Lin and Ahanotu, 1995). This capability makes it an excellent platform for modeling dynamic traffic. Moreover, as a dynamic formulation, time variant demands can be modeled readily.

This new signal control formulation is developed by embedding CTM through a mixed integer programming approach. Results show that this new formulation produces timing plans that are consistent with models that work for unsaturated conditions. In gridlock conditions, it produces timing plans that are better than conventional queue management practices. These results seem promising. The outline of this paper is the following. Section 2 depicts a brief summary of CTM together with the mixed integer programming formulation. Section 3 presents the numerical study and results. Finally, Section 4 discusses the concluding remarks.

2. Formulation

2.1. The cell transmission model

The Lighthill and Whitham (1955) and Richards (1956) (LWR) model can be stated by the following two conditions:

$$\frac{\partial f}{\partial x} + \frac{\partial k}{\partial t} = 0 \text{ and } f = F(k, x, t) \tag{1}$$

where f is the traffic flow; k is the density; x and t, respectively, are the space and time variables, and F is a function relating f and k. The first partial differential equation states the traffic flow conservation condition. The relation F between flow f and density k is a fundamental relationship in traffic flow theory. Given a set of well-posed initial conditions, one can determine f and k at any (x, t) by solving (1). This model is sometimes referred to as the hydrodynamic or kinematic wave model of traffic flow. Lighthill and Whitham (1955) and Newell (1991) developed two different solution approaches to this model. Daganzo (1994, 1995) further simplified the solution scheme by adopting the following relationship between traffic flow, f, and density, k:

$$f = \min\{Vk, Q, W(k_{\text{jam}} - k)\} \tag{2}$$

where k_{jam}, Q, V, W denote, respectively, the jam density, inflow capacity (or maximum allowable inflow), free-flow speed, and the speed of the backward shock wave (or the backward propagation speed of disturbances in congested traffic). Essentially, (2) approximates the flow-density relationship by a piece-wise linear model as shown in Fig. 1. By discretizing the road into homogenous sections (or cells) and time into intervals such that the cell length is equal to the distance traveled by free-flowing traffic in one time interval, Daganzo (1994, 1995) showed that the LWR results are approximated by this set of recursive equations:

$$n_j(t+1) = n_j(t) + f_j(t) - f_{j+1}(t) \tag{3}$$

$$f_j(t) = \min\{n_{j-1}(t), Q_j(t), (W/V)[N_j(t) - n_j(t)]\} \tag{4}$$

where the subscript j refers to a cell j, and $j+1(j-1)$ represents the cell downstream (upstream) of j. The variables $n_j(t), f_j(t), N_j(t)$ denote the number of vehicles, the actual inflow, and the maximum number of vehicles (or holding capacity) allowable in cell j at time t, respectively. The variables $Q_j(t)$, V, W follow the earlier definitions. It is important to differentiate between $Q_j(t)$ and $f_j(t)$: the former is the inflow capacity while the latter is the actual inflow. Basically, Eq. (3) describes the conservation of traffic in cell j: the number of vehicles in cell j at time $t+1$ is equal to the number of vehicles at time t plus those entered minus those left. Eq. (4) states that the number of vehicles entering (or the actual inflow to) cell j at time t is the minimum of three terms: vehicles at the upstream cell waiting to enter j, the inflow capacity of j, and a function of the available space in j.

Because (3) and (4) provide a numerical approximation to the LWR equations, all the traffic phenomena demonstrated in the LWR model, such as kinematic waves, can be replicated in CTM. Daganzo (1995) further extended the above recursive equations to handle network traffic with multiple origin–destination pairs. In this formulation, only the simplest form, expressed as (3)–(4), is used. Extending the present signal control formulation to include turning movements is possible by using the general CTM formulation.

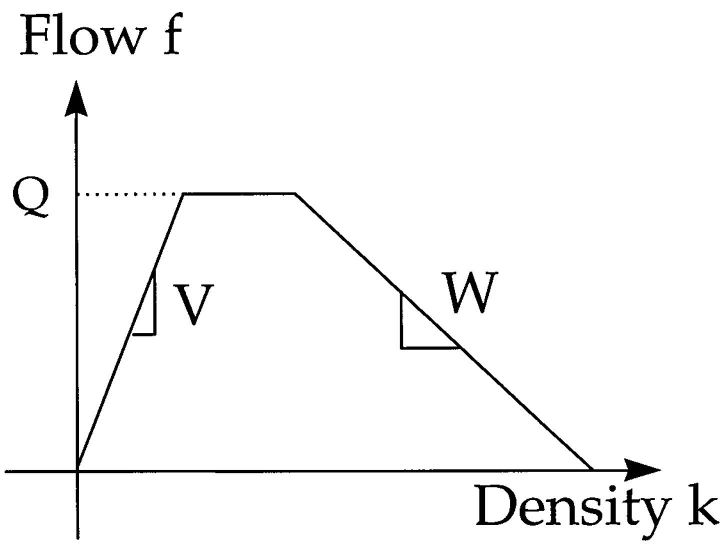

Fig. 1. The flow-density relationship used in CTM.

2.2. The cell-based traffic dynamics representation

Consider a network of one-way streets without turning movements. For ease of exposition, a simple network as shown in Fig. 2 is used for illustration. Links 1–2–3 represent a major arterial while links 4–5 and links 6–7 are minor cross streets. Links are classified into three mutually exclusive types:

1. source link (e.g. links 1, 4, 6 in Fig. 2), denoted by link set Ω—where exogenous traffic demands are introduced to the network,
2. exit link (e.g. links 3, 5, 7 in Fig. 2), denoted by link set E—where traffic flows terminate and exit the network,
3. intermediate link (e.g. link 2 in Fig. 2), denoted by link set Ψ—which is neither a source nor an exit link.

Each link belongs to only one of these three sets and is subdivided into cells. The cells are numbered from the upstream direction of traffic flow and follow this nomenclature: cell (i, j) represents the jth cell in link i. The first cell of a link, marked by "1" in Fig. 2 , serves as an entrance to the link. Except the cells with special characteristics as discussed below, all the cells are assigned a set of identical characteristics, including: $N_{ij}(t)$—holding capacity (in this study, this is constant over time, so the time dimension can be dropped); $Q_{ij}(t)$—inflow capacity; V—free-flow speed; W—backward shock wave speed. By carefully relating the neighboring cells, the difference equations Eqs. (3)–(4) can be written for each cell in the network, albeit a cell's neighbor may be in a different link. For example, for cell(2,1), the difference equations are:

$$n_{21}(t+1) = n_{21}(t) + f_{21}(t) - f_{22}(t) \tag{5}$$

$$f_{21}(t) = \min\{n_{13}(t), Q_{21}(t), (W/V)[N_{21} - n_{21}(t)]\} \tag{6}$$

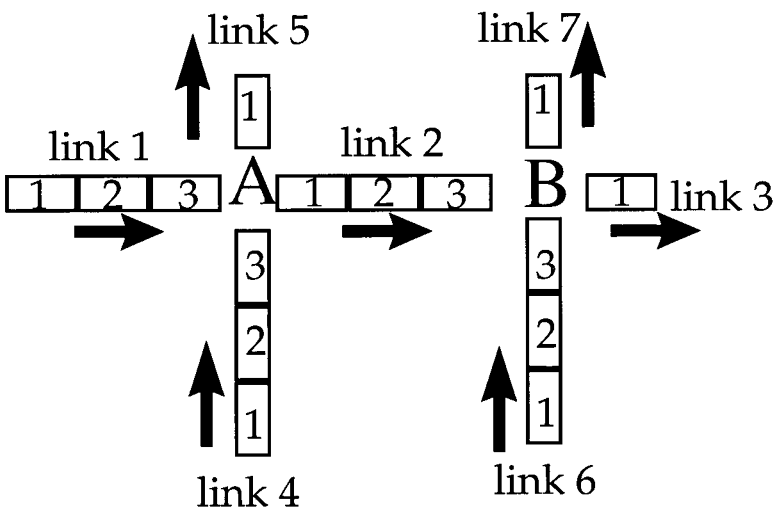

Fig. 2. The example network.

The first cell in a source link at time $t = 1$ stores the total demand that intends to enter the network—the effect of a big parking lot. The inflow capacity of the second cell, $Q_{i2}(t), i \in \Omega$, of a source link is set to the exogenous dynamic demand. Vehicles will enter the network according to the dynamic demand if space is available in the second cell. Otherwise, vehicles will wait in the big parking lot. As $Q_{i2}(t), i \in \Omega$ is time variant, one can model dynamic demand readily from this formulation. Mathematically,

$$n_{i1}(1) = \sum_t D_i(t), i \in \Omega \tag{7}$$

$$Q_{i2}(t) = D_i(t), i \in \Omega \tag{8}$$

where $n_{i1}(1)$ denotes the number of vehicles in cell(i,1) at time $t = 1$; $Q_{i2}(t)$ the inflow capacity of cell(i,2) at time t; and $D_i(t)$ the exogenous demand into source link i at time t.

The exit links have only one cell. This cell serves two purposes. Firstly, it simulates the action of a signal. Its inflow capacity is set to the saturation flow when t is in a green phase; zero otherwise. That is,

$$Q_{i1}(t) = \begin{cases} s & \text{if } t \in \text{ green phase}, i \in E \\ 0 & \text{if } t \in \text{red phase}, i \in E \end{cases} \tag{9}$$

where s is the saturation flow and $Q_{i1}(t)$ is the inflow capacity of cell (i,1), $i \in E$.

Secondly, it serves as a reservoir to store the vehicles that exit from the network—i.e. the arrival flow. The cell's holding capacity is set to infinity in order not to restrict the arrival of vehicles. (This holding capacity can assume a finite value to account for the case of a limited space at the exit cell.) That is,

$$N_{i1} = \infty, i \in E \tag{10}$$

For the intermediate links, the first cell also serves as a signal. That is,

$$Q_{i1}(t) = \begin{cases} s & \text{if } t \in \text{green phase}, i \in \Psi \\ 0 & \text{if } t \in \text{red phase}, i \in \Psi \end{cases} \tag{11}$$

2.3. The cell-based mixed integer programming formulation

Traffic signal control serves many purposes and hence many objective functions are possible. This study chooses to minimize the total network delay. CTM provides a convenient way to determine delay in a cell. This delay is defined to be the additional time beyond the nominal or free-flow travel time a vehicle stays in a cell. At the cell level, the delay is determined as:

$$d_{ij}(t) = n_{ij}(t) - f_{i,j+1}(t)$$

Basically, if the exit flow from cell(i, j) at time t is less than its current occupancy, those who cannot leave the cell will incur a delay of one time step. Once the delay has been determined at the cell level, it can be easily aggregated at the link or network level. Thus, the objective function is:

$$J = \min \sum_t \sum_i \sum_j d_{ij}(t) \tag{12}$$

The objective is to select the duration of the signal phases such as J is minimized. This minimization is subject to the constraints of exogenous dynamic demand, traffic dynamics prescribed by CTM, and the minimum and maximum durations of red and green times commonly adopted in practice.

The constraints include:

(A) Dynamic exogenous demand:

$$Q_{i2}(t) = D_i(t), \ i \in \Omega \tag{13}$$

where $D_i(t)$ is the exogenous dynamic demand to source link i.

(B) Traffic dynamics prescribed by CTM:

The difference Eq. (4) is nonlinear due to the embedded minimization. For simplicity, this formulation replaces Eq. (4) by three "less than or equal to" constraints as expressed in (15)–(17). With the objective of minimizing total network delay, it is anticipated that the actual inflow will take up the maximum allowed by (15)–(17). Hence, it is equivalent to (4). Mathematically, $f_{ij}(t)$ could assume a lower value, however, such as zero—a phenomenon often referred to as the vehicle holding problem. This problem can be strictly avoided by converting (4) to a set of mixed-integer linear constraints (Lo, 1999).

$$n_{ij}(t+1) = n_{ij}(t) + f_{ij}(t) - f_{i,j+1}(t) \tag{14}$$

$$f_{ij}(t) \leqslant n_{i,j-1}(t) \tag{15}$$

$$f_{ij}(t) \leqslant Q_{ij}(t) \tag{16}$$

$$f_{ij}(t) \leqslant \left(\frac{W}{V}\right)\left[N_{ij} - n_{ij}(t)\right] \tag{17}$$

The variables follow their earlier definitions.

(C) Control decisions:

This study simplifies the formulation by not including yellow time and assuming a constant cycle time. Methodologically, relaxing both simplifications would not pose any difficulties and is left to a future study. In fact, this dynamic formulation can readily work without a constant cycle time.

The decision is how long should each green duration last? As a dynamic formulation, the green duration could be different in each cycle. To include coordination between neighboring intersections, we also introduce an initial offset for the major approaches at the beginning of the planning horizon. Thus, the two decision variables are the initial offset and length of green time in each cycle. Offset is defined as the time between the start of the modeling horizon ($t = 1$) to the start of the first green, and is set to red (green) for the major (minor) street traffic. In the example shown in Fig. 2, the major approach is going eastbound. In this simple network, defining the green duration for one approach automatically defines the red duration for the competing approach at the same intersection. For ease of referencing, we number the phase consecutively as shown in Fig. 3.

For safety reasons and to avoid long red time that may provoke violation, both green and red times are restricted by upper and lower bounds:

$$R_{\min} \leqslant \theta_a < R_{\max} \tag{18}$$

where θ_a is the offset for the major approach at intersection a. The variables $R_{\max}$ and $R_{\min}$ denote the maximum and minimum red time. Similarly, the green times are bounded by:

$$G_{\min} \leqslant g_a(l) \leqslant G_{\max}, \ \forall l \tag{19}$$

where $G_{\max}, G_{\min}, g_a(l)$ denote, respectively, the maximum and minimum green time and the green time for the major approach at intersection a in cycle l. Since the cycle time C is fixed, the maximum and minimum green automatically define the minimum and maximum red, stated as:

$$\begin{aligned} R_{\max} &= C - G_{\min} \\ R_{\min} &= C - G_{\max} \end{aligned} \tag{20}$$

Choosing $\theta_a, g_a(l)$ as the decision variables, one must relate them to the inflow capacities, $Q_{i1}(t)$, of the corresponding signalized cells. To illustrate the procedure, we discuss the conditions for the offset period. The conditions for the other phases can be constructed in a similar way. In the offset period $1 \leq t \leq \theta_a$, the major approach at intersection a receives red, while the minor approach receives green. Mathematically, this can be written as:

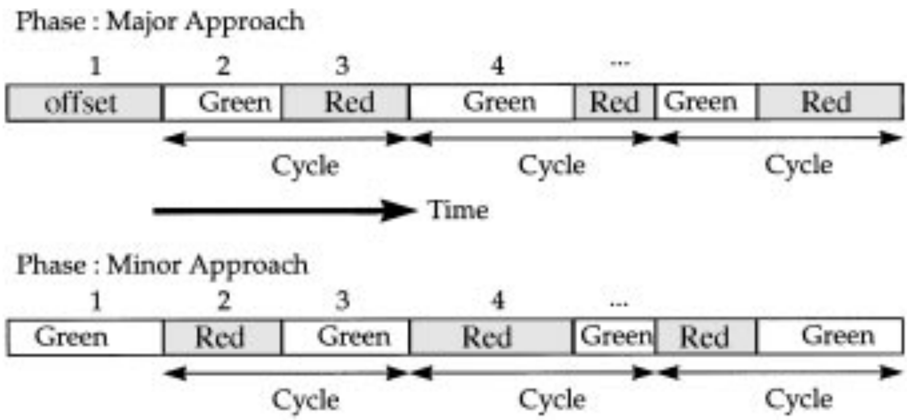

Fig. 3. The green/red periods of the competing approaches at an intersection.

$$\text{if } 1 \leqslant t \leqslant \theta_a, \text{ then} \begin{cases} Q_{m1}(t) = 0 \\ Q_{n1}(t) = s \end{cases} \tag{21}$$

where $m(n)$ represents the link on the major (minor) approach at intersection a.

As an "if-then" condition, (21) cannot be stated directly as part of the constraint set. All the constraints of a mathematical program must hold *simultaneously*; they do not allow for logic conditions such as (21). To capture the effect of (21), we use a mixed-integer programming technique. Conceptually, this involves two steps. Firstly, we identify for each time t whether it falls in this offset period or not. Secondly, according to the result of the first step, we set the inflow capacity of the signalized cell at time t to either the saturation flow or zero. This procedure is accomplished with two binary variables: $y_a(p, t), z_a(t)$. The variable p represents the consecutive phase number. For the offset period, $p = 1$, the variable $y_a(p, t)$ indicates whether time t lies within the period $t \leqslant \theta_a$. If it does, set $y_a(1, t) = 1$; otherwise, set $y_a(1, t) = 0$. The second binary variable $z_a(t)$ allocates the appropriate inflow capacity to the signalized cell according to the value of $y_a(1, t)$. The corresponding set of linear constraints that replicate (21) can be stated in the following:

$$L \cdot y_a(1, t) + \varepsilon \leqslant t - \theta_a \leqslant U \cdot [1 - y_a(1, t)] \tag{22}$$

$$y_a(1, t) + z_a(t) \leqslant 1 \tag{23}$$

$$Q_{m1}(t) = z_a(t) \cdot s \tag{24}$$

$$Q_{n1}(t) = [1 - z_a(t)] \cdot s \tag{25}$$

$$t \geqslant 1 \tag{26}$$

where U is a very large positive constant; L is a very large negative constant; and ε is a very small positive constant. For the purpose of illustration, one could consider these three constants as: $U \to \infty$; $L \to -\infty$; $\varepsilon \to 10^L$. If one substitutes the two possible values of $y_a(1, t)$ into (22), one would find the following result:

$$y_a(1, t) = 1 \Longleftrightarrow L + \varepsilon \leqslant t - \theta_a \leqslant 0 \Longleftrightarrow t \leqslant \theta_a$$
$$y_a(1, t) = 0 \Longleftrightarrow \varepsilon \leqslant t - \theta_a \leqslant U \Longleftrightarrow t > \theta_a$$

Thus, constraint (22) correctly captures the relationship between $y_a(1, t)$ and the time bound θ_a. In (23), (24), and (25), the value of $y_a(1, t)$ determined in (22) is used to derive the corresponding inflow capacities of the signalized cells. If one substitutes the two values of $y_a(1, t)$ into (23), (24), and (25), one would find the following result:

$$y_a(1, t) = 1 \Longleftrightarrow z_a(t) \leqslant 0 \Longleftrightarrow z_a(t) = 0 \Longleftrightarrow \begin{cases} Q_{m1}(t) = 0 \\ Q_{n1}(t) = s \end{cases}$$

$$y_a(1, t) = 0 \Longleftrightarrow z_a(t) \leqslant 1 \Longleftrightarrow z_a(t) = 1 \text{ or } 0 \Longleftrightarrow \begin{cases} Q_{m1}(t) = 0 \text{ or } s \\ Q_{n1}(t) = s \text{ or } 0 \end{cases}$$

The condition $y_a(1, t) = 1$ (i.e. $t \leqslant \theta_a$) together with (26) imply that t is in the offset period $1 \leqslant t \leqslant \theta_a$. For this case, we have the required outcome: $Q_{m1}(t) = 0$ for the major approach and $Q_{n1}(t) = s$ for the minor approach. On the other hand, with $y_a(1, t) = 0$ (i.e. time t is outside the offset period), $z_a(t)$ can be either 1 or 0. Therefore, the constraint set (22)–(26) does not dictate the inflow capacities outside the range $1 \leqslant t \leqslant \theta_a$. The variable $Q_{m1}(t)$ and $Q_{n1}(t)$ can be either s or zero dependent on whether the future t will fall in a green or red phase. In summary, the linear constraints (22)–(26) precisely replicate the "if-then" condition (21).

For each of the phases in the modeling horizon, we repeat this procedure and construct a set of mixed-integer linear constraints similar to (22)–(26). Each set of constraints transforms the signal states to the inflow capacities for t's that lie in that phase. Moreover, by constructing the sets of constraints such that each t lies in exactly one phase, there will be no conflict in assigning the value of $z_a(t)$.

3. Numerical study of performance

3.1. Study design

The objective of this numerical study is to demonstrate the properties of the model and verify its performance. Therefore, a small tractable example is preferable, although the model can be applied to larger networks and longer modeling horizons. To this end, the network shown in Fig. 1 is used. The parameter values selected include:

- Free-flow speed: 30 mile/h (or 48.3 km/h)
- Backward shock wave speed: 30 mile/h (or 48.3 km/h)
- Jam density: 200 vehicle/mile (or 124.3 vehicles/km)
- Saturated flow: 1800 vehicles/h (or 5 vehicles/time step)
- Time step: 10 s
- Cycle time: 40 s
- Minimum green and red time: 10 s
- Maximum green and red time: 30 s
- Modeling horizon: 6 cycles (or 240 s, 24 time intervals)

With the objective of demonstrating the properties of the model, we set the time step coarsely at 10 s. Each computer run took about 50 s on a Pentium II PC. Reducing the time interval would increase the solution time substantially due to the combinatorial nature of the problem. We are currently developing heuristics and more efficient codes to reduce the computational time, which is left to a future study.

This formulation is designed to handle the full range of traffic conditions. Three demand scenarios are set up to test its performance. In the first scenario, a light dynamic demand (720 vehicles/h/lane) is loaded to the network throughout the modeling horizon. The second scenario has a moderate demand (1080/h/lane) that lasts throughout the modeling horizon and starts with all the links at 1/4 the jam density. The third scenario represents the total gridlock situation: it has a heavy demand (1800/h/lane) that lasts for half of the modeling horizon and starts with all the links at jam density. The demands to the network are stopped after 12 time steps to see how the model dissipates the queues. The demand levels and initial conditions of the three scenarios are shown in Table 1. In general, this dynamic formulation can model any time-variant demand patterns.

Table 1
The dynamic loads for the three scenarios

Scenario	Initial network condition	Loading duration (time steps)	Demand/time interval in vehicles/time step (equivalent vehicles/h/lane)		
			Link 1	Link 4	Link 6
Light	Empty network	$1 \leqslant t \leqslant 24$	2.0 (720)	1.0 (360)	1.0 (360)
Moderate	All links at 1/4 jam density	$1 \leqslant t \leqslant 24$	3.0 (1080)	1.0 (360)	1.0 (360)
Gridlock	All links at jam density	$1 \leqslant t \leqslant 12$	5.0 (1800)	1.0 (360)	1.0 (360)
		$13 \leqslant t \leqslant 24$	0.0 (0)	0.0 (0)	0.0 (0)

Table 2
Vehicles in each cell (rounded to integer): light traffic

				Signal at A				Signal at B	
Time step	Cell(1,1)	Cell(1,2)	Cell(1,3)	A	Cell(2,1)	Cell(2,2)	Cell(2,3)	B	Cell(3,1)
1	48			XX				XX	
2	46	2		XX				XX	
3	44	2	2	XX				--	
4	42	2	4	--				XX	
5	40	2	2	--	4			XX	
6	38	2	2	XX	2	4		XX	
7	36	2	4	XX		2	4	--	
8	34	2	6	--			2	--	4
9	32	2	3	--	5			XX	6
10	30	2	2	XX	3	5		XX	6
11	28	2	4	XX		3	5	--	6
12	26	2	6	--			3	--	11
13	24	2	3	--	5			XX	14
14	22	2	2	XX	3	5		XX	14
15	20	2	4	XX		3	5	--	14
16	18	2	6	--			3	--	19
17	16	2	3	--	5			XX	22
18	14	2	2	XX	3	5		XX	22
19	12	2	4	XX		3	5	--	22
20	10	2	6	--			3	--	27
21	8	2	3	--	5			XX	30
22	6	2	2	--	3	5		XX	30
23	4	2	2	XX	2	3	5	--	30
24	2	2	4	--		2	3	--	35

"XX" represents red signal; "--" represents green.

3.2. Results

Tables 2–4 depict the number of vehicles (rounded to integer) in each cell in each time step. To facilitate discussion, cell(i,j) at time step t is referred to as $C(i,j,t)$. Columns 5 and 9 in these tables, respectively, show the signal states at intersections A and B in each time step. "XX" indicates red; "- -" indicates green. The second column shows the total demand intended to enter the network. The last column indicates the vehicles that left the network. For space limitations, we only present results for the major arterial.

The results demonstrate that the LWR results are captured in this formulation, which is important for capturing queue dynamics. For example, in Table 4, a forward starting wave is found at C(2,1,16), C(2,2,17), and C(2,3,18) after the start of green at $t = 15$. A backward shock wave is observed at C(2,3,6), C(2,2,7) and C(2,1,8) as the signal at intersection B turned red at $t = 5$. In fact, Tables 2–4 show the presence of many forward and backward kinematic waves.

For the light traffic scenario, Table 2 shows that the formulation determined a timing plan with forward progressive green time. The start of green time at intersection B is offset to 3 time steps from A which is equivalent to the travel time from A to B. The dash lines show the forward

Table 3
Vehicles in each cell (rounded to integer): moderate traffic

				Signal at A				Signal at B	
Time step	Cell(1,1)	Cell(1,2)	Cell(1,3)	A	Cell(2,1)	Cell(2,2)	Cell(2,3)	B	Cell(3,1)
1	72	4	4	XX	4	4	4	XX	
2	69	3	8	XX		4	8	--	
3	66	3	11	--			8	--	5
4	63	3	9	--	5		3	--	10
5	60	3	7	--	5	5		XX	13
6	57	3	5	XX	5	5	5	--	13
7	54	3	8	--		5	5	--	18
8	51	3	6	--	5		5	--	23
9	48	3	4	--	5	5		XX	28
10	45	3	3	XX	4	5	5	--	28
11	42	3	6	--		4	5	--	33
12	39	3	4	--	5		4	--	38
13	36	3	3	--	4	5		XX	42
14	33	3	3	XX	3	4	5	--	42
15	30	3	6	--		3	4	--	47
16	27	3	4	--	5		3	--	51
17	24	3	3	--	4	5		XX	54
18	21	3	3	XX	3	4	5	--	54
19	18	3	6	--		3	4	--	59
20	15	3	4	--	5		3	--	63
21	12	3	3	--	4	5		XX	66
22	9	3	3	XX	3	4	5	--	66
23	6	3	6	--		3	4	--	71
24	3	3	4	--	5		3	--	75

"XX" represents red signal; "--" represents green.

greenbands. This timing plan is consistent with models that work for unsaturated traffic, such as MAXBAND or PASSER.

The timing plan of the moderate traffic scenario is similar to that of the light traffic scenario (Table 3). The heavier flow on the major arterial demanded the wider greenband of 3 time steps, as shown by the dash lines.

For the gridlock scenario (Table 4), the maximum green time of 3 time steps was used in every cycle at intersection B. As long as the green time at B is fully utilized, it is not useful to maximize the green time at A. On the contrary, metering the traffic at A and hence returning green time to the minor approach will reduce its and hence the total system delay. The model recognized this and produced the above-stated timing plan. In fact, during the first three cycles, intersection A received the minimum green. This avoided the problem of defacto red (Abu-Lebdeh and Benekohal, 1997)—vehicles are blocked by the queue in front even though the signal is green. Traffic density on link 2 was gradually lowered by providing the maximum green at B and minimum green at A. When the density was sufficiently lowered to allow the saturation flow, the timing plan was switched to forward green progression. The dash lines in Table 4 show the greenbands.

Table 4
Vehicles in each cell (rounded to integer): gridlock traffic

				Signal at A				Signal at B	
Time step	Cell(1,1)	Cell(1,2)	Cell(1,3)	A	Cell(2,1)	Cell(2,2)	Cell(2,3)	B	Cell(3,1)
1	60	17	17	XX	17	17	17	XX	
2	60	17	17	XX	17	17	17	--	
3	60	17	17	--	17	17	12	--	5
4	60	17	17	XX	17	12	12	--	10
5	60	17	17	XX	12	12	12	XX	15
6	60	17	17	XX	7	12	17	--	15
7	60	17	17	--	2	17	12	--	20
8	60	17	12	XX	7	12	12	--	25
9	60	12	17	XX	2	12	12	XX	30
10	55	17	17	XX		8	17	--	30
11	55	17	17	--		8	12	--	35
12	55	17	12	XX	5	3	12	--	40
13	55	12	17	XX		5	10	XX	45
14	50	17	17	XX			15	--	45
15	50	17	17	--			10	--	50
16	50	17	12	--	5		5	--	55
17	50	12	12	--	5	5		XX	60
18	45	12	12	XX	5	5	5	--	60
19	40	12	17	--		5	5	--	65
20	35	17	12	--	5		5	--	70
21	35	12	12	--	5	5		XX	75
22	30	12	12	XX	5	5	5	--	75
23	25	12	17	--		5	5	--	80
24	20	17	12	--	5		5	--	85

"XX" represents red signal; "--" represents green.

Table 5
Vehicles in each cell (rounded to integer): Max Green

				Signal at A				Signal at B	
Time step	**Cell(1,1)**	**Cell(1,2)**	**Cell(1,3)**	**A**	**Cell(2,1)**	**Cell(2,2)**	**Cell(2,3)**	**B**	**Cell(3,1)**
1	60	17	17	XX	17	17	17	XX	
2	60	17	17	--	17	17	17	--	
3	60	17	17	--	17	17	12	--	5
4	60	17	17	--	17	12	12	--	10
5	60	17	17	XX	12	12	12	XX	15
6	60	17	17	--	7	12	17	--	15
7	60	17	12	--	7	17	12	--	20
8	60	12	12	--	12	12	12	--	25
9	55	12	12	XX	12	12	12	XX	30
10	50	12	17	--	7	12	17	--	30
11	45	17	12	--	7	17	12	--	35
12	45	12	12	--	12	12	12	--	40
13	40	12	12	XX	12	12	12	XX	45
14	35	12	17	--	7	12	17	--	45
15	30	17	12	--	7	17	12	--	50
16	30	12	12	--	12	12	12	--	55
17	25	12	12	XX	12	12	12	XX	60
18	20	12	17	--	7	12	17	--	60
19	15	17	12	--	7	17	12	--	65
20	15	12	12	--	12	12	12	--	70
21	10	12	12	XX	12	12	12	XX	75
22	5	12	17	--	7	12	17	--	75
23	0	17	12	--	7	17	12	--	80
24	0	12	12	--	12	12	12	--	85

"XX" represents red signal; "--" represents green.

Table 6
Average overall and corridor delays

Scenario	Average overall delay (s/exit vehicle)	Major corridor delay (Links: 1–2–3) (s/exit vehicle)	Minor approach delay (Link 4) (s/exit vehicle)	Minor approach delay (Link 6) (s/exit vehicle)
Light traffic	8.8	9.1	9.1	8.0
Moderate traffic	11.2	8.2	16.5	17.4
Gridlock traffic	88.0	110.6	2.5	15.0
Max green	108.8	135.3	15.0	15.0

Note: In scenarios with pre-existing traffic at $t = 1$, their prior delay is not known and assumed to be zero. This assumption will underestimate the average delay of these scenarios, which explains the lower average delay of the moderate traffic scenario as compared with the light traffic scenario. This boundary error, however, does not affect the last two scenarios since they have the same initial conditions.

For comparison purposes, we constructed the scenario called "Max Green", which simulated the effect of a common strategy used in gridlock situations—providing the maximum green at both A and B. "Max Green" has the same demand pattern and initial conditions as the gridlock scenario. To avoid the problem of defacto red, the green time of "Max Green" at A was deliberately set to start when the end of queue at cell(2,1) began to move. The results are shown in Table 5. This timing plan essentially kept link 2 at jam density for most of the time. With intersection B at the maximum green in every cycle, this timing plan could not improve the throughput of the major arterial. The provision of extra green at A, however, hurt the cross streets. To compare these scenarios, we constructed Table 6 , which shows the average overall and corridor delay per exit vehicle. The last two rows of Table 6 show the performance difference between "Max Green" and the gridlock scenario.

4. Concluding remarks

In this study, a novel traffic signal control formulation is developed through a mixed integer programming technique. The formulation considers dynamic traffic, uses dynamic traffic demand as input, and takes advantage of the newly developed traffic flow model—the cell transmission model (CTM) (Daganzo, 1994). CTM is a convergent numerical approximation to the Lighthill and Whitham (1955) and Richards (1956)—the LWR model. Because this signal control formulation embeds CTM, all the characteristics of the LWR model are inherited. As the LWR model covers the whole range of the fundamental relationship between speed, flow, and density, so does this cell-based signal control formulation.

For unsaturated or light to moderate traffic, this formulation produced a plan with progressive green time, similar to models that work for unsaturated traffic. For gridlock traffic, this new model determined a good plan to untie it. One does not need to tune or switch the model for the different traffic conditions; it "automatically" derives the best timing plan. The results seem promising. However, we note that these results are based on limited computational experiences. We plan to conduct more tests with the model to other network types and configurations.

This paper merely demonstrates the potential of this model with a small network. To make it suitable for more general networks, a number of improvements must be made. First and utmost, reducing solution time and memory requirements is an important aspect. Various heuristics are being developed at this moment. With faster solution time, one can apply the model to more complicated networks and examine whether similar conclusions can be drawn. It is also interesting to model a longer horizon to study the timing plans for transient traffic.

A second possible improvement is to add turning movements to the formulation. With CTM as a platform, it is anticipated that this improvement should be feasible without major difficulty. Ultimately, it is hoped that this new formulation will unify the earlier approaches and provide a realistic, robust platform to study traffic signal control.

Acknowledgements

This research is sponsored by the Hong Kong University of Science and Technology's Research Grant DAG97/98.EG03. The author is grateful for the helpful comments of two anonymous referees.

References

Abu-Lebdeh, G., Benekohal, R., 1997. Development of a Traffic Control and Queue Management Procedure for Oversaturated arterials. Transportation Research Board Annual Meeting Paper #970707.

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

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

D'Ans, G.C., Gazis, D.C., 1976. Optimal control of over-saturated store-and-forward transportation networks. Transportation Science 10, 1–19.

Eddelbuttel, J., Cremer, M., 1994. A new algorithm for optimal signal control in congested networks. Journal of Advanced Transportation 28 (3), 275–297.

Gazis, D.C., 1964. Optimum control of a system of over-saturated intersections. Operations Research 12, 815–831.

Gartner, N., 1985. Demand responsive traffic signal control. Transportation Research 19A, 369–373.

Gartner, N., Assmann, S., Lasaga, F., Hou, D., 1991. A multiband approach to arterial traffic signal optimization. Transportation Research 25B, 55–74.

Lighthill, M.J., Whitham, J.B., 1955. On kinematic waves. I. Flow movement in long rivers. II. A theory of traffic flow on long crowded road. Proceedings Of Royal Society A229, 281–345.

Lin, W.H., Daganzo, C.F., 1994. Technical description of NETCELL. General framework and data structure. University of California, Berkeley, Technical Note UCB-ITS-PATH-TN-95-3.

Lin, W.H., Ahanotu, D., 1995. Validating the basic cell transmission model on a single freeway link. University of California, Berkeley, Technical Note UCB-ITS-PATH-TN-95-3.

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

Lo, H., 1999. A dynamic traffic assignment formulation that encapsulates the cell-transmission model. Transportation and Traffic Theory. In press.

Michalopoulos, P., Stephanopoulos, G., 1977. Over-saturated signal systems with queue length constraints-I. Single intersection/II. Systems of intersections. Transportation Research 11, 413–428.

Newell, G.F., 1989. Theory of highway traffic signals. Course notes, University of California, Berkeley. UCB-ITS-CN-89-1.

Newell, G.F., 1991. A simplified theory of kinematic waves. Research report, University of California, Berkeley. UCB-ITS-RR-91-12.

Pignataro, L. et al., 1978. Traffic control in over-saturated street networks. NCHRP Report #194.

Quinn, D., 1992. A review of queue management strategies. Traffic Engineering and Control 33 (11), 600–605.

Rathi, A., 1988. A control scheme for high traffic density sectors. Transportation Research 22B, 81–101.

Richards, P.I., 1956. Shockwaves on the highway. Operations Research 4, 42–51.

Santiago, A., 1993. Federal highway administration's advanced traffic management systems research and development program. In: Large Urban Systems. pp. 1–14. FHWA, St Petersburg, Florida.

Shepherd, S.P., 1992. A review of traffic signal control. ITS Working Paper 349. University of Leeds, Institute for Transport Studies, UK.

Texas Transportation Institute (TTI), 1991. PASSER II-90 Microcomputer User's Guide. Texas A&M University, College Station, Texas.

United States General Accounting Office, 1994. Transportation infrastructure. Report No. GAO/RCED-94-105.

Wey, Wann-Ming, Jayakrishnan R., 1997. A network traffic signal optimization formulation with embedded platoon dispersion simulation. Transportation Research Board Annual Meeting Paper #971337.

Wood, K., 1993. Urban traffic control, systems review. Transport Research Laboratory Project Report 41.Courage, K., Wallace, C., 1991. TRANSYT/7F User's Guide. Federal Highway Administration, Washington, DC.