Acyclic Real-Time Traffic Signal Control Based on a Genetic Algorithm

Hairong Yang$^{1,2}$, Dayong Luo$^1$

$^1$ College of Information Science and Engineering, Central South University, Changsha 410012, Hunan, China
$^2$ School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410004, Hunan, China
Emails: yanghong@126.com dayongl@163.com

Abstract: This paper presents an acyclic real-time traffic signal control model with transit priority based on a rolling horizon process for isolated intersections. The developed model consists of two components, including: an Improved Genetic Algorithm (IGA)-based signal optimization module and a microscopic traffic simulation module. The acyclic real-time traffic signal control model optimizes the phase sequence and the phase length with the aim to minimize the total delay of both transit vehicles and general vehicles for the next decision horizon. Numerical results show that the proposed IGA signal optimization module could provide a more efficient search for optimal solutions. The results also show that the acyclic real-time traffic signal control model outperforms the fixed-time control model. It prioritizes transit vehicles while minimizing the impact on the general vehicles.

Keywords: Acyclic real-time control, traffic signal control, transit priority, genetic algorithm.

1. Introduction

Traffic signal control is the most important and efficient method for controlling traffic in urban areas. There are three categories of traffic signal control strategies: fixed-time control, traffic actuated control and traffic adaptive control. The state-of-the-art studies on a traffic signal are predominantly focused on adaptive strategies,
which are responsive to current traffic conditions through real-time optimization of the selected performance criteria. There are two real-time control schemes: a “cyclic” control scheme and an “acyclic” control scheme. One clear advantage of the acyclic control scheme in relation to the cyclic method is its flexibility in adjusting the traffic signal timing to changes in the traffic demand [1, 2].

Gartner first realized the necessity of an acyclic control scheme and suggested an optimization procedure, which is now referred to as the “rolling horizon” method. Yagär and Han [3] employed a rule-based optimization process to generate candidate signal timing plans. A heuristic decision-making process was used to generate phase plans. Dion and Hellinga [4] described the development and evaluation of a rule-based signal optimization procedure that explicitly considered the impacts of transit vehicles. Lee et al. [5] presented an innovative optimized strategy for integrated traffic and transit signal control. A Genetic Algorithm was adopted to resolve the model. Cai Chen et al. [6] presented a study on an adaptive traffic signal controller for real-time operation. The control algorithm was built on approximate dynamic programming. Aboudolas et al. [7] investigated the efficiency of a signal control methodology, in which the problem of the signal control was formulated as a quadratic-programming problem that aimed at minimizing and balancing the link queues. Ruchaj and Stanislawski [8] presented comparison of five algorithms used to control acyclic traffic lights at intersections of roads in an urban road network. Although these signal systems were operating successfully, there is still high potential for further improvements. Firstly, most of the past researches either failed to consider the impacts of transit vehicles or did not consider roundly. Secondly, the terminal delay experienced by all the vehicles left in a queue beyond the end of the decision horizon was not considered in the objective function in most of the past researches. Moreover, the optimization methods can be improved.

This paper presents an acyclic real-time traffic signal control model that is embedding transit priority in the optimization process. The model is unique in two ways. First, the paper proposes a more considerable objective function which explicitly considers the impacts of all transit vehicles in the intersection. In addition, the terminal delay is considered in the objective function. Second, an Improved Genetic Algorithm (IGA) which imported the ideas of Simulated Annealing extending is applied to optimize the problem.

2. Traffic signal control strategy

The developed acyclic real-time control model consists of two components including: an IGA-based signal optimization module and a microscopic traffic simulation module. The IGA-based signal optimization module is designed to optimize the phase sequence and phase length with the aim to minimize the delay of both transit and general vehicles. The primary function of the microscopic traffic simulation module is the delay evaluation of the candidate signal plans generated by the IGA module. Fig. 1 shows the architecture and signal optimization procedure within the system.
Fig. 1. The signal optimization procedure

The control model proposed provides real time signal control based on an acyclic rolling horizon process. As traffic demand changes, the rolling horizon approach uses a shorter horizon time to respond quickly to the demand changes. The acyclic real-time control model does not consider explicitly the traditional cyclic signal control concept, but rather determines the phases sequence and their switching times during the pre-determined horizon period. The term “horizon” differs from the cycle time in the fact that one signal phase may be provided more than once, or not provided at all. Fig. 2 illustrates an example of the acyclic rolling horizon signal control approach.

As shown in the above figure, the optimization period is divided into several decision horizons. From the beginning of every new time horizon, the IGA-based signal optimization module iteratively converges to an optimal signal timing plan to minimize the total delay for the next decision horizon. At the end of the process, only the first few seconds of the newly generated plan are implemented, so that the computation of signal timing for the next horizon must be completed while the implemented fraction of the current timing plan is running. As time goes on, the horizon also moves forward.

For this study, the decision horizon length is set to 60 and 10 s of the resulting plan are implemented.
3. Microscopic traffic simulation module

Most of the prediction models in past researches adopt a macroscopic model [9, 10]. As one core component of the presented signal model, a microscopic traffic simulation module is developed. This link-wise simulation module analyzes the flow of individual vehicles and calculates the total delays for given signal timing plans from the IGA-based signal optimization module. Since the module is used to calculate every single chromosome in the population and for all generations, it would be prohibitively slow to use commercial microscopic simulation. A simple simulator was developed, in which the behaviour of the traffic flow is simplified under four assumptions: 1) all traffic moves at a travel speed; 2) the traffic in the intersection is distributed evenly to each lane; 3) all vehicles do not change their driving lanes; 4) delay connected only with the influence of the traffic signal, is experienced by the vehicles.

The microscopic traffic simulation module calculates mainly two parts delays as follows.

1) The delays experienced by the general and transit vehicles leaving the stop-line within the decision horizon.

Define $d_{jh}^x = t_{lx} - c_{jh}^x$ as the delay that vehicle $x$ experiences on approach $j$ in the decision horizon $h$; $c_{jh}^x$ denotes the time of vehicle $x$ reaching the stop-line at a travel speed on approach $j$ in decision horizon $h$; $l_{jh}^x$ is the actual time of vehicle $x$ leaving the stop-line on approach $j$ in decision horizon $h$; $s_{jh}^x$ is the available service time for vehicle $x$ on approach $j$ in decision horizon $h$; $GE_{jh}$ is the Earliest Green time on approach $j$ in decision horizon $h$; $\gamma_{jh}$ is the saturation headway; $GN_{jh}$ is the starting time of the Next Green time on approach $j$ in decision horizon $h$.

Under the established assumption, the delay $d_{jh}^x (x)$ may be defined as the difference between the vehicles actual travel time and the travel time at a travel speed on the link. The delay can be expressed as

$$d_{jh}^x (x) = l_{jh}^x - c_{jh}^x$$

A vehicle $x$ may leave the stop-line either not affected by any delay or at the earliest available service time. $l_{jh}^x (x)$ can be expressed by equation

$$l_{jh}^x (x) = \max(s_{jh}^x (x), c_{jh}^x (x))$$

If $x$ is the first processed vehicle on the link, the starting time of the earliest green time on approach $j$ becomes the available service time. Otherwise, the service time for $x$ becomes a saturation headway later than the previous vehicle’s departure time.
According to (2) and (3), we obtain $l^j_h(x)$. If $l^j_h(x)$ is in the red time on approach $j$, the service time for vehicle $x$ becomes the starting time of the next green time on approach $j$ in decision horizon $h$. Define $x_r$ as the first vehicle facing a red signal in decision horizon $h$. $s^j_h(x)$ ($x \geq x_r$) is obtained according to equation (4):

$$s^j_h(x) = \begin{cases} \text{GE}^i_h & \text{if } x = 1, \\ l^j_h(x-1) + y^j_h & \text{otherwise.} \end{cases}$$

2) The delays experienced by all the vehicles left in a queue behind the end of the decision horizon,

$$d^j_r(y) = \max((t^j_e - c^j_r(y)), 0),$$

where $d^j_r(y)$ denotes the delay experienced by vehicle $y$ queuing after the stop-line on approach $j$ at the end of the horizon; $t^j_e$ is the end time of the decision horizon $h$; $c^j_r(y)$ is the time of vehicle $y$ reaching the queuing position at a travel speed on approach $j$.

4. IGA-based signal optimization module

GA is a biologically inspired method for function optimization that is loosely based on the theory of evolution [11]. As a global search method, GA has been successfully applied to combination optimization, machine learning, signal processing, adaptive controlling and artificial life. The signal control optimization problem in this paper is both complex and demanding. Conventional optimization methods, including hill-climbing, enumerative, and random search methods, lack both speed and robustness needed for such applications. This has led to the use of GA. Traditional GA suffer from two disadvantages in the evolution process of GA: one is the premature phenomenon because of the large difference among chromosomes at its early stage; another one is that the convergence efficiency of GA becomes lower with a smaller difference among chromosomes. We have improved GA to overcome the above shortcomings.

An IGA-based signal optimization module is designed to search for the signal timing plans that optimize the given objective function for the next decision horizon. The IGA module in this paper differs from earlier work in two aspects:

1) We have adopted a simulated annealing extending method to develop IGA.
2) In order to improve the calculation speed, a solution library is applied in the solving process. The solution library is mainly used for storage of the candidate signal scheme whose objective function value has been calculated.
4.1. Objective function

For a given decision horizon \( h \), the optimization goal is to determine the control variables which include the phase times \( G_h(k), \, k = 1, 2, \ldots, K_h \), and the sequence of each signal phase \( S_h(k), \, k = 1, 2, \ldots, K_h \), in order to minimize the objective function value \( F_h \) subject to various constraints. Here \( k \) is the phase index, \( F_h \) is the total delay of both transit and general vehicles. The objective function can be expressed as

\[
\text{Minimize } F_h = \sum_j \sum_n d_j^i (n) + \sum_j \sum_m \psi_m^j d_m^i (m) + \sum_j \sum_y w_y^j dr_y^j (y)
\]

subject to the following constraints:

\[
\sum_{k=1}^{K_h} G_h(k) = h \quad \text{for } k = 1, \ldots, K_h,
\]

\[
G_{\min}^i (k) \leq G_h^0 (k) \leq G_{\max} (k) \quad \text{for } k = 1, \ldots, K_h,
\]

\[
G_h^0 (k) = \begin{cases} G_h (1) + G_{h-1} (K_{h-1}) & \text{for } i = 1 \text{ and } S_h (1) = S_{h-1} (K_{h-1}), \\ G_h (k) & \text{otherwise}, \end{cases}
\]

\[
\text{Max} L_h^j < \text{Allow} L^j \quad \text{for } j = 1, \ldots, N_j,
\]

\[
\psi_m^j = f_m^j Q_m^j / Q_v,
\]

\[
f_m^j = \begin{cases} T_m^j - T_h^j & \text{if } T_m^j - T_h^j < T_{\max}, \\ 2 & \text{otherwise}. \end{cases}
\]

The objective function value \( F_h \) consists of three components: \( \sum_j \sum_n d_j^i (n) \) denotes the delay experienced by the general vehicles within the decision horizon; \( \sum_j \sum_m \psi_m^j d_m^i (m) \) denotes the delay experienced by the transit vehicles within the decision horizon; \( \sum_j \sum_y w_y^j dr_y^j (y) \) denotes the terminal delay experienced by all vehicles left in a queue beyond the end of the horizon. The purpose of the terminal delay is to counteract the bias that could cause the signal optimization process to select signal-switching decisions that yield a low cost in the near future but a high cost thereafter.

In (6), the weighting coefficients of \( \psi_m^j \) are assigned to the transit vehicles to provide transit priority in the optimization process. The weighting coefficient \( \psi_m^j \) can be calculated according to (11). \( Q_{m,v}^j \) is the passenger occupancy of the transit vehicle \( m \) on approach \( j \) during decision horizon \( h \); \( Q_v \) is the average passenger
occupancy of the general vehicles. \( f'_{h_j} \) is the adjustment factor for a transit vehicle schedule delay [12], which is assigned to every transit vehicle in real time; it can be obtained according to (12). \( T'_{h_j} \) is the transit schedule delay; \( T''_{h_j} \) is the acceptable schedule delay for the transit vehicles, set \( T''_{h_j} = 5 \) min; \( T_{\text{max}} \) is the difference of the maximum schedule delay and the acceptable schedule delay for transit vehicles, set \( T_{\text{max}} = 15 \) min.

The weighting coefficients of \( w'_{h} \) are assigned to the general or transit vehicles left in a queue; if vehicle \( y \) is a general vehicle, \( w'_{h} = 1 \). Otherwise, \( w'_{h} = \psi'_{h_j} \).

Equation (7) keeps the sum of all phase durations within the decision horizon of length \( h_L \).

The maximum and minimum phase duration constraints are set in equation (8). \( G_{\text{max}}(k) \), \( G_{\text{min}}(k) \) denote the maximum and minimum phase lengths of phase \( k \) respectively. \( G_k(k) \) can be obtained according to (9). If the first phase of the current horizon and the last phase of the previous horizon are the same, the phase time is their sum.

Equation (10) means that the maximum queue length on approach \( j \) must be less than the allowable queue accommodation lengths on approach \( j \).

4.2. Coding

A crucial step towards applying GA for the optimization problem is to formulate the genetic representation which can express the potential solutions. After the phase sequence is decided, the phase time is considered as the only variable, so that the chromosome is the string of the phase time. Binary coding is employed to code the phase time, each phase time is between the maximum phase time and the minimum phase time.

The specific implementation is as follows: A four-legs intersection is taken as an example, the defined phase sequence is: phase1 (EW), phase2 (NS), phase3 (EWL). Since the proposed signal control model optimizes the signal plans based on the time horizon length, a single signal phase can be provided more than once in one decision horizon. The maximum number of phases is 6 in a decision horizon of length 60 s. Considering the desired precision of the result, we adopt 6-bit binary to express one phase time, and then the string length is 36. Fig. 3 shows binary coding of a signal timing plan with four signal phases.

<table>
<thead>
<tr>
<th>A signal timing plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW 10sec</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Binary encoding</td>
</tr>
<tr>
<td>001010</td>
</tr>
<tr>
<td>A chromosome</td>
</tr>
</tbody>
</table>

Fig. 3. Binary encoding of a signal timing plan
4.3. Fitness function

Simulated Annealing is another heuristic search algorithm which is proposed by Paul. In order to overcome the shortcomings of GA, the idea of Simulated Annealing extending is imported. We compute the fitness function (13)

$$F_i = e^{-\frac{f_i}{T}}.$$  

where, $F_i$ is the fitness value of $i$-th individual; $f_i$ is the objective value of $i$-th individual based on (6); $T=T_0k^{g-1}$, $g$ is the generation number; $T_0$ and $T$ are the initial temperature and present temperatures respectively; $k$ are the cooling rates.

According to this method, when the temperature is high, the reproduction probability with a large fitness value and difference will be close. With decrease of the temperature, the difference of the reproduction probability will be enlarged, thereby; the prevalence of the excellent individual will be more obvious. Thus the method can solve not only the premature problem but also the low evolution efficiency problem.

4.4. Procedure of IGA

The procedures of the IGA proposed in this paper are designed according to Fig. 4.

**Step 1.** The input base data of the acyclic real-time traffic signal control model, the user parameters and algorithm parameters are defined: the population size $N$, initial temperature $T_0$, supreme genetic algebra $I$, cooling rates $k$, etc.

**Step 2.** Sensitivity analysis is conducted to select the most appropriate parameters: crossover probability $cp$ and mutation probability $mp$.

**Step 3.** Create a random initial population.
Step 4. Decode every single chromosome in the population. Verify if the solution is in the solution library, if yes, find the objective value in the solution library; otherwise, use a microscopic traffic simulation module to calculate the objective value, then save the signal timing plan and its objective value into the solution library.

Step 5. Simulated Annealing extending method is used to calculate the fitness value for each chromosome.

Step 6. Select a new population by the weighted wheel.

Step 7. Crossover procedure of IGA. Combine the chromosomes into pairs randomly and swap the characters of each pair between random positions with cp.

Step 8. Mutation procedure of IGA. Generate $N$ random numbers: $r_1, r_2, \ldots, r_k, \ldots, r_N$ within the interval $[0,1]$, if $r_k < mp$, then mutate this chromosome.

Step 9. If the stopping rule is met, then go to Step 10, else define $g=g+1$, $T=T_{0}k^{g-1}$, then return to Step 4.

Step 10. Stop the iterations and output the best object function value as a final result.

5. Case study

For this study, a typical four-legged isolated traffic intersection is modelled for the following purposes:

1) perform sensitivity analysis of IGA parameters;

2) the model performance is compared with respect to that of the fixed-time control to evaluate the ability of the proposed method;

3) the transit priority is embedded in optimal control to evaluate the performance of the system on transit priority.

The four-legged intersection has two lanes in the north and south approaches and three lanes in the east and west approaches. The east and west approaches are designed to have one exclusive left-turning lane, one exclusive through lane, and one shared through and right-turning lane. The north and south approaches have one exclusive through lane and one shared through and right turning lane. The left turns from the north and south bounds are not allowed. Each intersection approach is 350 m long. Fig. 5 illustrates the layout of the experimental intersection.

Fig. 5. Road condition of a four-legs intersection
The model is evaluated under traffic demand during the period [12:00, 12:30]. For this period, the traffic flow rate on east entrance is 928 pcu.h\(^{-1}\), the proportion of the left-turn, through and right-turn on east entrance is 15.9, 75.0, 9.1 % respectively; the traffic flow rate on west entrance is 795 pcu.h\(^{-1}\), the proportion of the left-turn, through and right-turn on west entrance is 16.4, 70.7, 12.9 % respectively; the traffic flow rate on north entrance is 512 pcu.h\(^{-1}\), the proportion of the through and right-turn on north entrance is 78.5, 21.5 % respectively; the traffic flow rate on south entrance is 231 pcu.h\(^{-1}\), the proportion of the through and right-turn on south entrance is 77.9, 22.1 % respectively.

Transit vehicles approaching the controlled intersection from four entrances are considered. The west entrance and south entrance arriving transits are given in Table 1.

<table>
<thead>
<tr>
<th>Entrans</th>
<th>Transit arriving</th>
</tr>
</thead>
<tbody>
<tr>
<td>West entrance</td>
<td>Through: 12:01 one is late for 6 min, 12:05 one arrives on time, 12:14 one is late for 3 min, 12:17 one arrives on time, 12:20 one is early for 3 min, 12:29 one arrives on time. Left turns: 12:03 one arrives on time, 12:17 one is late for 4 min, 12:29:20 one is late for 6 min</td>
</tr>
<tr>
<td>South entrance</td>
<td>Through: 12:02 one is late for 10 min, 12:06 one is late for 2 min, 12:16 one arrives on time, 12:25 one is early for 3 min</td>
</tr>
</tbody>
</table>

According to the above traffic flows, we have programmed this problem based on the acyclic real-time control model. In order to investigate the ability of the proposed method to effectively and efficiently provide signal control, its performance was compared with respect to that of a fixed-time signal operation. Two types of transit vehicles treatments are also experimented: with a TSP strategy and without a TSP strategy.

The functional components of the model consist of a number of parameters that may affect the efficiency of the model operation. \( Q_{a,h,j} \), \( Q \) is set to 35 and 3 respectively. The initial temperature \( T_0 = 100 \), the cooling rates \( k = 0.99 \); the population size \( N = 50 \), the supreme genetic algebra \( I = 100 \).

In the optimization process, the following three signal phases are imposed on the signal operation: 1) Phase 1: serving all eastbound and westbound traffic with a 15 s minimum duration, 60 s maximum duration; 2) Phase 2: serving all northbound and southbound traffic with a 15 s minimum duration, 60 s maximum duration; 3) Phase 3: serving both eastbound and westbound left-turners, with a 4 s minimum green interval, 10 s maximum green interval. Each phase is followed by 3 s of yellow interval and 2 s of all-red interval.

When applying IGA, several critical parameters should be carefully determined. Sensitivity analysis is realized to select the most appropriate parameters including the crossover probability \( cp \) and the mutation probability \( mp \) for IGA-based optimization module. The effect of the crossover probability on the performance is found by varying the \( cp \) values – 0.8, 0.85, 0.9. As given in Fig. 6a, 0.85 is found to be the best. The effect of the mutation probability is also found by varying its value – 0.04, 0.05, 0.06; Fig. 6b shows the effect. The results revealed that the value of 0.05 may be the best.
The simulating results are shown in Figs 7, 8 and Table 2.

Fig. 7 shows the convergence process compared in an IGA and standard GA algorithm. As it can be seen, IGA convergence results are better than GA during the whole evolution process. IGA method solves the premature problem and the low evolution efficiency problem of GA to a great extent.

Fig. 8 illustrates the signal timings implemented by the acyclic real-time traffic signal control model and the corresponding fixed-time cycle length. The timings illustrated in Fig. 8 indicate that the cycle times of an acyclic real-time traffic signal control model exhibit persistent variation, but the mean of the cycle times (46.8 s) is approximately equal to the optimal fixed-time cycle length (46 s).
The acyclic real-time traffic signal control model and the fixed-time optimization model are applied in practice respectively. Three types of vehicle delays, including the general vehicles delay, the transit vehicles delay, and the entire intersection vehicles delay are used to measure the performance of different control strategies. Table 2 shows the delays.

Table 2. Distribution of delays

<table>
<thead>
<tr>
<th>Control strategy</th>
<th>Entire vehicles</th>
<th>General traffics</th>
<th>Transit vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-time control (without TSP)</td>
<td>65.766</td>
<td>54.255</td>
<td>11.531</td>
</tr>
<tr>
<td>Fixed-time control (with TSP)</td>
<td>66.194</td>
<td>57.915</td>
<td>8279</td>
</tr>
<tr>
<td>Acyclic real-time traffic signal control</td>
<td>55.082</td>
<td>50.717</td>
<td>4365</td>
</tr>
</tbody>
</table>

As shown in Table 2 the results indicate that:

1. When compared with the fixed-time control model, the acyclic real-time traffic signal control model resulted in a significant delay improvement, it can provide more efficient real-time traffic signal control and provide significant benefits at individual intersections. The acyclic real-time traffic signal control model reduced the entire vehicles delays by as much as 16.2% when compared to the fixed-time control model without TSP and the reduced entire delays by 20.1% when compared to the fixed-time control model with TSP. The reason is that the acyclic real-time traffic signal control model can adjust the traffic signal timing of the decision horizon to respond to large and rapid changes in the traffic demand.

2. Acyclic real-time traffic signal control can reduce the transit vehicles delays by as much as 62.1% while reducing the general traffic delays by 6.5% when compared to an optimal fixed-time operation without TSP. The fixed-time (with TSP) resulted in 28.2% reduction in the transit vehicles delay and 6.8% increase in the general traffic delay. It can be concluded from the above results that the fixed-time (with TSP) can efficiently provide transit signal priority but has a large impact on the general vehicles, the acyclic real-time traffic signal control model benefits the transit vehicles efficiently while minimizing the impact of TSP on the general vehicles. The reason is that the fixed-time (with TSP) arbitrarily interrupts the normal signal operation to provide the TSP service when transit vehicles are detected, but the acyclic real-time traffic signal control model considers the system benefit when a transit vehicle is coming, it always tries to find the optimal system-wide performance.

6. Conclusions

This paper presented an acyclic real-time traffic signal control model based on a rolling horizon process. The idea of Simulated Annealing extending is imported to GA to design an IGA-based signal optimization module, which optimizes the phase sequence and phase length with the aim to minimize the total delay of both transit and general vehicles for the next decision horizon. The control model also involves a microscopic delay performance module that can evaluate the delays of signal timing plans generated by IGA. The numerical results have shown that the proposed
IGA optimization could provide a more efficient search for optimal solutions for the traffic signal control. The results have also shown that the proposed acyclic real-time traffic signal control can provide significant benefits when compared with the fixed-time control model. In addition, the acyclic real-time traffic signal control model can efficiently improve the transit vehicles operation in a single intersection while minimizing the impact on the general vehicles.

The application of the acyclic real-time traffic signal control model to a corridor containing several signalized intersections will be the aim of our future work.

Acknowledgements. This work was supported by the National Natural Science Foundation of China No 50808025 and by Fok Ying Tung Education Foundation No 122013.

References