Transportmetrica, Vol. 2, No. 1 (2006), 53-85

A LANE-BASED OPTIMIZATION METHOD FOR THE MULTI-PERIOD ANALYSIS OF ISOLATED SIGNAL-CONTROLLED JUNCTIONS

C.K. WONG[1], S.C. WONG[2] AND C.O. TONG[2]

Received 12 March 2005; received in revised form 4 October 2005; accepted 5 October 2005

Recently, a lane-based optimization method was developed for the design of isolated signal-controlled junctions. The lane markings are relaxed as binary-type control variables and are integrated into the design framework, in which the lane markings and signal settings can be optimized simultaneously to maximize the overall junction performance. To take into account the time-varying effects of traffic demand, the lane-based optimization method has been enhanced to cater for multi-period demand patterns. To eliminate ambiguity, only a single set of lane markings (also referred to as the permitted movements) can be established on the ground for operation throughout all of the design periods, which implies that during various design periods road users that approach a junction will be guided by the same set of lane markings. In accordance with the specific traffic conditions in different design periods, road users can make their own choice of traffic lane for turning, provided that they do not violate the permitted movement patterns. In the present formulation, the actual lane utilization patterns are referred to as the effective movements. A set of linear constraints is developed to relate all of the lane-based control variables and to prescribe the feasible solution region. The optimization for the usual objective function is formulated as a mathematical program. Standard and heuristic solution methods are derived, and numerical examples are also given for demonstration.

KEYWORDS: Lane-based optimization method, multi-period analysis, signal-controlled junctions, lane markings, isolated junctions

1. INTRODUCTION

In the literature, there are two well-established approaches to the systematic design of isolated signal-controlled junctions: the stage-based method (Webster, 1958; Allsop, 1971a, b, 1972, 1975, 1981; Tully, 1976, Burrow, 1987) and the group-based (or phase-based) method (Improta and Cantarella, 1984; Heydecker and Dudgeon, 1987; Gallivan and Heydecker, 1988; Heydecker, 1992; Silcock, 1997; Sang and Silcock, 1989). Both of these approaches take the lane marking layouts as exogenous inputs that are then excluded from the optimization procedures. The first attempt to combine the actual lane-use into the design framework to simultaneously optimize lane use and signal timings was conducted by Lam et al. (1997). Their findings showed that a proper match between the lane-use and signal settings certainly improved the overall performance of a junction. The aim was to maximize the junction capacity using a heuristic objective function, which was formulated to minimize the sum of the flow ratios together with the factored sum of the right-of-way variables. The designs of signal timings for pedestrian movement and the reordering of stage sequences to enhance the smoothness of the traffic flow were subject to a lower order of importance, and were arranged in the second and third levels of optimization, respectively. Relevant designs for the minimum junction delays and minimum cycle length still require further development.

More recently, a lane-based optimization method for isolated signal-controlled junctions was formulated to include lane-marking layouts as design variables for analysis (Wong et al., 2000; Wong and Wong, 2003a), in which the signal settings for

[1] Department of Building and Construction, City University of Hong Kong, Tat Chee Road, Kowloon, Hong Kong SAR, P.R. China. Corresponding author (E-mail: wongck@cityu.edu.hk).

[2] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, P.R. China.

traffic movements and pedestrian crossings were optimized in a unified framework. The lane markings, which were considered as discrete permitted movements in individual traffic lanes, were defined as binary variables to which only 0 or 1 could be assigned to represent their existence during the optimization process. To relax the lane markings as design parameters, allocated flows in the traffic lanes were also introduced into the lane-based formulation as continuous variables. Other group-based variables were preserved in the lane-based approach, including the successor functions, starts of green, durations of green, cycle length, and common flow multiplier. All of these variables are governed by and related to a set of linear constraints. The capacity maximization and cycle-length minimization problems were formulated as Binary-Mixed-Integer-Linear-Programs (BMILP), which are solvable by the standard branch-and-bound technique (Wong and Wong, 2003a). As delay functions are mostly non-linear in nature, the delay minimization problem was formulated as a Binary-Mixed-Integer-Non-Linear-Program (BMINLP). Conventionally, a piecewise linearization technique can be applied to break up the original non-linear problem into a series of BMILP sub-problems, but this works well only for convex functions (Improta and Cantarella, 1984). In the lane-based formulation, however, an examination of the property of the Hessian function proves that the delay function is no longer convex in the enlarged feasible solution region as a result of the relaxation of the turning flows as non-fixed parameters (in the delay function). However, it is found that a cutting plane algorithm offers promising solution characteristics for general non-convex BMINLP problems (Porn and Westerlund, 2000), and the solution of the lane-based delay minimization problem using Webster's delay function as the objective for optimization has also been tested (Wong and Wong, 2003b). In the cutting plane algorithm, the problem size of the BMILP sub-problems expands continuously until a sufficient number of hyper-planes (to be added in the form of linear constraints) are constructed to replicate the solution space as it is represented by the corresponding non-linear function. However, this means that vast computational effort is required to solve the problem, and for this reason a more efficient heuristic line search technique has been proposed and formulated (Wong et al., 2002) to intensively search over the feasible range of the cycle length for optimal signal settings, in which all of the binary variables including the lane permitted movements (lane markings) and successor functions are optimized in the lane-based module. All of the lane-based model outputs except for the signal timings and cycle length are then fed into a group-based module, called SIGSIGN, as fixed inputs for further optimization. A substantial reduction in computing time and a similar solution quality were obtained.

Mathematically, all of the lane-based model variables, including the permitted lane movements, can be varied to maximize the junction performances as long as they satisfy all of the governing constraints and the requirements that are listed by the relevant input data. Obviously, the demand flow pattern is one of the key inputs that influences the lane marking results, the allocated flow patterns, and the signal timings. In practice, traffic demands are time-variant and elastic due to different control policies (Szeto and Lo, 2005), and the approaching flow pattern may continue to vary across successive signal cycles. In this case, the lane-marking layout must be changed accordingly to suit the latest traffic demand pattern to maintain an effective signal-controlled system. If the physical layout of the lane markings is not compatible with the demand flow patterns, then the lane markings may prevent the signal-controlled junction from achieving an optimal performance. Hence, the design of a reliable set of lane markings for a wide range of traffic flow patterns is of paramount importance for the operation of signal-controlled junctions. Although the lane-based method can be used to determine the

optimal lane markings and signal settings almost instantly, only one set of demand flow patterns can be taken into consideration at a time (Wong et. al. 2000; Wong and Wong, 2003a). In a multi-period design, two or more sets of demand flow patterns are usually involved in the analysis. The existing lane-based method must be applied individually and independently for each of these design periods, and multiple sets of lane markings may be obtained if there are remarkable variations in the input demand flows, which is impractical for real applications. It cannot be assumed that ordinary road users can respond effortlessly and automatically to repeated changes in lane markings on the approach to a signal-controlled junction. In daily operations, only one set of lane markings can be chosen to channel the traffic, and once established on the ground, the permitted lane movements (as indicated by the lane markings) should not be frequently altered.

Conventionally, a series of manual trial-and-error tests on lane-marking sets must be conducted until a satisfactory junction performance is observed for all of the design periods. However, the results of this kind of lane marking are always biased toward one of the design periods. A set of lane markings that may have been found to be very robust in one design period may not be so well suited to others. This is an unsuitable situation for complex demand patterns during different time segments of a day, such as the morning peak and evening peak periods, in which the demand flow patterns may be totally reversed. This failing of the conventional approaches has motivated this study to explore the possibility of extending the lane-based method to a multi-period design by establishing one reliable set of lane markings that can cope with different demand patterns at an isolated signal-controlled junction in such a way that the usual objective functions can all be optimized.

2. THE LANE-BASED FORMULATION

In this section, the formulation of the lane-based optimization method for multi-period designs at isolated signal-controlled junctions is given, including the required input data, the relevant control variables, and the various sets of governing (linear) constraints.

2.1 Input data

The required data inputs for the analysis of multi-period designs at isolated signal-controlled junctions are summarized as follows.

- Number of traffic arms: N_T .
- Number of pedestrian crossings: N_P .
- Demand flows: $\mathbf{Q} = (Q_{i,j,\pi}, j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi)$.
- Number of approaching lanes: $\mathbf{L} = (L_i, i = 1,\ldots,N_T)$.
- Number of exit lanes: $\mathbf{E} = (E_i, i = 1,\ldots,N_T)$.
- Minimum (maximum) cycle length: $c_{\min}(c_{\max})$.
- Minimum duration of green: $\mathbf{g} = (g_{i,j}, \forall j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T$ and $g_{i,1}, \forall i = 1,\ldots,N_P)$.
- Clearance times: $\boldsymbol{\omega} = (\omega_{u,v}, \forall (u,v) \in \Psi)$ [#].

[#] Ψ refers to a set that contains all of the traffic movements on entering a signal-controlled junction.

- Lane-saturation flows for straight-ahead movement: $\bar{\mathbf{s}} = (s_{i,k}, \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T)$.
- Turning radius: $\mathbf{r} = (r_{i,j,k}, \forall j = 1,\ldots,N_T - 1; \forall k = 1,\ldots,L_i;\ i = 1,\ldots,N_T)$.
- Acceptable degree of saturation: $\mathbf{p} = (p_{i,k}, \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T)$.
- Time difference between actual & effective greens: $\mathbf{e} = (e_{i,k}, \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T)$.

N_T specifies the number of traffic arms and N_P specifies the number of pedestrian crossings at a signal-controlled junction. Let **Q** be the traffic demand for different design periods π and Π represents the total number of design periods to be considered. From different approaches, the numbers of traffic lanes for junction entry **L** and exit **E** are also required in this lane-based analysis. The minimum (or maximum) cycle length $c_{\min}(c_{\max})$ provides the lower (or upper) limit for the selection of an appropriate operating cycle length in the optimization process. To meet concerns about safety, the minimum green durations **g** should be assigned to the traffic movement. To ensure the safe separation of all of the conflicting movements across a junction at successive green signals, a clearance time matrix ω must be provided in the design calculations. The lane-saturation flows for straight-ahead movements $\bar{\mathbf{s}}$ are lane capacities, and are estimated based on the geometric factors of the site, such as lane widths, lane types, and road gradients. If turning movements are permitted in a traffic lane, then the turning radius **r** and the proportion of turning traffic will affect the lane-saturation flows $\bar{\mathbf{s}}$. To prevent excessive delay, the acceptable degree of saturation **p** for each traffic lane is set to limit the maximum attainable volume to capacity ratio. In addition, the time difference between the effective green and the actual (display) green **e** is also inputted.

2.2 Control variables

The lane-based control variables for a multi-period analysis can be specified as follows. Let $\Lambda = (\Lambda_b, \Lambda_c)$ be the set of control variables, where Λ_b is the subset of the binary variables and Λ_c the subset of the continuous variables.

The subset Λ_b consists of the following binary variables:

- Permitted movements: $\boldsymbol{\Delta} = (\Delta_{i,j,k}, j = 1,\ldots,N_T - 1; k = 1,\ldots,L_i; i = 1,\ldots,N_T)$.
- Effective movements: $\boldsymbol{\delta} = (\delta_{i,j,k,\pi}, j = 1,\ldots,N_T - 1; k = 1,\ldots,L_i; i = 1,\ldots,N_T;$ $\pi = 1,\ldots,\Pi)$.
- Successor functions: $\boldsymbol{\Omega} = (\Omega_{i,j,l,m,\pi}, ((i,j),(l,m)) \in \Psi_s; \pi = 1,\ldots,\Pi)$.[*]

The set of permitted movements $\boldsymbol{\Delta}$ represents the lane-marking layout, which is identical in all of the design periods so that road users will always be guided by the same set of permitted movements as marked on the ground. However, a set of effective movements is introduced into the formulation that represents the set of actual movements in different design periods. As the utilization pattern may be different across different periods, an additional time dimension is given for the set of effective movements. A set of successor functions $\boldsymbol{\Omega}$ is introduced to control the relative signal display sequences for every pair of incompatible traffic movements. As the successor functions are (design) period dependent, a time dimension has also been incorporated.

[*] Ψ_s, which is a subset of Ψ, refers to incompatible movements only.

The subset Λ_c consists of the following continuous variables.

- Allocated flows: $\mathbf{q} = (q_{i,j,k,\pi}, j = 1,\ldots,N_T - 1; k = 1,\ldots,L_i;\ i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi)$.
- Common flow multiplier: μ.
- Cycle length: c.
- Reciprocal of cycle length: $\zeta = 1/c$.
- Start of green for movements: $\boldsymbol{\theta} = (\theta_{i,j,\pi}, j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T;\ \pi = 1,\ldots,\Pi$ and $j = 1; i = 1,\ldots,N_P; \pi = 1,\ldots,\Pi)$.
- Duration of green for movements: $\boldsymbol{\varphi} = (\phi_{i,j,\pi}, j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T;\ \pi = 1,\ldots,\Pi$ and $j = 1; i = 1,\ldots,N_P; \pi = 1,\ldots,\Pi)$.
- Start of green for traffic lanes: $\boldsymbol{\Theta} = (\Theta_{i,k,\pi}, k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi)$.
- Duration of green for traffic lanes: $\boldsymbol{\Phi} = (\Phi_{i,k,\pi}, k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi)$.

The set of allocated flows $\mathbf{q}$ gives the actual turning flows in traffic lanes in different design periods. By definition, if there is an ineffective movement, then the corresponding allocated flow must be zero. The common flow multiplier μ is a flow-scaling factor, and also serves as an indicator of the junction capacity. The cycle length c gives the actual time duration for a complete signal cycle, and its reciprocal is defined as $\zeta\ (= 1/c)$. The start of green for traffic movements is represented by $\boldsymbol{\theta}$, and the duration of green (or the actual green) is given by $\boldsymbol{\varphi}$. In the lane-based formulation, the start of green and duration of green can also be specified on a lane basis as $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, respectively.

2.3 Governing constraints

2.3.1 Flow conservation

For each period π, the traffic demand is multiplied by a common flow-multiplier μ. This factor represents the level of the scaled traffic that can be attained, so that the junction can still perform reasonably well. With these factored demands, the flow conservation constraint can be set as

$$\mu Q_{i,j,\pi} = \sum_{k=1}^{L_i} q_{i,j,k,\pi}, \quad \forall j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi, \tag{1}$$

where $Q_{i,j,\pi}$ is the traffic demand from arm i to arm j in the design period π, and $q_{i,j,k,\pi}$ is the traffic flow from arm i to arm j to be allocated on lane k in the design period π. The junction capacity is then determined from the critical movement of (i, j, k) of the design period π.

2.3.2 Minimum effective movement in a lane

In the lane-based formulation for multi-period analysis, effective movements are defined to represent the traffic movements in the allocated lane flows. To ensure that traffic flows are allocated on every traffic lane that approaches a signal-controlled

junction in all of the design periods, a constraint set of the minimum effective movement in a lane is introduced:

$$\sum_{j=1}^{N_T-1} \delta_{i,j,k,\pi} \geq 1, \quad \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi . \tag{2}$$

2.3.3 Maximum permitted movements at the exit

Due to the geometric restrictions of the junction, in which the number of exit lanes (downstream traffic lanes) is less than the number of entry lanes (upstream traffic lanes), limitations on the number of permitted movements upstream have to be provided. If too many permitted movements are allowed on entering a particular traffic arm with fewer exit lanes, then unnecessary traffic merging occurs, which is an unfavorable engineering design. Hence, the following constraint set is derived to help prevent such a design:

$$E_{\Gamma(i,j)} \geq \sum_{k=1}^{L_i} \Delta_{i,j,k}, \quad \forall j = 1,\ldots,N_T-1; i = 1,\ldots,N_T . \tag{3}$$

$\Gamma(i,j)$ is a global arm conversion function that equals $i+j$ if $i+j \leq N_T$ or $i+j-N_T$ if $i+j \geq N_T$.

2.3.4 Permitted movements across adjacent lanes

More than one traffic movement may be involved in a traffic approach to enter a signal-controlled junction, such as left-turning, right-turning, and straight-ahead movements, depending on the input demand pattern. It is a typical signal setting that part of the duration of green of these movements shares the same time slot. Thus, the movements that are permitted across adjacent lanes within the same approach have to be strictly controlled through the following constraint set, so that internal conflicts among traffic movements within an approach are absolutely expunged:

$$\begin{aligned} 1-\Delta_{i,j,k+1} \geq \Delta_{i,m,k}, \quad & \forall m = j+1,\ldots,N_T-1; j = 1,\ldots,N_T-2; \\ & k = 1,\ldots,L_i-1; i = 1,\ldots,N_T . \end{aligned} \tag{4}$$

2.3.5 Cycle length

For the practical design of the signal settings, the cycle length constraint is applied to confine the operational cycle length ζ to a specific range between a minimum cycle length $c_{\min}$ and a maximum cycle length $c_{\max}$:

$$\frac{1}{c_{\min}} \geq \zeta \geq \frac{1}{c_{\max}} . \tag{5}$$

2.3.6 Lane signal settings

If a permitted movement $\Delta_{i,j,k}$ =1 exists, which implies that a lane marking for traffic movement from arm *i* to arm *j* on lane *k* is given, then for each of the design periods π, the start of green for that movement (*i*, *j*) and the start of green for that particular traffic lane (*i*, *k*) must be equalized for consistency:

$$M(1-\Delta_{i,j,k}) \geq \Theta_{i,k,\pi} - \theta_{i,j,\pi} \geq -M(1-\Delta_{i,j,k}), \quad (6)$$

and a similar condition is also applied for the duration of green:

$$M(1-\Delta_{i,j,k}) \geq \Phi_{i,k,\pi} - \phi_{i,j,\pi} \geq -M(1-\Delta_{i,j,k}). \quad (7)$$

If there is a shared lane design, $\Delta_{i,j,k} = \Delta_{i,j+1,k} = 1$ for instance, then the constraint sets put forward the equalities $\Theta_{i,k,\pi} = \theta_{i,j,\pi} = \theta_{i,j+1,\pi}$ and $\Phi_{i,k,\pi} = \phi_{i,j,\pi} = \phi_{i,j+1,\pi}$, which affirms that the start of green for the traffic lane (i, k) and the traffic movements (i, j) and (i, j+1) are all forced to be identical, as are the duration of green for each of the design periods π.

If traffic movements are permitted on traffic lanes other than lane k, then the equality property is extended to synchronize the signal settings, including the start and duration of green, for the associated traffic lanes. If these traffic lanes permit other movements, then identical signal settings are carried out due to the shared lane effects. The equality propagates until all of the related traffic movements and lanes are linked up.

However, it is also worth reasserting that if the movement (i, j) is not permitted on lane k, $\Delta_{i,j,k} = 0$, then the relevant equality restriction is removed, and there is no longer signal synchronization between the traffic lane (i, k) and the traffic movement (i, j). Their signal settings can therefore be different.

2.3.7 Start of green

In the lane-based approach, all signal timings operate within a fixed signal cycle, and the desirable cycle length can be determined by the optimization of different operational criteria. This set of constraints is constructed to ensure that all of the starts of green from all of the design periods fall within the range of a signal cycle

$$1 \geq \theta_{i,j,\pi} \geq 0, \quad \forall j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi \quad (8)$$

for traffic movements and

$$1 \geq \theta_{i,1,\pi} \geq 0, \quad \forall i = 1,\ldots,N_P; \pi = 1,\ldots,\Pi \quad (9)$$

for pedestrian crossings.

2.3.8 Duration of green

As all traffic signals work within a signal cycle, no duration of green for either vehicular or pedestrian movements can be assigned that exceeds one signal cycle. As a safety consideration, however, each traffic movement in all of the design periods should be given a minimum duration of green. The following constraints are then set to fulfill these two operational requirements:

$$1 \geq \phi_{i,j,\pi} \geq g_{i,j}\zeta, \quad \forall j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi \quad (10)$$

for traffic movements and

$$1 \geq \phi_{i,1,\pi} \geq g_{i,1}\zeta, \quad \forall i = 1,\ldots,N_P; \pi = 1,\ldots,\Pi \quad (11)$$

for pedestrian crossings, where $g_{i,j}$ is the minimum duration of green for a signal group (for traffic or for pedestrian crossings) for vehicular movements from arm i to arm j, $j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi$ or pedestrian crossings $i = 1,\ldots,N_P; \pi = 1,\ldots,\Pi$ (j = 1 only).

2.3.9 Order of signal displays

In each design period π, any two signal groups (i, j) and (l, m) are said to be incompatible if a movement that is controlled by (i, j) is incompatible with a movement that is controlled by (l, m). The set of incompatible signal groups, Ψ_s, can therefore be derived from Ψ, which is the set of incompatible movements. For any two incompatible signal groups (i, j) and (l, m) in the design period π, the order of signal displays is governed by a successor function (Heydecker, 1992), $\Omega_{i,j,l,m,\pi}$, where $\Omega_{i,j,l,m,\pi} = 0$ if the start of green of signal group (l, m) follows that of signal group (i, j), and = 1 if the opposite is true. The following constraints must be set for the successor functions:

$$\Omega_{i,j,l,m,\pi} + \Omega_{l,m,i,j,\pi} = 1, \quad \forall\big((i,j),(l,m)\big) \in \Psi_s; \pi = 1,\ldots,\Pi\,. \tag{12}$$

2.3.10 Clearance time

For any pair of incompatible vehicular movements *u* and *v* (when both are permitted) as given by $\Delta_{i,j,k}$ and $\Delta_{l,m,n}$, respectively, the following constraint set must be formulated to ensure that sufficient clearance time $\omega_{u,v}$ is given to effectively separate the right-of-way of each movement in a signal cycle for all of the design periods:

$$\begin{aligned}\theta_{l,m,\pi} + \Omega_{i,j,l,m,\pi} + M(2 - \Delta_{i,j,k} - \Delta_{l,m,n}) \\ \geq \theta_{i,j,\pi} + \phi_{i,j,\pi} + \omega_{u,v}\zeta, \quad \forall (u,v) \in \Psi, \pi \in \Pi.\end{aligned} \tag{13}$$

If $u = (i, j, k)$ is a vehicular movement and $v = (l,1)$ is a pedestrian crossing, then the required clearance time constraint becomes:

$$\theta_{l,1,\pi} + \Omega_{i,j,l,1,\pi} + M(1 - \Delta_{i,j,k}) \geq \theta_{i,j,\pi} + \phi_{i,j,\pi} + \omega_{u,v}\zeta, \quad \forall (u,v) \in \Psi, \pi \in \Pi\,. \tag{14}$$

However, if $u = (l,1)$ and $v = (i, j, k)$, then the constraint set is modified to:

$$\theta_{l,m,\pi} + \Omega_{i,1,l,m,\pi} + M(1 - \Delta_{l,m,n}) \geq \theta_{i,1,\pi} + \phi_{i,1,\pi} + \omega_{u,v}\zeta_{\pi}, \quad \forall (u,v) \in \Psi, \pi \in \Pi\,. \tag{15}$$

2.3.11 Prohibited movement

For each design period π, if the vehicular flow *q* of the traffic movement (*i*, *j*) is allocated on traffic lane *k,* which implies that $q_{i,j,k,\pi} > 0$, then the corresponding movement is considered to be an effective movement and, by definition, $\delta_{i,j,k,\pi} = 1$. To prevent a negative allocated flow, $q_{i,j,k,\pi} \geq 0$ is also a necessary condition that must be applied to all of the design periods. Conversely, if $\delta_{i,j,k,\pi} = 0$, then the movement is not an effective movement, and the corresponding lane-allocated flow $q_{i,j,k,\pi}$ is prohibited and forced to be zero:

$$M\delta_{i,j,k,\pi} \geq q_{i,j,k,\pi} \geq 0, \quad \forall k = 1,\ldots,L_i; j = 1,\ldots,N_T - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi\,, \tag{16}$$

where *M* is an arbitrarily large positive constant.

2.3.12 Flow factor

As the allocation of traffic flow is based on queuing theory, the degree of saturation on a pair of adjacent lanes that show a common effective movement must be identical. Whenever there is a pair of effective movements, there must also be a pair of permitted movements, and the signal settings for these pairs of adjacent lanes have been set to be identical, as described in Section 2.3.6. To ensure identical degrees of saturation, it suffices to equalize the flow factors, which are defined as the total allocated flow divided by the saturation flow, for these adjacent lanes. For each design period, the flow factor $y_{i,k,\pi}$ is expressed as:

$$y_{i,k,\pi} = \frac{\sum_{j=1,\ldots,N_T-1} q_{i,j,k,\pi}}{s_{i,k,\pi}}, \quad \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi\,. \tag{17}$$

The turning proportion $P_{i,j,k,\pi}$ is given by:

$$P_{i,j,k,\pi} = \frac{q_{i,j,k,\pi}}{\sum_{m=1,\ldots,N_T-1} q_{i,m,k,\pi}}, \quad \begin{array}{l}\forall j = 1,\ldots,N_T-1; k = 1,\ldots,L_i; \\ i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi,\end{array} \tag{18}$$

and it can be shown that

$$y_{i,k,\pi} = \frac{1}{\bar{s}_{i,k}} \sum_{j=1,\ldots,N_T-1} \left(1 + \frac{1.5}{r_{i,j,k}}\right) q_{i,j,k,\pi}, \quad \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi\,. \tag{19}$$

Therefore, the following set of constraints can be used to enforce the equalized flow factors for each of the design periods:

$$M(2 - \delta_{i,j,k,\pi} - \delta_{i,j,k+1,\pi}) \geq$$

$$\frac{1}{\bar{s}_{i,k}} \sum_{j=1,\ldots,N_T-1} \left(1 + \frac{1.5}{r_{i,j,k}}\right) q_{i,j,k,\pi} - \frac{1}{\bar{s}_{i,k+1}} \sum_{j=1,\ldots,N_T-1} \left(1 + \frac{1.5}{r_{i,j,k+1}}\right) q_{i,j,k+1,\pi} \tag{20}$$

$$\geq -M(2 - \delta_{i,j,k,\pi} - \delta_{i,j,k+1,\pi}), \quad \forall k = 1,\ldots,L_i - 1; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi,$$

where M is an arbitrary large positive constant, and k (left) and k+1 (right) are adjacent lanes in arm i.

2.3.13 Maximum acceptable degree of saturation

Given $p_{i,k}$ as the maximum allowable degree of saturation on lane k in arm i, the practical degree of saturation $\rho_{i,k,\pi}$ for each of the design periods has to fulfill the inequality

$$\rho_{i,k,\pi} = \frac{y_{i,k,\pi}}{\Phi_{i,k,\pi} + e\zeta} \leq p_{i,k}, \quad \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi\,. \tag{21}$$

From equation (19), the following constraint can be replaced to ensure that the degree of saturation cannot exceed the maximum acceptable limit:

$$\Phi_{i,k,\pi} + e\zeta \geq \frac{1}{p_{i,k}\bar{s}_{i,k}} \sum_{j=1,\ldots,N_T-1} \left(1 + \frac{1.5}{r_{i,j,k}}\right) q_{i,j,k,\pi}, \forall k = 1,\ldots,L_i; i = 1,\ldots,N_T; \quad \begin{array}{l}\\ \pi = 1,\ldots,\Pi.\end{array} \tag{22}$$

2.3.14 Maximum effective movements

One of the design objectives in the lane-based optimization for multi-period analysis is to establish a set of permitted movements that is shown identically in all design periods, so that road users will always be guided by the same set of lane markings on the ground when approaching a junction. According to the demand patterns that are given in various design periods, however, the actual lane-flow patterns can be allocated in a different way so that the overall junction performance is optimized. The effective movements for each design period may thus be different, but must be derived from the unique set of permitted movements. The following constraint set must be established to achieve this design purpose:

$$\Delta_{i,j,k} \geq \delta_{i,j,k,\pi}, \quad \forall j = 1,\ldots,N_T - 1; k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi \,. \tag{23}$$

Mathematically, the set of effective movements for different design periods is the subset of the set of permitted movements. A special case of this constraint set, in which the permitted movements equal the effective movements in all of the design periods, is given in equation (24). The flow patterns that are allocated in all of the design periods are then forced to be identical, but the flow magnitudes are still controlled by the flow inputs, and should be governed by the flow conservation constraints in Section 2.3.1. Optionally, this special form of the maximum effective movement constraint can be used to replace the constraint set of equation (23) to obtain a more stringent lane-marking design in a multi-period analysis. Equation (24) also serves as a compulsory substitution of equation (23) in the formulation when the problem is simplified to a single period design in which $\Pi = 1$:

$$\Delta_{i,j,k} = \delta_{i,j,k,\pi}, \quad \forall j = 1,\ldots,N_T - 1; k = 1,\ldots,L_i; i = 1,\ldots,N_T; \pi = 1,\ldots,\Pi \,. \tag{24}$$

2.3.15 Other signal group constraints

There may be other practicable constraints that can be applied to the relative timing of the starts and ends of green for different signal groups. These constraints can be set as follows. Let $z_{i,j,l,m,\pi}$ be the relative time (for the start or end of green) that is required to appear when measured from signal group (i,j) to group (l,m) in the design period π. For the starts of green

$$\theta_{i,j,\pi} + z_{i,j,l,m,\pi} = \theta_{l,m,\pi}, \tag{25}$$

and for the ends of green

$$\theta_{i,j,\pi} + \phi_{i,j,\pi} + z_{i,j,l,m,\pi} = \theta_{l,m,\pi} + \phi_{l,m,\pi} \,. \tag{26}$$

These constraints are mainly used to restrain two signal groups so that they start or end simultaneously (i.e., $z_{i,j,l,m,\pi} = 0$) for signal timing operations at a junction.

3. OPTIMIZATION CRITERIA FOR MULTI-PERIOD ANALYSIS OF ISOLATED JUNCTIONS

There are generally three optimization criteria for traffic signal designs in a multi-period analysis of isolated signal control junctions: capacity maximization, cycle length minimization, and delay minimization. The optimization problems of capacity maximization and cycle length minimization can be effectively formulated by setting up a linear objective function as a Binary-Mix-Integer-Linear-Program (BMILP). As the

delay minimization problem involves a non-linear delay objective function, the optimization problem must be formulated as a Binary-Mix-Integer-Non-Linear-Program (BMINLP). Detailed discussions of the mathematical framework for the capacity, cycle length, and delay optimization will be given in the following sections.

3.1 Capacity maximization

The junction capacity maximization problem for the multi-period design can be effectively formulated as the BMILP

$$\underset{\Lambda=(\Lambda_b,\Lambda_c)}{\text{Maximize}} \quad \mu \tag{27}$$

subject to the linear constraints in (1–16, 20, 23 [or 24], 25 and 26).

The mathematical program is linear in nature, and can be solved effectively by standard branch-and-bound routines.

3.2 Cycle length minimization

The cycle length minimization problem determines the shortest cycle length and corresponding signal settings that can handle the existing demand flow pattern. The problem is formulated to maximize the reciprocal of cycle length ζ with $\mu = 1.0$, subject to the set of constraints that has been specified in Section 2.3. The problem can be formulated as follows:

$$\underset{\Lambda=(\Lambda_b,\Lambda_c)}{\text{Maximize}} \quad \zeta \tag{28}$$

subject to the linear constraints in (1–16, 20, 23 [or 24], 25 and 26) and $\mu = 1.0$.

The cycle length minimization problem is again a Binary-Mix-Integer-Linear-Program (BMILP), and standard branch-and-bound techniques can be applied to solve the problem effectively.

3.3 Delay minimization

As the delay function is generally non-linear in nature, the problem must be formulated as a BMINLP problem:

$$\underset{\Lambda=(\Lambda_b,\Lambda_c)}{\text{Minimize}} \quad D \tag{29}$$

subject to the linear constraints in (1–16, 20, 23 [or 24], 25 and 26) and $\mu = 1.0$, where D defines the total delay of the junction.

4. SOLUTION ALGORITHMS

In this section, solution algorithms are developed to solve the capacity maximization, cycle length minimization, and delay minimization problems in the lane-based formulation for the multi-period analysis of isolated signal-controlled junctions.

4.1 Capacity maximization and cycle length minimization

The standard branch-and-bound technique has been found to be effective to solve capacity maximization and cycle length minimization problems that are formulated as

BMILPs. In this multi-period formulation, the capacity maximization and cycle length minimization problems are similar problem types and possess a linear objective function and a set of linear constraints, although the problem sizes are greater. The multi-period capacity maximization and cycle length minimization problems can also be formulated as BMILP problems, and the standard branch-and-bound algorithm is again employed as the solution method. Details of the implementation of the standard branch-and-bound routine can be found in Vanderbei, (2001) and Kolman and Beck (1995). A computer package – the MPL modeling system – that integrates a CPLEX solver (Maximal, 2002) is used to implement the branch-and-bound algorithm to solve the present BMILP for capacity maximization and cycle length minimization problems.

4.2 Delay minimization

As the feasible solution region of the delay minimization problems in the lane-based formulation is non-linear and non-convex, no standard method can be applied and no method guarantees a global optimum solution. It has been found that the line search algorithm provides a good balance between computational effort and the quality of the solution in resolving ordinary lane-based delay minimization problems (Wong et al., 2002; Wong and Wong, 2003b), and thus the line search technique will be explored further to solve the multi-period design problems.

The following optimizing modules are defined for the development of delay optimization heuristics.

$\mathrm{MLB1}[c;(\boldsymbol{\delta},\boldsymbol{\Delta},\mathbf{q},\boldsymbol{\Omega},\mu)]$ is the abstract form of the lane-based capacity maximization module for a multi-period analysis. MLB1 optimizes the common flow multiplier μ, the effective movements $\boldsymbol{\delta}$, the permitted movements $\boldsymbol{\Delta}$, the allocated flows $\mathbf{q}$, and the successor functions $\boldsymbol{\Omega}$ with a fixed cycle length c.

$\mathrm{MLB2}[\mu;(\boldsymbol{\delta},\boldsymbol{\Delta},\mathbf{q},\boldsymbol{\Omega},c)]$ is the abstract form of the lane-based cycle length minimization module for a multi-period analysis. MLB2 optimizes the cycle length c, the effective movements $\boldsymbol{\delta}$, the permitted movements $\boldsymbol{\Delta}$, the allocated flows $\mathbf{q}$, and the successor functions $\boldsymbol{\Omega}$ with a fixed common flow multiplier μ.

$\mathrm{GB}[(c_0,\boldsymbol{\delta}_\pi,\boldsymbol{\Delta}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi);(c_\pi,\boldsymbol{\psi}_\pi,D_\pi)]$ is the abstract form of the group-based delay minimization module for the evaluation of a single period π. GB optimizes the total rate of delay based on a set of lane-based variables $(\boldsymbol{\delta}_\pi,\boldsymbol{\Delta}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi)$ with an initial cycle length c_0, where $\boldsymbol{\delta}_\pi$ are the effective movements, $\boldsymbol{\Delta}_\pi$ the permitted movements, $\mathbf{q}_\pi$ the allocated flows and $\boldsymbol{\Omega}_\pi$ the successor functions for the period π. GB outputs the optimized cycle length c_π, the optimized group-based settings $\boldsymbol{\psi}_\pi$, and the minimized rate of delay D_π for the design period π.

The heuristic line search technique for optimizing the total rate of delay of a signal-controlled junction in a lane-based multi-period formulation can be described by the following step-by-step pseudo-codes.

Step 1: Set $\mu = 1.0$, $t = 0$, and evaluate $\mathrm{MLB2}[\mu;(\boldsymbol{\delta},\boldsymbol{\Delta},\mathbf{q},\boldsymbol{\Omega},\tilde{\chi})]$, where $\boldsymbol{\delta} = (\boldsymbol{\delta}_\pi, \pi = 1,\ldots,\Pi)$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\pi, \pi = 1,\ldots,\Pi)$, $\mathbf{q} = (\mathbf{q}_\pi, \pi = 1,\ldots,\Pi)$, and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_\pi, \pi = 1,\ldots,\Pi)$.

Step 2: For $\pi = 1$ to Π

Evaluate GB[$(\tilde{\chi},\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi);(c_{\tilde{\chi},\pi},\boldsymbol{\psi}_{\tilde{\chi},\pi},D_{\tilde{\chi},\pi})$] using the results from MLB2 $(\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi)$ as inputs, where $c_{\tilde{\chi},\pi}$ is the optimized cycle length, $\boldsymbol{\psi}_{\tilde{\chi},\pi}$ are the optimized group-based settings, and $D_{\tilde{\chi},\pi}$ is the minimized rate of delay for period π, corresponding to the initial cycle length $\tilde{\chi}$.

Store the solution results

$$\left(\boldsymbol{\delta}_{\tilde{\chi},\pi}=\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_{\tilde{\chi},\pi}=\boldsymbol{\Lambda}_\pi,\mathbf{q}_{\tilde{\chi},\pi}=\mathbf{q}_\pi,\boldsymbol{\Omega}_{\tilde{\chi},\pi}=\boldsymbol{\Omega}_\pi,c_{\tilde{\chi},\pi},\boldsymbol{\psi}_{\tilde{\chi},\pi},D_{\tilde{\chi},\pi}\right).$$

Next π.

Step 3: Set $\chi = c_{\max} - tS$.

Step 4: Check that $\chi \le \tilde{\chi}$, and then go to Step 8.

Step 5: Evaluate MLB1$[\chi;(\boldsymbol{\delta},\boldsymbol{\Lambda},\mathbf{q},\boldsymbol{\Omega},\mu)]$.

Step 6: For $\pi = 1$ to Π.

Evaluate GB[$(\chi,\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi);(c_{\chi,\pi},\boldsymbol{\psi}_{\chi,\pi},D_{\chi,\pi})$] using the results from MLB1 $(\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_\pi,\mathbf{q}_\pi,\boldsymbol{\Omega}_\pi)$ as inputs, where $c_{\chi,\pi}$, is the optimized cycle length, $\boldsymbol{\psi}_{\chi,\pi}$ are the optimized group-based settings, and $D_{\chi,\pi}$ is the minimized rate of delay for the period π, corresponding to the initial cycle length χ.

Store the solution results

$$\left(\boldsymbol{\delta}_{\chi,\pi}=\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_{\chi,\pi}=\boldsymbol{\Lambda}_\pi,\mathbf{q}_{\chi,\pi}=\mathbf{q}_\pi,\boldsymbol{\Omega}_{\chi,\pi}=\boldsymbol{\Omega}_\pi,c_{\chi,\pi},\boldsymbol{\psi}_{\chi,\pi},D_{\chi,\pi}\right).$$

Next π.

Step 7: Set $t = t+1$ and go to Step 3.

Step 8: Calculate the total delay $\tilde{D}_\chi = \sum_{\pi=1}^{\Pi} D_{\chi,\pi} W_\pi$ for all χ (including $\tilde{\chi}$).

Step 9: Determine the optimal result as that with smallest $\tilde{D}_\chi$, and the solution process is complete.

Initially, the lane-based cycle length minimization module MLB2 is used to determine the minimum cycle length $\tilde{\chi}$ (which also contributes one of the initial cycle lengths) with $\mu = 1.0$. The MLB2 output vector $(\boldsymbol{\delta},\boldsymbol{\Lambda},\mathbf{q},\boldsymbol{\Omega})$, which includes the effective movements, permitted movements, allocated flows, and successor functions, is considered as a fixed input in the group-based delay minimization module GB, which also forms part of the solution vector $\left(\boldsymbol{\delta}_{\tilde{\chi},\pi}=\boldsymbol{\delta}_\pi,\boldsymbol{\Lambda}_{\tilde{\chi},\pi}=\boldsymbol{\Lambda}_\pi,\mathbf{q}_{\tilde{\chi},\pi}=\mathbf{q}_\pi,\boldsymbol{\Omega}_{\tilde{\chi},\pi}=\boldsymbol{\Omega}_\pi\right)$. The optimized cycle length $c_{\tilde{\chi},\pi}$ and corresponding signal settings $\boldsymbol{\psi}_{\tilde{\chi},\pi}$ for each design period π are optimized individually using the GB module to achieve the optimal delay $D_{\tilde{\chi},\pi}$ in Step 3. The solution vector $\left(\boldsymbol{\delta}_{\tilde{\chi},\pi},\boldsymbol{\Lambda}_{\tilde{\chi},\pi},\mathbf{q}_{\tilde{\chi},\pi},\boldsymbol{\Omega}_{\tilde{\chi},\pi},c_{\tilde{\chi},\pi},\boldsymbol{\psi}_{\tilde{\chi},\pi},D_{\tilde{\chi},\pi}\right)$ becomes one of the possible optimal settings, and is recorded in a solution array for further comparison in Step 8 and Step 9. In the heuristic line search algorithm, solution candidates are derived from different initial cycle lengths. Once the minimum cycle length is located, other feasible initial cycle lengths can be determined by $\chi = c_{\max} - tS$ in Step 4, where $c_{\max}$ is the exogenously-defined maximum cycle length, t is a looping control parameter, and S is the step size that specifies the resolution of the trial initial cycle lengths. The initial cycle length is fixed and then applied to the lane-based capacity maximization module MLB1. Again, the output vector of the lane-based model

$(\boldsymbol{\delta},\boldsymbol{\Delta},\mathbf{q},\boldsymbol{\Omega})$ in Step 5 forms the necessary input of the group-based module GB if the corresponding initial cycle length is greater than the minimum cycle length. In Step 6, the group-based delay minimization module GB is evaluated individually for each of the design periods, and the subsequent model results are stored in the solution array. Set $t = t + 1$ in Step 7, and a new initial cycle length can be determined by $\chi = c_{\max} - tS$. Steps 5 and 6 are repeated until the initial cycle length is less than or equal to the minimum cycle length $\chi \le \widetilde{\chi}$, as given in Step 1. For every initial cycle length χ, the total delay that combines all of the design periods can be evaluated by $\widetilde{D}_{\chi} = \sum_{\pi=1}^{\Pi} D_{\chi,\pi} W_{\pi}$ in Step 9, where W_{π} is a weighting factor for the design period π to account for any non-uniform duration among the various design periods. By choosing the smallest $\widetilde{D}_{\chi}$ from the set of calculated total delays, the corresponding optimal signal settings for each of the design periods can be retrieved from the solution array $\left(\boldsymbol{\delta}_{\chi,\pi},\boldsymbol{\Delta}_{\chi,\pi},\mathbf{q}_{\chi,\pi},\boldsymbol{\Omega}_{\chi,\pi},c_{\chi,\pi},\boldsymbol{\psi}_{\chi,\pi},D_{\chi,\pi}\right)$ for all χ (including $\widetilde{\chi}$).

5. NUMERICAL EXAMPLES

In this section, the capacity maximization, cycle length minimization, and delay minimization problems are solved to demonstrate the effectiveness and efficiency of the lane-based optimization method for the multi-period analysis of isolated signal-controlled junctions. A four-arm junction with two pedestrian crossings is studied, which is the same junction that is shown in Figure 1. The relevant input data and assumptions are given as follows. There are four approaching lanes on each arm of the junction example. The number of exit lanes on each arm is equal to the number of approaching lanes on that arm, except for Arm 4, where there is only one exit lane for traffic from the other arms. Two pedestrian crossings are located on Arm 3. The saturation flows for straight-ahead movements on the kerbside and offside lanes are taken to be 1,965 and 2,105 pcu/h/lane, respectively. The saturation flows have not yet been fixed, but will vary according to the presence and intensity of turning traffic. It is assumed that all of the turning radii are 12 meters. The maximum cycle length is set to be 120 seconds, and the maximum acceptable degree of saturation is 90% on all lanes. The minimum green duration is 5 seconds for all traffic movements, and the minimum green for the pedestrian crossings is 20 seconds. The effective green is always 1 second longer than the actual green in the calculation. Three design periods are studied, and the traffic demands divided into morning peak, off-peak, and evening peak periods are given in Table 1. The required clearance time for any two conflicting movements (including both traffic and pedestrian movements) is set to be 6 seconds. A 2-second reduction in the clearance time is set for the following conflicting pairs: traffic movement following a pedestrian movement or a pair of traffic and pedestrian movements that belong to the same approach. All of the left-turning and straight-ahead timings are set to end at the same time for all cases (a usual practice in Hong Kong).

5.1 Capacity maximization

The problem of capacity maximization is that of maximizing the common flow multiplier μ subject to a set of linear constraints in (1–16, 20, 23, 25 and 26). The problem is formulated as a Binary-Mixed-Integer-Linear-Program (BMILP), and the

standard branch-and-bound technique has been applied to solve it. The model results are given as follows. The maximized common flow multiplier is 1.078, which implies that the junction reserve capacity is 7.8%. The corresponding optimal cycle length is found to be located at the maximum limit of 120 seconds. Figure 2 plots the patterns of the permitted movements and the effective movements for the three design periods. The associated optimal allocated flows and signal settings are given in Tables 2 to 4.

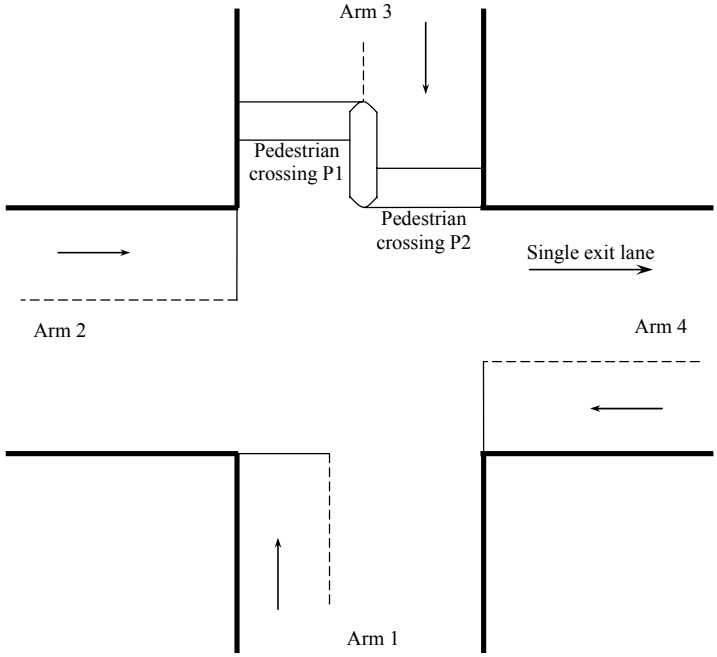


FIGURE 1: Layout of the example four-arm junction

TABLE 1: Traffic demands of a three-period design for the four-arm junction example

From arm	To arm*	Morning peak demand in pcu/h ($\pi = 1$)	Off-peak demand in pcu/h ($\pi = 2$)	Evening peak demand in pcu/h ($\pi = 3$)
1	2	400	400	550
	3	500	200	300
	4	200	100	150
2	1	500	100	200
	3	200	100	150
	4	600	400	650
3	1	400	300	50
	2	350	300	400
	4	350	300	450
4	1	50	100	600
	2	500	400	700
	3	450	400	500

* Represents the global destination arms.

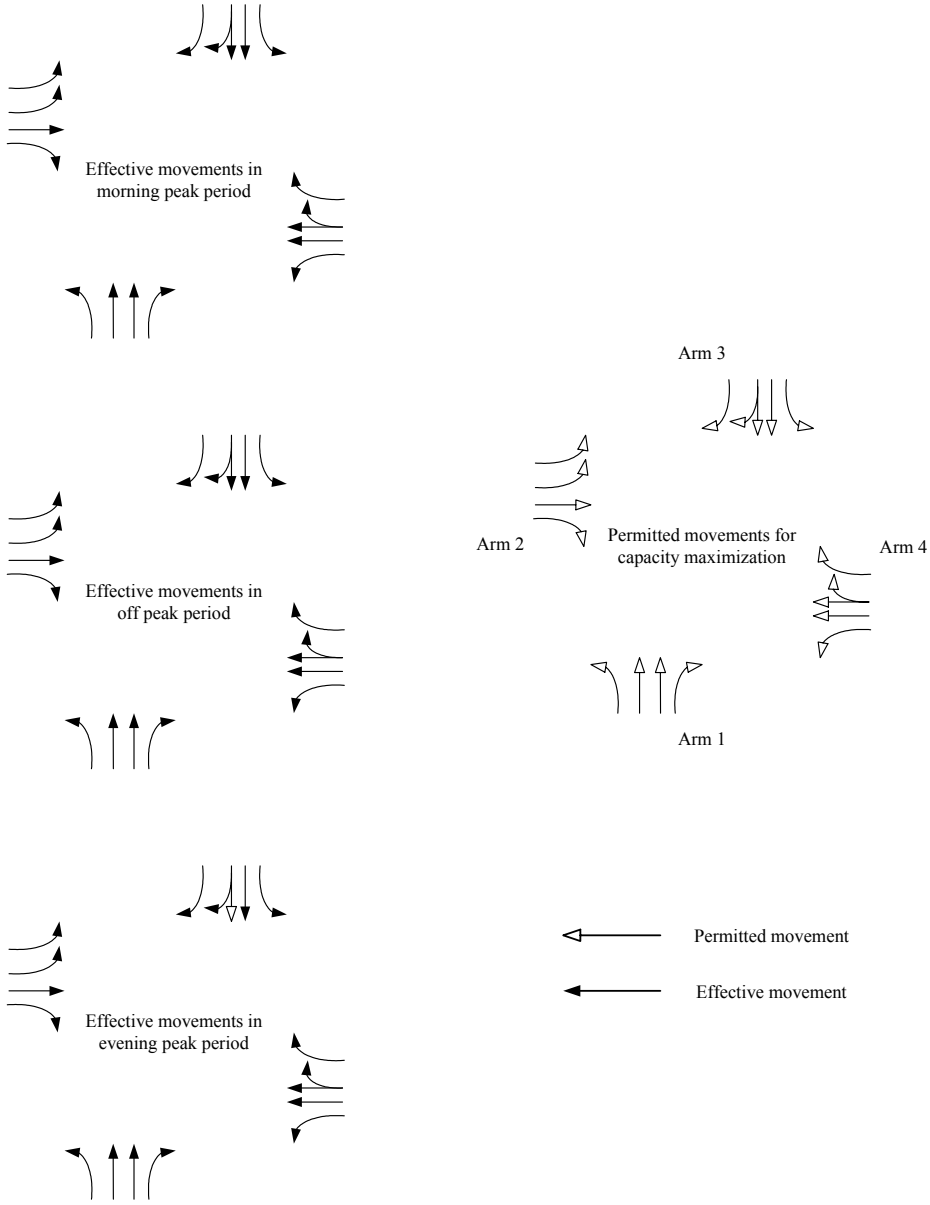


FIGURE 2: Arrangements of the permitted and effective movements for capacity maximization in the three-period design problem

TABLE 2: Lane-based model results for capacity maximization in the morning peak period (μ = 1.078 and cycle length = 120.00 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		400.0			400.0	1.00	1746.7	0.2290	64.0	0.4292	0.0
1	2			250.0		250.0	0.00	2105.0	0.1188	17.1	0.8349	47.0
1	3			250.0		250.0	0.00	2105.0	0.1188	17.1	0.8349	47.0
1	4				200.0	200.0	1.00	1871.1	0.1069	15.4	0.8349	46.0
2	1			96.6		96.6	1.00	1746.7	0.0553	40.0	0.1657	0.9
2	2			103.4		103.4	1.00	1871.1	0.0553	40.0	0.1657	0.9
2	3				600.0	600.0	0.00	2105.0	0.2850	41.0	0.8349	0.0
2	4	500.0				500.0	1.00	1871.1	0.2672	38.4	0.8349	0.0
3	1				350.0	350.0	1.00	1746.7	0.2004	48.7	0.4941	66.3
3	2	264.6				264.6	0.00	2105.0	0.1257	18.1	0.8349	96.9
3	3	135.4	114.8			250.2	0.46	1990.8	0.1257	18.1	0.8349	96.9
3	4		235.2			235.2	1.00	1871.1	0.1257	18.1	0.8349	96.9
4	1	50.0				50.0	1.00	1746.7	0.0286	6.0	0.5725	85.9
4	2		335.4			335.4	0.00	2105.0	0.1593	22.9	0.8349	69.0
4	3		164.6	151.9		316.4	0.48	1985.9	0.1593	22.9	0.8349	69.0
4	4			298.1		298.1	1.00	1871.1	0.1593	22.9	0.8349	69.0
	P1									20.0*		96.9
	P2									20.0*		42.3

* Represents the actual green times.

TABLE 3: Lane-based model results for capacity maximization in the off-peak period (μ = 1.078 and cycle length = 120.00 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		400.0			400.0	1.00	1746.7	0.2290	32.9	0.8349	0.0
1	2			100.0		100.0	0.00	2105.0	0.0475	6.8	0.8349	26.1
1	3			100.0		100.0	0.00	2105.0	0.0475	6.8	0.8349	26.1
1	4				100.0	100.0	1.00	1871.1	0.0534	7.7	0.8349	114.1
2	1			48.3		48.3	1.00	1746.7	0.0276	22.8	0.1453	61.9
2	2			51.7		51.7	1.00	1871.1	0.0276	22.8	0.1453	61.9
2	3				400.0	400.0	0.00	2105.0	0.1900	27.3	0.8349	57.4
2	4	100.0				100.0	1.00	1871.1	0.0534	7.7	0.8349	6.8
3	1				300.0	300.0	1.00	1746.7	0.1718	45.7	0.4514	6.8
3	2	212.5				212.5	0.00	2105.0	0.1010	14.5	0.8349	37.9
3	3	87.5	111.1			198.6	0.56	1967.4	0.1010	14.5	0.8349	37.9
3	4		188.9			188.9	1.00	1871.1	0.1010	14.5	0.8349	37.9
4	1	100.0				100.0	1.00	1746.7	0.0573	51.7	0.1330	57.4
4	2		283.3			283.3	0.00	2105.0	0.1346	19.3	0.8349	89.7
4	3		116.7	148.1		264.8	0.56	1967.4	0.1346	19.3	0.8349	89.7
4	4			251.9		251.9	1.00	1871.1	0.1346	19.3	0.8349	89.7
	P1									20.0*		37.9
	P2									20.0*		55.4

* Represents the actual green times.

TABLE 4: Lane-based model results for capacity maximization in the evening peak period (μ = 1.078 and cycle length = 120.00 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		550.0			550.0	1.00	1746.7	0.3149	45.3	0.8349	0.0
1	2			150.0		150.0	0.00	2105.0	0.0713	10.2	0.8349	35.0
1	3			150.0		150.0	0.00	2105.0	0.0713	10.2	0.8349	35.0
1	4				150.0	150.0	1.00	1871.1	0.0802	11.5	0.8349	33.7
2	1			72.4		72.4	1.00	1746.7	0.0415	28.7	0.1731	120.0
2	2			77.6		77.6	1.00	1871.1	0.0415	28.7	0.1731	120.0
2	3				650.0	650.0	0.00	2105.0	0.3088	44.4	0.8349	104.4
2	4	200.0				200.0	1.00	1871.1	0.1069	15.4	0.8349	120.0
3	1				450.0	450.0	1.00	1746.7	0.2576	37.0	0.8349	62.3
3	2	50.0				50.0	0.00	2105.0	0.0238	15.4	0.1855	84.0
3	3	0.0	200.0			200.0	1.00	1871.1	0.1069	15.4	0.8349	84.0
3	4		200.0			200.0	1.00	1871.1	0.1069	15.4	0.8349	84.0
4	1	600.0				600.0	1.00	1746.7	0.3435	58.6	0.7031	20.4
4	2		420.8			420.8	0.00	2105.0	0.1999	28.7	0.8349	50.3
4	3		279.2	125.9		405.1	0.31	2026.3	0.1999	28.7	0.8349	50.3
4	4			374.1		374.1	1.00	1871.1	0.1999	28.7	0.8349	50.3
	P1									20.0*		96.0
	P2									20.0*		102.4

* Represents the actual green times.

In Figure 2, the optimal pattern of the permitted movements for maximizing the junction capacity is represented by the set of hollow arrows (shown on the right-hand side of the diagram). The three sets of effective movements for the different demand profiles of the three design periods are marked as solid arrows. It is found that the patterns of the effective movements in the morning peak and off-peak periods are identical to the set of permitted movements, and thus all of the hollow arrows in these two design periods are filled in. It can also be observed that there is one hollow arrow (the movement from Arm 3 to Arm 1 on lane 3 (3.1.3) in the evening peak period design, and the associated movement is not considered to be effective, although the permitted movement still exists as a road marking for guidance.

In Tables 2, 3, and 4, which provide the lane flow and signal timing details for the morning peak, off-peak, and evening peak design periods, respectively, column (1) gives the origin arms and column (2) specifies the traffic lanes. In columns (3-6), which represent the different destination arms, the resulting allocated flows on the lanes are distributed accordingly. Summing the allocated flows vertically for each arm gives the input demand flows, and summing them horizontally for each traffic lane gives the sum of the lane flows, which is shown in column (7). The resulting turning proportions are collected in column (8). The revised lane-saturation flows, which take the turning proportions into consideration, are displayed in column (9). The flow factor (ratio) in column (10) can be easily deduced by dividing the sum of the lane flows by the saturation flows. With the effective green times in column (11) optimized from the models, the degree of saturation for each traffic lane can be calculated, as shown in column (12). The starts of green for each traffic lane are provided in column (13). All of the end of green times can be obtained by adding the start of green times to the actual green times, which are 1 second less than the effective green times in this study. The allocated lane-flow patterns for the effective movements are identically matched. The magnitudes of all of the allocated flows are controlled by the input demand flows for different design periods, and all of the effective movements from different design periods are established from and confined by the set of permitted movements.

It is worth noting from Figure 2 that the movement from Arm 3 to Arm 1 on lane 3 (3,1,3) is an effective movement in the morning peak and off-peak periods, but only a permitted movement in the evening peak period. Correspondingly, Tables 2 and 3 show that traffic lanes 2, 3, and 4 from Arm 3 belong to the same traffic stream due to the shared-lane effects for the movements of (3,1,3) and (3,2,3). Their resulting flow factors, degree of saturation, and signal timings must therefore be identical. However, during the evening peak period, the movement (3,1,3) is a permitted movement, but is not effectively used (the allocated flow equals zero), and thus the related traffic lanes 2 and 3 from Arm 3 need not be forced to form a single traffic stream. In Table 4, it can be seen that the flow factors and degree of saturation of these lanes become disparate. As the permitted movement (3,1,3) is still present as a shared lane-marking with the movement (3,2,3), the signal settings for these two movements, including the start and duration of green (= effective greens) have to be constrained to operate in a synchronized manner for safety reasons.

To obtain a better account of the performance of the proposed algorithms, some model statistics are given in the following. The lane-based capacity maximization model for the junction example in a three-period analysis possesses 1,856 linear constraints, 323 continuous variables, and 466 integer variables. Obtaining the required solution takes around 14 minutes on a Pentium-4 1.8 GHz processor.

5.2 Cycle length minimization

The same four-arm junction with two pedestrian crossings is studied for the cycle length minimization problem. All of the input data and assumptions as given in the capacity maximization problems are duplicated. The optimization objective is to maximize the reciprocal of the cycle time (= minimize the cycle length), subject to a set of linear constraints in (1–16, 20, 23, 25 and 26) and $\mu = 1.0$ (i.e., analyzed according to the existing demand flow patterns). Again, the problem is a Binary-Mixed-Integer-Linear-Program (BMILP) and the standard branch-and-bound technique can be applied to solve the problem. The model results for the cycle length minimization are given as follows. The minimum cycle length is 88.13 seconds. The patterns of the permitted movements and the effective movements for the three design periods are plotted in Figure 3. The associated optimal allocated flows and signal settings are given in Tables 5 to 7.

Obviously, the pattern of the permitted movements in a cycle length minimizing setting is quite different from those in a capacity maximizing setting, as a comparison of Figures 2 and 3 demonstrates. One more shared lane-marking for the straight-ahead and right-turn movements has been established in Arm 1, and two right-turn markings are given, instead of two left-turn markings, in Arm 2 for the cycle length minimization. The demand flow patterns that are used are identical for the capacity maximization and cycle length minimization problems, and the different lane-usage pattern is merely the result of switching the objective function for optimization. It is shown that only the effective movements in the morning peak period are identical to the permitted movements. Two different ineffective movements are produced in the off-peak and evening peak periods, as indicated by the hollow arrows. Due to the different demand patterns, the ineffective movements are shown in a different way for the off-peak and evening peak periods. This simply demonstrates the extraordinary capability of the present lane-based optimization model to handle very different demand patterns in a multi-period design. The model is able to develop a single set of permitted movements that can entertain very different demand flow patterns in different design periods.

Tables 5 to 7 tabulate the allocated flows and signal settings for the three design periods of the cycle length minimization problem. The table formats are the same as those that are given in Tables 2 to 4, and detailed explanations of the table entries can be found in the previous section. The size of the cycle length minimization model is exactly the same as the capacity maximization model. There are 1,856 linear constraints, 323 continuous variables, and 466 binary variables. As the model adopts a different objective function, the optimal solution region may be shifted, and the solution takes a little longer at around 52 minutes on a Pentium-4 1.8 GHz processor.

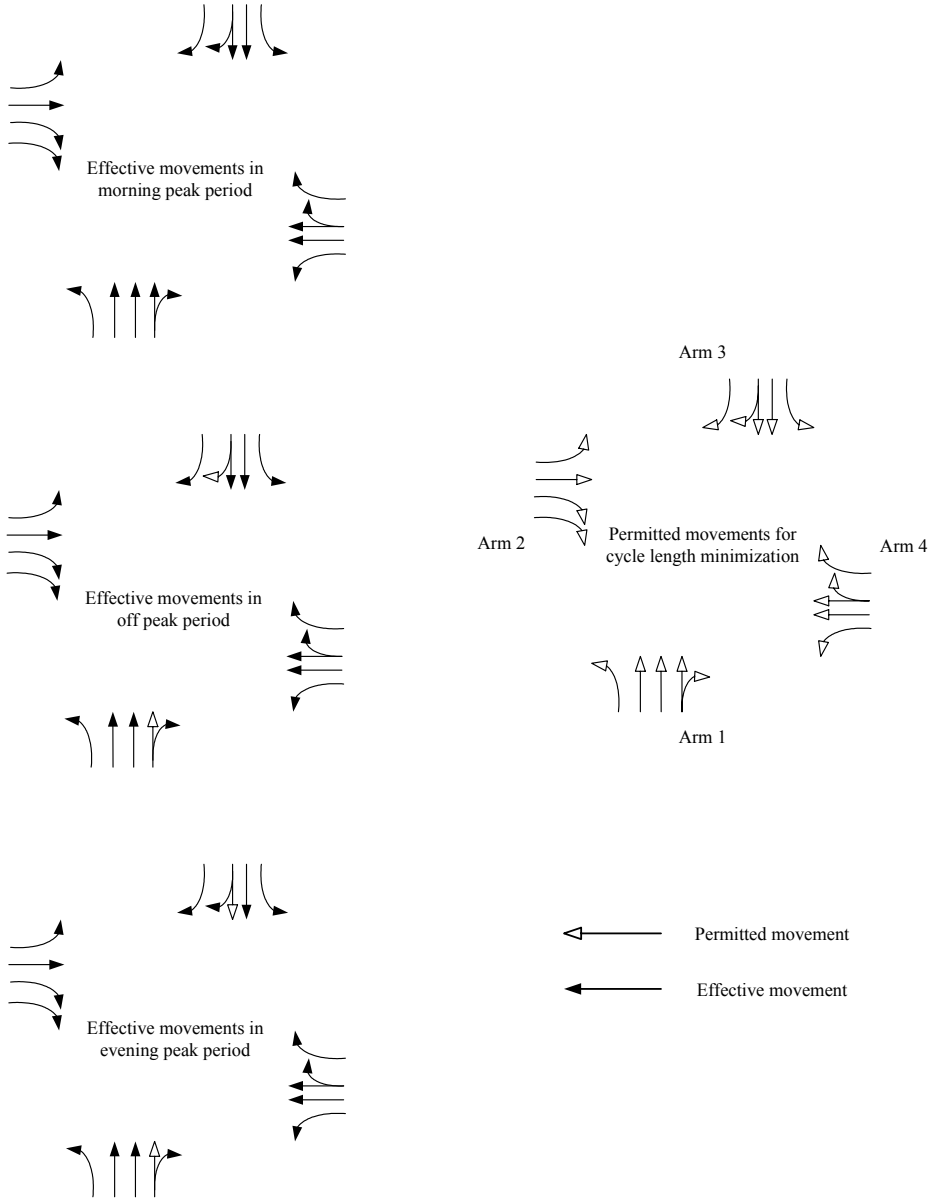


FIGURE 3: Arrangement of the permitted and effective movements for cycle length minimization in the three-period design problem

TABLE 5: Lane-based model results for cycle length minimization in the morning peak period (μ = 1.00 and cycle length = 88.13 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		400.0			400.0	1.00	1746.7	0.2290	44.2	0.4571	0.0
1	2			241.7		241.7	0.00	2105.0	0.1148	11.2	0.9000	32.9
1	3			241.7		241.7	0.00	2105.0	0.1148	11.2	0.9000	32.9
1	4			16.7	200.0	216.7	0.92	1887.2	0.1148	11.2	0.9000	32.9
2	1			200.0		200.0	1.00	1746.7	0.1145	22.3	0.4528	5.6
2	2				600.0	600.0	0.00	2105.0	0.2850	27.9	0.9000	0.0
2	3	250.0				250.0	1.00	1871.1	0.1336	13.1	0.9000	0.0
2	4	250.0				250.0	1.00	1871.1	0.1336	13.1	0.9000	0.0
3	1				350.0	350.0	1.00	1746.7	0.2004	19.6	0.9000	62.4
3	2	264.6				264.6	0.00	2105.0	0.1257	12.3	0.9000	69.8
3	3	135.4	114.8			250.2	0.46	1990.8	0.1257	12.3	0.9000	69.8
3	4		235.2			235.2	1.00	1871.1	0.1257	12.3	0.9000	69.8
4	1	50.0				50.0	1.00	1746.7	0.0286	6.0	0.4205	58.8
4	2		335.4			335.4	0.00	2105.0	0.1593	15.6	0.9000	49.2
4	3		164.6	151.9		316.4	0.48	1985.9	0.1593	15.6	0.9000	49.2
4	4			298.1		298.1	1.00	1871.1	0.1593	15.6	0.9000	49.2
	P1									20.0*		69.8
	P2									20.0*		85.1

* Represents the actual green times.

TABLE 6: Lane-based model results for cycle length minimization in the off-peak period (μ = 1.00 and cycle length = 88.13 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		400.0			400.0	1.00	1746.7	0.2290	29.6	0.6817	0.0
1	2			100.0		100.0	0.00	2105.0	0.0475	6.0	0.6978	23.6
1	3			100.0		100.0	0.00	2105.0	0.0475	6.0	0.6978	23.6
1	4			0.0	100.0	100.0	1.00	1871.1	0.0534	6.0	0.7850	23.6
2	1			100.0		100.0	1.00	1746.7	0.0573	6.0	0.8410	12.6
2	2				400.0	400.0	0.00	2105.0	0.1900	18.6	0.9000	0.0
2	3	50.0				50.0	1.00	1871.1	0.0267	6.0	0.3925	0.0
2	4	50.0				50.0	1.00	1871.1	0.0267	6.0	0.3925	0.0
3	1				300.0	300.0	1.00	1746.7	0.1718	33.9	0.4468	34.6
3	2	150.0				150.0	0.00	2105.0	0.0713	15.7	0.4000	52.8
3	3	150.0	0.0			150.0	0.00	2105.0	0.0713	15.7	0.4000	52.8
3	4		300.0			300.0	1.00	1871.1	0.1603	15.7	0.9000	52.8
4	1	100.0				100.0	1.00	1746.7	0.0573	36.8	0.1372	11.0
4	2		283.3			283.3	0.00	2105.0	0.1346	13.2	0.9000	34.6
4	3		116.7	148.1		264.8	0.56	1967.4	0.1346	13.2	0.9000	34.6
4	4			251.9		251.9	1.00	1871.1	0.1346	13.2	0.9000	34.6
	P1									20.0*		52.8
	P2									20.0*		0.0

* Represents the actual green times.

TABLE 7: Lane-based model results for cycle length minimization in the evening peak period (μ = 1.00 and cycle length = 88.13 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h)				(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start
		1	2	3	4							
1	1		550.0			550.0	1.00	1746.7	0.3149	43.1	0.6441	0.0
1	2			150.0		150.0	0.00	2105.0	0.0713	7.9	0.8000	35.2
1	3			150.0		150.0	0.00	2105.0	0.0713	7.9	0.8000	35.2
1	4			0.0	150.0	150.0	1.00	1871.1	0.0802	7.9	0.9000	35.2
2	1			150.0		150.0	1.00	1746.7	0.0859	21.7	0.3487	8.5
2	2				650.0	650.0	0.00	2105.0	0.3088	30.2	0.9000	0.0
2	3	100.0				100.0	1.00	1871.1	0.0534	6.0	0.7850	0.0
2	4	100.0				100.0	1.00	1871.1	0.0534	6.0	0.7850	0.0
3	1				450.0	450.0	1.00	1746.7	0.2576	25.2	0.9000	57.9
3	2	50.0				50.0	0.00	2105.0	0.0238	10.5	0.2000	72.7
3	3	0.0	200.0			200.0	1.00	1871.1	0.1069	10.5	0.9000	72.7
3	4		200.0			200.0	1.00	1871.1	0.1069	10.5	0.9000	72.7
4	1	600.0				600.0	1.00	1746.7	0.3435	33.6	0.9000	34.0
4	2		420.8			420.8	0.00	2105.0	0.1999	19.6	0.9000	48.1
4	3		279.2	125.9		405.1	0.31	2026.3	0.1999	19.6	0.9000	48.1
4	4			374.1		374.1	1.00	1871.1	0.1999	19.6	0.9000	48.1
	P1									20.0*		72.7
	P2									20.0*		86.1

* Represents the actual green times.

5.3 Delay minimization

Again, the four-arm junction with two pedestrian crossings is used to demonstrate the evaluation of the overall junction delay. For the problem of delay minimization in a multi-period design, the Webster's delay function,* which is given in equation (30), is set as the objective function for optimization (Webster, 1958; Allsop, 1972b). The formula assumes that the average stream arrival rates are constant.

$$D_{i,k} = \frac{9}{10}\left(\frac{\sum_j q_{i,j,k}(1-\gamma_{i,k})^2}{2\zeta(1-y_{i,k})} + \frac{(y_{i,k}/\gamma_{i,k})^2}{2(1-y_{i,k}/\gamma_{i,k})} \right), \tag{30}$$

where $D_{i,k}$ is the rate of delay and $\gamma_{i,k}$ is the effective duration of green for the traffic lane k on arm i of a signal junction. The total rate of delay of the junction, D, is the sum of the delays for all of the traffic lanes on all of the arms, $D = \sum_{i=1}^{N_T} \sum_{k=1}^{L_i} D_{i,k}$.

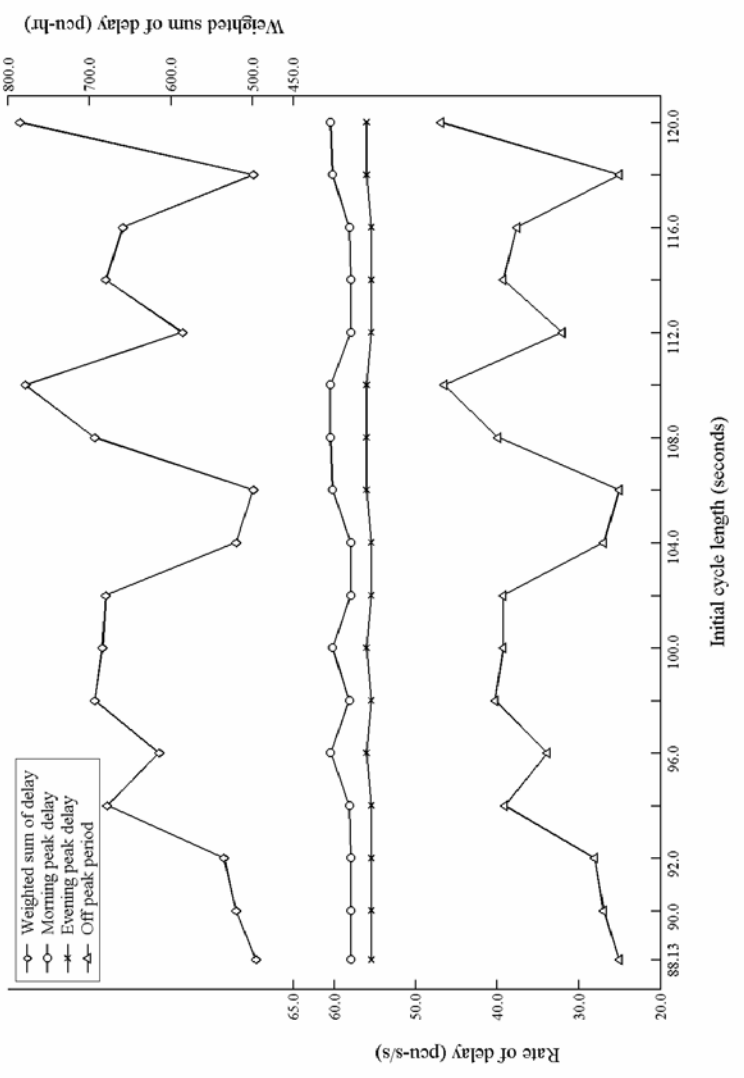
As has been discussed, the delay minimization problem in the lane-based formulation is non-linear and non-convex, and is defined in the feasible solution region. No standard solution method can be applied. A heuristic line search technique has been developed to solve the lane-based delay minimization problem, and it has been demonstrated that the solution method optimizes the Webster's delay expression effectively. Various lane-based modules are responsible for generating a set of solution candidates for the permitted movements, effective movements, allocated flows, and successor functions. All of these lane-based model outputs are fixed in the group-based model, in which further optimization of the signal settings on behalf of the junction delay is conducted.

The heuristic line search algorithm that was derived in Section 4.2 starts with the solution of a set of capacity maximization problems with fixed cycle lengths, the choice of which covers a range from the minimum cycle length to the maximum limit of a cycle length, defined externally. The optimal cycle length is always pushed to its maximum limit in a capacity maximization problem. In other words, the two initial cycle-length boundaries have already been determined in the capacity maximization (120.00 seconds) and cycle length minimization (88.13 seconds) problems that are given in Sections 5.1 and 5.2, respectively. In this study, a 2-second step size is used to contribute 17 evaluation points in the line search space, where the maximum and minimum boundaries are inclusive.

The optimized delays in various design periods as derived by SIGSIGN, the group-based module, are plotted in Figure 4. Each initial cycle length ultimately serves three group-based optimizations for the total junction delays, in addition to the other signal settings. The problem of a three-period design is similar to undertaking three separate junction designs. For one initial cycle length, three individual junction delays are optimized for the various design periods, as given in Figure 4 (the morning peak delay is shown with a circular marker, the off-peak delay with a triangular marker, and the evening peak delay with a cross marker). Among the 17 evaluations, the average delay for the morning peak is 58.96 pcu-s/s, for the off-peak period is 34.74 pcu-s/s, and for the evening peak period is 55.67 pcu-s/s. From Figure 4, it can be observed that the

* The sheared delay formula can be adopted in the present approach to replace the Webster's delay function in case of junction over-saturation. The maximum degree of saturation on each lane may correspondingly be increased from the current limit of 90.0% to over 100.0%, depending on the subsequent design criteria.

FIGURE 4: Optimized junction delay by the group-based module



delays for the morning peak and evening peak periods vary only narrowly within a range of ± 2.5% from the mean values across the 17 data points. In contrast, the off-peak delays change quite significantly, with maximum delays of 46.94 pcu-s/s (deviating over +35% from the mean) and minimum delays of 25.00 pcu-s/s (around a -28% deviation from the mean).

To assess the overall junction performance, the actual time intervals of the three design periods have to be taken into consideration. In Hong Kong, most traffic studies make the general assumption that the morning peak and evening peak periods last for only 1.5 hours, but that the off-peak period normally runs for 13 hours. The remaining 8-hour study period is always in a light flow condition that starts just before midnight, and is thus excluded from this study. The delay-weighted factors in the morning and evening peak periods take on the values of 1.5 and 13.0 in the off-peak period. For better presentation, the weighted sum of delay (the line with a diamond marker) is also plotted on the same graph in Figure 4. It is found that the minimum weighted sum of delay is 495.61 pcu/hr, and the corresponding initial cycle length is located at the minimum cycle length, which is 88.13 seconds. The optimized arrangements of the permitted and effective movements for the delay minimization are exactly the same as those that are given by the cycle length minimization in the three-period design problem that is plotted in Figure 3. Details of the optimized allocated flows from the lane-based module and other signal settings from the group-based module are tabulated in Tables 8 to 10.

The formats of these tables are similar to those that are reported in the previous sections. An additional column (14) is attached to Tables 8 to 10 that shows the actual lane-based delay values, and the bold figures give the total junction delay by summing up the delays for all of the traffic lanes in the junction example. The tabulated results in Tables 8 to 10 are divided into two parts. The first part comprises the lane-based module that is given in columns 3 to 6, which includes the patterns of the permitted and effective movements and the allocated lane flows. The results in columns 7 to 10 are also calculated based on the allocated flow values, such as the total lane flows, the lane-turning proportions, the revised lane-saturation flows, and the lane-flow factors. All of these lane-based results are fixed and are then input into the group-based model for further optimization. The group-based model mainly optimizes the overall junction delay (column 14) and other signal timings, such as the optimized cycle length (given in the first row of the table), effective green (in column 11), and start of green (in column 13). With all of these results, the actual degree of saturation for each traffic lane can be evaluated, and is placed in column 12. The group-based module is also used to optimize the starts of green and actual duration of green, which are given in the lower part of the tables, for the two pedestrian crossings.

The total computing time for the 17 evaluations, including all of the MLB1, MLB2, and GB module computations, takes about 380 minutes on a Pentium-4 1.8 GHz processor.

6. CONCLUSION

In this paper, a lane-based optimization method for the multi-period design of lane markings and signal settings in isolated signal-controlled junctions has been presented. Capacity maximization and cycle length minimization have been formulated as BMILP problems that can be solved by the standard branch-and-bound technique. The delay minimization problem has been formulated as a BMINLP, and the Webster's delay function has been applied as the objective for optimization. A heuristic algorithm has

TABLE 8: Lane-based model results for delay minimization in the morning peak period (Optimized cycle length = 105.41 seconds)

(1)	(2)	(3–6)				(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
From arm	Lane	To arm, allocated flows (pcu/h)				Sum of lane	Turning	Saturation	Flow	Effective	Degree of	Green	Total
		1	2	3	4	flow	proportion	flow	factor	green	sat.	start	delay
1	1		400.0			400.0	1.00	1746.7	0.2290	53.5	0.4509	0.0	1.82
1	2			241.7		241.7	0.00	2105.0	0.1148	15.1	0.8005	38.4	4.09
1	3			241.7		241.7	0.00	2105.0	0.1148	15.1	0.8005	38.4	4.09
1	4			16.7	200.0	216.7	0.92	1887.2	0.1148	15.1	0.8007	38.4	3.82
2	1			200.0		200.0	1.00	1746.7	0.1145	30.9	0.3912	2.6	1.60
2	2				600.0	600.0	0.00	2105.0	0.2850	33.4	0.9000	0.0	8.81
2	3	250.0				250.0	1.00	1871.1	0.1336	33.4	0.4219	0.0	1.91
2	4	250.0				250.0	1.00	1871.1	0.1336	33.4	0.4219	0.0	1.91
3	1				350.0	350.0	1.00	1746.7	0.2004	41.9	0.5042	58.6	2.33
3	2	264.6				264.6	0.00	2105.0	0.1257	16.5	0.8058	84.0	4.35
3	3	135.4	114.8			250.2	0.46	1990.8	0.1257	16.5	0.8037	84.0	4.16
3	4		235.2			235.2	1.00	1871.1	0.1257	16.5	0.8040	84.0	4.01
4	1	50.0				50.0	1.00	1746.7	0.0286	40.6	0.0744	38.4	0.26
4	2		335.4			335.4	0.00	2105.0	0.1593	20.4	0.8216	58.6	5.11
4	3		164.6	151.9		316.4	0.48	1985.9	0.1593	20.4	0.8214	58.6	4.92
4	4			298.1		298.1	1.00	1871.1	0.1593	20.4	0.8222	58.6	4.75
	P1									20.0*		84.0	
	P2									56.5*		103.4	
												Total delay	57.95

* Represents the actual green times.

TABLE 9: Lane-based model results for delay minimization in the off-peak period (Optimized cycle length = 70.08 seconds)

(1) From arm	(2) Lane	(3–6) To arm, allocated flows (pcu/h) 1	2	3	4	(7) Sum of lane flow	(8) Turning proportion	(9) Saturation flow	(10) Flow factor	(11) Effective green	(12) Degree of sat.	(13) Green start	(14) Total delay
1	1		400.0			400.0	1.00	1746.7	0.2290	27.4	0.5859	6.1	2.06
1	2			100.0		100.0	0.00	2105.0	0.0475	6.0	0.5549	27.5	1.08
1	3			100.0		100.0	0.00	2105.0	0.0475	6.0	0.5549	27.5	1.08
1	4			0.0	100.0	100.0	1.00	1871.1	0.0534	6.0	0.6242	27.5	1.24
2	1			100.0		100.0	1.00	1746.7	0.0573	11.5	0.3503	11.0	0.74
2	2				400.0	400.0	0.00	2105.0	0.1900	16.4	0.8128	6.1	4.13
2	3	50.0				50.0	1.00	1871.1	0.0267	16.4	0.1143	6.1	0.27
2	4	50.0				50.0	1.00	1871.1	0.0267	16.4	0.1143	6.1	0.27
3	1				300.0	300.0	1.00	1746.7	0.1718	32.7	0.3681	38.5	1.00
3	2	150.0				150.0	0.00	2105.0	0.0713	14.0	0.3564	57.1	0.99
3	3	150.0	0.0			150.0	0.00	2105.0	0.0713	14.0	0.3564	57.1	0.99
3	4		300.0			300.0	1.00	1871.1	0.1603	14.0	0.8021	57.1	3.47
4	1	100.0				100.0	1.00	1746.7	0.0573	24.7	0.1625	27.5	0.40
4	2		283.3			283.3	0.00	2105.0	0.1346	13.7	0.6885	38.5	2.54
4	3		116.7	148.1		264.8	0.56	1967.4	0.1346	13.7	0.6899	38.5	2.43
4	4			251.9		251.9	1.00	1871.1	0.1346	13.7	0.6897	38.5	2.34
	P1									19.9*		57.1	
	P2									30.4*		4.1	
												Total delay	25.04

* Represents the actual green times.

TABLE 10: Lane-based model results for delay minimization in the evening peak period (Optimized cycle length = 106.94 seconds)

(1)	(2)	(3–6)				(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
From arm	Lane	To arm, allocated flows (pcu/h)				Sum of lane	Turning	Saturation	Flow	Effective	Degree of	Green	Total
		1	2	3	4	flow	proportion	flow	factor	green	sat.	start	delay
1	1		550.0			550.0	1.00	1746.7	0.3149	52.3	0.6437	0.2	3.32
1	2			150.0		150.0	0.00	2105.0	0.0713	10.6	0.7184	41.9	2.58
1	3			150.0		150.0	0.00	2105.0	0.0713	10.6	0.7184	41.9	2.58
1	4			0.0	150.0	150.0	1.00	1871.1	0.0802	10.6	0.8082	41.9	3.30
2	1			150.0		150.0	1.00	1746.7	0.0859	31.6	0.2910	5.4	1.14
2	2				650.0	650.0	0.00	2105.0	0.3088	36.7	0.9000	0.2	9.07
2	3	100.0				100.0	1.00	1871.1	0.0534	13.3	0.4304	0.2	1.23
2	4	100.0				100.0	1.00	1871.1	0.0534	13.3	0.4304	0.2	1.23
3	1				450.0	450.0	1.00	1746.7	0.2576	44.6	0.6171	57.5	3.20
3	2	50.0				50.0	0.00	2105.0	0.0238	13.9	0.1834	88.3	0.54
3	3	0.0	200.0			200.0	1.00	1871.1	0.1069	13.9	0.8249	88.3	4.02
3	4		200.0			200.0	1.00	1871.1	0.1069	13.9	0.8249	88.3	4.02
4	1	600.0				600.0	1.00	1746.7	0.3435	64.8	0.5667	18.5	2.23
4	2		420.8			420.8	0.00	2105.0	0.1999	25.8	0.8296	57.5	5.87
4	3		279.2	125.9		405.1	0.31	2026.3	0.1999	25.8	0.8292	57.5	5.71
4	4			374.1		374.1	1.00	1871.1	0.1999	25.8	0.8291	57.5	5.41
	P1									20.0*		88.3	
	P2									55.3*		105.2	
												Total delay	55.44

* Represents the actual green times.

been developed to solve the problem that incorporates the lane-based and group-based optimization modules. The binary variables include the effective movements and permitted movements on traffic lanes and the successor functions to govern the order of the signal displays. The continuous variables include the lane-allocated flows, the common flow multiplier, the cycle length, and the starts and duration of green for the traffic movements, lanes, and pedestrian crossings. A set of linear constraints has been set up to relate all of the variables for feasibility and safety requirement designs. Numerical examples have been given to demonstrate the practical applicability of the proposed methodology.

ACKNOWLEDGEMENTS

The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. HKU7018/01E and HKU7187/05E).

REFERENCES

Allsop, R.E. (1971a) Delay-minimising settings for fixed-time traffic signals at a single road junction. Journal of the Institute of Mathematics and its Applications, 8, 164-185.

Allsop, R.E. (1971b) SIGSET: a computer program for calculating traffic signal settings. Traffic Engineering and Control, 13, 58-60.

Allsop, R.E. (1972) Estimating the traffic capacity of a signalized road junction. Transportation Research, 6, 245-255.

Allsop, R.E. (1975) Computer program SIGCAP for assessing the traffic capacity of signal-controlled road junctions – description and manual for users. Transportation Operations Research Group Working Paper, 11, University of Newcastle upon Tyne.

Allsop, R.E. (1981) Computer program SIGSET for calculating delay-minimising traffic signal timings – description and manual for users. Transport Studies Group Research Report, University College London.

Burrow, I.J. (1987) OSCADY: a computer program to model capacities, queues and delays at isolated traffic signal junctions. TRRL Report, RR 105, Transport and Road Research Laboratory, Crowthorne.

Gallivan, S. and Heydecker, B.G. (1988) Optimising the control performance of traffic signals at a single junction. Transportation Research Part B, 22, 357-370.

Heydecker, B.G. (1992) Sequencing of traffic signals. In J.D. Griffiths (ed.) Mathematics in Transport and Planning and Control, Clarendon Press, Oxford, pp. 57-67.

Heydecker, B.G. and Dudgeon, I.W. (1987) Calculation of signal settings to minimise delay at a junction. Proceedings of 10th International Symposium on Transportation and Traffic Theory, MIT, July, Elsevier, New York, pp. 159-178.

Improta, G. and Cantarella, G.E. (1984) Control system design for an individual signalized junction. Transportation Research Part B, 18, 147-167.

Kolman, B. and Beck, R.E. (1995) Elementary Linear Programming with Applications. Academic Press, San Diego.

Lam, W.H.K, Poon, A.C.K. and Mung, G.K.S. (1997) Integrated model for lane-use and signal-phase designs. Journal of Transportation Engineering, ASCE, 123, 114-122.

Maximal Software, Inc. (2002) MPL Modeling System. Release 4.2.

Porn, R. and Westerlund, T. (2000) A cutting plane method for minimizing pseudo-convex functions in the mixed integer case. Computers and Chemical Engineering, 24, 2655-2665.

Sang, A.P. and Silcock, J.P. (1989) SIGSIGN user manual. Steer Davies and Gleave Ltd and Transport Studies Group, University College London.

Silcock, J.P. (1997) Designing signal-controlled junctions for group-based operation. Transportation Research Part A, 31, 157-173.

Szeto, W.Y. and Lo, H.K. (2005) Strategies for road network design over time: Robustness under uncertainty. Transportmetrica, 1, 47-63.

Tully, I.M.S.N.Z. (1976) Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. PhD Thesis, OUEL Report, 1189/77, University of Oxford.

Vanderbei, R.J. (2001) Linear Programming: Foundations and Extensions. Kluwer Academic, Boston.

Webster, F.V. (1958) Traffic signal settings. Road Research Technical Paper, No. 39, HMSO, London.

Wong, C.K. and Wong, S.C. (2003a) Lane-based optimization of signal timings for isolated junctions. Transportation Research Part B, 37, 291-312.

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

Wong, C.K., Wong, S.C. and Tong, C.O. (2000) Lane-based optimization method for maximizing reserve capacity of isolated signal-controlled junctions. Proceeding of the Fifth Conference of Hong Kong Society for Transportation Studies, 2 December, Hong Kong, pp. 176-184.

Wong, C.K., Wong, S.C., Tong, C.O. and Lam, W.H.K. (2002) Lane-based optimization method for minimizing delay of isolated signal-controlled junctions. Proceedings of the 7th International Conference on Applications of Advanced Technology in Transportation, 5-7 August, Cambridge, Massachusetts, USA, pp. 199-206.