Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5, pp. 1358 - 1366, 2005

THE ADVANTAGES OF THE MULTI-PERIOD LANE-BASED OPTIMIZATION METHOD FOR TRAFFIC SIGNAL DESIGN

C.K. WONG
Research Associate
Department of Civil Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong SAR
P.R. China
Fax: +852-2559-5337
E-mail: a9595583@graduate.hku.hk

S.C. WONG
Associate Professor
Department of Civil Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong SAR
P.R. China
Fax: +852-2559-5337
E-mail: hhecwsc@hkucc.hku.hk

C.O. TONG
Associate Professor
Department of Civil Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong SAR
P.R. China
Fax: +852-2559-5337
E-mail: cotong@hku.hk

Abstract: This paper presents a case study of applying the multi-period lane-based optimization method to the design of isolated signal-controlled junctions. A site survey was conducted to collect the traffic demand data in three different time periods, capturing the morning peak, off peak and evening peak traffic conditions at one of the busiest junctions in Hong Kong. The details of the existing design parameters, including the lane markings and signal settings and the junction geometries, were also recorded. Junction performance in terms of reserve capacity and total delay is optimized by the multi-period lane-based model, and the results are compared to the existing situations. Significant improvements are observed if the resultant lane markings and signal settings of the multi-period lane-based model are adopted.

Key Words: lane-based method, traffic signal, multi-period, signal optimization

1. INTRODUCTION

The lane-based optimization method has recently been developed to introduce lane markings as binary variables in the optimization framework for the design of isolated signal-controlled junctions (Wong et al., 2002; Wong and Wong, 2003a, b). The lane-based method is formulated as a direct extension of the conventional group (phase) based approach in which lane markings are considered as exogenous inputs (Allsop, 1992). In reality, traffic demand varies at different times of the day, and the flow patterns approaching a signal-controlled junction may not be identical across successive signal cycles. The lane-based optimization method can produce multiple sets of lane markings if the different demand patterns are analyzed individually and independently. It is not practical to change lane marking patterns so frequently in daily operations due to safety concerns. Indeed, lane markings are not subject to repetitive alternations once physically established on ground. This raises the practical design problem of producing a reliable set of lane markings that is suitable for different demand cases in different time periods. In this study, the original lane-based optimization method is refined to deal with this multi-period design problem (Wong, 2004). The following sections provide a brief outline of the multi-period lane-based optimization method, including the required input data, control variables, and governing constraints. The usual optimization criteria for the design of traffic signal settings is also reviewed. These practical design problems can be formulated as different mathematical programs, and the relevant solution algorithms are discussed. A case study of a typical signal-controlled junction in Hong Kong will be used to demonstrate the method.

2. THE MULTI-PERIOD LANE-BASED OPTIMIZATION METHOD

2.1 Input Data

The required data inputs for the analysis of multi-period design isolated signal-controlled junctions are summarized as follows. Traffic demands in various design periods have to be given. The numbers of traffic arms and pedestrian crossings at a signal-controlled junction also have to be specified. From different approaches, the numbers of traffic lanes for junction entries and exits are required in the lane-based analysis. Maximum and minimum cycle lengths provide the upper and lower limits for the selection of an appropriate operating cycle length in the optimization process. To meet concerns about safety, the minimum green durations should be assigned for both traffic and pedestrian movements. To ensure a safe separation of all conflicting movements across a junction at successive green signals, a clearance time (or intergreen) matrix has to be defined for the design calculations. Lane saturation flows for straight-ahead movement that give the lane capacities are estimated based on site geometric factors such as lane widths, lane types, and road gradients. The turning radius has to be given to revise the lane saturation flows whenever there is turning traffic. To prevent excessive delays, the acceptable degree of saturation for each traffic lane is set to limit the maximum attainable volume to the capacity ratio. In addition, time differences between the effective greens and the actual (display) greens are included in the input data set.

2.2 Control Variables

The lane-based model variables can be grouped into two sets. The first set consists of the control variables and is binary in nature. It includes the permitted movements, effective movements, and successor functions. The permitted movements represent the lane-marking layout, which is identical in all design periods so that road users are always guided by the same set of permitted movements. However, the effective movements are introduced into the formulation to represent the set of actual movements in different design periods. The successor functions are established to control the relative signal display sequences for every pair of incompatible traffic movements, which are also (design) period dependent.

The other set of control variables is continuous in nature. It is comprised of the allocated flows, common flow multiplier, cycle length/reciprocal of cycle length, start of green for movements, duration of green for movements, start of green for traffic lanes, and duration of green for traffic lanes. The set of allocated flows 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 is a flow-scaling factor, and also serves an indicator for the junction capacity. The cycle length gives the actual time duration for a complete signal cycle, and its reciprocal is also required to simplify the mathematical formulation. For traffic movements, the start of green and the duration of green (or the actual green) are defined to represent the time slot in which the right-of-way is granted. In the lane-based formulation, the start of green and duration of green can also be specified on a lane basis.

2.3 Governing Constraints

(1) Flow conservation: For each period, the traffic demand is multiplied by a common flow multiplier which 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 constraints have to be set to equalize the allocated lane flows and the demand flows. (2) Minimum effective movement in a lane: In the lane-based formulation, effective movements are defined to represent the traffic movements with non-zero allocated lane flows. To ensure that traffic flows are allocated on every traffic lane approaching a signal-controlled junction, the constraint set of the minimum effective movement in a lane has to be established. (3) Maximum permitted movements at the exit: Due to the geometric restriction of the junction, appropriate control of the number of permitted movements (smaller or equal to the number of exit lanes) has to be provided to avoid unnecessary traffic merging. (4) Permitted movements across adjacent lanes: Within the same traffic approach, the movements, such as left-turning, right-turning, and straight-ahead movement, that are permitted across adjacent lanes have to be strictly controlled so that internal conflicts among traffic movements are absolutely expunged. (5) Cycle length: For the practical design of signal settings, the cycle length constraint is applied to confine the operational cycle length by a specific range between a minimum cycle length and a maximum cycle length. (6) Lane signal settings: For permitted traffic movements that share a traffic lane, the signal settings, including the starts and durations of green, should be set identical for consistency. (7) Start of green: This constraint set is established to ensure that all of the starts of green times fall within the range of a signal cycle. (8) Duration of green: Again, all traffic signals work within a signal cycle so that no duration of green for both vehicular and pedestrian movements can be assigned to exceed one signal cycle. For safety considerations, however, every traffic movement in all of the design periods should be given a minimum green duration. (9) Order of signal displays: For any two incompatible movements, the order of the signal displays is controlled by a pair of successor functions (Heydecker, 1992) so that their sequences of receiving right-of-way within a signal cycle can be arranged. (10) Clearance time: This constraint set is constructed to ensure that sufficient clearance time is given to separate the right-of-way of any two conflicting movements within a signal cycle. (11) Prohibited movement: This set of constraints is established to force the allocated lane flows to be zero if the corresponding movement is not effective. (12) Flow factor: Based on queuing theory, the degrees of saturation on a pair of adjacent lanes that possess a common effective movement must be identical. Apart from the identical signal settings in (6), the identical degrees of saturation are obtained by equalizing the flow factors (defined as the total allocated flow divided by the saturation flow) for these adjacent lanes. (13) Maximum acceptable degree of saturation: To prevent excess delay, the degrees of saturation for traffic lanes are always restricted by a maximum acceptable limit. (14) Maximum effective movements: According to the demand patterns that are given in various design periods the actual lane-flow patterns can be allocated in a different way on the condition that the effective movements can only be derived from the unique set of permitted movements. Optionally, a special form of maximum effective movement constraint can be used to obtain a more stringent lane-marking design in a multi-period analysis, and also serve as a compulsory substitution in the formulation single period design by equalizing the effective movements and permitted movements. (15) Other

signal group constraints: There may be other constraints in setting up the relative timing of the starts and ends of green for different signal groups to meet practical operations.

3. OPTIMIZATION CRITERIA

For the multi-period analysis of isolated signal control junctions, there are generally three optimization criteria for traffic signal designs: (1) capacity maximization, (2) cycle length minimization, and (3) delay minimization. The optimization problems of (1) and (2) can be effectively formulated by setting up a linear objective function that maximizes the common flow multiplier and the reciprocal of cycle length as a Binary-Mixed-Integer-Linear-Program (BMILP). As the delay minimization problem (3) usually involves a non-linear delay objective function, the optimization problem has to be formulated as a Binary-Mixed-Integer-Non-Linear-Program (BMINLP). In the present lane-based optimization framework, the formulation is quite flexible so that either the Webster's delay formula or the sheared delay expression can be adopted as the objective for optimization depending on the actual congestion level of the junction. All of these mathematical programs for optimizing different design objectives are evaluated subject to the above sets of linear constraints.

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. The multi-period capacity maximization and cycle length minimization problems can be formulated as Binary-Mixed-Integer-Linear-Programming (BMILP) problems, and a standard branch-and-bound algorithm can be applied. The MPL modeling system, which integrates a CPLEX solver, is used to implement the branch-and-bound algorithm to solve the BMILP for the capacity maximization and cycle length minimization problems.

As the delay function for estimating the overall junction delay is mostly non-linear, the delay minimization problem has to be formulated as a Binary-Mixed-Integer-Non-Linear-Programming (BMINLP) problem. Conventionally, a piecewise linearization technique can be applied to the delay function, and the BMINLP problem can be divided into a series of BMILP sub-problems. The standard branch-and-bound algorithm can still be used as the solution method for solving the conventional delay minimization problems. However, in the present lane-based formulation, the objective delay function is proven to be non-linear and non-convex defined in the feasible solution region. 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 the computational effort and the quality of the solution (Wong and Wong, 2003b). Based on the capacity maximization and cycle length minimization modules, the upper and lower boundaries of the feasible range of the cycle length for the delay minimization can be identified. The line search technique can then be developed to intensively explore along the feasible region of the cycle length so that a set of period-specific signal settings and lane markings can be obtained to minimize the overall junction delays. Detailed outlines of the line search technique can be found in the work of Wong et al. (2002) and Wong and Wong (2003b). A flow chart is also given in Figure 1 for illustrating the solution steps.

5. A CASE STUDY IN HONG KONG

To demonstrate how the multi-period lane-based model works for a practical design problem, a site survey has been conducted to collect all relevant input data and also the existing lane markings and signal settings for evaluation and comparison. The junction is the intersection of Hennessy Road and Fleming Road in Wanchai (Hong Kong). There are four entry and three exit lanes along Hennessy Road and two entry and exit lanes along Fleming Road. The geometric layout and existing lane markings of the example junction are given in Figure 2. Traffic demands, existing signal settings, and junction performances are shown in Table 1.

Assume that the maximum cycle length is limited to 120 seconds (which is common practice in Hong Kong), the maximum acceptable degrees of saturation for all traffic lanes are 90%, and the minimum greens are 5 and 20 seconds for traffic and pedestrian movements respectively. The required intergreens for any pair of conflicting movements are set to be 6 seconds and 4 seconds if pedestrian movements are involved. All left-turning and straight-ahead signals are set to end at the same time (an usual practice for safety concerns). Then, the multi-period lane-based model refines the set of lane markings and designs a better set of signal settings over the existing designs such that the overall junction performances in various design periods in terms of reserve capacity (R.C.) and total delay are improved. The optimized junction reserve capacities are 39.35%, 26.33%, and 43.02% for the morning peak, off peak, and evening peak periods. The minimum cycle length is 74.62 seconds. Using a 2-second step size in the line search algorithm, there are 24 candidate solutions for the capacity maximizing (module) evaluations at fixed cycle lengths. The optimized junction delays are 28.03 pcu, 33.47 pcu, and 26.85 pcu for the morning peak, off peak, and evening peak periods.

The optimized lane marking pattern is given in Figure 3. Two newly added lane markings are shown for the right-turn along Hennessy Road: [2,1,3] (from arm 2 to arm 1 on lane 3) and [4,3,3] (from arm 4 to arm 3 on lane 3). All filled arrows (lane markings) represent the permitted and effective movement patterns in all of the design periods. The hollow arrow [4,3,3] is a lane marking that is only effective in the off peak period. Thus, the right-turn traffic from arm 4 on lane 4 forms to a separate traffic stream in the morning and evening peak periods. Nevertheless, due to safety reasons, the signal settings of the right-turn movement are forced to synchronize with those of the left-turn and straight-ahead movements. Table 2 summarizes the key modeling results of the multi-period lane-based model. Two different signal settings, including the vehicular and pedestrian movements for maximizing the junction capacity and minimizing the total delay, are also provided in Table 2.

6. CONCLUSIONS

This paper has presented a brief outline of the multi-period lane-based model for designing isolated signal-controlled junctions. The model takes the multi-period traffic demand pattern into consideration and gives a unique set of lane markings for practical operations. The model can also adjust the signal settings, including the cycle lengths, to optimize the overall junction performances in various design periods. A real junction test has been conducted, and comparison with the existing design has been made. The multi-period lane-based model optimizes and improves junction performance in terms of the reserve capacity and total delay in all of the design periods by implementing new sets of lane markings and signal settings with affordable computation effort.

ACKNOWLEDGEMENTS

The work that is described in this paper was jointly supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU7018/01E and HKU7031/02E).

REFERENCES

Allsop, R.E. (1992) Evolving application of mathematical optimisation in design and operation of individual signal-controlled road junctions. In J.D. Griffiths (ed.) **Mathematics in Transport and Planning and Control**, Clarendon Press, Oxford, pp. 1-24.

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.

Wong, C.K. (2004) Lane-based optimization method for traffic signal design. PhD Thesis, Department of Civil Engineering, The University of Hong Kong, Hong Kong.

Wong, C.K. and Wong, S.C. (2003a) Lane-based optimization of signal timings for isolated junctions, **Transportation Research**, **Vol. 37B, No. 1**, 63-84.

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**, **Vol. 2, No. 4**, 379-406.

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**, Cambridge, Massachusetts, USA, 5-7 August, pp. 199-206.

Figure 1 Solution procedures for delay optimization.

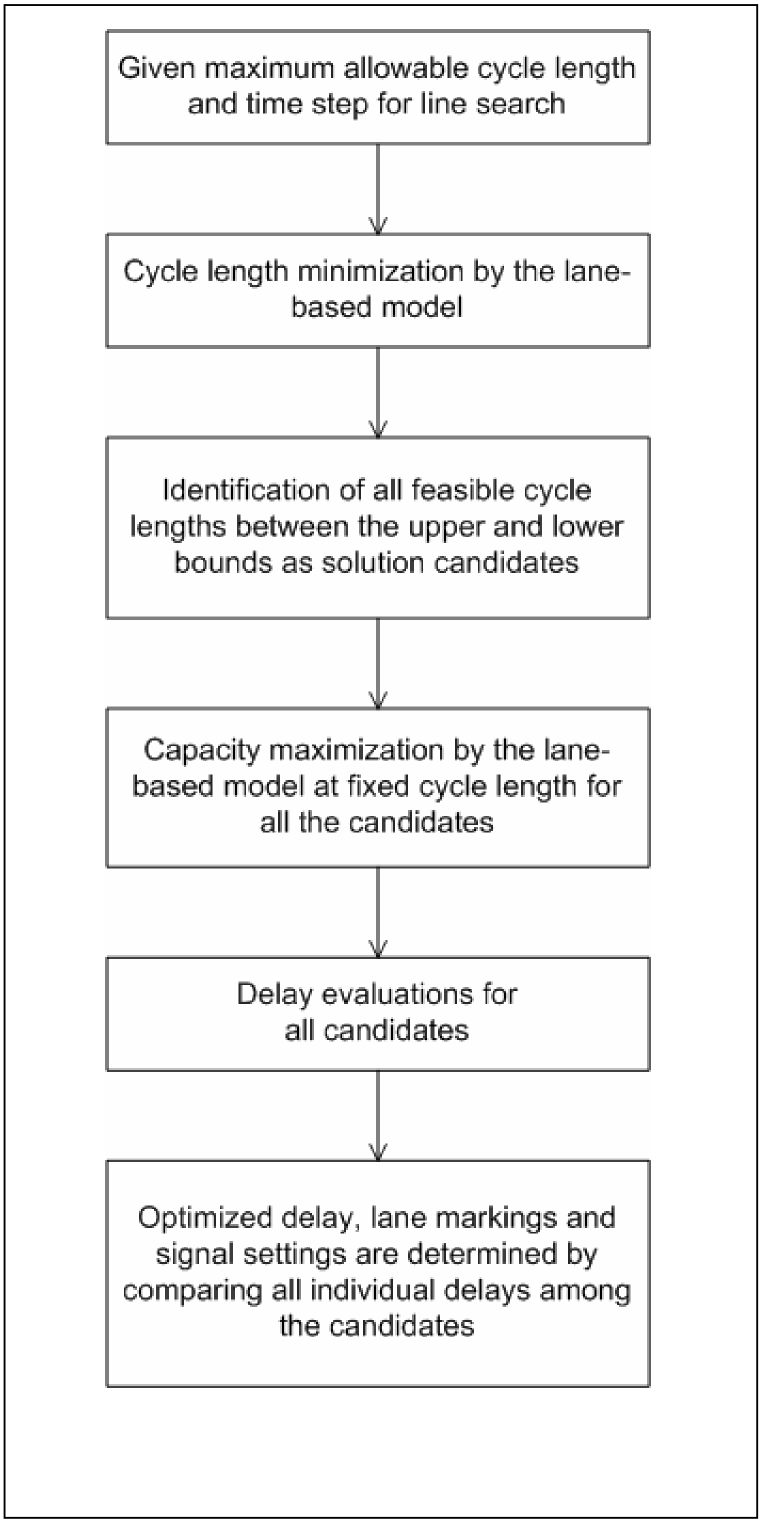

Figure 2 Geometric layout and existing lane markings of the example junction.

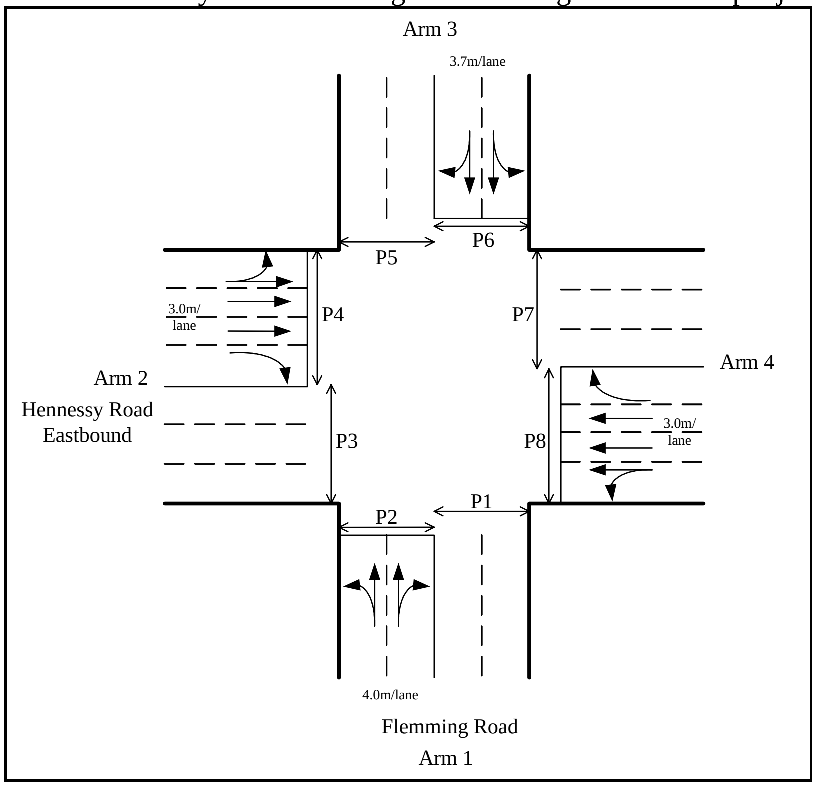

Figure 3 Optimized lane marking pattern of the example junction.

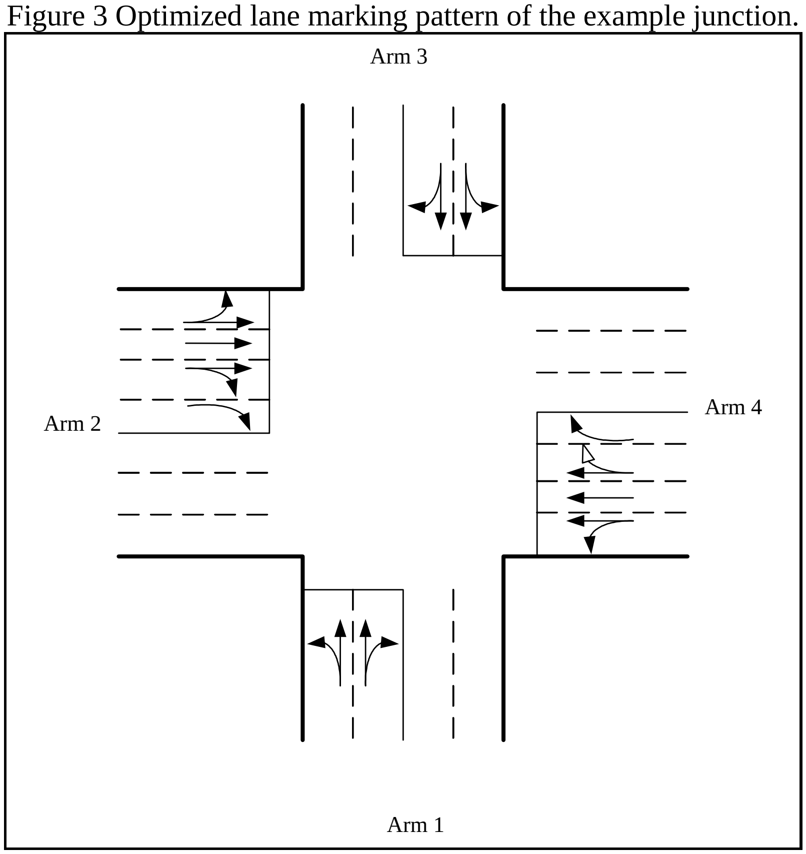

Table 1 Traffic demand pattern, existing signal settings and junction performances.

From arm	To arm	Morning peak		Off peak		Evening peak	
		Demand (pcu/h)	Green start/ duration (s)	Demand (pcu/h)	Green start/ duration (s)	Demand (pcu/h)	Green start/ duration (s)
1	2	180	0.0/23.0	118	0.0/23.0	131	0.0/23.0
	3	305		265		219	
	4	199		227		203	
2	1	245	29.0/15.0	291	29.0/17.0	202	31.0/18.0
	3	103		152		140	
	4	543		645		475	
3	1	235	50.5/17.0	253	53.0/21.0	282	57.0/19.0
	2	73		128		98	
	4	171		175		198	
4	1	203	74.0/25.0	178	81.0/18.0	172	84.0/23.0
	2	514		703		564	
	3	150		282		188	
Existing cycle length		105.0 s		105.0 s		115.0 s	
Existing R.C.		17.24%		7.74%		4.75%	
Existing total delay[1]		31.45 pcu		39.54 pcu		35.02 pcu	

[1] Webster's two-term delay expression is adopted.

Table 2 Key modeling results from the multi-period lane-based model.

Arm	Lane	Morning peak		Off peak		Evening peak	
		Lane flow (pcu/h)	[2]Green start /duration (s)	Lane flow (pcu/h)	Green start /duration (s)	Lane flow (pcu/h)	Green start /duration (s)
1	1	330.9	0.0/31.5; 2.3/19.0	300.8	0.0/25.4; 3.3/16.0	271.0	0.0/26.1; 1.4/16.0
	2	353.1		309.2		282.0	
2	1	208.6	65.7/20.4; 49.3/15.7	252.0	60.7/22.8; 47.3/16.0	186.3	61.4/19.2; 45.4/16.0
	2	237.7		290.8		218.7	
	3	233.5		286.7		217.7	
	4	211.3		258.5		194.4	
3	1	224.7	92.1/21.9; 71.3/15.7	264.9	31.4/23.3; 25.3/16.0	272.3	86.6/27.4; 67.4/16.0
	2	254.3		291.1		305.7	
4	1	210.6	37.5/22.2; 27.3/16.0	267.0	89.6/24.4; 69.3/16.0	219.3	32.1/23.3; 23.4/16.0
	2	253.2		310.4		258.4	
	3	253.2		309.7		258.4	
	4	150.0		275.9		188.0	
P1			118.0/35.5; 0.3/23.0		118.0/29.4; 1.3/20.0		118.0/30.1; 87.4/20.0
P2			35.5/80.3; 25.3/63.7		29.4/86.6; 23.3/64.0		30.1/85.7; 21.4/64.0
P3			63.7/24.4; 47.3/20.0		58.7/26.8; 45.3/20.0		59.4/23.2; 43.4/20.0
P4			90.1/91.4; 69.1/67.0		87.6/89.2; 67.3/64.0		84.6/92.8; 65.4/64.0
P5			90.1/25.9; 69.1/20.0		29.4/27.3; 23.3/20.0		84.6/31.4; 65.4/20.0
P6			118.2/89.9; 0.3/67.0		58.7/88.7; 45.3/64.0		118.0/84.6; 87.4/64.0
P7			35.5/25.8; 25.3/20.0		87.6/28.4; 67.3/20.0		30.1/27.3; 21.4/20.0
P8			63.7/89.8; 47.3/66.7		118.0/87.6; 1.3/64.0		59.4/88.7; 43.4/64.0
Optimized R.C.		39.35% [120.0 s][3]		26.33% [120.0 s]		43.02% [120.0 s]	
Optimized total delay		28.03 pcu [90.7 s]		33.47 pcu [88.0 s]		26.85 pcu [88.0 s]	

[2] Green start/duration for maximizing R.C.; Green start/duration for minimizing total delay.
[3] [.] Numeric values inside represent the operating cycle lengths.