Contents lists available at ScienceDirect

Transportation Research Part C

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

An arterial signal optimization model for intersections experiencing queue spillback and lane blockage

Yue Liu [a],[*], Gang-Len Chang [b]

[a] *Department of Civil Engineering and Mechanics, University of Wisconsin at Milwaukee, P.O. Box 784, Milwaukee, WI 53201-0784, USA*
[b] *Department of Civil Engineering, University of Maryland at College Park, College Park, MD 20740, USA*

ARTICLE INFO

ABSTRACT

This paper presents an arterial signal optimization model that features its effectiveness on: (1) explicitly modeling physical queue evolution on arterial links by lane-group to account for shared-lane traffic interactions; and (2) capturing the dynamic interactions of spillback queues among lane groups and between neighboring intersections due to high demand, geometric constraints, or signal settings. Depending on the detected traffic patterns, one can select the control objective to be either minimizing the total travel time or maximizing the total throughput over the target area. The solution procedures developed with the Genetic Algorithm (GA) have been tested with an example arterial of four intersections under different demand scenarios. Extensive experimental analyses in comparison with results from TRANSYT-7F (version 8) reveal that the proposed model and solution method are quite promising for use in design of arterial signals, especially under congested, high demand traffic conditions.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Contending with excessive delays at signalized intersections due to either high demand or ineffective control has long been one of the imperative issues for traffic researchers and practitioners. Over the past several decades, a large body of literature has been devoted on this vital issue, and most of such models fall into the following two classes: mathematical programming approach and simulation-based approach.

In the mathematical programming approach, a set of mixed integer linear programming (MILP) formulations have been proposed in the literature (Gartner et al., 1975a,b; Little et al., 1981; Cohen and Liu, 1986; Gartner et al., 1991; Chaudhary and Messer, 1993) aiming to maximize the bandwidth or to minimize the intersection delays (Wong, 2003). Despite the contribution of those studies, most of such models have not addressed the issue of having heavy or unbalanced turning movements that may disrupt the progression bandwidth. Also, most existing models for bandwidth maximization do not account for the impact of overflow turning queue length on traffic flow thus limit their applicability during over-saturated traffic conditions.

To take into account traffic flow interactions, some researchers have proposed the simulation-based approach, in which mathematical models are developed to represent the complex interactions between traffic state evolution and key control parameters so that signal timings can be optimized based on the performance indices generated from the underlying traffic flow model. Various versions of TRANSYT (Robertson, 1969), TRANSYT-7F (Wallace et al., 1988), and extended models (Wong, 1996) are perhaps the most widely used signal timing optimization packages within this category due to its

---

* Corresponding author. Tel.: +1 414 229 3857; fax: +1 414 229 6958.
E-mail address: yueliu1980@gmail.com (Y. Liu).

relatively realistic representation of traffic flow evolution. Through proper integration with real-time data from an embedded adjustment mechanism, such models have been extended to comprise the kernel of the real-time adaptive traffic control systems that react to actual traffic conditions on-line. Examples of such systems includes SCOOT (Split, Cycle and Offset Optimization Technique) (Robertson and Bretherton, 1991), SCATS (Sydney Coordinated Adaptive Traffic System) (Lowrie, 1982), OPAC (Gartner, 1983), and RHODES (Mirchandani and Head, 2001). Other simulation-based approaches include store-and-forward models (D'Ans and Gazis, 1976; Papageorgiou, 1995), queue-and-dispersion models (Kashani and Saridis, 1983; Wu and Chang, 1999; Van den Berg et al., 2003), stochastic models (Yu and Recker, 2006), and discrete-time kinematic models (Lo et al., 2001). Despite their effectiveness in capturing the interrelations between traffic dynamics and control variables, those models have not yet addressed the critical issue of frequently incurring blockages among lane-groups due to spillback queues at congested intersections.

To contend with the impacts of over-saturated conditions on the signal design, Michalopoulos and Stephanopolos (1977, 1978) proposed the so-called "bang–bang" control model that attempts to find an optimal point for switching the timing among intersection approaches during the over-saturated period. Chang and Lin (2000) improved the Michalopoulos and Stephanopolos model into a discrete type, and further extend it to cover an over-saturated network by introducing a traffic flow propagation model between two adjacent intersections (Chang and Sun, 2004). Abu-Lebdeh and Benekohal (1997, 2003) presented a set of dynamic formulations for signal optimization to manage the formation and dissipation of queues on over-saturated network links. A discrete-time signal-coordination model was formulated and solved using Genetic Algorithms (GA) by Girianna and Benekohal (2004) to generate optimal (real-time) signal timings that distribute queues over a number of over-saturated signalized intersections and over a number of cycles. However, only limited efforts have been made by the above studies to explicitly model the dynamic spillback and blocking effects among lane groups and neighboring intersections, which are critical for designing meaningful multi-cycle signal control in congested conditions.

To contend with this critical issue, TRANSYT-7F on Release 8 (Li and Gan, 1999) has enhanced its capability by introducing a queue penalty function as the component of the disutility index if queue blocking occurs. Multiple cycle and stepwise simulation options are also employed to account for spillback effects. Abu-Lebdeh et al. (2007) recently presented models that capture traffic output of intersections in congested interrupted flow conditions with explicit consideration of interactions between traffic streams at neighboring signals. The most recent version 13 of TRANSYT by TRL (Binning et al., 2008) employs the cell transmission model in signal optimization as an alternative to the existing platoon dispersion model, which offers relatively more realistic modeling of the time-varying blocking effects between neighboring intersections in multiple cycles.

Despite the promising results from those enhanced macroscopic models, some critical issues remain to be addressed. First, most studies model the dynamic queue evolution either at a link-based level or at an individual movement-based level, which could result in either difficulty in integrating with multiple signal phases or in modeling the queue discharging rates when there exist shared lanes in the target intersection approach. Secondly, the interaction of queues among neighboring lane groups in a link due to spillback has not been explicitly modeled during congested conditions. For example, left turn traffic with insufficient left-turn pocket capacity could block the through traffic, and vice versa. Although some researchers have attempted to address this issue by developing mesoscopic or microscopic traffic-simulation-based signal optimizer (Park et al., 1999; Yun and Park, 2006; Stevanovic et al., 2007), however it would be difficult for such models to accommodate initial traffic states and explicitly model their evolution without using traffic density as a state variable. Besides, concerns are often raised regarding the computing efficiency and efforts needed to calibrate various behavioral parameters for such microscopic-simulation methods.

In response to the aforementioned concerns, this paper presents a traffic signal optimization model embedded with a set of enhanced macroscopic traffic flow formulations which can: (1) satisfy both computational efficiency and modeling accuracy; (2) represent the dynamic evolution of physical queues with respect to the signal status, arrivals, departures, and lane channelization; and (3) capture explicitly the dynamic interaction of queues among neighboring lanes and intersections experiencing spillback and blockage under congested traffic conditions.

2. The dynamic network flow formulation

In this section, this study has proposed a lane-group-based model to capture the arterial traffic dynamics. The key concept is to decompose the evolution process of traffic flow on a given arterial link into six subsets, including demand entries, upstream arrivals, joining the end of queue, merging into lane groups, departing process, and flow conservation (shown in Fig. 1). To facilitate the model presentation, the notations used hereafter are summarized in Table 1.

2.1. Demand entries

Arterial demand entries are modeled as follows:

$$IN_r[k] = \min\left[D_r[k] + \frac{w_r[k]}{\Delta t}, Q_i, \frac{s_i[k]}{\Delta t} \right] \quad (1)$$

$$w_r[k+1] = w_r[k] + \Delta t[D_r[k] - IN_r[k]] \quad (2)$$
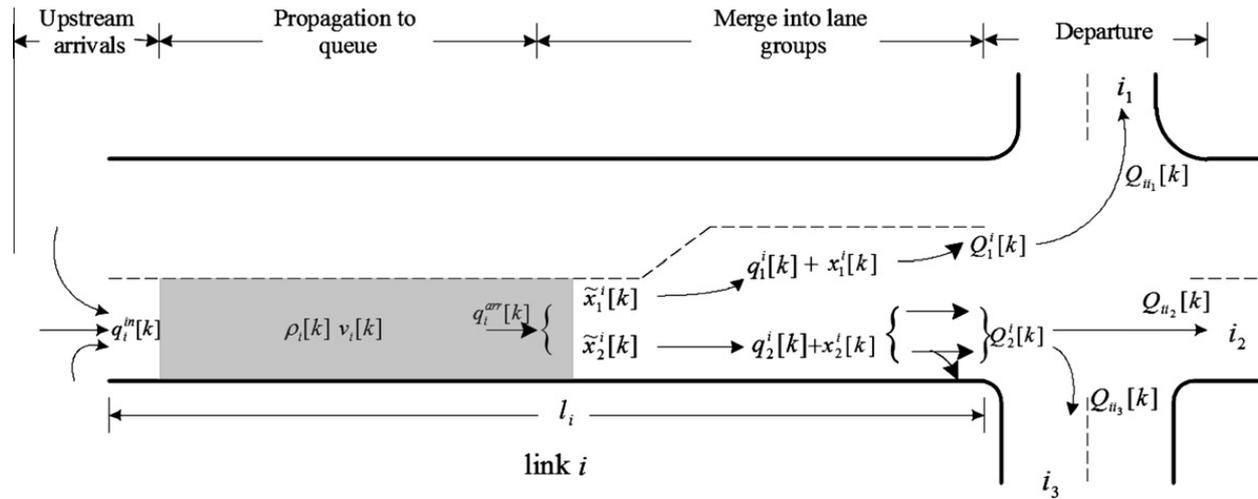
Fig. 1. Traffic flow dynamics along an arterial link.

Table 1
List of key variables used in the arterial network flow formulations.

Δt	Update interval of arterial dynamics (in seconds)
k	Time step index corresponds to time $t = k\Delta t$
T	The time horizon under consideration (in #. of Δt)
$n, n \in S_N$	Index of arterial intersections
$i, i \in S^U$	Index of arterial links
S^{OUT}	Set of outgoing boundary links in the arterial
$p, p \in P_n$	Index of signal phases at the intersection n
S_r	Set of traffic demand entries
$\Gamma(i), \Gamma^{-1}(i)$	Set of upstream and downstream links of link i
l_i, n_i, N_i, Q_i	Length (in meters), # of lanes, storage capacity (in vehs), and discharge capacity (in vph) of link i
$m, m \in S_i^M$	Index of lane groups at link i
$\delta_m^{ij}, j \in \Gamma^{-1}(i)$	A binary indicating whether the movement from link i to j uses lane group m
N_m^i, Q_m^i	Storage capacity (in vehs) and discharge capacity (in vph) for lane group m
$\Omega^i[k]$	Blocking matrix between lane groups at link i
$\omega_{m'm}^i[k] \in \Omega^i[k]$	Blocking coefficient between lane group m' and m at step k
$D_r[k], r \in S_r$	Flow rate generated at demand entry r at step k (in vph)
$IN_r[k], r \in S_r$	Flow rate entering the link from demand entry r at step k (in vph)
$w_r[k], r \in S_r$	Queue waiting on the entry r at step k (in vehs)
$q_i^{in}[k]$	Number of upstream inflow vehicles of link i at step k (in vehs)
$\gamma_{ij}[k], j \in \Gamma^{-1}(i)$	Relative turning proportion of arterial traffic from link i to j
$N_i[k]$	Number of vehicles at link i at step k (in vehs)
$q_i^{arr}[k]$	Number of vehicles arriving at end of queue of link i at step k (in vehs)
$s_i[k]$	Available space of link i at step k (in vehs)
$x_i[k]$	Number of vehicles in queue at link i at step k (in vehs)
$q_m^{i,pot}[k]$	Number of vehicles potentially to merge into lane group m of link i at step k (in vehs)
$q_m^i[k]$	Number of vehicles join the queue of lane group m at step k (in vehs)
$x_m^i[k]$	Queue length of lane group m at link i at step k (in vehs)
$\tilde{x}_m^i[k]$	Number of arrival vehicles with destination to lane group m queued outside the approach lanes due to blockage at step k (in vehs)
$\lambda_m^{ij}[k], j \in \Gamma^{-1}(i)$	Percentage of movement from link i to j in lane group m
$Q_m^i[k]$	Number of vehicles depart from lane group m at link i at step k (in vehs)
$Q_{ij}^{pot}[k]$	Number of vehicles potentially depart from link i to link j at step k (in vehs)
$Q_{ij}[k]$	Flows actually depart from link i to link j at step k (in vehs)
$g_n^p[k]$	Binary value indicating whether signal phase p of intersection n is green or not at step k

Eq. (1) indicates that the flow enters downstream link i from demand entry r depends on the demand distribution at r, existing flows queuing at r, discharge capacity of the link i, and the available space in the link i. Eq. (2) updates the queue waiting at the demand entry during each time step.

2.2. Upstream arrivals

Upstream arrival equations depict the evolution of flows arriving at the upstream of the link over time. Eqs. (3) and (4) define the upstream arrival flows for different types of links.

For internal links (with both sets of upstream and downstream links), inflows to link i can be formulated as the sum of actual departure flows from all upstream links:

$$q_i^{in}[k] = \sum_{j \in \Gamma(i)} Q_{ji}[k] \qquad (3)$$

For source links (connected with demand entry r), inflows can be stated as:

$$q_i^{in}[k] = IN_r[k] \cdot \Delta t \qquad (4)$$

2.3. Propagation to the end of queue

This set of dynamic equations represents the evolution of upstream inflows to the end of queue with the average approaching speed. The mean speed of vehicles, $v_i[k]$, depending on the density of the segment between the link upstream and the end of queue, $\rho_i[k]$, can be described with the following equation (Ben-Akiva, 1996):

$$v_i[k] = \begin{cases} v_i^{free}, & \text{if} \quad \rho_i[k] < \rho^{\min} \\ v^{\min} + \left(v_i^{free} - v^{\min}\right) \cdot \left[1 - \left(\frac{\rho_i[k] - \rho^{\min}}{\rho^{jam} - \rho^{\min}}\right)^\alpha\right]^\beta, & \rho_i[k] \in [\rho^{\min}, \rho^{jam}] \\ v^{\min}, & \text{if} \quad \rho_i[k] > \rho^{jam} \end{cases} \qquad (5)$$

where $v_i[k]$ represents the mean approaching speed of vehicles from upstream to the end of queue at link i at step k; $\rho^{\min}$ is the minimum critical density below which traffic at link i moves at the free flow speed $\left(v_i^{free}\right)$; $v^{\min}$ is the minimum traffic flow speed corresponding to the jam density (ρ^{jam}); and α, β are constant model parameters to be calibrated. The density of the segment from link upstream to the end of queue, $\rho_i[k]$, is computed with the following equation:

$$\rho_i[k] = \frac{N_i[k] - x_i[k]}{n_i \left(l_i - \frac{x_i[k]}{n_i \rho^{jam}}\right)} \qquad (6)$$

where $N_i[k] - x_i[k]$ represents the number of vehicles moving at the segment between the link upstream and the end of queue, and $l_i - x_i[k]/(n_i \cdot \rho^{jam})$ depicts the length of that segment over time. Then, the number of vehicles arriving at the end of queue at link i can be dynamically updated with:

$$q_i^{arr}[k] = \min\{\rho_i[k] \cdot v_i[k] \cdot n_i \cdot \Delta t, N_i[k] - x_i[k]\} \qquad (7)$$

where $\rho_i[k] \cdot v_i[k] \cdot n_i \cdot \Delta t$ represents the flows potentially arriving at the end of queue at time step k, which is limited by $N_i[k] - x_i[k]$.

2.4. Merging into lane groups

After vehicles arrive at the end of queue at a link, they will try to change lanes and merge into different lane groups based on their destinations. Most previous studies assume that the arriving vehicles could always merge into their destination lanes without being blocked. However, such an assumption may not be realistic under the following scenarios: (a) the intended lane group has no more space to accommodate vehicles (e.g., a fully occupied left-turn bay); and (b) the overflowed queues from other lane groups are blocking the target lane group (shown in Fig. 2). Therefore, arriving vehicles that could not merge into their destination lane group m due to either overflows or blockage will form queues outside, denoted by $\tilde{x}_m^i[k]$.

To illustrate such scenarios, it should be noted that the number of vehicles allowed to merge into lane group m at time step k depends on the available storage capacity of the lane group, given by:

$$\max\left\{N_m^i - x_m^i[k], 0\right\} \qquad (8)$$

Further, the aforementioned blocking impacts between different lane groups can be classified as complete blockage and partial blockage (shown in Fig. 2). In order to model such queue interactions between every pair of lane groups in a dynamic manner, we define a blocking matrix for each arterial link i, denoted by $\Omega^i[k]$. The dimension of the blocking matrix is $M_i \times M_i$, and M_i is the number of lane groups at link i. The matrix element, $\omega_{m'm}^i[k]$, takes a value between 0 and 1 to depict the blocking effect on lane group m due to the queue spillback at lane group m' at time step k. In this paper, we modeled $\omega_{m'm}^i[k]$ as follows:

$$\omega_{m'm}^i[k] = \begin{cases} 1 & x_{m'}^i[k] > N_{m'}^i, \ \text{complete blockage} \\ \phi_{m'm} \cdot \dfrac{q_{m'}^{i,pot}[k]}{\sum\limits_{m \in S_i^M} q_m^{i,pot}[k]} & x_{m'}^i[k] > N_{m'}^i, \ \text{partial blockage} \\ 0 & \text{no blockage or } x_{m'}^i[k] \leqslant N_{m'}^i \end{cases} \qquad (9a)$$

$$q_m^{i,pot}[k] = \tilde{x}_m^i[k] + \sum_{j \in \Gamma^{-1}(i)} q_i^{arr}[k] \cdot \gamma_{ij}[k] \cdot \delta_m^{ij} \qquad (9b)$$

(i) Left-turn lane group partially blocks the right-through lane group

(ii) Right-through lane group completely blocks the left-turn lane group
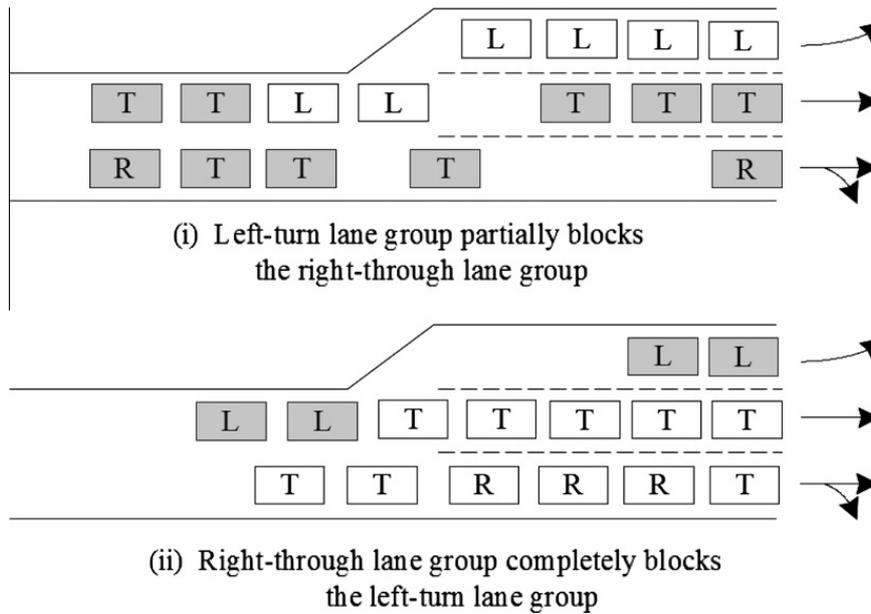
Fig. 2. Blockages between lane groups.

where $q_i^{arr}[k]$ is the total flow arriving at the end of queue of link i at time step k; $\gamma_{ij}[k]$ is the turning fraction going from link i to j; δ_m^{ij} is a binary value indicating whether traffic going from link i to j uses lane group m. Hence, one can approximate $\left[\tilde{x}_m^i[k] + \sum_{j \in \Gamma^{-1}(i)} q_i^{arr}[k] \cdot \gamma_{ij}[k] \cdot \delta_m^{ij}\right]$ as the potential level of flows that may merge into lane group m at time step k, denoted as $q_m^{i,pot}[k]$ (see Eq. (9b)).

To ensure the blocking matrix effectively discriminate the case of complete blockage from the case of partial blockage, the type of blocking impact between any given pair of lane groups is pre-determined based on the geometric features on a target intersection approach. For example in Fig. 2, the impact of left-turn lane group on the right-through lane group is easily observed to be a partial blockage, while the impact of right-through lane group on the left-turn lane group is an obvious complete blockage. Thus, at each time step, the model will be able to "understand" the blocking types among lane group queues and properly evaluate each element in the blocking matrix once a queue spillback occurs in a lane group. As shown by Eq. (9a), for the complete blockage or no blockage cases, $\omega_{m'm}^i[k]$ can be easily determined to be 1 or 0, based on the geometric features of the approach (shown in Fig. 2-ii). For the partial blockage case, $\omega_{m'm}^i[k]$ can be approximated by $\phi_{m'm} \cdot q_{m'}^{i,pot}[k] / \sum_{m \in S_i^M} q_m^{i,pot}[k]$. Where $\phi_{m'm}$ is a constant parameter between 0 and 1 that is related to driver's response to lane blockage and geometry features, and $q_{m'}^{i,pot}[k] / \sum_{m \in S_i^M} q_m^{i,pot}[k]$ approximates the fraction of merging lanes occupied by the overflowed traffic from lane group m' at time step k.

Taking the link shown in Fig. 2 as an example, there are two lane groups in the link: left-turn and right-through (named as L and R-T, respectively). Therefore, the blocking matrix is in 2×2 dimension, constructed as $\begin{bmatrix} \omega_{L,L}^i[k] & \omega_{L,R-T}^i[k] \\ \omega_{R-T,L}^i[k] & \omega_{R-T,R-T}^i[k] \end{bmatrix}$. where the elements will be updated as follows:

$$\omega_{L,L}^i[k] = \begin{cases} 1 & if \quad x_L^i[k] > N_L^i \\ 0 & if \quad x_L^i[k] \leqslant N_L^i \end{cases} \quad , L \text{ blocks itself;}$$

$$\omega_{L,R-T}^i[k] = \begin{cases} \phi_{L,R-T} \cdot \dfrac{q_L^{i,pot}[k]}{q_L^{i,pot}[k] + q_{R-T}^{i,pot}[k]}, & if \quad x_L^i[k] > N_L^i \\ 0 & if \quad x_L^i[k] \leqslant N_L^i \end{cases} \quad , L \text{ partially blocks } R-T;$$

$$\omega_{R-T,L}^i[k] = \begin{cases} 1 & if \quad x_{R-T}^i[k] > N_{R-T}^i \\ 0 & if \quad x_{R-T}^i[k] \leqslant N_{R-T}^i \end{cases} \quad , R-T \text{ completely blocks } L;$$

$$\omega_{R-T,R-T}^i[k] = \begin{cases} 1 & if \quad x_{R-T}^i[k] > N_{R-T}^i \\ 0 & if \quad x_{R-T}^i[k] \leqslant N_{R-T}^i \end{cases} \quad , R-T \text{ blocks itself;}$$

Considering the impact of blocking matrix, the number of vehicles allowed to merge into lane group m at time step k is given by:

$$\max\left\{ q_m^{i,pot}[k] \cdot \left[1 - \sum_{m' \in S_i^M} \omega_{m'm}^i[k]\right], 0 \right\} \tag{10}$$

where according to the definition of $\omega_{m'm}^i[k], 1 - \sum_{m' \in S_i^M} \omega_{m'm}^i[k]$ is the residual fraction of capacity to accommodate those vehicles merging to lane group m.

Finally, the number of vehicles allowed to merge into lane group m at time step k should be the minimum value of Eqs. (8) and (10), given by:

$$q_m^i[k] = \min\left\{\max\left\{N_m^i - x_m^i[k], 0\right\}, \max\left\{q_m^{i,pot}[k] \cdot \left[1 - \sum_{m' \in S_i^M} \omega_{m'm}^i[k]\right], 0\right\}\right\} \qquad (11)$$

2.5. Departing process

The number of vehicles potentially departing from link i to link j at time step k is given by:

$$Q_{ij}^{pot}[k] = \sum_{m \in S_i^M} \min\left\{q_m^i[k] + x_m^i[k], Q_m^i \cdot g_n^p[k]\right\} \cdot \lambda_m^{ij}[k] \qquad (12)$$

$$\lambda_m^{ij}[k] = \frac{\delta_m^{ij} \cdot \gamma_{ij}[k]}{\sum_{j \in \Gamma^{-1}(i)} \delta_m^{ij} \cdot \gamma_{ij}[k]} \qquad (13)$$

where $\min\{q_m^i[k] + x_m^i[k], Q_m^i \cdot g_n^p[k]\}$ depicts the potential departing flows from lane group m at time step k; $\lambda_m^{ij}[k]$ is the percentage of traffic in lane group m going from link i to j. Therefore, $\min\{q_m^i[k] + x_m^i[k], Q_m^i \cdot g_m^p[k]\} \cdot \lambda_m^{ij}[k]$ is the flow potentially departing from link i to j in lane group m. Then its summation over all lane groups in link i comes to Eq. (12). Assuming that a total of one unit flow is to depart from link i at time step k, $\delta_m^{ij} \cdot \gamma_{ij}[k]$ will be the amount of flows going to link j from lane group m within that one unit, and $\sum_{j \in \Gamma^{-1}(i)} \delta_m^{ij} \cdot \gamma_{ij}[k]$ will be the total amount of flows departing from lane group m. Hence, we have Eq. (13) holds.

Note that the actual number of vehicles departing from link i to link j at time step k is also constrained by the available storage space at destination link j. Since the total flow towards one destination link j may consist of several flows from different upstream links, this study assumes that the free storage space at downstream link j allocated to accommodate the departing flow from upstream link i is proportional to link i's potential departing flow to link j. Therefore, the actual departing flows from link i to link j at time step k is given by the following equation:

$$Q_{ij}[k] = \min\left\{Q_{ij}^{pot}[k], \frac{Q_{ij}^{pot}[k]}{\sum_{i \in \Gamma(j)} Q_{ij}^{pot}[k]} \cdot s_j[k]\right\} \qquad (14)$$

where $s_j[k]$ is the available space in link j at time step k, and $Q_{ij}^{pot}[k]/\sum_{i \in \Gamma(j)} Q_{ij}^{pot}[k]$ is the proportion of the available space in link j allocated to accommodate flows from link i.

Then, the actual departing flow from lane group m at link i can be obtained by:

$$Q_m^i[k] = \sum_{j \in \Gamma^{-1}(i)} Q_{ij}[k] \cdot \delta_m^{ij} \qquad (15)$$

2.6. Flow conservation

The lane group based queues are advanced as follows:

$$x_m^i[k+1] = x_m^i[k] + q_m^i[k] - Q_m^i[k] \qquad (16)$$

Queues outside the approach lanes due to overflows or blockages are advanced as follows:

$$\tilde{x}_m^i[k+1] = \tilde{x}_m^i[k] - q_m^i[k] + \sum_{j \in \Gamma^{-1}(i)} q_i^{arr}[k] \cdot \gamma_{ij}[k] \cdot \delta_m^{ij} \qquad (17)$$

Then, the total number vehicles queued at link i is computed as:

$$x_i[k+1] = \sum_{m \in S_i^M} (x_m^i[k+1] + \tilde{x}_m^i[k+1]) \qquad (18)$$

The evolution of total number of vehicles present at link i can be stated as:

$$N_i[k+1] = N_i[k] + \sum_{j \in \Gamma(i)} Q_{ji}[k] - \sum_{j \in \Gamma^{-1}(i)} Q_{ij}[k] \qquad (19)$$

Finally, one can compute the available storage space of link i as follows:

$$s_i[k+1] = N_i - N_i[k+1] \qquad (20)$$

Note that to properly compute all the traffic state variables in the proposed model, one should pay attention to the order in which those traffic flow formulations are determined. The order for implementing those formulations can be divided into two phases: the "update" phase and the "advance" phase. The "update" phase calculates the traffic state variables at the current time step k (Eqs. (1)–(15)), then the "advance" phase moves forward the "clock" into the next time step $k + 1$ and update the traffic flow status based on the flow conservation (Eqs. (16)–(20)).

3. The signal optimization model

With the above enhanced network flow formulations, we can construct the following model to optimize signal timings along the target arterial, including the cycle length, offsets of the coordinated phases, and splits for each intersection. In this study, we assume that the set of intersections along the target arterial share the same cycle length, and the phase sequence for each intersection is preset.

3.1. Objective function

Given the time period T of analysis, Eq. (21) represents the primary objective of the control model for minimizing the total time spent by all vehicles in the control area (including vehicles already entering the network and vehicles waiting in the virtual queue at the demand entry). The control model also aims at maximizing the total throughput, i.e., the total number of vehicles that can go through the control area under over-saturated conditions. Since the throughput equals the total number of vehicles entering the outgoing links, one can also state the control objective as Eq. (22).

$$\min \quad \sum_{k=1}^{T} \left[\sum_{i \in S^U} N_i[k] + \sum_{r \in S_r} w_r[k] \right] \cdot \Delta t \tag{21}$$

$$\max \quad \sum_{k=1}^{T} \sum_{i \in S^{OUT}} q_i^{in}[k] \tag{22}$$

3.2. Control parameters and decision variables

C_{min}, C_{max}	minimal and maximal cycle length;
G_{np}^{min}	minimal green time for phase p of intersection n;
C	common cycle length of the target arterial for the given period T;
Δ_n	offset of intersection n for the given period T;
NP_n	number of phases of intersection n;
G_{np}	green time for phase p of intersection n for the given period T;
I_{np}	inter-green time for phase p of intersection n;

3.3. Constraints

Eqs. (1)–(7), (11)–(20), representing the dynamic traffic state evolution along the arterial, are the principal constraints for the control model. Moreover, the following constraints are common restrictions for the signal control parameters:

$$C \geqslant C_{min}, \quad C \leqslant C_{max} \tag{23}$$

$$G_{np} \geqslant G_{np}^{min}, \quad G_{np} < C, \quad n \in S_N \tag{24}$$

$$\sum_{p \in P_n} G_{np} + \sum_{p \in P_n} I_{np} = C, \quad n \in S_N \tag{25}$$

$$\Delta_n \geqslant 0, \quad \Delta_n < C, \quad n \in S_N \tag{26}$$

Eq. (23) restricts the common cycle length to be between the minimal and maximal values. Eq. (24) requires that the green time for each phase should satisfy the minimal green time, but not exceed the cycle length. And, the sum of green times and inter-greens for all phases at intersection n should be equal to the cycle length illustrated by Eq. (25). Moreover, the offset of intersection n is constrained by Eq. (26) to lie between 0 and the cycle length.

Fig. 3 shows an example of phase diagram employed in this paper to describe the split information. To represent the signal status of phase p at each time step k, the binary variable $g_n^p[k]$ is defined to indicate whether or not the corresponding phase p is green. It shall include the following relations between the phase status and signal control parameters:

$$g_n^p[k] = \begin{cases} 1 & if \quad \sum_{j=1}^{p-1}(G_{nj} + I_{nj}) < \mod(k - \Delta_n, C) \leqslant \sum_{j=1}^{p-1}(G_{nj} + I_{nj}) + G_{np} \quad p \in P_n, n \in S_N \\ 0 & o.w. \end{cases} \tag{27}$$

Fig. 3. A signal controller with a set of phases P_n.

To facilitate the model formulation, we further introduce two sets of auxiliary 0–1 variables:

$$\delta_n'^p[k] = \begin{cases} 1 & if \quad \mod(k - \Delta_n, C) \leqslant \sum_{j=1}^{p-1}(G_{nj} + I_{nj}) \\ 0 & o.w. \end{cases} \quad p \in P_n, n \in S_N \tag{28}$$

$$\delta_n''^p[k] = \begin{cases} 1 & if \quad \mod(k - \Delta_n, C) > \sum_{j=1}^{p-1}(G_{nj} + I_{nj}) + G_{np} \\ 0 & o.w. \end{cases} \quad p \in P_n, n \in S_N \tag{29}$$

Then Eq. (27) can be converted into the following constraints:

$$(\delta_n'^p[k] - 0.5) \cdot \mod(k - \Delta_n, C) \leqslant (\delta_n'^p[k] - 0.5) \cdot \sum_{j=1}^{p-1}(G_{nj} + I_{nj}), \quad p \in P_n, n \in S_N \tag{30}$$

$$(\delta_n''^p[k] - 0.5) \cdot \mod(k - \Delta_n, C) > (\delta_n''^p[k] - 0.5) \cdot \left(\sum_{j=1}^{p-1}(G_{nj} + I_{nj}) + G_{np}\right), \quad p \in P_n, n \in S_N \tag{31}$$

$$g_n^p[k] + \delta_n'^p[k] + \delta_n''^p[k] = 1, \quad p \in P_n, n \in S_N \tag{32}$$

To provide a realistic range for the optimized solution, the proposed model also includes nonnegative constraints for control parameters. In summary, the optimal control model with respect to C, $\{G_{np}\}$, and $\{\Delta_n\}$ can be recapitulated below:

$$\min \quad \sum_{k=1}^{T} \left[\sum_{i \in S^U} N_i[k] + \sum_{r \in S_r} w_r[k] \right] \cdot \Delta t, \text{ or}$$

$$\max \quad \sum_{k=1}^{T} \sum_{i \in S^{OUT}} q_i^{in}[k] \tag{33}$$

s.t. Eqs. (1)–(7), (11)–(20), (23)–(26), (and), (31), (32)

4. Solution algorithm

The proposed optimization model consists of complex formulations, including binary parameters as well as non-linear system constraints. It is thus difficult to find the global optimal solution through traditional non-linear programming approaches. This study explores a Genetic Algorithm (GA)-based heuristic to yield efficient model solutions for signal settings.

To generate control parameters that satisfy the traffic signal optimization constraints, the following decoding scheme is developed based on the phase structure as shown in Fig. 3. A total number of $NP_n + 1$ fractions ($\lambda_j, j = 1 \dots NP_n + 1$) are generated for the controller at intersection n from decomposed binary strings by converting the binary string to a decimal number and dividing the number by the maximum possible decimal number represented by the binary string. The $NP_n + 1$ fractions are used to code the green times, cycle length, and offsets as shown by the following equations:

$$G_{np} = G_{np}^{\min} + \left(C - \sum_{j \in P_n} G_{nj}^{\min} - \sum_{j \in P_n} I_{nj}\right) \cdot \lambda_p \cdot \prod_{j=1}^{p}(1 - \lambda_{j-1}), \quad p = 1 \dots NP_n - 1, n \in S_N \tag{34}$$

$$G_{np} = G_{np}^{\min} + \left(C - \sum_{j \in P_n} G_{nj}^{\min} - \sum_{j \in P_n} I_{nj}\right) \cdot \prod_{j=1}^{p}(1 - \lambda_{j-1}), \quad p = NP_n, n \in S_N \tag{35}$$

$$C = C_{min} + (C_{max} - C_{min}) \cdot \lambda_{NP} \tag{36}$$

$$\Delta_n = (C - 1) \cdot \lambda_{NP+1} \tag{37}$$

Eq. (36) constrains the random cycle lengths generated through the binary string within the maximum and minimum allowable cycle lengths. Using the cycle length generated above, Eq. (37) would result in an offset value that lies between 0 and the cycle length minus one. The green times are assigned to each phase within a feasible range with Eqs. (34) and (35), in which λ_0 was set to zero to accommodate the case when $j = 1$.

The first population of GA is generated randomly and each individual is decoded to a set of signal timing plans based on the aforementioned scheme. Then, the proposed network flow model will compute the objective value for the given analysis period T. The corresponding fitness measure can be obtained from the objective function value. Based on the fitness evaluation, the crossover and mutation procedure are processed. This procedure will continue until the stop criterion is satisfied. The solution algorithm will first maximize the total system throughput, and then switch to minimize the total time spent in the system if the arterial network is in normal, under-saturated conditions (see Fig. 4) due to the fact that two under-saturated traffic scenarios may have quite different total time spent in the network even though they have approximately the same level of total throughput.

5. Numerical tests

5.1. Experiment design

To illustrate the applicability of the proposed model, this study employs an example arterial consisting of four intersections, similar to the one used by Park et al. (1999), for numerical tests. Basic layouts of the arterial and phase configurations are given in Fig. 5. The spacing between intersections in the arterial is set to be 121.92 m (400 ft). To test the capability of the proposed model with respect to capturing blockages between different lane-groups under over-saturated conditions, all links along the arterial direction are designed to have one full lane for right-through traffic and an exclusive left-turn pocket of 30.48 m (100 ft).

The turning fractions for all intersection approaches are set to be 30% left-turn, 60% through, and 10% right-turn. This numerical test includes 10 demand entries (A–J) and three volume levels (low, medium, and high) designed to test the performance of proposed control model. Table 2 summarizes all experimental scenarios:
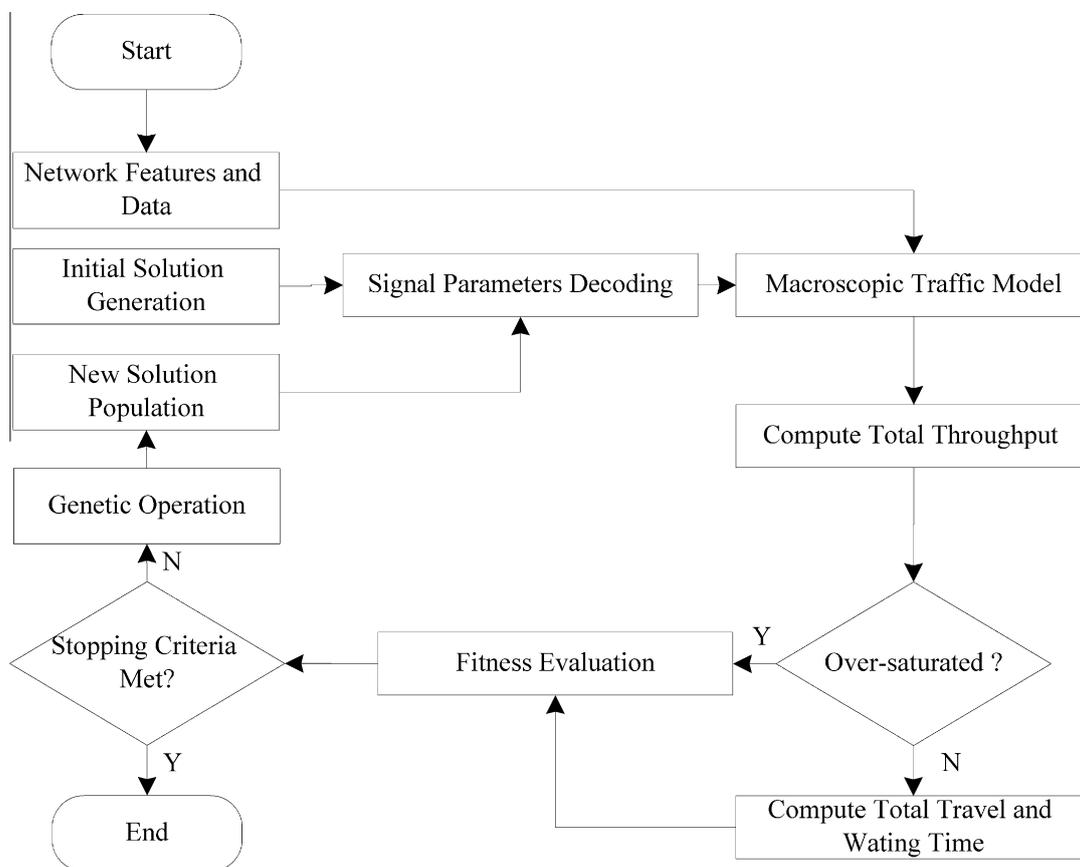
Fig. 4. Flowchart of the solution algorithm for signal optimization.
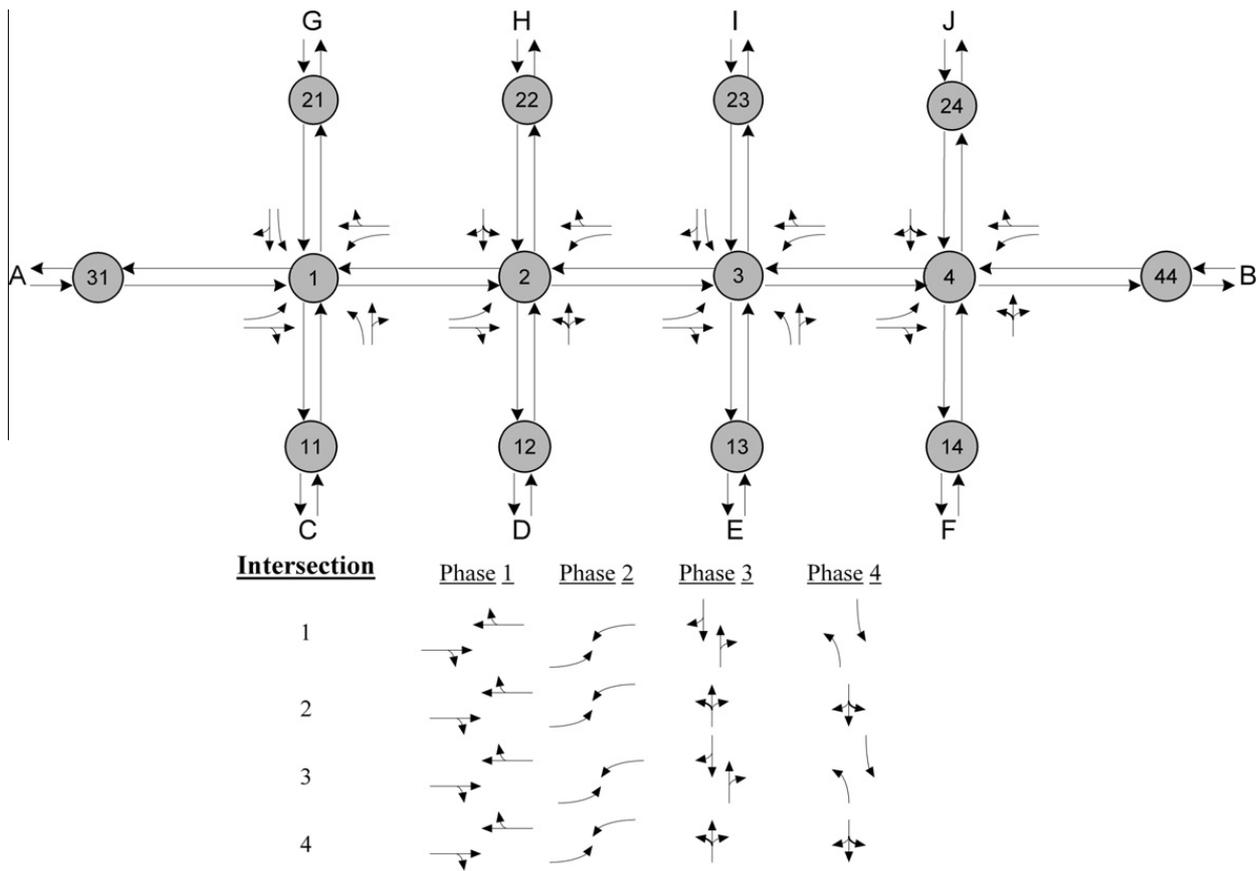
Fig. 5. Experimental arterial layout and phase settings.

Table 2
Experimental scenarios for model evaluation.

Demand scenario	Degree of saturation	Demand entries (in vph)									
		A	B	C	D	E	F	G	H	I	J
Low	0.4	400	400	300	200	300	200	300	200	300	200
Medium	0.8	600	600	450	300	450	300	450	300	450	300
High	1.2	1000	1000	750	500	750	500	750	500	750	500

5.2. Optimization model settings

The network flow model parameters are given below:

- The free flow speeds are set to be 64.4 km/h (40 mph) for arterial and 48.3 km/h (30 mph) for side streets, and the minimum density is set to be 12.4 veh/km/lane (20 veh/mile/lane);
- Jam density is set to be 130.4 veh/km/lane (210 veh/mile/lane), and the minimum speed is 8.05 km/h (5 mph);
- An average vehicle length of 7.62 m (25 ft) is used to calculate the storage capacity of left-turn bay; and
- Only the case of complete blockage exists (no need to calibrate $\phi_{m'm}$ for partial blockage) between the right-through and left-turn lane-groups since there is only one full lane in the link. However, for other geometric configurations of intersection approaches, $\phi_{m'm}$ needs to be calibrated before the implementation of the proposed model. A detailed calibration process can be found in Liu et al. (2008).

The signal timing constraints are given as:

- Common network cycle length is between C_{min}, C_{max} = 48 s, 150 s;
- Minimal green time $G_{np}^{min} = 7$ s; and
- Inter-green time $I_{np} = 5$ s.

The GA optimization is performed with the following parameters:

- The population size is 30;
- The maximum number of generation is 200;
- The crossover probability is 0.5; and
- The mutation probability is 0.03.

5.3. Experimental results and analysis

The proposed model was coded in C++ and tested on an AMD Turion 2.2 GHz processor and 2.0 G RAM, running under Windows. The GA-based heuristic was implemented with the MIT GA C++ Library v.2.4.6 (Wall, 1999). For the case study network, it takes around 0.5–1 min for the GA to converge to a good solution to design signal plans in a 1-h period under three designed levels of traffic demand. For a large-scale network, however, one can integrate the proposed model with the rolling time horizon scheme to achieve the acceptable computing efficiency.

In this section, the optimized signal plans obtained from the proposed model will be compared with plans from TRANSYT-7F using CORSIM as an unbiased evaluator. With TRANSYT-7F, the phase settings shown in Fig. 5 were set as the input, and the network cycle length was optimized over a range of 48–150 s. Stepwise simulation options with default disutility and performance indices are selected. Default run-control parameters were used for the optimization process, and network parameter values were set to reflect the experimental arterial features. Since the hill-climbing algorithm in TRANSYT (version 8) does not necessarily reach a global optimum, this study has specified different optimization node sequences within the input file to avoid the local optimal solution. The best signal timing plans obtained from this process were selected as the final candidate for comparison. Since TRANSYT-7F release 8 can model the turn-bay spillover effects, left-turn bays on arterial links were coded in the Record Type 291 (Link Data Further Continuation) in order to obtain a fair comparison between the proposed model and TRANSYT-7F in terms of capturing queue interaction. For the high-demand scenario (degree of saturation above 1.0), the spillback penalty functions were used to accommodate queue blocking or spillback effects. The optimization processes for both TRANSYT-7F and the proposed optimization model were implemented for one hour with a 5-min initialization interval. The optimized signal timing plans were then input into CORSIM for comparison.

To overcome the stochastic nature of a microscopic simulation system, an average of 20 simulation runs has been used. For the MOE comparison, since CORSIM calculates total delays or average delays only for departed vehicles, it is not computationally convenient to use delay as the MOE for over-saturated conditions. Hence, in this study we use total delay as the MOE for under-saturated conditions, and employ throughput and total queue time for over-saturated conditions, as suggested by previous studies (Park et al., 1999).

Tables 3 and 4 show the optimization and comparison results from the proposed model and TRANSYT-7F under different demand levels defined in Table 2.

As indicated in Table 3, one can reach the following findings:

- For the low- and medium-demand scenarios, a shorter cycle length was obtained from the proposed model with the objective of minimizing the total time spent in the network. Therefore, the proposed model outperforms TRANSYT-7F in terms of total system delay and queue times, but yields less system throughput (shown in Table 4) due to the relatively larger percentage of lost time in the cycle length.

Table 3
Signal timings from the proposed model and TRANSYT-7F.

Demand Scenarios	Intersection	Cycle length (s)		Offset (s)		Start of green (s)							
						Phase I		Phase II		Phase III		Phase IV	
		I[a]	II[b]	I	II	I	II	I	II	I	II	I	II
Low	1	52	65	17	7	0	0	16	22	28	35	40	53
	2			11	8	0	0	13	21	27	33	39	49
	3			18	3	0	0	13	23	25	35	39	53
	4			9	2	0	0	15	21	27	33	39	49
Medium	1	70	80	10	6	0	0	21	30	33	44	58	67
	2			5	6	0	0	24	27	36	40	53	60
	3			9	3	0	0	22	29	34	43	55	66
	4			4	0	0	0	21	25	34	40	53	60
High	1	93	111	7	25	0	0	24	42	45	61	77	94
	2			2	0	0	0	26	38	43	55	72	83
	3			6	24	0	0	26	40	45	58	77	93
	4			1	1	0	0	29	36	45	55	72	83

[a] The proposed dynamic model.
[b] TRANSYT-7F.

Table 4
Comparison of CORSIM simulation results.

Scenarios	MOEs	Simulation results from CORSIM (1 h)		
		Proposed model	TRANSYT-7F	Improvement[a] (%)
Low-demand	Total delay (veh-min)	1728.6	1794.6	−3.7
	Total queue time (veh-min)	1327	1410.4	−5.9
	Total throughput (veh)	2798	2807	−0.3
Medium-demand	Total delay (veh-min)	3307.2	3358.2	−1.5
	Total queue time (veh-min)	2607.8	2680.7	−2.7
	Total throughput (veh)	4206	4219	−0.3
High-demand	Total queue time (veh-min)	10625.9	13089.6	−18.8
	Total throughput (veh)	5737	5574	+2.9

[a] Improvement is calculated by $(MOE_{proposedmodel} - MOE_{TRANSYT-7F})/MOE_{TRANSYT-7F}$.

● For the high-demand scenario, severe blockages between lane groups (right-through and left-turn) and upstream-downstream links in the network can be observed from CORSIM simulation animations. Even though TRANSYT-7F tries to select longer cycle lengths to maximize the phase capacity for this scenario, it may adversely increase the chance of blockages due to the higher arrival rates to downstream links. In contrast, the proposed approach can model those blockage effects explicitly and dynamically, and then select the most suitable cycle length to accommodate it. As shown in Table 4, the proposed model yields less queue time and larger throughput than TRANSYT-7F under the high-demand scenario.

To investigate the performance of the offset and split settings generated by the proposed model, this study has compared the queue time for each intersection approach under different demand scenarios, as shown in Fig. 6 through Fig. 8. Under the low- and medium-demand scenarios (see Figs. 6 and 7), the proposed model can achieve the arterial progression performance comparable to TRANSYT-7F, as most of the Eastbound and Westbound approaches (arterial direction) produced less queue time. For the side streets, the proposed model outperforms TRANSYT-7F since all Northbound and Southbound approaches produced less queue time. This is due to the fact that the proposed model minimizes not only the total system travel time but also the total system queue time, which results in the reduction of side street queue time without significantly impacting the arterial progression.

Under the high-demand scenario (see Fig. 8), the proposed model provides better arterial progression than does TRANSYT-7F (all Eastbound and Westbound approaches produce less queue time) due to its embedded dynamic traffic flow equations which are capable of handling blocking effects under over-saturated conditions. For side streets, the proposed model's performance is also comparable to that of TRANSYT-7F because the control objective is to maximize the total system throughput, which may cause longer waiting time on the side streets.
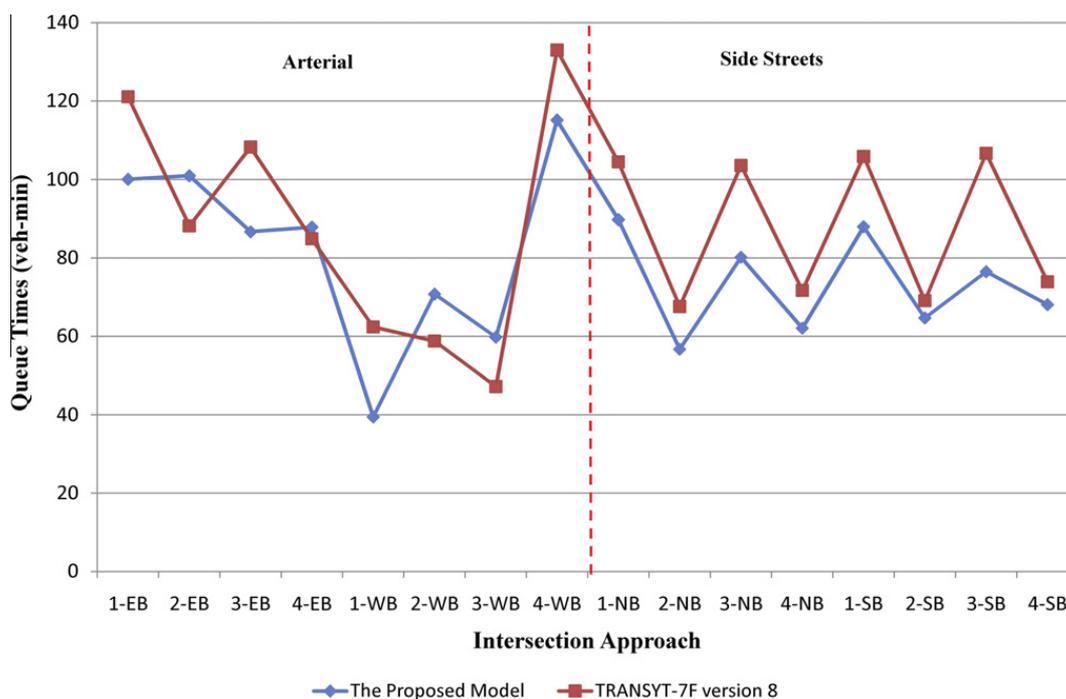
Fig. 6. Queue times of approaches under the low-demand scenario.
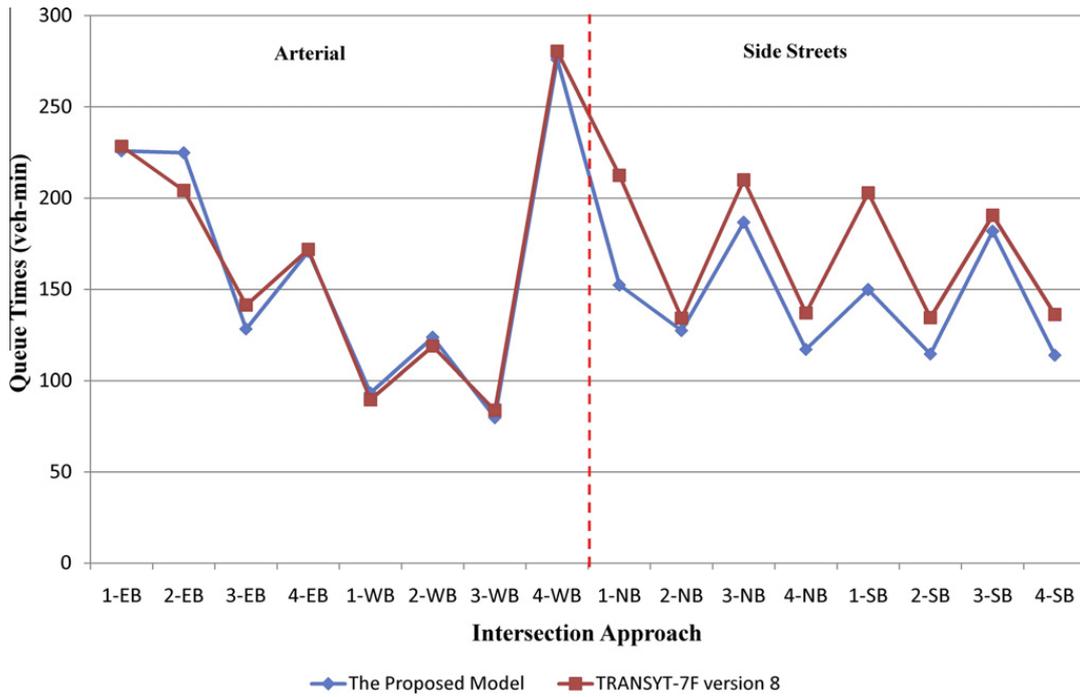
Fig. 7. Queue times of approaches under the medium-demand scenario.
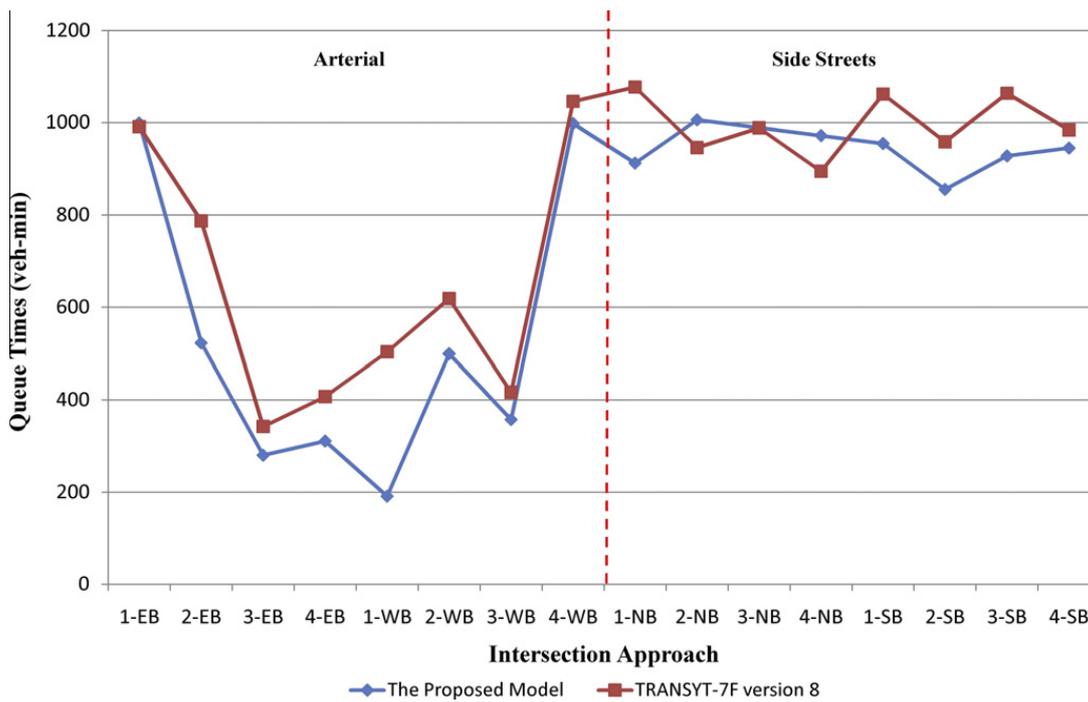
Fig. 8. Queue times of approaches under the high-demand scenario.

In summary, one can reach the following conclusions by comparing the performance measures of the proposed model and TRANSYT-7F under different demand scenarios.

- The proposed model outperforms TRANSYT-7F in terms of total system queue time for all experimental demand scenarios.
- For under-saturated traffic conditions (low- and medium-demand scenarios), the proposed model can produce better signal timings than TRANSYT-7F with respect to total system delay and total system queue time. In particular, the proposed model not only obtain a comparable arterial progression performance with TRANSYT-7F, but also effectively mitigate the congestion at the side streets, which is evidenced by the lower queue time at the Northbound and Southbound approaches;

- For over-saturated traffic conditions, in terms of the total system queue time and total system throughput, the proposed model can mitigate the congestion and blockage more effectively than TRANSYT-7F due to the use of enhanced dynamic traffic flow equations. In addition, compared to TRANSYT-7F, the proposed model does not incur excessive waiting time or queues on the side streets.

6. Conclusions

This study has presented an optimization model for design of arterial signal timings. To capture the critical operational issues at signalized intersections as well as to ensure computing efficiency, this study has proposed a set of enhanced macroscopic traffic flow equations which can precisely model the traffic evolution along the arterial link, especially the blocking effects between different lane groups under over-saturated conditions. Aiming at maximizing the efficiency of the arterial with the limited roadway capacity, the proposed model can adapt to different traffic demand patterns with the most suitable control objectives. The numerical results have confirmed the effectiveness of the proposed model in comparison with TRANSYT-7F for design of arterial signal timings under different traffic demand patterns.

Note that this paper has presented preliminary evaluation results for the proposed model. More extensive numerical experiments or field tests will be conducted to assess the effectiveness of the proposed model under various demand patterns, turning proportions, and geometry configurations. Another possible extension to this study is to integrate phasing sequence optimization. With the proposed model's capability to capture intersection approach queue interactions, it is expected that the impact of phasing sequence on signal control performance will be better captured and factored in the control objective function for proposed model to further prevent the occurrence of blockages and spillbacks.

References

Abu-Lebdeh, G., Benekohal, R.F., 1997. Development of a traffic and queue management procedure for oversaturated arterials. Transportation Research Record 1603, 119–127.

Abu-Lebdeh, G., Benekohal, R.F., 2003. Design and evaluation of dynamic traffic management strategies for congested conditions. Transportation Research Part A 37, 109–127.

Abu-Lebdeh, G., Chen, H., Benekohal, R.F., 2007. Modeling traffic output for design of dynamic multi-cycle control in congested conditions. Journal of Intelligent Transportation Systems 11 (1), 25–40.

Ben-Akiva, M., 1996. Development of a Deployable Real-time Dynamic Traffic Assignment System, Task D Interim Report: Analytical Developments for DTA System. MIT, Cambridge, MA.

Binning, J.C., Burtenshaw, G., Crabtree, M., 2008. TRANSYT 13 User Guide. Transport Research Laboratory, UK.

Chang, T.-H., Lin, J.-T., 2000. Optimal signal timing for an oversaturated intersection. Transportation Research Part B 34, 471–491.

Chang, T.-H., Sun, G.-Y., 2004. Modeling and optimization of an oversaturated signalized network. Transportation Research Part B 38, 687–707.

Chaudhary, N.A., Messer, C.J., 1993. PASSER-IV: a program for optimizing signal timing in grid networks. In: 72nd Annual Meeting of the Transportation Research Board, Washington, DC.

Cohen, S.L., Liu, C.C., 1986. The bandwidth-constrained TRANSYT signal optimization program. Transportation Research Record 1057, 1–7.

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

Gartner, N.H., Little, J.D.C., Gabbay, H., 1975a. Optimization of traffic signal settings by mixed integer linear programming. Part I: the network coordination problem. Transportation Science 9, 321–343.

Gartner, N.H., Little, J.D.C., Gabbay, H., 1975b. Optimization of traffic signal settings by mixed integer linear programming. Part II: the network synchronization problem. Transportation Science 9, 321–343.

Gartner, N., 1983. OPAC: a demand-responsive strategy for traffic signal control. Transportation Research Record 906, 75–81.

Gartner, N.H., Assmann, S.F., Lasaga, F.L., Hou, D.L., 1991. A multi-band approach to arterial traffic signal optimization. Transportation Research Part B 25 (1), 55–74.

Girianna, M., Benekohal, R.F., 2004. Using genetic algorithms to design signal coordination for oversaturated networks. Journal of Intelligent Transportation Systems 8 (2), 117–129.

Kashani, H.R., Saridis, G.N., 1983. Intelligent control for urban traffic systems. Automatica 19, 191–197.

Li, M.-T., Gan, A.C., 1999. Signal timing optimization for oversaturated networks using TRANSYT-7F. Transportation Research Record 1683, 118–126.

Little, J.D.C., Kelson, M.D., Gartner, N.H., 1981. MAXBAND: a program for setting signals on arterials and triangular networks. Transportation Research Record 795, 40–46.

Liu, Y., Yu, J., Chang, G.L., Rahwanji, S., 2008. A lane-group based macroscopic model for signalized intersections account for shared lanes and blockages. In: Proceeding of 11th IEEE Conference on Intelligent Transportation Systems, pp. 639–644.

Lo, H., Chang, E., Chan, Y., 2001. Dynamic network traffic control. Transportation Research Part A 35, 721–744.

Lowrie, P., 1982. The Sydney coordinated adaptive control system – principles, methodology, algorithms. IEE Conference Publication, p. 207.

Michalopoulos, P.G., Stephanopolos, G., 1977. Oversaturated signal system with queue length constraints – I. Transportation Research 11, 413–421.

Michalopoulos, P.G., Stephanopolos, G., 1978. Optimal control of oversaturated intersections: theoretical and practical considerations. Traffic Engineering and Control, 216–221.

Mirchandani, P., Head, L., 2001. A real-time traffic signal control system: architecture, algorithms, and analysis. Transportation Research Part C 9, 415–432.

Papageorgiou, M., 1995. An integrated control approach for traffic corridors. Transportation Research Part C 3, 19–30.

Park, B., Messer, C.J., Urbanik, T., 1999. Traffic signal optimization program for oversaturated conditions: genetic algorithm approach. Transportation Research Record 1683, 133–142.

Robertson, D.I., 1969. TRANSYT: a traffic network study tool. RRL Report LR 253, Road Research Laboratory, England.

Robertson, D.I., Bretherton, R.D., 1991. Optimizing networks of traffic signals in real-time: the SCOOT method. IEEE Transactions on Vehicular Technology 40, 11–15.

Stevanovic, A., Martin, P.T., Stevanovic, J., 2007. VisSim-based genetic algorithm optimization of signal timings. Transportation Research Record 2035, 59–68.

Van den Berg, A., Hegyi, B., De Schutter, Hellendoorn, J., 2003. A macroscopic traffic flow model for integrated control of freeway and urban traffic networks. In: 42nd IEEE International Conference on Decision and Control, Hawaii, pp. 2774–2779.

Wall, M., 1999. GAlib, a C++ library of genetic algorithm components. Massachusetts Institute of Technology. <http://lancet.mit.edu/ga/> (retrieved on 04.16.2009).

Wallace, C.E., Courage, K.G., Reaves, D.P., Shoene, G.W., Euler, G.W., Wilbur, A., 1988. TRANSYT-7F User's Manual: Technical Report. University of Florida, Gainesville, FL.

Wong, S.C., 1996. Group-based optimisation of signal timings using the TRANSYT traffic model. Transportation Research Part B 30 (3), 217–244.

Wong, S.C., 2003. A lane-based optimization method for minimizing delay at isolated signal-controlled junctions. Journal of Mathematical Modeling and Algorithms 2, 379–406.

Wu, J., Chang, G.L., 1999. An integrated optimal control and algorithm for commuting corridors. International Transactions on Operations Research 6, 39–55.

Yu, X.H., Recker, W.W., 2006. Stochastic adaptive control model for traffic signal systems. Transportation Research Part C 14, 263–282.

Yun, I., Park, B., 2006. Application of stochastic optimization method for an urban corridor. Proceedings of the Winter Simulation Conference 3–6, 1493–1499.