Publish Information: Liu, H., K. N. Balke, and W. Lin (2008). "A Reverse Causal-effect Modeling Approach for Real Time Signal Control of An Oversaturated Intersection." *Transportation Research, Part C*, **16**(6), 742-754.

A reverse causal-effect modeling approach for signal control of an oversaturated intersection

Hongchao Liu[1], Kevin N. Balke[2], and Wei-Hua Lin[3]

[1]Assistant Professor, Dept. of Civil & Environmental Engineering, Texas Tech Univ., Box 41023, Lubbock, TX 79409. E-mail: hongchao.liu@ttu.edu)
[2]Research Engineer, Texas Transportation Institute, The Texas A&M Univ. System, 3135 TAMU College Station, TX 77843. E-mail: k-balke@tamu.edu
[3]Associate Professor, Dept. of Systems and Industrial Engineering, Univ. of Arizona, Tucson, AZ 85721. E-mail: weilin@sie.arizona.edu

Abstract

A novel approach is presented in which signalized intersections are treated as normal highway bottlenecks for improved computational efficiency. It is unique in two ways. First, it treats the signalized intersections as common freeway bottlenecks by a reversed cause and effect modeling approach. Both traffic arrivals and departures are modeled by smooth continuous functions of time as if there were no interruptions to traffic flows from signals. The use of smooth continuous functions for departure curves instead of commonly used step functions makes it easy to apply differential calculus in optimization and future extension to a system of intersections. Second, a dynamic linear programming (LP) model is then developed to maximize the total vehicular output from the intersection during the entire period of congestion subject to prevailing capacity and other operational constraints. The continuous optimal departure flow rate (the effect) is then converted to signal timing parameters (the cause) that can be readily implemented. Two numerical examples are presented to demonstrate the properties of the proposed algorithm and examine its performance.

Key Words: Traffic signals; Optimization models

1. Background

Signalized intersections frequently become oversaturated due to the temporal and spatial variation in traffic flow. Under oversaturated traffic conditions, steady-state models may break down when the vehicle arrival rate exceeds the intersection capacity, leading to the carryover of queues from one cycle to another. The design of an effective traffic signal timing plan for oversaturated traffic is more intricate than that for undersaturated traffic.

Mathematical models on control variables in the signal time plan for oversaturated intersections have been proposed by many researchers. Some of the remarkable contributions in early days include: the semi-graphical approach by Gazis (1964), Gazis and Potts (1965); the work on verification of Dunne-Potts's phase switching policy for oversaturated flow conditions conducted by Green (1967); the so-called bang-bang two-stage timing method proposed by Michalopoulos and Stephanopoulos (1977); and Newell's (1989) monograph on traffic signal control theory.

Recently Park *et al*. (1999), Abu-Lebdeh and Benekohal (2000) proposed their genetic algorithms for optimal signal timing and queue management of oversaturated intersections. Lieberman *et al.* (2000) developed a mixed-integer linear programming approach for queue length control. Chang and Lin (2000), Chang and Sun (2004) developed a discrete dynamic model and performance index approach to optimize signal parameters during the entire period of oversaturated conditions. Lo and Chow (2004) developed a dynamic intersection signal control optimization method for oversaturated condition based on cell-transmission model. Li and Prevedouros (2004) presented a hybrid optimization and rule-based oversaturated control algorithm for isolated signals.

Despite the substantial difference in details, the existing models can be roughly categorized into two groups. The first group calculates stochastic delay by assuming the overflow queue in a cycle follows a certain distribution (e.g., normal distribution). The second group considers time varying

demands and aims at minimizing the total intersection delay subject to queue length constraints and looking for best phase switching strategies. The development of theoretical models has led to many on-line (e.g. SCOOT, OPAC, ROHDES, SCATS) and off-line signal timing tools (e.g. TRANSYT-7F, PASSER, and SYNCHRO). Shepherd (1992) and Wood (1993) have summarized some of these models from both theoretical and practical point of view and concluded that most of them serve well for undersaturated or slightly congested flow conditions and break down when demand exceeds capacity. Manufacturers have improved these models since then to make them adaptive to time-varying traffic conditions. Few studies, however, have been conducted assessing the enhanced models with respect to their performance under oversaturated conditions since the work of Shepherd and Wood.

Unlike undersaturated intersections, for which many well-established theories and signal timing policies have been developed, consensus has not been reached with regard to the control philosophy and timing strategy of an oversaturated system. For an isolated intersection, a common belief is that timely and efficient allocation of green time among the intersection approaches is essential to optimize traffic flows. Since queues could not discharge and may exist for successive cycles under oversaturated conditions, describing the dynamics of traffic flow as well as the queue formation and dissipation process is believed to be essential. Hence, a formulation that is computationally effective is considered necessary.

Intersection delay, which can be calculated by cumulative curves of vehicular arrivals and departures, has been arguably the sole performance measure of signal timing plans. Since traffic flow exhibits unique regulated feature at signalized intersections due to frequent interruptions by traffic signals, modeling and optimization of interrupted flow becomes critical for delay optimization. This paper presents a unique modeling and optimization approach for dynamic computation of signal timing parameters of an oversaturated intersection. A reversed cause and effect modeling approach is designed to model interrupted flows at signalized intersections for

enhanced computational efficiency. Both arrival and departure flows at the intersection are modeled as smooth time-dependent functions as if there were no interruptions from signals. One main advantage of working with smooth functions is that we can use differential calculus in the optimization process. It also provides us with the flexibility to scale the optimization process so that the methodology can be extended to cover a network of many intersections in the future study. With the proposed approach, the intersection is treated as a common highway bottleneck. A dynamic linear programming (LP) model is formulated to maximize the total vehicular output or the throughput from the intersection subject to prevailing capacity and other operational constraints. The continuous optimized departure flow rates (the effect) are then utilized to derive signal timing parameters (the cause). Two numerical examples are presented to demonstrate the properties of the proposed control algorithm.

2. Description of the proposed reverse causal-effect modeling approach

At signalized intersections, the departure flow $\mu(t)$ follows a step function which can be expressed as:

$$\mu(t) = \begin{cases} s(t) \text{ or } \lambda(t) & \text{if } t \text{ is in a green phase} \\ 0 & \text{if } t \text{ is in a red phase} \end{cases}$$

where $s(t)$ is the saturation flow and $\lambda(t)$ is the arrival rate, both being time dependent. The shape of the departure curve is a sole function of the signal timing plan if the traffic demand and the saturation flow rate are assumed to be known. A most common objective for both fixed-time and real-time signal timing design is to develop traffic signal plans that minimize total vehicle delay, which can be achieved through the use of cumulative curves for arrivals and departures.

Fig. 1 illustrates the queue formation and dissipation process during a congested time period. *A(t)* represents the cumulative number of vehicles on the approach over time, while *D(t)* (the solid

curve) represents the cumulative number of vehicles departing the intersection over time. The vertical distance between the two curves at time t_j represents the number of vehicles stored (or queue length) at the time when the *j*th vehicle arrives in the system. The horizontal distance between *A(t)* and *D(t)* represents the waiting time (or delay) for the *j*th vehicle before it discharges from the intersection. The area between *A(t)* and *D(t)* during the queuing time interval (t_1, t_n) represents the total vehicle delay during the period of congestion. In oversaturated conditions, the objective is to minimize the area between *A(t)* and *D(t)* in Fig. 1 for all intersection approaches. This is equivalent to optimizing the total vehicular output from the intersection during the congested period.

Place Fig.1 about here

If we treat a signalized intersection as a normal highway bottleneck, both arrival and departure flows can be represented by smooth time-dependent functions as if there were no traffic signals. Consequently, the vehicular departure curve represented by the solid line can be represented by the dotted line as shown in Fig. 1, which approximates the real departure curve and the objective becomes to minimize the area between the arrival and the hypothesized departure curves. As the departure curve of each signalized approach is in fact governed by a specific signal timing plan, the optimal departure curve must be the result of the optimal signal setting. Therefore, attention can be directed to optimize the hypothesized departure curve and then convert the optimized departure flow rate (the effect) to corresponding signal control parameters (the cause). The proposed methodology expressed in a conceptual form is illustrated in Fig. 2 for a single approach of an oversaturated intersection.

Place Fig. 2 about here

Since the departure flow is smoothed over time, the slope of cumulative departure curve is an approximation of the real value over time and no longer restricted to the three values, i.e., 0, λ, and *s*. Instead, it is a smooth function of time *t*. The value of the slope can be in a broader range,

determined by traffic demand, saturation flow, and allocation or split of the green time at the intersection. If traffic demand and the saturation flow rate are given, the departure flow depends solely on the strategy of green time allocation or the phase switching policy. In order to identify the best phase switching strategy, a dynamic linear program is formulated on the basis of smooth traffic flows. Within each discrete time interval Δt (one cycle in this study), the objective of the LP is to optimize the total vehicular output from the intersection. The optimized departure flow rates are then converted to signal control parameters according to their interrelationship. In depth description of the formulation and approximation process is provided along with the first example for ease of illustration. In the following section, we will address the details for the dynamic LP, including the formulation of the objective function, constraints, and the decision variables.

3. A dynamic linear program

3.1. Objective function

If the arrival flow rate at link (*i*,*j*) at time *t*, which travels through node *i*, *j* to *k* is $\lambda_{ijk}(t)$, then the cumulative number of arrivals by time *t* can be given by $A_{ijk}(t)$, where $A_{ijk}(t) = \int_t \lambda_{ijk}(t)dt$. Similarly, the departure flow rate and number of departures which travels through node *i*, *j* to *k* at time *t* are $\mu_{ijk}(t)$ and $D_{ijk}(t)$ respectively and $D_{ijk}(t) = \int_t \mu_{ijk}(t)dt$.

Since the departure curve is hypothesized as a smooth time-dependent function, the objective of minimizing the total delay during the congested period can be expressed as:

minimize: $$F = \sum_{ijk} \int_t (A_{ijk}(t) - D_{ijk}(t))dt = \sum_{ijk} \int_t \{\int_u \lambda_{ijk}(u)du - \int_u \mu_{ijk}(u)du\}dt. \quad (1)$$

For both on-line and off-line signal control models, traffic demands need to be given. The difference is that off-line models use historical data while real-time control requires time-dependent

traffic demand measured from the field. Assuming that the real-time traffic demands are known, the objective then becomes to maximize the total vehicular departure:

$$\text{maximize: } F = \sum_{ijk} \int_t D_{ijk}(t)dt = \sum_{ijk} \int_t \int_u \mu_{ijk}(u)dudt \ . \tag{2}$$

A significant difference between the objective of this approach and some traditional methods is that the former aims to optimize total output from the entire intersection rather than from some predetermined "critical" approaches. Under the situation in which traffic demand is approaching the capacity level, any approach of the intersection may become saturated during some specific time periods due to the random fluctuation in traffic demand. In this model, the "critical" approaches that influence the objective function are identified by the model automatically.

At this point, the control factors in the objective function, expressed as a set of time-dependent decision variables, are continuous departure flows rather than a set of signal timing parameters. The optimal flow rates, at which the hypothesized vehicular departures are discharged from the intersection, are subject to traffic demand, saturation flow, and the splits of green. The relationship among these parameters will be investigated in the next section and integrated into the optimization process.

3.2. Turning movements and grouped movements/streams

Signal timing parameters are calculated on the basis of grouped movements or lane groups. In order to establish a relationship between the smoothed departure flow curve and the associated signal timing parameters, it is necessary to correlate the turning movements with grouped movements or traffic streams.

If we define the arrival and departure flow rate of stream m of phase p at time t as $\lambda_m^p(t)$ and $\mu_m^p(t)$, respectively, the number of cumulative arrivals and departures by time t are denoted by

$A_m^p(t)$ and $D_m^p(t)$ accordingly. Fig. 3 illustrates an example with respect to the relationship between the turning movements and the streams/grouped movements.

Place Fig. 3 about here

The relationship between turning movements and grouped flows can be constructed by defining a set *M* for associated link flows according to the given phase pattern:

$$\sum_{m \in M} \lambda_m^p(t) = \sum_{(i,j,k) \in M} \lambda_{ijk}(t), \tag{3}$$

$$\sum_{m \in M} \mu_m^p(t) = \sum_{(i,j,k) \in M} \mu_{ijk}(t).$$

For example, for the turning movements and grouped flows illustrated in Fig. 3, the set *M* can be defined as:

$$\lambda_m^p(t), \lambda_{m'}^p(t), \mu_m^p(t), \mu_{m'}^p(t), \lambda_{ijk'}(t), \mu_{ijk'}(t), \lambda_{ijk}(t), \mu_{ijk}(t) \in M.$$

The purposes of establishing such a relationship are two folds: 1) to eliminate signal timing parameters from the linear program; and 2) to find the upper bound of the flow rates of the hypothesized departure curves (the lower bound is zero). This is described in the following section.

3.3. Constraint of the capacity of lane groups/streams

During the period of oversaturation, traffic from different approaches to the intersection competes for green lights. Timely and efficient allocation of green time becomes essential to relieve congestion. The green split (the percentage of the effective green time allocated to each of the various phases in a signal cycle) is the factor reflecting the allocation of green time. The green ratio η is the ratio of total effective green to the signal cycle which is determined by the cycle length *C* and total lost time *L*:

$$\eta = 1 - \frac{L}{C}. \tag{4}$$

The summation of the green splits of all signal phases equals to the green ratio:

$$\sum_p g_p = \eta , \tag{5}$$

where g_p is the green split of phase *p*.

The grouped movements are served by individual lanes or lane groups with certain maximum discharge rates, i.e., saturation flow rates. For each stream of various signal phases, the hypothesized departure flow rate $\mu_m^p(t)$ is subject to:

$$\mu_m^p(t) \le s_m^p g_p(t) \text{ for every } (p,m), \tag{6}$$

where s_m^p is the saturation flow rate of the lane or lane group accommodating stream *m* of phase *p*. This relationship is constructed based on the understanding that the continuous departure flow averaged over time must be a percentage of the saturation flow at any specific time and their value is governed by the percentage of green time allocated to the specific signal phase.

If we define μ / s as the "departure flow ratio" to differentiate from the "flow ratio", which is the actual or design flow rate for the critical movement divided by the saturation flow rate for that movement, the sum of the departure flow ratios from the critical approaches of the entire intersection yields:

$$\sum_p Max_m \left[\frac{\sum_{m \in M} \mu_m^p(t)}{\sum_{m \in M} s_m^p} \right] \le \eta . \tag{7}$$

Since the relationship between the turning movements and the streams were established, the maximum departure flow ratio can be represented in the form of directional flows to keep it consistent with the objective function:

$$\sum_p Max_m \left[\frac{\sum_{m \in M} \mu_m^p(t)}{\sum_{m \in M} s_m^p} \right] = \sum_p Max_m \left[\frac{\sum_{(i,j,k) \in M} \mu_{ijk}(t)}{\sum_{m \in M} s_m^p} \right] \le \eta . \tag{8}$$

As can be seen from the above expression, signal timing parameters are totally eliminated from the formula. Not only does it constitute an important constraint of the dynamic LP, but the departure flow ratio constraint also functions as an automatic "search engine" to identify the critical approaches in real-time in conjunction with the objective function during the formulation. The property of the constraint will be discussed in detail through the description of a numerical example given later.

3.4. Constraint of the maximum directional departure flow

Since in the oversaturated condition queues may exist for successive cycles, the dynamics of queue formation and dissipation is essential in our approach. By assuming that the intersection approach has enough space to store queued vehicles, we can use the point queue (or vertical queue) model to capture the dynamics of queues. Following the assumption, if we denote $X_{ijk}(t)$ the number of queued vehicles accumulated on the direction *(i,j,k)* of link (*i,j*) by time *t*, the state equation for $X_{ijk}(t)$, in the form of the normal arrival and the hypothesized departure, can be expressed as:

$$X_{ijk}(t+dt) = X_{ijk}(t) + \lambda_{ijk}(t)dt - \mu_{ijk}(t)dt \geq 0 \ , \tag{9}$$

where,

$$X_{ijk}(t) = A_{ijk}(t) - D_{ijk}(t) . \tag{10}$$

Thus, the maximum directional departure flow rate is subject to:

$$\mu_{ijk}(t) \leq X_{ijk}(t)/dt + \lambda_{ijk}(t) . \tag{11}$$

Since the departure flow cannot be negative, the constraint of non-negativity applies:

$$\mu_{ijk}(t) \ \geq 0 \ . \tag{12}$$

3.5. Summary of the dynamic linear program

In summary, the dynamic linear program reads

maximize: $F = \sum_{ijk} \int_t D_{ijk}(t)dt = \sum_{ijk} \int_t \int_u \mu_{ijk}(u)dudt$

$$\text{s.t.} \sum_p Max_m \left[\frac{\sum_{m \in M} \mu_m^p(t)}{\sum_{m \in M} s_m^p} \right] = \sum_p Max_m \left[\frac{\sum_{(i,j,k) \in M} \mu_{ijk}(t)}{\sum_{m \in M} s_m^p} \right] \leq \eta ,$$

$$\sum_{m \in M} \lambda_m^p(t) = \sum_{(i,j,k) \in M} \lambda_{ijk}(t) ,$$

$$\sum_{m \in M} \mu_m^p(t) = \sum_{(i,j,k) \in M} \mu_{ijk}(t) ,$$

$$\mu_{ijk}(t) \leq X_{ijk}(t)/dt + \lambda_{ijk}(t) ,$$

$$\mu_{ijk}(t) \geq 0 .$$

The control factors of the LP are the directional departure flows from the entire intersection, which are combined to grouped movements or streams by a predefined set *M*. By investigating the relationship between the smoothed departure flow rates and some common signal timing parameters especially the split of green and the green ratio, a regulation to the maximum "departure flow ratio" is established. Since the allocation of green time is determined by the critical approaches, the constraint on the "departure flow ratio" also serves as a real-time "search engine" to look for the crucial approaches that influence the number of total vehicular outputs from the intersection. Another important constraint is established through the queue state equation to define the maximum possible departure rates.

Signal timing parameters are totally eliminated from the LP and the objective of the formulation becomes to maximize total vehicular output from the intersection subject to prevailing capacity and control constraints. The optimized flow rates can then be converted to signal timing parameters according to the relationship between the split of green and departure flow as defined in Eq. (6).

3.6. Solution algorithm

Given a discrete time interval, the dynamic linear optimization program can be decomposed with respect to discrete time so that the optimization procedure can be carried out from the beginning of saturation or any given initial conditions. As illustrated in Fig. 4, the arrival rate $\lambda_{ijk}(t)$ and the

departure rate $\mu_{ijk}(t)$ are assumed to stay constant during the time interval [t, $t+\Delta t$) for $\forall(i, j, k)$. Within each time interval, the linear program searches for the critical approaches of each signal phase and optimizes the departure rates of all directional flows from the intersection. Since traffic demands are assumed to be given, once the departure flow rate was calculated, the departure curve can be obtained by time t and thus $X_{ijk}(t)$.

Place Fig. 4 about here

The calculation algorithm is summarized step-by-step as follows:

Step 1: Initialize variables, including present time, flow rates, queue, and cumulative numbers:

t := 0,

$\lambda_{ijk}(t)$:= time dependent turning movement data $\forall(i, j, k)$

$\mu_{ijk}(t)$:= 0 $\forall(i, j, k)$,

$A_{ijk}(t)$:= 0 $\forall(i, j, k)$,

$D_{ijk}(t)$:= 0 $\forall(i, j, k)$,

$X_{ijk}(t)$:= 0 $\forall(i, j, k)$;

Step 2: Calculate optimal $\mu_{ijk}(t)$ by solving the linear program *F*;

Step 3: Group directional flows into grouped movements or streams;

Step 4: Convert the critical $\mu_m^p(t)$ to $g_p(t)$ by the relationship between μ and split of green:

$$g_p(t) = \frac{\max_{m \in p} \mu_m^p(t)}{s_m^p};$$

Step 5: Check with $g_{p\min}$ (minimum green of *p*). If $g_p(t) < g_{p\min}$, then adjust $\mu_m^p(t)$ accordingly;

Step 6: Check the time horizon, go to step 2 or exit if the end of the time horizon is reached.

4. Numerical examples

4.1. The formulation process and properties of the algorithm

A hypothetical signalized intersection with a typical lag-lag protected left-turn phase is shown in Fig. 5. The saturation flow of two major through movements and left-turns are assumed to be 1800 *veh/h/lane*, the cross street has one lane on each direction with a lower saturation flow of 1200 *veh/h/lane* in consideration of the conflicts incurred by the opposing movements.

Place Fig. 5 about here

Table 1 shows the turning movement data used in the first numerical example, while Fig. 6 shows graphically the turning movements of each approach of the hypothetical intersection.

Place Table 1 about here
Place Fig. 6 about here

The objective of this example is to demonstrate the formulation process as well as the major properties of the proposed model. Calculation of a certain time interval, for example, one cycle is sufficient for this purpose because the calculation is recursive. A multi-cycle example is introduced in next section to demonstrate the model's performance over an entire congested period.

The ratio of traffic flow rate to capacity, i.e., v/c ratio, is used by traffic engineers to represent the degree of traffic saturation. The critical v/c ratio is the v/c ratio for the intersection as a whole, which is determined by

$$X_c = \sum (\frac{v}{s})_{ci} (\frac{C}{C-L}), \qquad (13)$$

where $\sum (\frac{v}{s})_{ci}$ is summation of flow ratios for all critical lane groups. The critical v/c ratio of the sample intersection is 1.41, which indicates a typical oversaturated condition. In addition, we assume the following variables are known at time *t* and are kept constant during calculation:

- Cycle length: 110 s
- Loss time: 10 s
- $\eta = 1 - L/C$: 0.9
- minimum green: g_{1min} = 35 s, g_{2min} = 12 s, g_{3min} = 8 s

Saturation flows: $s_j^1 = 1800$ (vehicles/h), $j \in [1,4]$,

$s_j^2 = 1800$ (vehicles/h), $j \in [1,2]$,

$s_j^3 = 1200$ (vehicles/h), $j \in [1,2]$.

Based on the given phase pattern, the relationship between turning movements and grouped flows are defined as follows:

$$\lambda_1^1(t) + \lambda_2^1(t) = \lambda_{153}(t) + \lambda_{154}(t),$$
$$\mu_1^1(t) + \mu_2^1(t) = \mu_{153}(t) + \mu_{154}(t),$$

$$\lambda_3^1(t) + \lambda_4^1(t) = \lambda_{351}(t) + \lambda_{352}(t),$$
$$\mu_3^1(t) + \mu_4^1(t) = \mu_{351}(t) + \mu_{352}(t),$$

$$\lambda_1^2(t) = \lambda_{152}(t),$$
$$\mu_1^2(t) = \mu_{152}(t),$$

$$\lambda_2^2(t) = \lambda_{354}(t),$$
$$\mu_2^2(t) = \mu_{354}(t),$$

$$\lambda_1^3(t) = \lambda_{451}(t) + \lambda_{452}(t) + \lambda_{453}(t),$$
$$\mu_1^3(t) = \mu_{451}(t) + \mu_{452}(t) + \mu_{453}(t),$$

$$\lambda_2^3(t) = \lambda_{251}(t) + \lambda_{254}(t) + \lambda_{253}(t),$$
$$\mu_2^3(t) = \mu_{251}(t) + \mu_{254}(t) + \mu_{253}(t).$$

Thus, the constraint of the maximum departure flow ratio is represented by:

$$\max\left[\frac{\mu_1^1(t)}{s_1^1},\frac{\mu_2^1(t)}{s_2^1},\frac{\mu_3^1(t)}{s_3^1},\frac{\mu_4^1(t)}{s_4^1}\right]+\max\left[\frac{\mu_1^2(t)}{s_1^2},\frac{\mu_2^2(t)}{s_2^2}\right]+\max\left[\frac{\mu_1^3(t)}{s_1^3},\frac{\mu_2^3(t)}{s_2^3}\right]\le\ \eta\ .$$

In terms of turning movements, the above expression can be rewritten as:

$$\max\left[\frac{\mu_{154}(t)+\mu_{153}(t)}{s_1^1+s_2^1},\frac{\mu_{352}(t)+\mu_{351}(t)}{s_3^1+s_4^1}\right]+\max\left[\frac{\mu_{152}(t)}{s_1^2},\frac{\mu_{354}(t)}{s_2^2}\right]+\max\left[\frac{\sum_m\mu_{45m}(t)}{s_1^3},\frac{\sum_k\mu_{25k}(t)}{s_2^3}\right]\le\ \eta\ .$$

The maximum departure ratio is governed by the "critical" movements of each phase. This constraint not only confines the maximum departure ratio but also serves as a "search engine" for the best combination of traffic streams that influences the rates of the departure curve. The left side of the formula is in fact an abbreviation of every possible combination of grouped movements/streams under the prevailing phase pattern. It contains eight pieces as shown in the generalized linear program below:

$$\text{maximize:}\quad F=\sum_{ijk}\mu_{ijk}(t)\qquad\qquad i,j,k\in[1,5]$$

subject to:

$$\left[\frac{\mu_{154}(t)+\mu_{153}(t)}{s_1^1+s_2^1}\right]+\left[\frac{\mu_{152}(t)}{s_1^2}\right]+\left[\frac{\sum_m\mu_{45m}(t)}{s_1^3}\right]\le\ \eta\ ,$$

$$\left[\frac{\mu_{154}(t)+\mu_{153}(t)}{s_1^1+s_2^1}\right]+\left[\frac{\mu_{152}(t)}{s_1^2}\right]+\left[\frac{\sum_k\mu_{25k}(t)}{s_2^3}\right]\le\ \eta\ ,$$

$$\left[\frac{\mu_{154}(t)+\mu_{153}(t)}{s_1^1+s_2^1}\right]+\left[\frac{\mu_{354}(t)}{s_2^2}\right]+\left[\frac{\sum_m\mu_{45m}(t)}{s_1^3}\right]\le\ \eta\ ,$$

$$\left[\frac{\mu_{154}(t)+\mu_{153}(t)}{s_1^1+s_2^1}\right]+\left[\frac{\mu_{354}(t)}{s_2^2}\right]+\left[\frac{\sum_k\mu_{25k}(t)}{s_2^3}\right]\le\ \eta\ ,$$

$$\left[\frac{\mu_{352}(t)+\mu_{351}(t)}{s_3^1+s_4^1}\right]+\left[\frac{\mu_{152}(t)}{s_1^2}\right]+\left[\frac{\sum_m \mu_{45m}(t)}{s_1^3}\right]\le \eta,$$

$$\left[\frac{\mu_{352}(t)+\mu_{351}(t)}{s_3^1+s_4^1}\right]+\left[\frac{\mu_{152}(t)}{s_1^2}\right]+\left[\frac{\sum_k \mu_{25k}(t)}{s_2^3}\right]\le \eta,$$

$$\left[\frac{\mu_{352}(t)+\mu_{351}(t)}{s_3^1+s_4^1}\right]+\left[\frac{\mu_{354}(t)}{s_2^2}\right]+\left[\frac{\sum_m \mu_{45m}(t)}{s_1^3}\right]\le \eta,$$

$$\left[\frac{\mu_{352}(t)+\mu_{351}(t)}{s_3^1+s_4^1}\right]+\left[\frac{\mu_{354}(t)}{s_2^2}\right]+\left[\frac{\sum_k \mu_{25k}(t)}{s_2^3}\right]\le \eta,$$

$$\mu_{ijk}(t)\le X_{ijk}(t)/dt+\lambda_{ijk}(t) \qquad i,j,k\in[1,5],$$

$$\mu_{ijk}(t)\ge 0 \qquad i,j,k\in[1,5].$$

If we express the dynamic LP in its standard form (i.e., maximize*:* $Z = CX$. s.t. $AX \le B$ and $X \ge 0$), vector B is in the following form in the computation:

$$B=\left[.9\ \ .9\ \ .9\ \ .9\ \ .9\ \ .9\ \ .9\ \ .9\ \ .9\ \ \lambda_{ijk}(t)+X_{ijk}(t)/\Delta t\right]^T.$$

The direct result from the LP is the optimized directional departure flows maximizing the total vehicular output from the intersection. The calculated departure rates along with the arrival demands are depicted in Fig. 7, from which we can see that all of the approaches were cleared except for the northbound of the minor street indicating that the priority was given to the major street under the given traffic condition. The result coincides with the well known policy that giving priority to the oversaturated approaches with higher saturation flow rates will give lower intersection delay. However, it is worthy of note that in the condition where the saturation flow

rates of every approach are exactly the same, which is a very unlikely situation, different optimization methods (other than LP) may come up with different results under the same objective.

Place Fig. 7 about here

Clearly, the optimized departure flow rate should be the result of the optimal signal timing strategy, which can be obtained as follows:

$g_1 = 2000/3600 = 0.56$;

$g_2 = 400/1800 = 0.22$;

$g_3 = (24+98+24)/1200 = 0.12$;

and, $\sum g_p = \eta = 0.9$.

The effective green splits for each phase are computed by converting the departure flows using the following relationship:

$$g_p = \frac{\max_{m \in p} \mu_m^p}{s_m^p}.$$

Consequently, the optimal green time of each phase for the given traffic demand and saturation flow are calculated by:

$$G_i = g_i \times C,$$

where,

G_i = the green time for phase i; and

C = the cycle length (in seconds).

$$G_1 = g_1 \times C = 0.56 \times 110\,\text{sec} = 61.6\,\text{sec}$$
$$G_2 = g_2 \times C = 0.22 \times 110\,\text{sec} = 24.2\,\text{sec}$$
$$G_3 = g_3 \times C = 0.12 \times 110\,\text{sec} = 13.2\,\text{sec}$$

$$\sum G_i = 99.0\,\text{sec}$$
$$L = 10.0\,\text{sec}$$
$$C = 109.0\,\text{sec}$$

The cycle length and the total green time are 1.0 second off because the value of η was approximated to 0.9 (from 0.91) in the calculation.

The numerical example presented a heavily congested case with a critical v/c ratio of 1.41. Traffic demands were extremely high on both of the westbound major street and the northbound minor street. The result shows the priority was given to phase one and phase two of the major street. As a result, queues accumulated on the northbound of the minor street. A simple verification is performed below to demonstrate the effectiveness of the result in terms of maximizing total departure from the intersection.

Many prevalent signal timing tools (e.g., SYNCHRO) use the Highway Capacity Manual method to calculate green split for undersaturated conditions. The first example compares the proposed algorithm with the HCM method. Highway Capacity Manual 2000 suggests green time be allocated according to the flow ratios of the critical approaches. In this numerical example, the critical flow ratios of phases one, two, and three are: $(\lambda_c / s_c)_1 = 0.56$, $(\lambda_c / s_c)_2 = 0.22$, and $(\lambda_c / s_c)_3 = 0.5$, respectively (the subscript *c* denotes "critical"). As a result, the green time of each phase can be calculated as: G_1 = 43.7 seconds, G_2 = 17.2 seconds, and G_3 = 39.1 seconds. As the HCM method tries to equalize the v/c ratio along various phases (1.41 for all three critical approaches in this case), queue will exist at every critical approach of the intersection.

Assuming that the traffic condition remains unchanged in one hour, the approximate hourly departure from the westbound through and right-turn approach is 1430 out of 2000 arrivals (or queued vehicles = 570 vehicles), the departures from the westbound left-turn and the northbound approaches are 282 (queued vehicles = 118 vehicles) and 427 vehicles (queued vehicles = 173 vehicles) respectively. The total number of hourly departures from the critical approaches is 2139 vehicles, which is about 400 less than the optimized value of 2546 vehicles.

In such a case, the situation cannot be improved solely by the design of the signal-timing plan. Geometric modification of the intersection needs to be considered (e.g., add an additional left-turn

bay to increase the saturation flow rate of the northbound) in order to avoid the queue on the northbound spilling back to the upstream intersection.

4.2. Examination of the model over an entire congested period

The purpose of the second example is to examine the performance of the proposed algorithm by comparing with the TRANSYT-7F model. TRANSYT-7F was the model of choice for three reasons: 1) it provides a reliable reference of the global optimum with regard to the queue clearance time; 2) SCOOT has been described as an "online" TRANSYT-7F because it uses a similar optimization method; and 3) the control algorithm of TRANSYT-7F is open to public. Through the test, our interest is to examine whether the proposed model can clear queue in a time comparable to that of TRANYSTY-7F, while at the same time reduce intersection delay because of its real-time nature. A time-varying traffic demand was designed to replicate an entire period of congestion including the whole queue formation and dissipation process. As described earlier, the departure flow is a significant approximation of the real departure, this example also examines the convergence of the model over a long period of time. As shown in Table 2, nine traffic demands are given to each of the turning movements at a time step of 120 seconds.

Place Table 2 about here

We used the "Genetic Algorithm optimization" with the "Queuing Ratio" option provided with TRANSYT-7F V.10.0 for the best solution. As can be seen from Table 3, the green time plan given by TRANSYT-7F remains constant during the calculation period, which is 25 s, 65 s, and 21 s, for phases 1, 2, and 3, respectively, while the dynamic linear program generates time-varying plans according to the real-time traffic demand.

Place Table 3 about here

Also as shown by Fig. 8, despite the different green time allocation patterns, the queue was cleared almost at the same time by the two models, which verified the LP model's performance in terms of the queue clearance time.

Place Fig. 8 about here

Although the two model's queue clearance time is comparable, the total intersection delay from the proposed LP model and the TRANSYT-7F model is different. The total intersection delay resulted from the proposed approach is much less than that of the TRANSYT-7F model. This can be further verified by Fig. 9, which is the cumulative curve diagram obtained from the calculation. In a cumulative curve diagram, the area between the curve of the total arrival and the curve of the total departure represents the total intersection delay. The departure curve from the LP approach is time dependent and adaptive to the real-time traffic demand, while the result from TRANSYT-7F shows a linear departure curve which is not sensitive to the time-varying traffic demand. As is evident, the proposed LP model is much efficient in delay reduction.

Place Fig. 9 about here

It is worthy of note that the model's advantage in intersection delay may be solely due to its real-time nature and does not necessarily mean it outperforms the on-line version of TRANSYT-7F, i.e., SCOOT. Field test is needed for this purpose which is beyond the scope of this study. The notable advantage of the model lies in its computational efficiency due to the application of differential calculus and LP in formulation.

5. Concluding Remarks

For signalized intersections, it is well known that traffic signal timing plans developed with steady-state optimization approaches may fail when intersections become oversaturated. This paper presented a reversed cause and effect modeling approach specifically for optimizing traffic signal timing parameters under oversaturated traffic conditions. This approach treats an oversaturated

signalized intersection as a common roadway bottleneck for improved computational efficiency. By approximating the cumulative arrival and departure curves with continuous functions of time, the optimal solution in terms of system throughput or the optimal departure flow rate for the entire time period can be easily calculated with a dynamic linear programming model proposed in the paper. The optimal green split ratios, resulting from the optimized departure flow (the effect), were then utilized to identify the optimal signal control parameters (the cause). Two numerical examples were presented in the paper. The result showed that the proposed method is promising in that it can automatically identify the "critical" movements of a phase and generate results with maximum vehicular throughput during congested time period. As a result, it clears the queue as fast as the TRANSYT-7F model and advantages in the total intersection delay and computation time. Moreover, in the proposed method, traffic flows were treated as turning movements and grouped movements or streams, making it easier for future extension of the proposed methodology to handle a network with multiple intersections.

Future research may include the relaxation of the fixed cycle length constraint so that the cycle length can be updated dynamically in accordance with the change in arrivals during the entire period of oversaturation, especially during the transient periods between the peak and off-peak hours; taking into account the effect of queue blocking; and eventually extending the methodology to system wide optimization for large scale networks.

Acknowledgement

The authors benefited from many discussions with Prof. Masao Kuwahara of University of Tokyo from which the ideas of this paper were developed.

References

Abu-Lebdeh, G., Benekohal, R.F., 2000. Genetic Algorithm for Traffic Signal Control and Queue Management of Oversaturated Two-Way Arterials. *Transportation Research Record*, no. 1727, 61-67.

Chang, T.H., Lin, J.T., 2000. Optimal Signal Timing for an Oversaturated Intersection. *Transportation Research Part B*, vol. 34, no. 6, 471-491.

Chang, T.-H., Sun, G.Y., 2004. Modeling and Optimization of an Oversaturated Signalized Network. *Transportation Research Part B*, vol. 38, no. 8, 687-707.

Gartner, N.H., 1983. OPAC: A Demand Responsive Strategy for Traffic Signal Control. *Transportation Research Record*, no. 906, 75-81.

Gazis, D.C., 1964. Optimal Control of a System of Oversaturated Intersections. *Operations Research*, vol. 12, 815-831.

Gazis, D.C., Potts, R.B., 1965. The Oversaturated Intersections. *Proceedings of the Second International Symposium on the Theory of Road Traffic Flow*, 221-237.

Green, D.H., 1967. Control of Oversaturated Intersections. *Operations Research*, vol. 18, no. 2, 161-173.

Hale, D., 2006. Traffic Network Study Tool – TRANSYT-7F, United States Version. McTrans Center in the University of Florida, Gainesville, Florida.

Hasan, M., 1999. Evaluation of Ramp Control Algorithms Using a Microscopic Traffic Simulation Laboratory, MITSIM. Master Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

Head, K.L., Mirchandani, P.B., Sheppard, D., 1992. Hierarchical Framework for Real-Time Traffic Control. *Transportation Research Record*, no. 1360, 82-88.

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Winton, R.I., 1981. SCOOT – A Traffic Responsive Method of Coordinating Signals. Laboratory Report 1014, Transport and Road Research Laboratory, Crowthorne, Berkshire, U.K.

Li, H., Prevedouros, P. D., 2004. Traffic Adaptive Control for Oversaturated Isolated Intersections: Model Development and Simulation Testing. *Journal of Transportation Engineering*, vol. 130, no. 5, 594-601.

Lieberman, E.B., Chang, J., Prassas, E.S., 2000. Formulation of Real-time Control Policy for Oversaturated Arterials. *Transportation Research Record*, no. 1727, 68-78.

Lo, H. K., Chow A. H., 2004. Control Strategies for Oversaturated Traffic. *Journal of Transportation Engineering*, vol. 130, no. 4, 466-478.

Luk, J.Y.K., Sims, A.G., Lowrie, P.R., 1982. SCATS – Application and Field Comparison with a TRANSYT Optimised Fixed Time System. *Proceedings of International Conference on Road Traffic Signaling, Institution of Electrical Engineers*, 71-74.

Michalopoulos, P.G., Stephanopoulos, G., 1977. Oversaturated Signal Systems with Queue Length Constraints-I single Intersections. *Transportation Research*, vol. 11, 413-421.

Michalopoulos, P.G., Stephanopoulos, G., 1977. Oversaturated Signal Systems with Queue Length Constraints-II Systems of Intersections. *Transportation Research*, vol. 11, 423-428.

Newell, G.F., 1989. Theory of Highway Traffic Signals. Research Report UCB-ITS-CN-89-1, University of California at Berkeley.

Park, B., Messer, C.J., Urbanik II, T., 1999. Traffic Signal Optimization Program for Oversaturated Conditions: Genetic Algorithm Approach. *Transportation Research Record*, no. 1683, 133-142.

Shepherd, S. P., 1992. A Review of Traffic Signal Control, ITS Working Paper 349, University of Leeds, Institute for Transport Studies, UK.

Transportation Research Board (TRB), 2000. Highway Capacity Manual. Special Report, No. 209, National Research Council, Washington, D.C.

Wood, K., 1993. Urban Traffic Control, Systems Review. Project Report 41, Transport and Road Research Laboratory, Crowthorne, Berkshire, U.K.

List of Figures

List of Tables

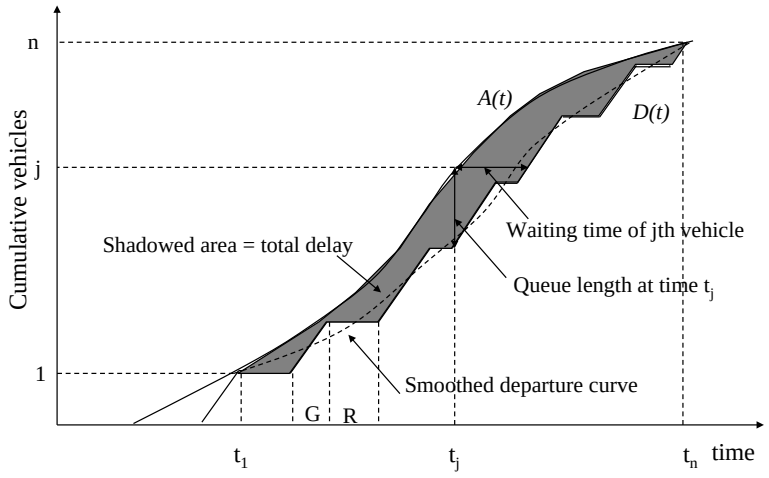

Fig. 1. Cumulative curves at an oversaturated approach.

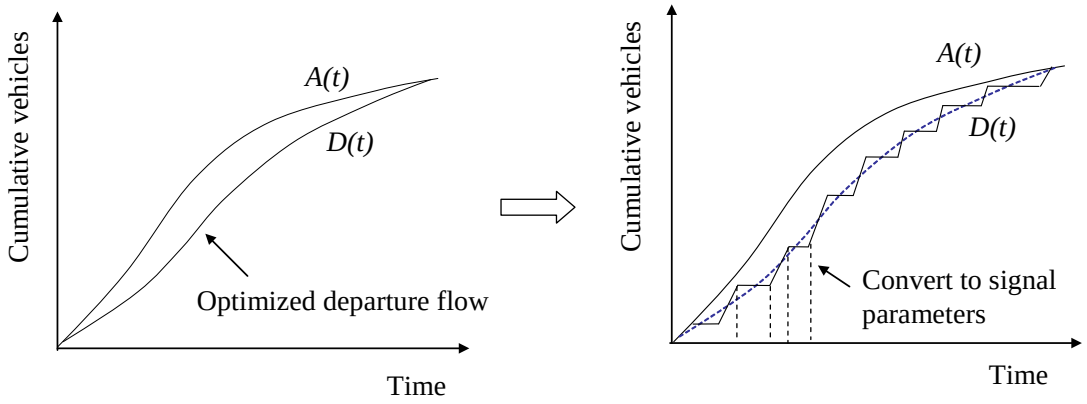

Fig. 2. A conceptual illustration of the proposed approach.

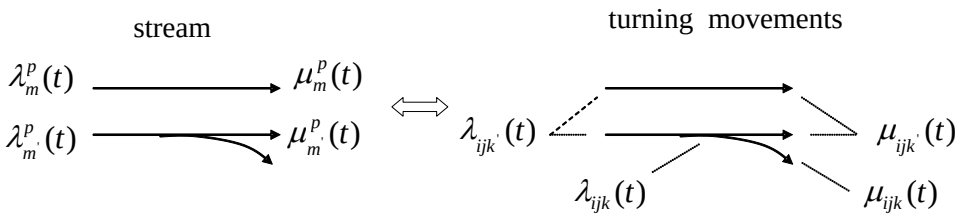

Fig. 3. Turning movements and grouped movements/streams.

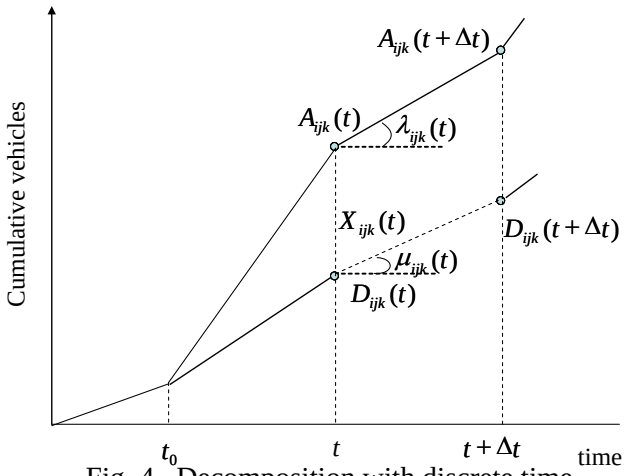

Fig. 4. Decomposition with discrete time.

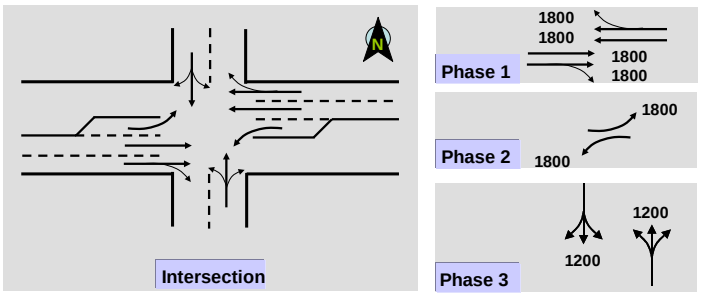

Fig. 5. A hypothetical intersection.

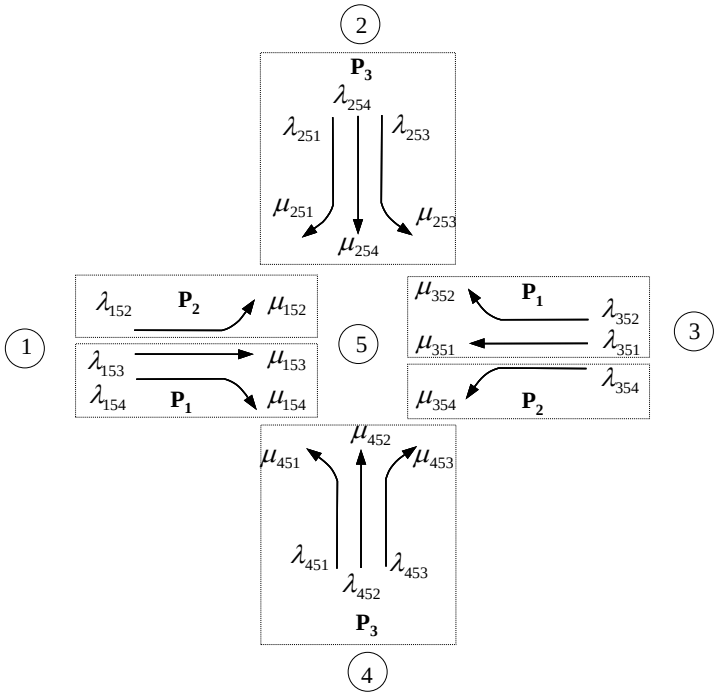

Fig. 6. Turning movements and signal phases.

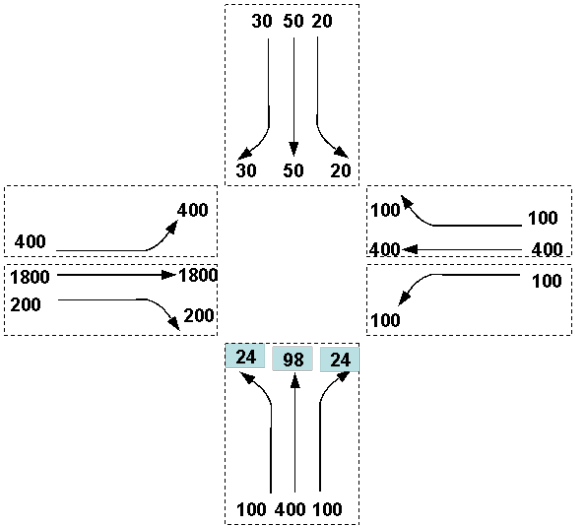

Fig. 7. Optimized departure flows.

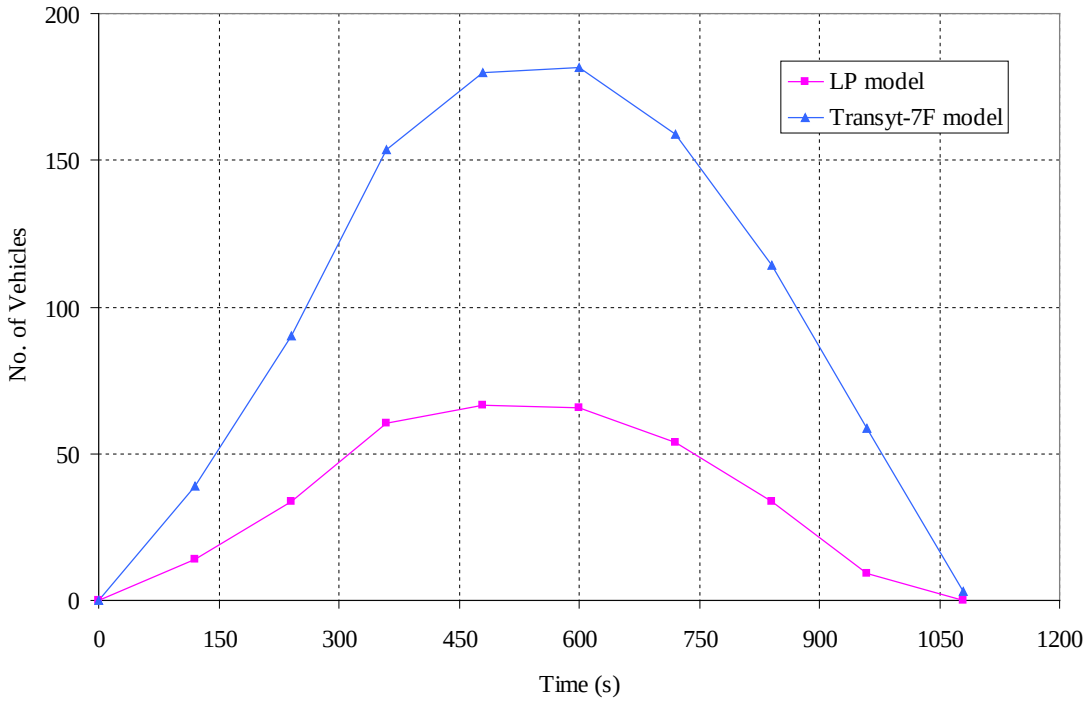

Fig. 8. Total queue length of the Intersection.

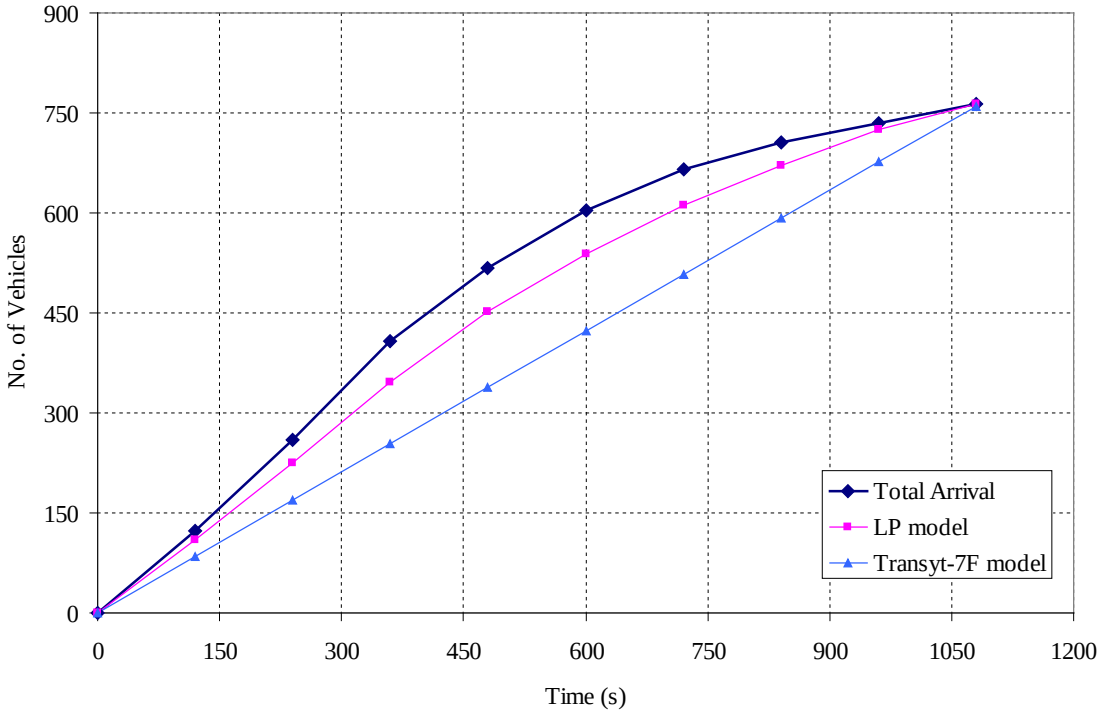

Fig. 9. The total intersection delay.

Table 1

Turning movements of the sample intersection (in veh/h)

Turning movements	L	TH	R	Total
N. Bound	100	400	100	600
S. Bound	20	50	30	100
E. Bound	400	1800	200	2400
W. Bound	100	400	100	600

Table 2

The dynamic traffic demand

Time Steps		1	2	3	4	5	6	7	8	9
North Bound	L	100	110	120	90	70	50	30	20	20
	TH	400	440	480	360	280	200	120	80	80
	R	100	110	120	90	70	50	30	20	20
South Bound	L	20	22	24	18	14	10	10	10	10
	TH	50	55	60	45	35	25	25	25	25
	R	30	33	36	27	21	15	15	15	15
East Bound	L	400	440	480	360	280	200	180	180	180
	TH	1800	1980	2160	1620	1260	900	540	360	360
	R	200	220	240	180	140	100	60	40	40
West Bound	L	100	110	120	90	70	50	30	20	20
	TH	400	440	480	360	280	200	120	80	80
	R	100	110	120	90	70	50	30	20	20

Table 3

Comparison with TRANSYT-7F

The proposed approach					Transyt-7F				
P1	P2	P3	Departure	Queue	P1	P2	P3	Departure	Queue
0	0	0	0	0	0	0	0	0	0
27	67	18	109	14	25	65	21	85	39
27	73	11	225	34	25	65	21	169	90
19	80	12	347	60	25	65	21	254	154
40	60	11	451	67	25	65	21	338	180
19	47	46	539	65	25	65	21	423	182
13	33	64	612	54	25	65	21	507	159
12	20	79	672	34	25	65	21	592	114
12	13	86	726	9	25	65	21	676	59
36	40	36	764	0	25	65	21	761	3