Group-based Hierarchical Adaptive Traffic-Signal Control Part I: Formulation

Seunghyeon Lee[1], S.C. Wong[2], Pravin Varaiya[3]

[1]*Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand*

[2]*Department of Civil Engineering, the University of Hong Kong, Pokfulam Road, Hong Kong, China*

[3]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720-1722, USA*

ABSTRACT

A group-based adaptive traffic-control method for isolated signalized junctions is developed that includes a hierarchical structure comprising tactical and local levels of signal timing optimization. The control method optimizes the signal timings in adaptive traffic-control systems, and takes full advantage of flexible new technologies to incorporate the most up-to-date traffic information, as collected in real time. The definitions, combinations, and sequencing of the cycle structure stages are generated automatically using a procedure for optimizing the signal-timing plans in response to online data from traffic detectors. This new method provides a wider search space and improves the efficiency of the signal-control systems, thus improving the junction performance, minimizing delays, and maximizing capacity in real time. A multi-resolution strategy is proposed for updating the elements of the signal plans cycle-by-cycle and adjusting the current green signal timing second-by-second. The group-based variables and parameters for the proactive global-optimization method utilize lane-based predictive traffic-flow information, such as arrival and discharge rates, expressed as the slopes of polygonal delay formulas. Therefore, there is a high degree of flexibility in the tactical identification of the optimal signal plan in response to the real-time predicted traffic information, the objective function of the polygonal delay formula, and the direct differential equations for the adaptive group-based variables. The reactive local signal-control policy, which is formed based on the max-pressure strategy, is developed to locally adjust the current green signal time and to accommodate delicate demand fluctuations second-by-second at the fine-resolution level. The most appropriate cycle-structure for the tactical level of control is identified using a group-based global-optimization procedure that takes advantage of the latest available information. In part II of this study (Lee et al. 2017), the effectiveness of the proposed methods is validated based on the actualized mathematical frameworks, computer simulations, and a case study, using the appropriate computer programs.

Keywords: Global proactive-optimization scheme, Local reactive-control policy, Lane-based control delay, Group-based derivatives, Time windows

1. INTRODUCTION

Traffic signals are used in many places, and are among the most effective ways of reducing traffic congestion at junctions. There are generally two classes of traffic-signal control: a) fixed-time control, in which the signal settings are derived offline and based on historical traffic information, and b) adaptive control, in which the signal settings are determined online based on the most up-to-date traffic information collected in real time by traffic detectors. Despite its higher cost, adaptive signal control has been shown to be more effective and efficient than fixed-time control. Most signal-control systems are currently specified by a stage-based method, in which the signal settings are defined by a sequence of signal stages with pre-set inter-green periods, irrespective of the control methods used.

Recent advances in the technology of microprocessor controllers have facilitated the development of the group-based method, in which the signal settings are defined by the start and duration of the green signals among the signal groups. The group-based method takes full advantage of the flexibility provided by advanced microprocessor controllers to independently regulate each of the signalized junctions of a signal group. The definitions, combinations, and sequencing of stages of the cycle-structures need not be pre-specified, and can be automatically generated during the optimization of the signal-timing plan. This method allows a wider search space for setting more efficient signal controls and thus improves the junction performance, minimizes delay, and maximizes capacity. This group-based method has been implemented in isolated junctions, and extended to the area traffic control of signalized networks. However, these networks are all based on fixed-time control. Therefore, the application of the group-based method to the adaptive signal-control arena is worth investigating because it may provide a better control strategy for real-time traffic control in both signal-controlled intersections and signalized networks, and thereby substantially improve the network performance.

There are several challenges in developing a group-based adaptive signal-control strategy. First, although traffic data can be directly observed by traffic detectors at isolated signalized junctions, this traffic information is still mostly incomplete. For example, the turning proportions, prevalent arrival patterns, and queue-length distributions are all essential inputs for adaptive control. However, the fixed detectors cannot directly measure these important traffic features. Therefore, it is necessary to develop efficient procedures to estimate these quantities from the data collected by the detectors. Second, the traffic flows in junctions and networks are highly stochastic, and an effective strategy is needed to determine the signal timings at all levels of resolution. For example, it is extremely difficult, if not impossible, to optimize some characteristics of the cycle-structure, such as the cycle length and stage sequence, because it is computationally challenging to update these features of the signal plan second-by-second in real time. Even with a high-power computer processor, it is still operationally difficult to abruptly change the cycle-structure in the middle of a cycle. Moreover, due to the high variability of the traffic flow, any cycle structure that is determined based on past information may not be effective when the traffic pattern deviates from that information, as can occur locally during the course of operation. Therefore, in this paper, a multi-resolution strategy is developed to simultaneously determine the global parameters (such as the cycle length and the cycle-structure) at a coarse-resolution level, and to allow for local adjustments in factors such as the stage-switching times in response to real-time-traffic information at the fine-resolution level.

The basic concept of adaptive traffic-signal systems is that the signal settings are determined online using the most up-to-date traffic information collected in real time by traffic detectors. The group-based adaptive traffic-signal-control logic consists of a proactive global-optimization procedure for determining group-based signal settings. This procedure involves a group-based optimization method for signal timings introduced by Wong (1993), and a local reactive-control policy based on max-pressure theory (Varaiya 2013). The proposed method uses techniques to estimate the turning proportions and queue lengths in individual lanes, a predictive model for control delay, and approximate formulas of the derivatives of the control delay as defined by the Incremental Queue Accumulations (IQAs) developed in Lee et al. (2015b), Lee et al. (2015a), Lee and Wong (2017b), and Lee and Wong (2017a). The fundamental concept of the control delay, IQA, has been continuously updated in the U.S. Highway Capacity Manual (TRB, 2010) to improve the accuracy and robustness of the delay estimation methods. The outputs from the aforementioned studies are used to optimize the signal timings cycle-by-cycle and to adjust the green phase second-by-second with group-based variables. The predicted future traffic flows are used to establish the constraints for maximizing the reserve capacity. The final successor function and the initial guess for the nonlinear mathematical programs are achieved by maximizing the reserve capacity, and are then used to minimize the control delay for the developed objective function and the gradient. Signal timings can be reactively adjusted in response to transient fluctuations in the traffic flow. These adjustments can be made based on the cycle-structure, which includes the signal sequence, cycle time, and duration of green time, as constructed by the global proactive-optimization scheme cycle-by-cycle. This study of the group-based adaptive traffic-signal-control logic is presented separately in two papers, with the formulation provided in part I (this paper) and the implementation presented in part II (Lee et al. 2017). The literature review and comprehensive mathematical frameworks for each sub-part of the proposed group-based adaptive traffic-signal system are described in this paper. The implementation of the system, including the detailed data-processing methods based on a simulation platform, is presented in Lee et al. (2017). The two papers are integrated in that the mathematical frameworks in the first part could not be accomplished without the detailed practical data-processing methods provided in the second part, and the detailed technical components in the second part are materialized and simulated using the formulations proposed in the first part.

The key contribution of the proposed method is the ability to capture and manage the transient and persistent variations in traffic flows in the proposed hierarchical group-based formulation for adaptive signal-controls, which combines a local reactive-control policy and a global proactive-optimization method. At the global level of control, a proactive global-optimization procedure is developed for group-based signal settings. The total delay at the junction for a given signal plan and its derivatives are estimated according to changes in timings following Lee and Wong (2017b) and Lee and Wong (2017a), respectively. These estimates serve as the objective function and the gradient information for the formulation of a mathematical program for the proactive global-optimization procedure. In this mathematical program, the group-based timing variables, that is, the start and duration of the green phase for each signal group, and the successor functions between conflicting signal groups are defined as the control variables. The set of constraints includes limits on the duration of the green signal, clearance time, capacity, and various other constraints that govern the safety and operational requirements specified by the users.

The main advantage of the group-based optimization method is that the most effective cycle structure (or the best overall pattern of signal stages) can be obtained automatically. The definitions of the stages, their combinations and sequences, and the structures of the inter-green periods between consecutive stages do not need to be prespecified. This freedom from predetermined controls allows a high degree of flexibility in tactically identifying the optimized signal plan in response to real-time-traffic information. However, despite these advantages, and although the group-based optimization procedure is highly efficient (Wong, 1996, 1997), it is still anticipated that a significant amount of computing time will be required to obtain the optimized solution. Therefore, the global approach is to compute the group-based solution during the current cycle and to implement the optimized solution in the next cycle, which provides a window for the computational work. Nevertheless, the signal control in the current cycle is adjusted reactively in real time at the local level of control. The resolution of this global control level is in minutes, because the cycle length and cycle-structures are allowed to vary from cycle to cycle.

At the global level of control, the cycle length and the cycle structure of the signal plan for a junction are obtained before the commencement of the current cycle. However, traffic can vary locally and can deviate from the patterns used for computation at the global level. Therefore, a reactive local signal-control policy is developed to locally adjust the signal timings in response to second-by-second variations in traffic patterns. From the global control plan, the group-based optimization procedure identifies the most appropriate cycle-structure according to the latest available information for which the number of stages, sequences, and durations have been automatically generated. At the local level of control, a transition zone is set before and after each switching time from one stage to the next (e.g., three seconds for each zone). During the real-time operation of the current cycle, the queuing information is monitored by the real-time, second-by-second queue-length estimation procedure. The trigger switching occurs in the subsequent stage, when the criteria specified by the local control policy are met, or when the end of the transition zone time is reached. Therefore, the stage-switching time can occur before or after the time set by the global plan. The resolution of this local control level is in seconds, because changes to the signal features are decided second-by-second.

The literature is reviewed in the next section. The hierarchical structure of a group-based adaptive signal-control logic is introduced in Section 3. The mathematical framework for the global proactive-optimization method is described in Section 4. The data-processing methods and procedures for maximizing the reserve capacity and minimizing the control delay are explained in Section 4. In Section 5, the logical framework for the local reactive-control policy is provided, in which a transition zone is set to adjust the green phase according to the traffic pressure. The conclusions are given in Section 6. In part II: Implementation (Lee et al. 2017), detailed methods are provided for implementing the proposed formulations and the performance of the proposed group-based adaptive traffic-signal control is validated in relation to the existing methods.

2. LITERATURE REVIEW

Traffic signals are the most effective means of controlling the conflicting traffic flows at isolated signalized junctions. Adaptive traffic-control systems (ATCS), in which the signal settings are established in real time using the most up-to-date traffic detector information, have shown better performance than fixed-time controls in many places, despite their higher construction and operational costs. To construct an adaptive traffic-control logic, the control system must be capable of estimating the control delays incurred by queued vehicles for both long- and short-term analysis periods. The control delay is commonly used as the performance index in ATCS. The main features of the control delay are derived from the theories and equations in Webster (1958), stochastic queueing theory (Saaty 1961), the concept of queued vehicular delay (Miller 1963b), and the time-dependent delay formula (Akçelik 1980).

2.1 Optimization methods for signal timings

Due to its simplicity, the stage-based method, as defined by a predetermined sequence of stages and inter-green periods, is currently the most commonly used technique for constructing signal-timing plans for isolated signalized junctions and networks. Allsop (1971, 1972) constructed mathematical programs to minimize the delay and maximize capacity at isolated junctions.

To pursue a more flexible structure of signal timings with greater applicability to diverse urban traffic and road geometric conditions, Heydecker and Dudgeon (1987) and Heydecker (1992) introduced a group-based method in which the signal settings were defined by the start and duration of the green signal groups. To address the problems caused by increasing numbers and volumes of traffic and pedestrian streams at junctions, Improta and Cantarella (1984) proposed a method for directly considering the time domain in which each group of streams has a right-of-way, and then specifying the cycle-structure with a set of binary integer variables related to the incompatible signal groups. Similarly, Gallivan and Heydecker (1988) and Heydecker and Dudgeon (1987) developed a method that specified the cycle-structure according to one of the stage sequences generated by Tully (1976). This type of stage sequence does not need to be maintained during optimization, and thus this method provided a high degree of flexibility for the optimization of signal plans through applying mathematical programming formulations. To reduce the large number of stage sequences produced by Tully's method, Heydecker (1992) introduced a procedure to group all of the possible stage sequences into a much smaller number of equivalence classes. The cycle-structure of each equivalence class was represented by a successor function, which made this method more computationally attractive. Most importantly, this method was able to include the structure of the inter-green periods and several additional characteristics of the cycle length and sequence selection into the optimization of the timings.

Because such methods deal directly with groups of traffic streams without needing to maintain the stage-structure during optimization, these methods are called group-based approaches (or "phase-based approaches" in British terminology). A more detailed mathematical framework for applying the group-based method at isolated junctions was reported by Silcock (1997). The group-based method has also been extended to linked signals (Wong 1995, 1996, 1997), combined signal control problems, and assignment problems (Wong and Yang, 1999, Wong et al. 2002). Allsop (1992) provided a comprehensive comparison of the stage-based and group-based methods. The group-based method was also extended to incorporate the design of road markings into the signal optimization procedure in a unified and systematic manner (Wong and Wong 2003a, 2003b, Wong et al. 2006). The advantages of the group-based method include its high degree of flexibility in specifying signal plans (without being restricted to a predetermined set of stages) and its ability to directly optimize the cycle-structure. These advantages make the group-based method capable of dealing with a wide range of traffic patterns in a systematic way.

2.2 Adaptive traffic-control systems

With respect to the adaptive control logic, one of the first contributions toward achieving the goal of real-time responsive signal-control systems was the two-phase heuristic signal-control algorithm introduced by Dunne and Potts (1964) and Green (1967). The fundamental concept of real-time responsive signal-control systems was developed for application in traffic-actuated control and ATCS. Morris and Pak-Poy (1967) and Gordon et al. (2005) proposed traffic-actuated control systems that used simple logic to extend or terminate the current state of a signal controller on the basis of real-time traffic information collected from stop-line or upstream detectors. Despite their simplicity and relatively widespread availability in the field, these systems have several challenges, such as rigid cycle structures, reactive logics that consider only the current traffic conditions, an absence of optimal solutions for long-term or large-scale networks, and a lack of effectiveness in over-saturated traffic situations.

To overcome the above-mentioned critical issues, SCATS has been proposed as an adaptive traffic-control strategy that matches the current traffic states to the best signal plan among the available predetermined signal plans (Lowrie 1982). Miller (1963a) devised an algorithm to decide whether to extend or terminate the current state of a signal controller according to the vehicular delay on a network. This binary choice approach is widely used in current ATCS. As an alternative, Smith (1979a, 1979b, 1981), Smith and Ghali (1990), and Smith and Van Vuren (1993) developed and tested a useful concept for adaptive signal control called the P0 signal-control policy. Moreover, the store-and-forward modeling approach, which is based on Miller's algorithm, has become one of the most widely used mathematical models for optimizing signal timings in real time. This approach has been applied in both under-saturated and over-saturated traffic conditions by Gazis (1964), Gazis and Potts (1965), Grafton and Newell (1967), Ross et al. (1971), Rosdolsky (1973), D'Ans and Gazis (1976), and Michalopoulos and Stephanopoulos (1977a, 1977b). The store-and-forward modeling approach has become the foundation of many traffic-signal optimization methods, such as those developed by Bang and Nilsson (1976) and Aboudolas et al. (2009).

In addition, several methods have been proposed that can optimize or adjust signal timings at both the tactical and local levels. These methods have been applied in numerous practical ATCS, such as PRODY (Farges et al. 1984), SCOOT (Hunt et al. 1981), OPAC (Gartner 1983), RHODES (Mirchandani and Head 2001), ACS-Lite (Luyanda et al. 2003), and CRONOS (Boillot et al. 2006). These systems combine the store-and-forward approach with other models for predicting queue lengths and turning movements. The store-and-forward approach has also been combined with dynamic programming methods.

More recently, Varaiya (2013) proposed the max-pressure traffic-signal-control policy for isolated signalized junctions and signalized networks. Max-pressure control only requires data on the turning flows and queue lengths (at each junction and the adjacent junctions) to achieve network optimization without prior knowledge of the traffic demand. An approximate traffic dynamics approach has also been considered by Smith (2011), Ge and Zhou (2012), Han et al. (2014), and Han and Gayah (2015), and has shown good performance in reducing the computational effort needed for the real-time calculation of signal timings with flexible analysis periods. Moreover, advanced artificial intelligent techniques are used to design real-time traffic signal control logic in Murat et al. (2005), Srinivasan et al. (2006), Liu et al. (2015), and Li et al. (2016). The advantages of adaptive control include the use of sensor technologies to update the traffic patterns in real time and its flexibility in adjusting the control strategies to accommodate the most up-to-date traffic patterns.

3. HIERARCHICAL STRUCTURE OF GROUP-BASED ADAPTIVE SIGNAL-CONTROL LOGIC

This section describes the formulation of the proposed adaptive traffic-control logic, which uses a real-time method of estimating the lane-based queue lengths, a global proactive method of optimizing the signal timings, and a local reactive-control policy. The proposed hierarchical structure for the real-time recursive traffic-control logic is illustrated in Figure 1, followed by detailed explanations of the features involved.

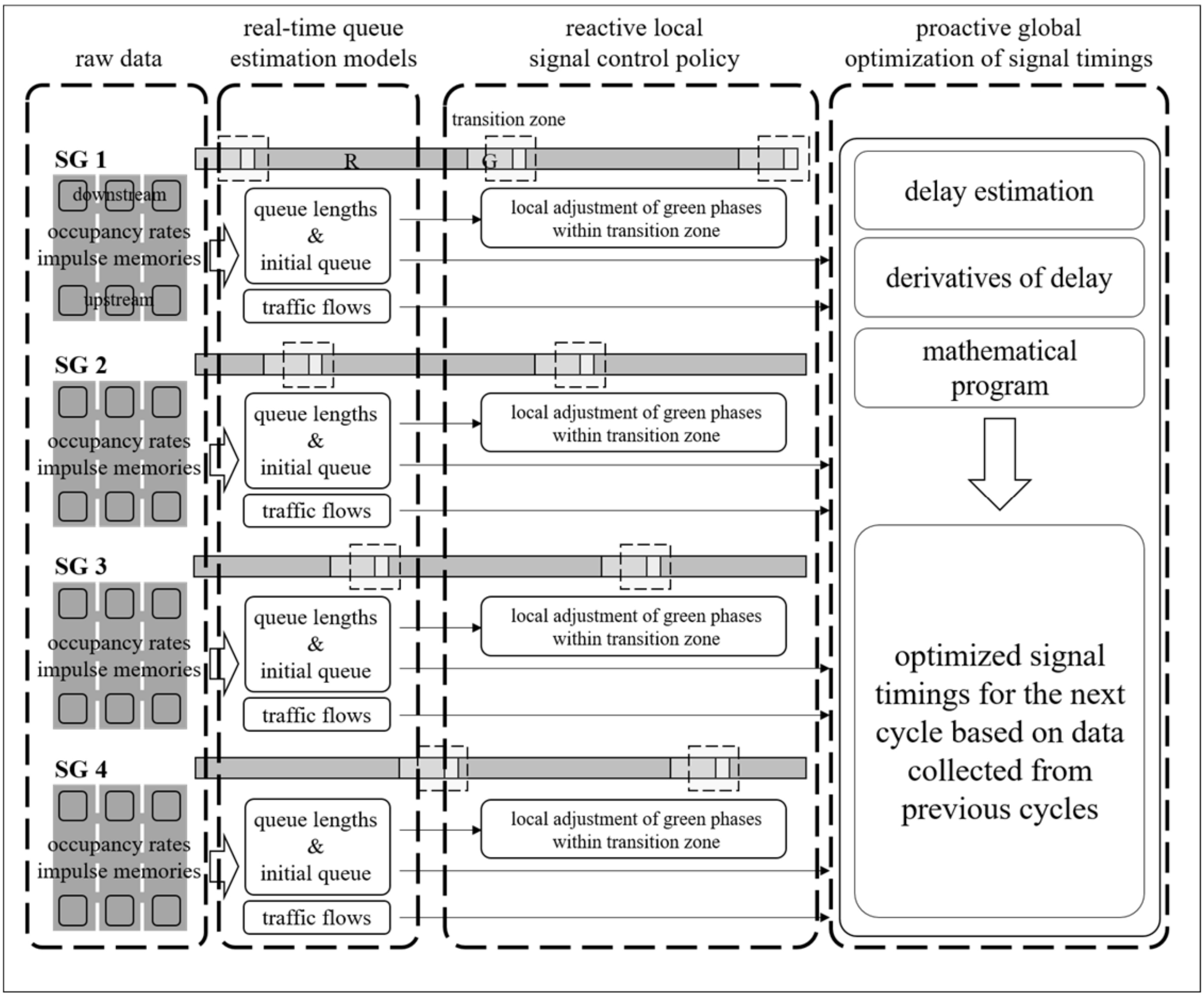

Figure 1. Hierarchical structure for real-time recursive traffic-control logic

As shown in Figure 1, the occupancy rates and impulse memories are collected from upstream and downstream detectors, and then used to estimate the queue lengths and calculate the rates of traffic flow. The real-time estimation method of lane-based queue lengths introduced in Lee et al. (2015a) is modified to dynamically estimate the number of queued vehicles. The following assumptions are applied in this estimation method: 1) upstream and downstream detectors are installed in each lane; 2) the occupancy rates and impulse memories can be collected from this set of detectors; and 3) the data-collection and data-processing systems for real-time recursive logic are sufficiently reliable. The estimated number of queued vehicles is based on discriminant models and the estimated downstream arrivals. The green phases are adjusted every second in response to the estimated queue lengths and the local reactive-control policy. The concept of delay informs the construction of the signal plans for the next cycle in the global proactive-optimization method, which involves a delay-estimation method and a set of variable derivatives for the nonlinear mathematical programs. The impulse memories collected from the upstream detectors are used to estimate the rates of traffic flow. The traffic flow is estimated by vehicles per second for a reasonable period of analysis, and all of these estimates enable a group-based optimization of the signal timings.

4. THE MATHEMATICAL FRAMEWORK FOR THE GLOBAL PROACTIVE-OPTIMIZATION SCHEME

The proactive global-optimization procedure for group-based signal settings is determined based on the method for optimizing signal timings introduced by Wong (1993). This optimization method is applied along with the methods developed in the authors' previous studies, including the methods for estimating the turning proportions and queue lengths in individual lanes, the predictive model for control delay, and the approximate formulas for derivatives of the control delays on the IQAs, in Lee et al. (2015b), Lee et al. (2015a), Lee et al. (2017b), and Lee et al. (2017a), respectively. The time windows for the proactive-optimization method consist of the next, current, and previous cycle. The start and duration of the green time and the cycle time for the next cycle are determined for the current cycle using the mathematical predictive and optimization frameworks applied to each cycle. These frameworks include the procedures for maximizing the reserve capacity and minimizing the control delay. The prediction of future patterns is based on the detector data collected from the previous cycle. The recursive process for predicting the future traffic flows, along with the temporal and spatial information for the control delays on the IQAs, as influenced by the projected future queue lengths, is realized by the rolling horizon approach in Lee et al. (2017). The data-processing method is introduced in sub-section 4.1. Next, the procedures for maximizing the reserve capacity and minimizing the control delay are explained in sub-sections 4.2 and 4.3, respectively.

4.1 Data-processing method

The methods developed in the authors' previous studies, including the methods for estimating the lane-based turning proportions and queue lengths, the proposed predictive models for lane-based control delay, and the approximate expressions of the derivatives for lane-based control delay, all play significant roles in the group-based global-optimization scheme. The proposed data-processing method for this scheme is illustrated in Figure 2.

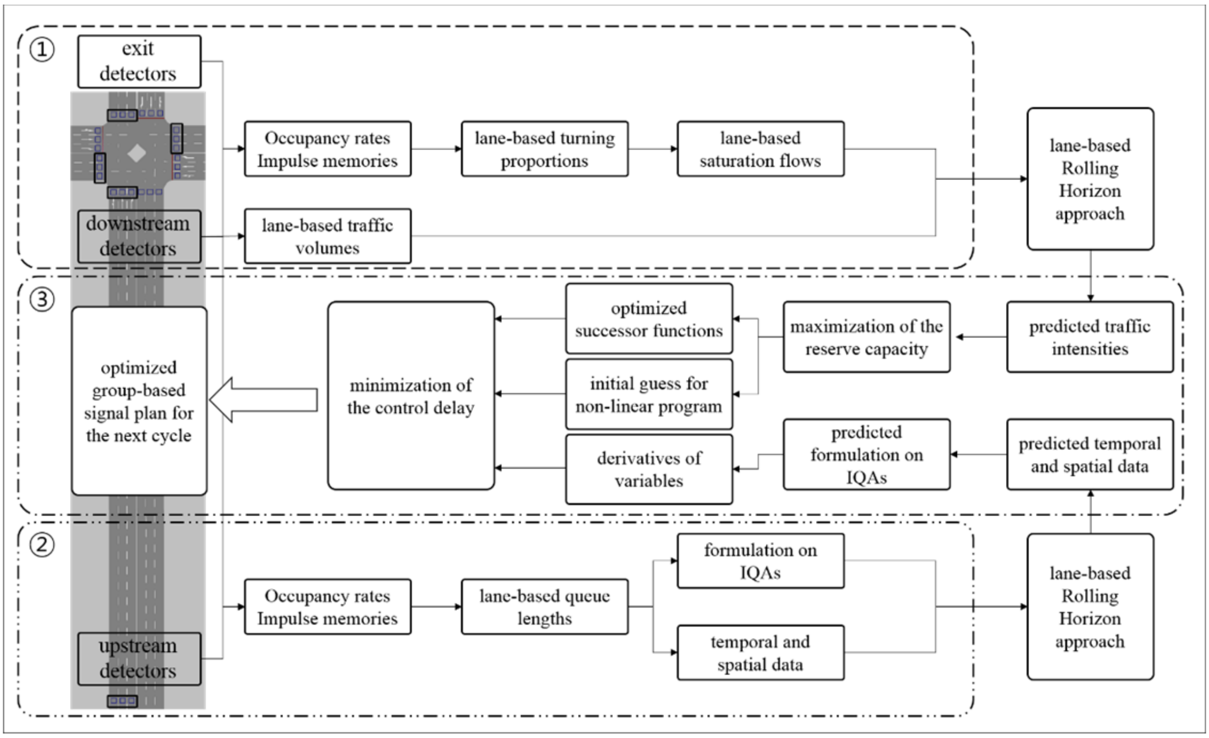

Figure 2. The proposed data-processing method for the global-optimization scheme

In section ① of Figure 2, the occupancy rates and impulse memories collected from the downstream and exit detectors on the previous cycle are used to calculate both the lane-based traffic volumes and the turning proportions, as explained in Lee et al. (2015b). The estimated lane-based turning proportions are then used as input data to compute the lane-based saturation flows in relation to the geometric information on the target intersection. In section ② of Figure 2, the lane-based queue lengths are estimated based on the occupancy rates and impulse memories, as collected from the downstream and upstream detectors using the set of procedures developed in Lee et al. (2015a). The procedure for computing the IQAs and the temporal and spatial information on the lane-based queue lengths in the previous cycle is explained in Lee and Wong (2017b). The outputs of sections ① and ② are then used as the inputs for the lane-based rolling-horizon approach, which predicts the traffic flows, adjustment factors, and queue lengths at the inflection points on the IQAs for the next cycle, as shown in section ③ of Figure 2. The detailed technical implementation of the lane-based rolling horizon approach is illustrated in Lee et al. (2017). The predicted future traffic flows are used to establish the constraints for maximizing the reserve capacity. The final successor function and the initial guess for the nonlinear mathematical programs are achieved through maximizing the reserve capacity, and these outcomes are then used to minimize the control delay in the next cycle. Furthermore, the predicted formulations of the IQAs are established based on the predicted temporal and spatial information on the lane-based queue lengths, and these IQAs are used to construct the objective function and the gradient for minimizing the control delay developed in Lee and Wong (2017a). The traditional framework for optimizing the group-based variables and the results from all of the previous steps are used to minimize the control delay for the future cycle. The optimized group-based variables, including the start points and durations of the green time for each signal group and the cycle time, serve to specify the signal plan for the next cycle. These three computational procedures are applied to the current cycle, based on the data collected from the previous cycle, to construct the signal structure for the next cycle. This data-processing method is performed in each cycle for each individual lane. The practical implementation of the proposed method is described in Lee et al. (2017).

To help explain the mathematical frameworks for optimizing the group-based variables, the definitions of the common indices, parameters, and variables used in this paper are given below.

K	number of lanes
N	number of cycles
T	number of time steps in a cycle
$q_k^n(t)$	number of queued vehicles in lane *k* at time *t* in the *n*th cycle
θ_k^n	time from cycle origin to the start of an actual green signal for control group *k*, divided by the cycle time in the *n*th cycle
ϕ_k^n	duration of the actual green signal for control group *k*, divided by the cycle time in the *n*th cycle
ζ^n	reciprocal value of the cycle time in the *n*th cycle
v_k^n	traffic volumes in lane *k* in the *n*th cycle
s_k^n	saturation flows in lane *k* in the *n*th cycle
μ^n	reserve capacity on the *n*th cycle

$$\boldsymbol{\Gamma}^n = \{\zeta^n, \theta_1^n \ldots \theta_k^n, \phi_1^n \ldots \theta_k^n\}, \forall k \in K, \forall n \in N$$

$$\mathbf{\Omega}_k^n = \left\{\alpha_k^n, \beta_k^n, \gamma_k^n, \delta_k^n\right\}, \forall k \in K, \forall n \in N$$

$$\mathbf{\Psi}_k^n = \left\{\hat{q}_k^n\left(\alpha_k^n\right), \hat{q}_k^n\left(\frac{\theta_k^n}{\zeta^n} + \beta_k^n\right), \hat{q}_k^n\left(\frac{\theta_k^n + \phi_k^n}{\zeta^n} + \gamma_k^n\right), \hat{q}_k^n\left(\frac{\theta_k^n + \phi_k^n}{\zeta^n} + \gamma_k^n + \delta_k^n\right), \hat{q}_k^n\left(\frac{1}{\zeta^n}\right)\right\},$$
$$\forall k \in K, \forall n \in N$$

$$\mathbf{\Phi}_k^n = \left\{\lambda_{\mathrm{A},k}^n, \mu_{\mathrm{B},k}^n, \lambda_{\mathrm{C},k}^n\right\}, \forall k \in K, \forall n \in N$$

K, *N*, and *T* are the common indices used in the steps outlined in this section, and $q_k^n(t)$ represents the final output of the lane-based queue estimation models introduced in Lee et al. (2015a). θ_k^n, ϕ_k^n, and ζ^n are commonly used temporal group-based variables introduced by Heydecker and Dudgeon (1987). v_k^n and s_k^n represent the traffic volumes and saturation flows, respectively, which are used to compute the traffic flows in the individual lanes on each cycle. $\mathbf{\Gamma}^n$ is the set of group-based variables, which are optimized on the basis of the predicted parameters. $\mathbf{\Omega}_k^n$ is the set of adjustment factors used to identify the status of the queue-formation patterns in the current cycle. $\mathbf{\Psi}_k^n$ is the set of queue lengths at an inflection point. $\mathbf{\Omega}_k^n$ and $\mathbf{\Psi}_k^n$ are used as seed parameters to predict future traffic patterns in a rolling-horizon procedure. $\mathbf{\Phi}_k^n$ is the set of slopes, including the arrival rates before an effective green time, and the discharge and arrival rates after an effective green time. $\mathbf{\Phi}_k^n$ is computed on the basis of the predicted $\mathbf{\Omega}_k^n$ and $\mathbf{\Psi}_k^n$ on a rolling-horizon step. The computed $\mathbf{\Phi}_k^n$ and the predicted $\mathbf{\Omega}_k^n$ are then used to estimate the future control delays in the appropriate time windows with the optimized $\mathbf{\Gamma}^n$.

4.2 Maximization of the reserve capacity

The concept of reserve capacity has been widely applied in the design of optimized signal structures and in measuring the performance of signal-control systems for isolated signalized junctions and signalized road networks. Webster and Cobbe (1966) specified the mathematical formula for computing the reserve capacity at simple signalized junctions. Allsop (1972) devised a linear mathematical program to create the most appropriate signal timings for maximizing the reserve capacity of the target intersection, given static traffic conditions. Heydecker and Dudgeon (1987) used the concept of reserve capacity in a set of group-based methods to optimize the signal timings. For network problems, Wong and Yang (1997, 1999) considered the reserve capacity of a signal-controlled road network, and formulated mathematical programs to maximize the capacity. They calculated the reserve capacity based on the advanced mathematical frameworks developed in the literature. In this section, the constraints are given based on real-time-collected traffic information. In part II: Implementation, these constraints provide the most appropriate signal structure using the estimated saturation flow rates. The initial guess for the nonlinear mathematical programming with delay minimization is also provided for each cycle. The procedure for maximizing the reserve capacity is then solved using the approach in Lee et al. (2017).

4.2.1 *Signal structures*

In group-based optimization procedures, the signal timing plans, including the signal sequences, durations of green time, and cycle times, are defined by group-based variables, successor functions, and an inter-green matrix. To determine the signal sequence and structure, Improta and

Cantarella (1984) introduced a set of binary-integer variables related to incompatible signal groups. These variables directly specify the cycle-structure based on the time domain for each group of streams having the right-of-way. Gallivan and Heydecker (1988) and Heydecker and Dudgeon (1987) devised a method for defining the cycle-structure as specified by one of the stage sequences generated by Tully (1976). Heydecker (1992) introduced a procedure for grouping all of these possibilities into a much smaller number of equivalence classes, thereby reducing the large number of stage sequences produced by Tully's method.

In this paper, the signal sequence and structure are defined by a successor function, which is a more attractive computational method and is widely used in group-based approaches. This method deals directly with groups of traffic streams without requiring a predetermined signal stage-structure in the optimization process. The successor functions have been widely applied in diverse group-based approaches to optimize the signal timings (Silcock 1997, Wong 1995, 1996, 1997, Wong and Yang 1999, Wong et al. 2001). The successor function, the corresponding clearance time matrix, and the initial guesses of the group-based variables for the nonlinear mathematical programs (which involve maximization methods for identifying the greatest reserve capacity among the possible set of successor functions) are all determined under the given traffic conditions for each cycle. The successor function Λ_{kj}^{n} is set as 0 if the start of the green phase for lane *k* precedes that for lane *j*, and is otherwise set as 1. The first signal group is assumed to be fixed in every cycle because the optimization procedure is repeated for each cycle, and an increase in the number of successor functions (all of which should be considered in the optimization procedures at the end of each cycle) becomes meaningless with higher computational demand. A detailed practical description is provided in Lee et al. (2017).

4.2.2 *Control variables and constraints*

To maximize the reserve capacity using linear mathematical programming, a set of constraints is developed for the cycle time, duration of green time, clearance time, and traffic flows. The set of group-based variables, $\Gamma^n = \{\zeta^n, \theta_1^n .. \theta_k^n, \phi_1^n .. \phi_k^n\}, \forall k \in K, \forall n \in N$, is designed to maximize the reserve capacity at the target intersection μ^n on the *n*th cycle. The constraints are established using the following equations.

$$\frac{1}{C_{\min}} \geq \zeta^n \geq \frac{1}{C_{\max}}, \forall n \in N \tag{1}$$

$$G_{k,\min}\zeta^n \leq \phi_k^n \leq G_{k,\max}\zeta^n, \forall k \in K, \forall n \in N \tag{2}$$

$$\hat{\phi}_k^n = \phi_k^n + e_k\zeta^n, \forall k \in K, \forall n \in N \tag{3}$$

$$\theta_k^n + \phi_k^n - \theta_j^n + C_{kj}\zeta^n \leq \Lambda_{kj}^n, \forall k \in K, \forall n \in N \tag{4}$$

$$\theta_k^n + \phi_k^n - \theta_j^n + C_{kj} = 0, \forall n \in N \tag{5}$$

$$\phi_k^n + e_k\zeta^n - \mu^n \frac{v_k^n}{\tilde{s}_k^n p_k} \geq 0, \forall k \in K, \forall n \in N \tag{6}$$

The cycle time is restricted to a certain appropriate range in Equation (1). $C_{\min}$ and $C_{\max}$ are the minimum and maximum cycle times, respectively. In Equation (2), the duration of green time is within the acceptable upper and lower limits, $G_{k,\min}$ and $G_{k,\max}$, which are the minimum and maximum durations of green time, respectively. $G_{k,\min}$ depends on the duration of green time for the compatible pedestrian streams. The effective duration of green time, $\hat{\phi}_k^n$, is confined in Equation (3). The effective green duration is generally one second longer than the actual green time, and e_k is set as 1. In Equation (4), the cycle-structure and signal sequence are specified to define the relationship between incompatible pairs of signal groups, k and j, on the basis of the successor function, Λ_{kj}^n, and clearance time, C_{kj}. If two incompatible signal groups border each other in terms of the clearance time, then Equation (5) specifies the relationship between the closed adjacent green signal groups because the starting signal group is fixed in this example, and Equation (4) is inadequate for dealing with the clearance time between two adjoining signal groups. The traffic flow in lane k on the nth cycle, $v_k^n / \tilde{s}_k^n$, is used in Equation (6) to keep the traffic streams below a specific acceptable congestion level by setting maximum acceptable levels for the saturation flow rates, p_k, which are the capacity constraints.

4.2.3 *Linear programming formulation for the reserve capacity maximization problem*

The constraints are expressed in the form of a matrix for implementing the developed optimization procedure in traffic-simulation programs and computer programs. This matrix of constraints retains the possibility of extending the devised algorithm to a signal-controlled network. The following equation illustrates the form of the matrix used to maximize the reserve capacity.

$$\max \mu^n \tag{7}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & \cdots & a_{1K} \\ a_{21} & a_{22} & & & & & & a_{2K} \\ a_{31} & & a_{33} & & & & & a_{3K} \\ a_{41} & & & a_{44} & & & & a_{4K} \\ a_{51} & & & & a_{55} & & & a_{5K} \\ a_{61} & & & & & a_{66} & & a_{6K} \\ \vdots & & & & & & \ddots & \vdots \\ a_{N_C 1} & a_{N_C 2} & a_{N_C 3} & a_{N_C 4} & a_{N_C 5} & a_{N_C 6} & \cdots & a_{N_C K} \end{bmatrix} \begin{bmatrix} \theta_1^n \\ \vdots \\ \theta_K^n \\ \phi_1^n \\ \vdots \\ \phi_K^n \\ \zeta^n \\ \mu^n \end{bmatrix} \begin{matrix} = \\ \le \end{matrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Lambda_{11}^n \\ \vdots \\ \Lambda_{KJ}^n \end{bmatrix}, \forall n \in N \tag{8a}$$

$$A^n \begin{Bmatrix} \Gamma^n \\ \mu^n \end{Bmatrix} \begin{matrix} = \\ \le \end{matrix} \begin{Bmatrix} 0 \\ \Lambda^n \end{Bmatrix}, \forall n \in N \tag{8b}$$

The objective function to maximize the reserve capacity on the nth cycle is given in Equation (7). In Equation (8a), the indices K, N_C, N, and J are the number of lanes, total number of constraints, number of cycles, and number of lanes for which the exit lane corresponds to lane k, respectively. The coefficients in the aforementioned constraints are given on the left-hand side of the matrix in the form of equality and inequality constraints, along with the group-based variables and the reserve-capacity variable. The constant vector (which includes the zero column and the successor functions) is described on the right-hand side in the set of equality and inequality constraints. The coefficients on the left-hand side vary according to the set of successor functions. For a typical layout of a signalized junction that includes three-lane and four-arm approaches, there are 32 sets of coefficients, which correspond to the set of successor functions in the constraints on the signal structures in Lee et al. (2017). Equation (8b) illustrates the general matrix form of the constraints, with the group-based variables and the reserve capacity variable. In Lee et al. (2017), this matrix form is established in computer programs using Fortran to identify the recursive process cycle-by-cycle.

4.3 Minimization of the control delay

The concept of control delay, which is commonly used as the performance index in ATCS, is derived from the theories and equations in Webster (1958), stochastic queueing theory (Saaty 1961), queued vehicular delay (Miller 1963b), and the time-dependent delay formula (Akçelik 1980). The methods for estimating the control delay are integrated with major queue-estimation methods, including the conservation equation (Lindley 1952) and shockwave theory (Lighthill and Whitham 1955, Richards 1956). In addition to the aforementioned cornerstones for the development of the control delay and queue-estimation methods, a diverse set of delay-estimation methods has been developed, which includes methods for assessing both the deterministic and stochastic characteristics of control delay.

The developed control delay formulation given in Lee and Wong (2017b) and the approximate expressions of the derivatives explained in Lee and Wong (2017a), are used as the objective functions and gradients, respectively, in the nonlinear mathematical program for minimizing delay, as based on $\mathbf{\Gamma}_0^n$ and $\hat{\Lambda}^n$, which are defined in the capacity maximization procedure of the previous step. A discrete directional search method (introduced in the IMSL Fortran Numerical Library), which is illustrated in Lee et al. (2017), is used to find the solution. In this case, nonlinear mathematical programs are used to assess the lane-based control delay and approximate the expressions of the derivatives induced from the IQAs with the adjustment factors and group-based variables. The detailed formulations of these programs are provided in the following sub-sections.

4.3.1 *Control variables and constraints*

The set of constraints for a cycle time, namely the duration of green time, the clearance time, and the traffic flows used in the capacity maximization procedure, are partially applied in the nonlinear mathematical programs to minimize the control delay with a certain modification. For the future time windows, the control delays derived from the IQAs in the future three cycles are considered as a performance index. This index prevents a preposterous solution caused by a transient drastic fluctuation in traffic demand in a particular cycle. The set of group-based variables in the future three cycles, $\overline{\Gamma}^n \supset \Gamma^n, \Gamma^{n+2}, \Gamma^{n+3}$, is considered in the modified set of constraints developed in the previous step, with additional constraints derived from the concept of IQAs. The modified set of constraints is constructed as in the following equations.

$$\overline{\mathbf{\Gamma}}^n = \left\{\zeta_1^n, \theta_{1,1}^n \ldots \theta_{1,k}^n, \phi_{1,1}^n \ldots \theta_{1,k}^n, \zeta_2^n, \theta_{2,1}^n \ldots \theta_{2,k}^n, \phi_{2,1}^n \ldots \theta_{2,k}^n, \zeta_3^n, \theta_{3,1}^n \ldots \theta_{3,k}^n, \phi_{3,1}^n \ldots \theta_{3,k}^n\right\}, \quad \forall k \in K, \forall n \in N \tag{9}$$

$$\frac{1}{C_{\min}} \geq \zeta_c^n \geq \frac{1}{C_{\max}}, \forall c \in C, \forall n \in N \tag{10}$$

$$G_{k,\min}\zeta_c^n \leq \phi_{c,k}^n \leq G_{k,\max}\zeta_c^n, \forall c \in C, \forall k \in K, \forall n \in N \tag{11}$$

$$\hat{\phi}_{c,k}^{n} = \phi_{c,k}^{n} + e_k \zeta_c^n, \forall k \in K, \forall c \in C, \forall n \in N \tag{12}$$

$$\theta_{c,k}^{n} + \phi_{c,k}^{n} - \theta_{c,j}^{n} + C_{kj}\zeta_c^n \le \Lambda_{c,kj}^{n}, \forall c \in C, \forall k \in K, \forall n \in N \tag{13}$$

$$\theta_{c,k}^{n} + \phi_{c,k}^{n} - \theta_{c,j}^{n} + C_{c,kj} = 0, \forall c \in C, \forall n \in N \tag{14}$$

$$\phi_{c,k}^{n} + e_k \zeta_c^n - \mu_c^n \frac{v_{c,k}^{n}}{\tilde{s}_{c,k}^{n} p_k} \ge 0, \forall c \in C, \forall k \in K, \forall n \in N \tag{15}$$

The group-based variables in the three cycles, $\overline{\Gamma}^n$, are defined by Equation (9) for all of the nth cycles and all of the lanes in a signal-controlled intersection. In Equation (10), the cycle times in the future three cycles are within the maximum and minimum cycle lengths, $C_{\min}$ and $C_{\max}$. $G_{k,\min}$ and $G_{k,\max}$, which are the minimum and maximum durations of green time, are applied in Equation (11) to restrict the duration of green time in lane k for the three future cycles, c+1 to c+3. The effective duration of green time, $\hat{\phi}_{c,k}^{n}$, is defined in Equation (12), and e_k is generally set as one second to extend the actual duration of green time to the effective duration of the green signal. e_k is the extra effective green time for Stream k in a time unit and is usually taken as 1 second in the group-based optimization procedure for signal timings. In Equation (13), the clearance time, C_{kj}, and the final successor function derived from the capacity maximization, $\Lambda_{c,kj}^{n}$, define the relationship between incompatible pairs of signal groups, k and j. These factors specify the cycle-structure and the signal sequence of the group-based variables, which are similarly applied in the future three cycles.

For two incompatible signal groups that are adjacent to each other, Equation (14) specifies the precise clearance time between them, without successor functions on the right-hand side in the equality constraints. These equality constraints are essential for clearly defining the cycle-structure in cases where the starting signal group is unchangeable. Equation (15) determines that the effective duration of green time can deal with the traffic flow in lane k in the three cycles, $v_{c,k}^{n} / \tilde{s}_{c,k}^{n}$, considering the maximum acceptable degree of saturation in flow rates, p_k, needed to operate the traffic streams below a specific acceptable congestion level. μ_c^n is set as 1 in the delay minimization problem to find the solution under the given capacity constraints. In addition to the aforementioned basic constraints, the constraints for defining the lane-based polygons in the IQAs are specified, as illustrated in the following figure.

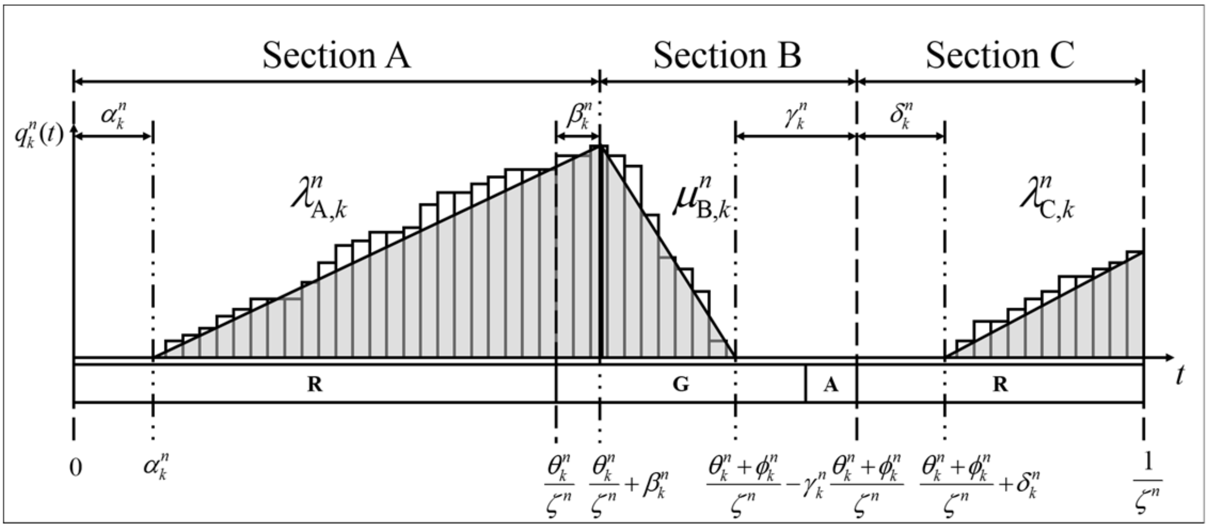

Figure 3. Lane-based polygons in the IQAs

In Figure 3, the polygons in the IQAs for describing the control delay in a cycle should be larger than 0, and the constraints are established to maintain the areas of the trapezoids and triangles as positive in sections A, B, and C. To reduce the sensitivity of the group-based variables in the final solution with respect to transient fluctuations in traffic demand, the group-based variables for the future three cycles are considered with the additional constraints given in the following equations.

$$-\theta_{c,k}^n+\left(\alpha_{c,k}^n-\beta_{c,k}^n\right)\zeta_c^n\leq 0 \tag{16}$$

$$-\lambda_{A,c,k}^n\theta_{c,k}^n+\left\{\lambda_{A,c,k}^n\alpha_{c,k}^n-2\hat{q}_{c,k}^n\left(\alpha_{c,k}^n\right)-\lambda_{A,c,k}^n\beta_{c,k}^n\right\}\zeta_c^n\leq 0 \tag{17}$$

$$-\phi_{c,k}^n+\left(\beta_{c,k}^n-\gamma_{c,k}^n\right)\zeta_c^n\leq 0 \tag{18}$$

$$\begin{aligned}&-2\lambda_{A,c,k}^n\theta_{c,k}^n-\mu_{B,c,k}^n\phi_{c,k}^n\\&-\left\{2\hat{q}_{c,k}^n\left(\alpha_{c,k}^n\right)+2\lambda_{A,c,k}^n\left(\beta_{c,k}^n-\alpha_{c,k}^n\right)+\mu_{B,c,k}^n\left(\gamma_{c,k}^n-\beta_{c,k}^n\right)\right\}\zeta_c^n\leq 0\end{aligned} \tag{19}$$

$$-\lambda_{C,c,k}^n\theta_{c,k}^n-\lambda_{C,c,k}^n\phi_{c,k}^n+\left\{2\hat{q}_{c,k}^n\left(\frac{1}{\zeta_c^n}\right)-\lambda_{C,c,k}^n\left(\gamma_{c,k}^n+\delta_{c,k}^n\right)\right\}\zeta_c^n\leq-\lambda_{C,c,k}^n \tag{20}$$

$$\theta_{c,k}^n+\phi_{c,k}^n+\left(\gamma_{c,k}^n+\delta_{c,k}^n\right)\zeta_c^n\leq 1 \tag{21}$$

Equations (16) and (17), which are derived from the first formula in Equation (25), confine the temporal and spatial coordinates, respectively, to being positive for the polygons in section A. For each future cycle c, the set of adjustment factors, $\mathbf{\Omega}_k^n$, and the set of queue lengths at an inflection point $\mathbf{\Psi}_k^n$ are used as $\mathbf{\Omega}_{c,k}^n$ and $\mathbf{\Psi}_{c,k}^n$, respectively. $\mathbf{\Omega}_{c,k}^n$ and $\mathbf{\Psi}_{c,k}^n$ are predicted by the recursive process of the Kalman filters (as described in Lee and Wong (2017a)) to calculate the set of slopes $\mathbf{\Phi}_{c,k}^n$ and the areas of the polygons in the IQAs for the future cycles. The second formula in Equation (25), which calculates the area of the polygons in each section, is used to derive Equations (18) and (19), in which the temporal and spatial coordinates are restricted, respectively. In Equations (20) and

(21), the x- and y-coordinates of the polygons in section C (which are estimated by the last formula in Equation (25)) are constrained to maintain the temporal and spatial information as positive. The new constraints follow the basic form of the constraints that are required in computer programming because of the extended applicability needed for implementation in the simulation programs.

4.3.2 *Objective function and gradients*

The performance index is the control delay on the future three cycles, which is derived from the developed polygons of the IQAs as described in Lee and Wong (2017b). This index is based on the predicted parameters for the future three cycles in the delay minimization program. The approximate expressions of the derivatives in the IQAs are derived in Lee and Wong (2017a). These expressions are used as substitutes for the gradients in the nonlinear mathematical programs. The optimization procedure is described by the following equations.

$$D_c^n\left(\Gamma_c^n\right)=\frac{1}{2}\sum_{k=1}^{K}\left[\begin{array}{l}\left\{2\hat{q}_{c,k}^n\left(\alpha_{c,k}^n\right)+\lambda_{\mathrm{A},c,k}^n\left(\frac{\theta_k^n}{\zeta_c^n}+\beta_{c,k}^n-\alpha_{c,k}^n\right)\right\}\left(\frac{\theta_k^n}{\zeta_c^n}+\beta_{c,k}^n-\alpha_{c,k}^n\right)\\+\left\{2\hat{q}_{c,k}^n\left(\alpha_{c,k}^n\right)+2\lambda_{\mathrm{A},c,k}^n\left(\frac{\theta_{c,k}^n}{\zeta_c^n}+\beta_{c,k}^n-\alpha_{c,k}^n\right)+\mu_{\mathrm{B},c,k}^n\left(\frac{\phi_{c,k}^n}{\zeta_c^n}+\gamma_{c,k}^n-\beta_{c,k}^n\right)\right\}\left(\frac{\phi_{c,k}^n}{\zeta_c^n}+\gamma_{c,k}^n-\beta_{c,k}^n\right)\\+\left\{2\hat{q}_{c,k}^n\left(\frac{1}{\zeta_c^n}\right)-\lambda_{\mathrm{C},c,k}^n\left(\frac{1}{\zeta_c^n}-\frac{\theta_{c,k}^n+\phi_{c,k}^n}{\zeta_c^n}-\gamma_{c,k}^n-\delta_{c,k}^n\right)\right\}\left(\frac{1}{\zeta_c^n}-\frac{\theta_{c,k}^n+\phi_{c,k}^n}{\zeta_c^n}-\gamma_{c,k}^n-\delta_{c,k}^n\right)\end{array}\right] \tag{22}$$

$$D_c^n=\frac{1}{2}\sum_{k=1}^{K}\left[D_{\mathrm{A},c,k}^n\left(\Gamma^n,\hat{\Omega}_{c,k}^n,\hat{\Phi}_{c,k}^n\right)+D_{\mathrm{B},c,k}^n\left(\Gamma^n,\hat{\Omega}_{c,k}^n,\hat{\Phi}_{c,k}^n\right)+D_{\mathrm{B},k}^n\left(\Gamma^n,\hat{\Omega}_{c,k}^n,\hat{\Phi}_{c,k}^n\right)\right] \tag{23}$$

$$\frac{\partial D_c^n\left(\Gamma_c^n\right)}{\partial \Gamma_c^n}=\left(\frac{\partial D_c^n\left(\Gamma_c^n\right)}{\partial \theta_{c,k}^n},\ldots,\frac{\partial D_C^n\left(\Gamma_C^n\right)}{\partial \theta_{C,K}^n},\frac{\partial D_c^n\left(\Gamma_c^n\right)}{\partial \phi_{c,k}^n},\ldots,\frac{\partial D_C^n\left(\Gamma_C^n\right)}{\partial \phi_{C,K}^n},\frac{\partial D_c^n\left(\Gamma_c^n\right)}{\partial \zeta_c^n},\ldots,\frac{\partial D_C^n\left(\Gamma_C^n\right)}{\partial \zeta_C^n}\right) \tag{24}$$

$,\forall n \in N$

The control delay for a particular cycle among the future cycles can be calculated using modified versions of Equations (22) and (23), which include the index c for the number of cycles in the optimization procedure and the area of polygons in the IQAs (see Lee and Wong 2017b). The set of gradients for the group-based variables for the three cycles derived in Lee and Wong (2017a) is provided in Equation (24).

In Part II: Implementation (Lee et al. 2017), the above-mentioned control variables and constraints, the matrix form of the delay minimization problem, and the objective function and gradients are integrated into the optimization procedure according to the rolling-horizon approach and the nonlinear mathematical programs. The optimization procedure for minimizing the control delay on the consecutive future cycles (the (n + 1), (n + 2), and (n + 3)th cycles) is performed on the nth cycle to predict the cycle-structure, as designed by the optimum group-based variables for the (n + 1)th cycle. These variables, in turn, are determined on the basis of the predicted temporal and spatial traffic information derived from data collected on the (n − 1)th cycle. The time needed for data processing (including the time for the data collection and the application of the estimation methods, predictive models, recursive processes, and optimization methods) is given for each cycle, to provide the optimized cycle structure for the future cycle given the fluctuations in traffic conditions, such as the turning proportions and traffic demand, which are measured in real time.

4.3.3 *Nonlinear programming for the delay minimization problem*

The form of the nonlinear programming for the delay minimization problem is a modified form of the programming used in the capacity maximization problem to integrate the objective function, gradients, coefficients, variables, and constants in the three future cycles. The modified form of the nonlinear programming for the delay minimization problem is given in the following equations.

$$\min \sum_{c=1}^{3} D_c^n\left(\boldsymbol{\Gamma}_c^n\right) \tag{25}$$

$$\text{subject to} \begin{bmatrix} A_1^n & & & & & \\ & A_1'^n & & & & \\ & & A_2^n & & & \\ & & & A_2'^n & & \\ & & & & A_3^n & \\ & & & & & A_3'^n \end{bmatrix} \begin{Bmatrix} \boldsymbol{\Gamma}_1^n \\ \mu_1^n \\ \boldsymbol{\Gamma}_2^n \\ \mu_2^n \\ \boldsymbol{\Gamma}_3^n \\ \mu_3^n \end{Bmatrix} \begin{matrix} = \\ \le \end{matrix} \begin{Bmatrix} 0 \\ \Lambda_1^n \\ B_1^n \\ 0 \\ \Lambda_2^n \\ B_2^n \\ 0 \\ \Lambda_3^n \\ B_3^n \end{Bmatrix}, \forall n \in N \tag{26}$$

$$\rightarrow \begin{bmatrix} A_c^n & \\ & A_c'^n \end{bmatrix} \begin{Bmatrix} \boldsymbol{\Gamma}_c^n \\ \mu_c^n \end{Bmatrix} \begin{matrix} = \\ \le \end{matrix} \begin{Bmatrix} 0 \\ \Lambda_c^n \\ B_c^n \end{Bmatrix}, \forall n \in N, \forall c \in C \tag{27}$$

$$\frac{\partial D_c^n\left(\boldsymbol{\Gamma}_c^n\right)}{\partial \boldsymbol{\Gamma}_c^n}, \forall n \in N \tag{28}$$

In Equation (25), the objective function is expressed as $\sum_{c=1}^{3} D_c^n\left(\boldsymbol{\Gamma}_c^n\right)$, which is the total control delay for the next consecutive three cycles, where μ_c^n is set as 1, and Λ_c^n, A_c^n, and the initial guess for the delay minimization problem are given in the capacity maximization procedure. To consider the newly added constraints for confining the area of the polygons in the IQAs, the coefficient vector on the left-hand side and the coefficient vector of constraints on the right-hand side are applied as $A_c'^n$ and B_c^n, respectively, to the matrixes in Equations (26) and (27), which are developed from Equations (7) and (8). $A_c'^n$ and B_c^n are applied differently in the delay minimization program cycle-by-cycle, according to the most appropriate successor function chosen in the capacity maximization procedure and the predicted parameters of the IQAs.

In Equation (26), the vectors of the coefficients on the constraints, the group-based variables on the left-hand side, and the vectors of the constants on the right-hand side are derived from Equations (7) and (8). To consider the next three consecutive cycles, the index c provides the number of cycles to be applied in the delay optimization procedure given by Equation (27). The coefficients described in the constraints from Equations (10) to (21) are on left-hand side of the matrix along with the group-based variables for the next three consecutive cycles and the capacity variables, which are set as 1. However, the vectors of the constants, including the zero column, the successor functions, and the right-hand-side vector in Equations (16) to (21), are given on the right-hand side of the set of constraints. The sets of equality and inequality constraints are described in Equation (27). The common constraints for a cycle time, namely, the duration of green time, the cycle-structure, and the capacity, are included in Equation (27), along with the newly developed constraints to guarantee that the areas of the polygons in each subsection of the IQAs are positive. The set of gradients for the group-based variables for the three cycles derived in Lee and Wong (2017a) is provided in Equation (28).

5. THE LOGICAL FRAMEWORK FOR THE LOCAL REACTIVE-CONTROL POLICY

The signal timings can be reactively adjusted with respect to the transient fluctuations of traffic flows based on the cycle-structure (which includes the signal sequence, cycle time, and duration of green time), as constructed by the global proactive-optimization scheme cycle-by-cycle. Although the real-time adjustment of signal timings can be easily attained in a stage-based approach (in which the cycle structure is predetermined), the local reactive-control policy cannot be included as a set of simple methods for adjusting the green phase in a group-based ATCS. Instead, the flexible cycle structures are designed in each cycle in response to the time-varying traffic flows. The pairs of incompatible green phases and the sequences of green times in a cycle (as expressed by the successor functions in the group-based approach for the given cycle structure) all play significant roles in the local reactive-control policy. In this paper, the local adjustment of green time is based on the max-pressure control policy developed by Varaiya (2013). This policy is chosen because it involves a decentralized property, provides stability in accommodating real-time queue lengths, and requires minimum prior knowledge about the traffic flows. Varaiya (2013) introduced the max-pressure controller for arbitrary networks of signalized intersections in the book *Advances in Dynamic Network Modeling in Complex Transportation Systems*. According to the max-pressure control policy, the sum of queue lengths multiplied by the saturation flows for each signal group is the trigger for changing the current signal group to the next incompatible signal group.

In addition to the original max-pressure adaptive-controller problem, the concepts of a transition zone and a cycle-varied pair of incompatible signal groups are developed to facilitate the group-based local reactive-control policy second-by-second. In the global proactive-optimization method, the local reactive-control policy for each signal group or each lane is activated for a transition zone on the basis of comparisons between the real-time queue lengths and the saturation flows between a pair of incompatible signal groups in the given cycle structure. The stage-based and group-based approaches to the max-pressure control policy are introduced along with several additional processes, including a transition zone and a cycle-varied pair of incompatible signal groups. These approaches apply a max-pressure controller to the flexible group-based signal structure, as explained in the following sub-sections. The stage-based max-pressure control policy is introduced in sub-section 5.1, and the group-based approach to the max-pressure control policy is described in sub-section 5.2. The transition zones for the group-based approach are described in sub-section 5.3. The programs and the installment of the local reactive-control policy are explained in the second part of this study (Lee et al. 2017).

5.1 Stage-based max-pressure control policy

The development of the group-based local reactive-control policy serves as a stepping-stone to construct a stage-based approach for developing a max-pressure control policy. This policy then informs a local reactive-control policy for the hierarchical adaptive traffic-signal control. The stage-based max-pressure control policy was introduced as a component of stage-based hierarchical adaptive traffic-signal control (Lee et al. 2016). In the stage-based approach, the pressure on the current stage is continuously and dynamically compared to the pressure on the next stage within a transition zone. A transition zone is the permitted amount of time in which the max-pressure signal-control policy can decide whether to extend or finish the current stage. This zone is set as a pivot of several seconds before and after the end time of each stage (as established by the global-optimization scheme). If the pressure on the current stage is greater than that on the next stage, then a signal controller extends the current stage until the very end of the current time slot. Otherwise, the current stage finishes at the effective end time of the current time slot, and the next stage starts after the duration of the inter-stage. This modified Varaiya's max-pressure signal-control policy can be described with respect to a stage-based approach as follows.

1) When the current state of a signal controller is a stage for the traffic stream *j* and the current time slot is within a transition zone, compare the pressure on the current stream *j*, $\sum_{k=1}^{K_j} s_j^n q_k^n(t)$, to the pressure on the next stream $j + 1$, $\sum_{k=1}^{K_{j+1}} s_{j+1}^n q_k^n(t)$, in each time slot during a transition zone.
2) If $\sum_{k=1}^{K_j} s_j^n q_k^n(t) \geq \sum_{k=1}^{K_{j+1}} s_{j+1}^n q_k^n(t)$, the current stage for the traffic stream *j* is continued until the end time of the current time slot.
3) If $\sum_{k=1}^{K_j} s_j^n q_k^n(t) < \sum_{k=1}^{K_{j+1}} s_{j+1}^n q_k^n(t)$, the current stage for the traffic stream *j* finishes at the effective end time of the current time slot, and the next stage for traffic stream $j + 1$ starts after the inter-stage time.
4) If the current time slot reaches the end time of the transition zone, the current stage for the stream *j* finishes at the end time of the current time slot, and the next stage for the stream $j + 1$ starts after the inter-stage time.

5.2 Group-based approach to the max-pressure control policy

The evident difference between the stage-based and group-based approaches to the max-pressure control policy relates to the varying pairs of incompatible signal groups in each cycle. These flexible pairs of signal groups induce a number of combinations in the set of comparisons between incompatible groups. Fortunately, the start signal group for each cycle is fixed in this paper, in that at least two pairs of incompatible groups are constant. The basic concept of the group-based max-pressure control policy is similar to that of the stage-based approach, except for its flexible cycle structure. The successor function Λ_{kj}^n (which maximizes the reserve capacity through the capacity maximization problem in the global proactive-optimization method) is used to define the sets of pairs and the cliques of incompatible signal groups for the future cycle.

The pairs of incompatible signal groups illustrate that the couples of signal groups serve mutually exclusive traffic movements. The traffic movements for going straight make a pair with the permissive, protected left-turning traffic movements on the opposite side. The clique of incompatible signal groups indicates that the pairs of incompatible signal groups can be duplicated simultaneously. For the layout of the general signal-controlled intersection consisting of four arms and three movements on each approach with LTOR (as represented in Table 1), two cliques and four pairs of incompatible signal groups are used to define the cycle structure for the group-based local adjustment of the green time in each cycle. The structure of the signal groups is expressed by the sets of pairs and cliques illustrated in Figure 4.

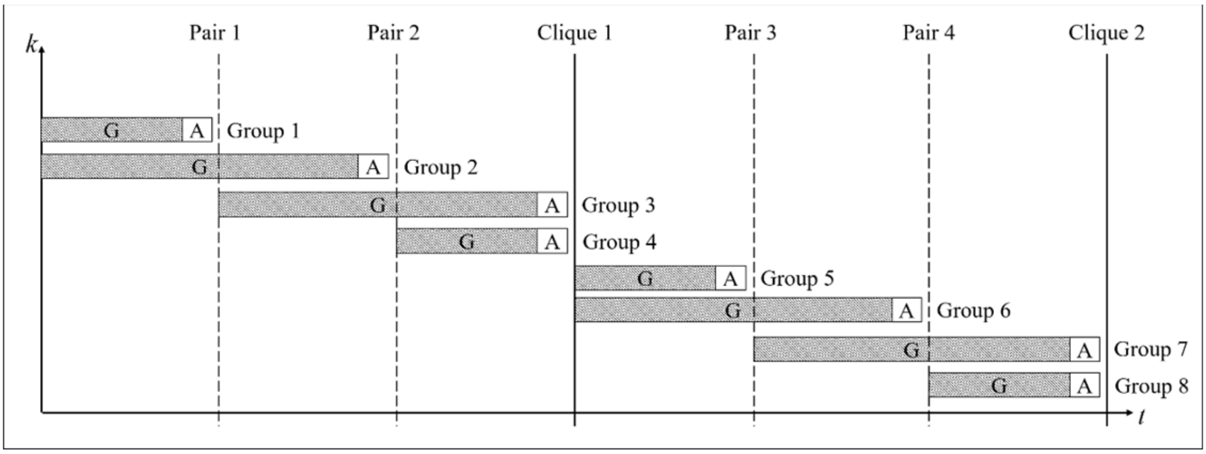

Figure 4. The pairs and cliques of incompatible signal groups

Twelve kinds of traffic movements can be served by the eight signal groups in the typical layout of a signalized intersection in Lee et al. (2017). One signal group can either follow or precede the other signal group in each pair of incompatible signal groups. Then, each defined pair of signal groups coexists with the other pair of signal groups in the same clique of signal group pairs. Two mutually exclusive cliques follow or precede each other in each cycle. At the top of Figure 4, the borders of the pairs and cliques are shown by broken lines and solid lines, respectively.

Signal groups 1 and 3 are in pair 1, and groups 2 and 4 are in pair 2. Moreover, pair 3 includes signal groups 5 and 7, and pair 4 contains signal groups 6 and 8. Pairs 1 and 2 and pairs 3 and 4 belong to cliques 1 and 2, respectively. All of the groups that belong to the same pairs serve mutually exclusive, simultaneous traffic flows. These flows are separated by the broken lines with inter-green times, including the amber and all red signals. The traffic movements that are served by the groups in one clique are incompatible with those of the other clique. The defined hierarchical structure of the signal groups is used to formulate the local reactive-control policy in the flexible cycle structure, as represented in Figure 5.

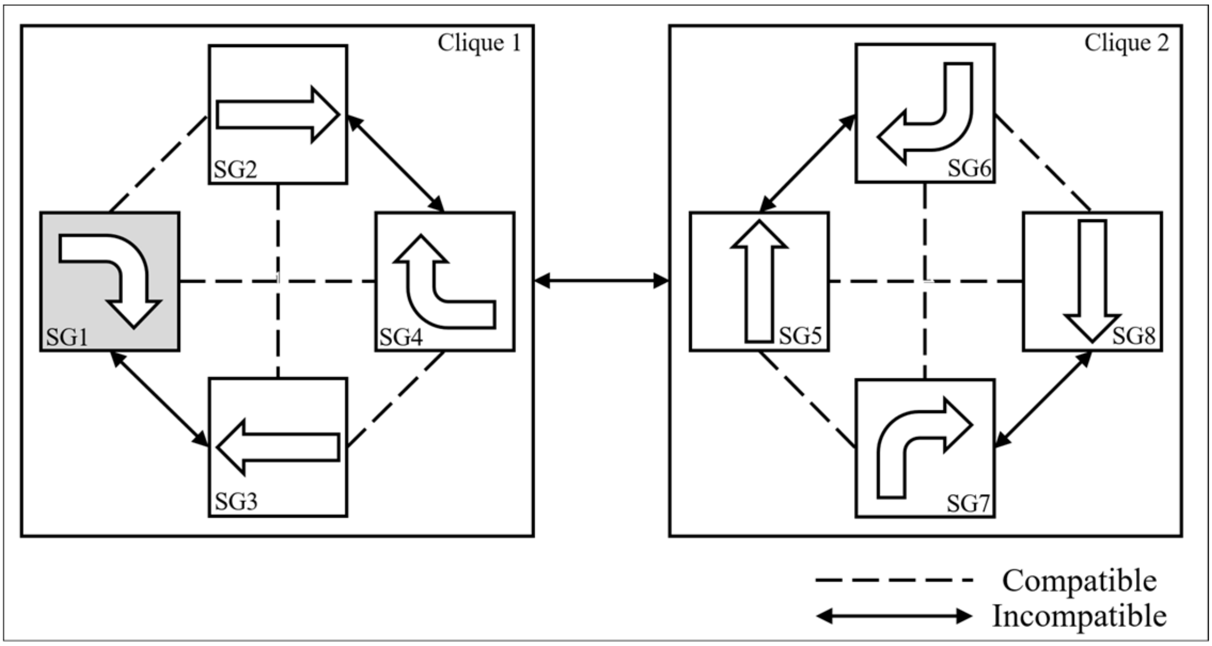

Figure 5. The defined hierarchical structure of signal groups

To operate the local reactive-control policy every few seconds, the flexible borders of the pairs and cliques should be mathematically formulated by the group-based variables and the successor functions that are used in the global proactive-optimization method. In this paper, an assumption is made that one signal group is fixed as the first group in each cycle. This assumption implies that the order of one pair in the first clique is unchanged for all cycles. The following equations are used to compare the pressure between incompatible signal groups according to the optimized cycle structure, and to either maintain the current signal group or to trigger the inter-green time for the opposite incompatible signal group. The preceding signal group is on the left-hand side, and the following signal group is on the right-hand side in each pair or clique. If the left-hand side comes later than the right-hand side, the current signal group is maintained. Otherwise, the inter-green time between the incompatible signal groups is activated using the following equation.

$$\sum_{k=1}^{K_1} s_k^n q_k^n(t) \geq \sum_{k=1}^{K_3} s_k^n q_k^n(t) \tag{29}$$

In Equation (29), signal group 1 continuously deals with the set of traffic flows in lane k. K_1, is served by signal group 1 if the pressure on signal group 1 is larger than the pressure on signal group 3. However, the inter-green time between signal groups 1 to 3 is activated if the pressure on signal group 3 is larger than that on group 1. The relationship between the pressure on groups 1 and 3 is expressed by a relatively simple inequality formula, because the start signal group in each cycle is fixed as signal group 1. The pressure on each signal group is illustrated by the estimated lane-based queue lengths in lane k at time t in the nth cycle multiplied by the estimated saturation flows in lane k in the nth cycle.

$$\left(1-\Lambda_{2,4}^n\right)\sum_{k=1}^{K_2} s_k^n q_k^n(t) + \Lambda_{2,4}^n \sum_{k=1}^{K_4} s_k^n q_k^n(t) \geq \left(1-\Lambda_{2,4}^n\right)\sum_{k=1}^{K_4} s_k^n q_k^n(t) + \Lambda_{2,4}^n \sum_{k=1}^{K_2} s_k^n q_k^n(t) \tag{30}$$

$$\left(1-\Lambda_{5,7}^{n}\right)\sum_{k=1}^{K_5} s_k^n q_k^n(t) + \Lambda_{5,7}^{n}\sum_{k=1}^{K_7} s_k^n q_k^n(t) \geq \left(1-\Lambda_{5,7}^{n}\right)\sum_{k=1}^{K_7} s_k^n q_k^n(t) + \Lambda_{5,7}^{n}\sum_{k=1}^{K_5} s_k^n q_k^n(t) \tag{31}$$

$$\left(1-\Lambda_{6,8}^{n}\right)\sum_{k=1}^{K_6} s_k^n q_k^n(t) + \Lambda_{6,8}^{n}\sum_{k=1}^{K_8} s_k^n q_k^n(t) \geq \left(1-\Lambda_{6,8}^{n}\right)\sum_{k=1}^{K_8} s_k^n q_k^n(t) + \Lambda_{6,8}^{n}\sum_{k=1}^{K_6} s_k^n q_k^n(t) \tag{32}$$

In Equation (30), the pressure on signal groups 2 and 4 is combined with the successor functions from groups 2 to 4 in the *n*th cycle, $\Lambda_{2,4}^{n}$, which allows the optimized cycle structure from the global optimization of the signal timings to be considered proactively. The successor function Λ_{kj}^{n} is set as 0 if the start of the green signal in lane *k* precedes that for lane *j*. Otherwise, the successor function is set as 1. Accordingly, $\Lambda_{2,4}^{n}$ is set as 0 if the start time of signal group 2 precedes the start time of signal group 4, and otherwise it is set as 1. The role of the cycle-varied successor function, as multiplied by the pressure on each signal group, is to assign the preceding and following signal groups to the left-hand side or the right-hand side, respectively, in each cycle. The current preceding signal group continuously serves its own traffic movements for each pair of groups, including signal groups 5 and 7 and 6 and 8, if the traffic conditions for these pairs satisfy the given inequalities in Equations (31) and (32). The above-mentioned three inequality equations are performed within the same clique, and the pressure on each of the inter-cliques is defined in the following equations.

$$\begin{aligned}&\left(1-\Lambda_{2,4}^{n}\right)\sum_{k=1}^{K_4} s_k^n q_k^n(t) + \Lambda_{2,4}^{n}\sum_{k=1}^{K_2} s_k^n q_k^n(t) + \sum_{k=1}^{K_3} s_k^n q_k^n(t) \geq \\ &\left(1-\Lambda_{5,7}^{n}\right)\sum_{k=1}^{K_5} s_k^n q_k^n(t) + \Lambda_{5,7}^{n}\sum_{k=1}^{K_7} s_k^n q_k^n(t) + \left(1-\Lambda_{6,8}^{n}\right)\sum_{k=1}^{K_6} s_k^n q_k^n(t) + \Lambda_{6,8}^{n}\sum_{k=1}^{K_8} s_k^n q_k^n(t)\end{aligned} \tag{33}$$

$$\begin{aligned}&\left(1-\Lambda_{5,7}^{n}\right)\sum_{k=1}^{K_7} s_k^n q_k^n(t) + \Lambda_{5,7}^{n}\sum_{k=1}^{K_5} s_k^n q_k^n(t) + \left(1-\Lambda_{6,8}^{n}\right)\sum_{k=1}^{K_8} s_k^n q_k^n(t) + \Lambda_{6,8}^{n}\sum_{k=1}^{K_6} s_k^n q_k^n(t) \geq \\ &\sum_{k=1}^{K_2} s_k^n q_k^n(t) + \sum_{k=1}^{K_1} s_k^n q_k^n(t)\end{aligned} \tag{34a}$$

$$\begin{aligned}&\left(1-\Lambda_{5,7}^{n}\right)\sum_{k=1}^{K_7} s_k^n q_k^n(t) + \Lambda_{5,7}^{n}\sum_{k=1}^{K_5} s_k^n q_k^n(t) + \left(1-\Lambda_{6,8}^{n}\right)\sum_{k=1}^{K_8} s_k^n q_k^n(t) + \Lambda_{6,8}^{n}\sum_{k=1}^{K_6} s_k^n q_k^n(t) \geq \\ &\sum_{k=1}^{K_4} s_k^n q_k^n(t) + \sum_{k=1}^{K_1} s_k^n q_k^n(t)\end{aligned} \tag{34b}$$

To compare the pressure between two cliques, the pressure incurred by each of their adjacent signal groups is defined with each successor function. In Equation (33), the signal groups adjacent to the first clique and the second clique consist of signal group 3 and the signal group that follows the signal group in pair 2. These adjacent signal groups are defined by the successor function between groups 2 and 4. If the pressure caused by the last groups in the first clique is larger than the pressure that occurs in the signal groups preceding the couple of pairs in the second clique, then the first clique continuously serves its own traffic flows. In Equations (34a) and (34b), the pressure on the last signal groups in each pair of the second clique is compared with the pressure on the possibly preceding signal groups in each pair of the first clique in the next cycle. This comparison determines whether the current green signal time is extended or terminated. The boundary between the end of the second clique and the start of the first clique is at the end of the cycle, and the optimum successor functions in the next cycle do not yet exist at that time. Therefore, the current last clique is continuously activated when the pressure induced by the current last signal group in the last clique is larger than the pressure induced by the possible simultaneous combinations of the first signal groups in the next cycle, namely signal groups 1 and 2, or signal groups 1 and 4.

The current signal groups are continuously activated if the current conditions of pressure satisfy the aforementioned inequality equations in each pair or clique. Otherwise, the current signal groups are terminated. The temporal borders of the signal groups in each pair or clique are determined flexibly to cope with transient fluctuations of traffic flows, and to achieve this based on the cycle structures derived from the global-optimization procedure. To prevent overly high sensitivity to a transient fluctuation in traffic flows in the local reactive-control policy, the activation of the local control policy should be restricted by a temporal permissible range, which is called a transition zone.

5.3 Transition zones for a group-based approach

The local reactive-control policy, as based on the group-based max-pressure theory, is largely dependent on the saturation flows and the queue lengths in each lane, as measured second-by-second. If the local control policy is allowed to be activated without temporal limitations on each pair or each clique, then the signal controller becomes too sensitive to consider the global tendencies in traffic conditions that arise beyond the micro-fluctuations in traffic flows. The concept of a transition zone in the group-based adaptive traffic-signal-control system is similar to the concept used in the stage-based approach.

A transition zone is the permitted time section in which the local reactive-control policy can determine whether to extend or to terminate the current signal groups. The zone is set as a pivot of several seconds before and after the end times of the corresponding signal groups in each pair (or between cliques), as established by the global proactive-optimization scheme. If the pressure on the current signal groups is greater than that on the next signal groups in a pair, or between the cliques in a previous time slot, then the signal controller extends the current signal groups until the end of the current time slot. Otherwise, the current signal groups finish at the effective end of the current time slot, and the next signal groups start after the duration of the inter-green time. The distinctive advantage that the group-based local reactive-signal control has over the stage-based approach is that a transition zone (expressed by group-based variables and temporal factors) restricts the permitted temporal area in which the current signal group is either continuously activated or immediately terminated in each pair (or inter-clique) of incompatible signal groups. The transition zone is therefore defined differently in each cycle. The concept of a transition zone is illustrated in the following figure.

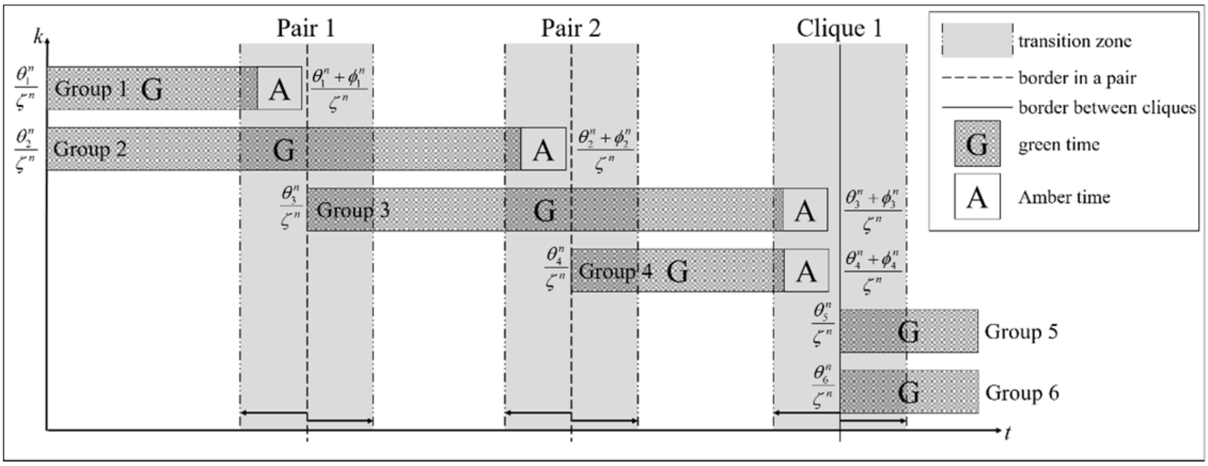

Figure 6. The defined hierarchical structure of signal groups

In Figure 6, each shaded area shows a transition zone for each pair or inter-clique, and the local reactive-control policy for each signal group is only activated within this temporal range. The start and duration of each green time, as optimized by the global-optimization method, is adjusted with respect to the transient fluctuations in traffic flows, which are served by the corresponding signal group and governed by the group-based max-pressure control policy in each transition zone. In Lee et al. (2017), the local reactive control policy is actualized using the practical implementation plan and the corresponding programs.

6. CONCLUSION

In this paper, a proactive global-optimization procedure is developed for group-based signal settings based on the traditional group-based optimization process proposed by Wong (1993). These new methods include procedures for estimating the lane-to-lane turning flows and lane-based queue-lengths, a model for predicting the control delay on the IQAs, and approximate formulas for the derivatives of the polygons on the IQAs. The key challenges are first to identify the most appropriate cycle-structure for minimizing the control delay on a cycle for the given predicted traffic conditions. A second challenge is to implement the developed sub-modules using an adaptive traffic-control logic based on a group-based proactive global-optimization method for the signal timings in each cycle. The contributions of the authors' previous studies (i.e., the proposed mathematical methods for analyzing lane-based turning proportions and lane-based queue lengths, the predictive models of lane-based control delay, and the approximate expressions of derivatives for lane-based control delay) all play significant roles in the development of the group-based global-optimization scheme. The successor function, the corresponding clearance-time matrix, and the initial guesses of the group-based variables for the nonlinear mathematical programs all serve to identify the greatest reserve capacity among the possible successor functions through the capacity maximization methods. These methods are applied under the given traffic conditions as they arise in each cycle. The control delay formulation developed in Lee and Wong (2017b) and the approximate expressions of the derivatives developed in Lee and Wong (2017a) are used as the objective functions and the gradients in the nonlinear mathematical programming method for delay minimization, based on $\boldsymbol{\Gamma}_0^n$ and $\hat{\Lambda}^n$ as defined in the capacity maximization procedure from the previous step. A discrete directional search method introduced in the IMSL Fortran Numerical Library is used to find solutions for nonlinear mathematical programs involving lane-based control delay and the approximate expressions of the derivatives induced from IQAs using the adjustment factors and group-based variables in part II: implementation (Lee et al. 2017).

In addition to presenting the global proactive-optimization method, this paper contributes to the literature by showing that the signal timings can be reactively adjusted with respect to transient fluctuations in the traffic flows. These adjustments include modifications of the signal sequence, cycle time, and duration of green time. These developments enable a global proactive-optimization scheme to be constructed cycle-by-cycle. The pairs of incompatible green phases and sequences of green time in each cycle structure, as expressed by the successor functions in the group-based approach, play significant roles in determining the local reactive-control policy. The local adjustment of green time is based on the max-pressure control policy developed by Varaiya (2013), which enables decentralized analysis, provides stability in real-time queue lengths, and has minimal requirements for pre-knowledge of the traffic flows. The concepts of the signal transition zone and the cycle-varied pair of incompatible signal groups are developed to facilitate a group-based local reactive-control policy that can be modified second-by-second. The local reactive-control policy for each signal group or each lane is activated for a transition zone based on comparisons of the real-time queue lengths and saturation flows for a pair of incompatible signal groups in a given cycle structure. The results are combined to provide an optimized system using the global proactive-optimization method.

In part II: Implementation (Lee et al. 2017), the two proposed methods are installed in a simulation platform along with the detailed practical mathematical and logical components of the implementation. For the implementation of the integrated adaptive traffic-signal-control logic, the global proactive-optimization scheme is actualized based on a rolling-horizon approach, the plan for the signal structures, the saturation flow rates, and a mathematical programming approach, which are illustrated in Lee et al. (2017). In addition, the local reactive-control policy is installed using the detailed implementation scheme and the appropriate computer programs. In Lee et al. (2017), the simulation platform and a case study are used to validate the performance of the proposed method in relation to the existing methods.

Acknowledgements

The work described in this paper was jointly supported by a Research Postgraduate Studentship and a grant from the University Research Committee of the University of Hong Kong (201411159005).

References

Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2009. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C: Emerging Technologies* 17, 163-174.

Akçelik, R., 1980. Time-dependent expressions for delay, stop rate and queue length at traffic signals. Australian Road Research Board, Melbourne, Australia.

Allsop, R.E., 1971. Delay-minimizing settings for fixed-time traffic signals at a single road junction. *IMA Journal of Applied Mathematics* 8, 164-185.

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

Allsop, R.E., 1992. Evolving application of mathematical optimisation in design and operation of individual signal-controlled road junctions. In *Mathematics in Transport and Planning and Control, 1-24*, Clarendon Press, Oxford, U.K.

Bang, K.L., Nilsson, L.E., 1976. Optimal control of isolated traffic signals. *Proceedings of the International Federation of Automatic Control/International Foundation for Information*

Process/International Federation of Oper Research Society International Symposium: Control in Transportation Systems, Columbus, Ohio.
Boillot, F., Midenet, S., Pierrelée, J.C., 2006. The real-time urban traffic control system CRONOS: Algorithm and experiments. *Transportation Research Part C: Emerging Technologies* 14, 18-38.
D'ans, G.C., Gazis, D.C., 1976. Optimal control of oversaturated store-and-forward transportation networks. *Transportation Science* 10, 1-19.
Dunne, M.C., Potts, R.B., 1964. Algorithm for traffic control. *Operations Research* 12, 870-881.
Farges, J.L., Henry, J.J., Tuffal, J., 1984. The PRODYN real-time traffic algorithm. In *4th IFAC Symposium on Control of Transportation Systems*, Baden Baden, Germany.
Gallivan, S, Heydecker, B.G., 1988. Optimising the control performance of traffic signals at a single junction. *Transportation Research Part B: Methodological* 22, 357-370.
Gartner, N.H., 1983. OPAC: A demand-responsive strategy for traffic signal control. *Transportation Research Record: Journal of the Transportation Research Board* 906, 65-72.
Gazis, D.C., 1964. Optimum control of a system of oversaturated intersections. *Operations Research* 12, 815-831.
Gazis D.C., Potts R.B., 1965. The oversaturated intersection. *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, OECD, Paris, 221-237.
Ge, Y.E, Zhou, X., 2012. An alternative definition of dynamic user optimum on signalised road networks. *Journal of Advanced Transportation* 46, 236-253.
Gordon, R.L., Tighe, W., Siemens ITS, 2005. *Traffic Control Systems Handbook*. US Department of Transportation, Federal Highway Administration, Office of Operations.
Grafton, R.B., Newell, G.F., 1967. Optimal policies for the control of an undersaturated intersection. In *Proceedings of the Third International Symposium on the the Theory of Traffic Flow*. New York.
Green, D.H., 1967. Control of oversaturated intersections. *Operational Research Quarterly* 18, 161-173.
Han, K., Gayah, V.V., 2015. Continuum signalized junction model for dynamic traffic networks: Offset, spillback, and multiple signal phases. *Transportation Research Part B: Methodological* 77, 213-239.
Han, K., Gayah, V.V., Piccoli, B., Friesz, T.L., Yao, T., 2014. On the continuum approximation of the on-and-off signal control on dynamic traffic networks. *Transportation Research Part B: Methodological* 61, 73-97.
Heydecker, B.G., 1992. Sequencing of traffic signals. In Griffiths, J.D. (Ed.), *Mathematics in Transport and Planning and Control*. Clarendon Press, Oxford, 57-67.
Heydecker, B.G., Dudgeon, I.W., 1987. Calculation of signal settings to minimise delay at a junction. In *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, MIT, July, Elsevier, New York, 159-178.
Hunt, P.B., Robertson, D.I., Bretherton, R.D., Winton, R.I., 1981. *SCOOT – A Traffic Responsive Method of Coordinating Signals*. LR1014, Transport and Road Research Laboratory, Crowthorne, UK.
Improta, G., Cantarella, G.E., 1984. Control system design for an individual signalized junction. *Transportation Research Part B: Methodological 18*, 147-167.
Lee, S., Wong, S.C., Li, Y.C., 2015a. Real-time estimation of lane-based queue lengths at isolated signalized junctions. *Transportation Research Part C: Emerging Technologies* 56, 1-17.
Lee, S., Wong, S.C., Pang, C.C.C., Choi, K., 2015b. Real-time estimation of lane-to-lane turning flows at isolated signalized junctions. *Intelligent Transportation Systems, IEEE Transactions on* 16, 1549-1558.
Lee, S., Wong, S.C., 2017a. Group-based approach to derivatives of delay models based on incremental queue accumulations for adaptive traffic control systems. *Transportmetrica B*, under review.
Lee, S., Wong, S.C., 2017b. Group-based approach to predictive delay model based on incremental queue accumulations for adaptive traffic control systems. *Transportation Research Part B: Methodological 98*, 1-20.

Lee, S., Wong, S.C., Varaiya, P., 2017. Group-based hierarchical adaptive traffic-signal control part II: Implementation. *Transportation Research Part B: Methodological*, under review.

Lee, S., Wong, S.C., Varaiya, P., 2016. Stage-based max pressure adaptive traffic control policy at isolated signalized junctions. In *Presented at the Transportation Research Board 93rd Annual Meeting*. Washington, D.C., United States.

Lighthill M.J., Whitham G.B., 1955. On kinematic waves. I: Flood movement in long rivers. II: A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, London, United Kingdom, A229, 281-345.

Li, L., Lv, Y., Wang, F., 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3(3), 247-254.

Lindley, D.V., 1952. The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge Univ. Press, 277-289.

Liu, H., Han, K., Gayah, V., Friesz, T. L., Yao, T, 2015. Data-driven linear decision rule approach for distributionally robust optimization of on-line signal control. *Transportation Research Part C: Emerging Technologies* 59, 260-277.

Lowrie, P., 1982. The Sydney coordinated adaptive traffic system-principles, methodology, algorithms. In *International Conference on Road Traffic Signalling*, London, United Kingdom.

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D., Mirchandani, P., 2003. ACS-Lite algorithmic architecture: Applying adaptive control system technology to closed-loop traffic signal control systems. *Transportation Research Record: Journal of the Transportation Research Board* 1856, 175-184.

Mirchandani, P.B., Head, L., 2001. A real-time traffic signal control system: Architecture, algorithms, and analysis. *Transportation Research Part C: Emerging Technologies* 9, 415-432.

Michalopoulos, P.G., Stephanopoulos, G., 1977a. Oversaturated signal systems with queue length constraints-I: Single intersection. *Transportation Research* 11, 413-421.

Michalopoulos, P.G., Stephanopoulos, G., 1977b. Oversaturated signal systems with queue length constraints-II: Systems of intersections. *Transportation Research* 11, 423-428.

Miller, A.J., 1963a. A computer control system for traffic networks. In *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, London, 200-220.

Miller, A.J., 1963b. Settings for fixed-cycle traffic signals. *Operational Research Quarterly* 14, 373-386.

Morris, R.W., Pak-Poy, P.G., 1967. Intersection control by vehicle actuated signals. *Traffic Engineering and Control* 9, 288-293.

Murat, Y. S., Gedizlioglu, E., 2005. A fuzzy logic multi-phased signal control model for isolated junctions. *Transportation Research Part C: Emerging Technologies* 13 (1), 19-36.

Richards, P.I., 1956. Shock waves on the highway. *Operations Research* 4, 42-51.

Rosdolsky, H.G., 1973. A method for adaptive traffic control. *Transportation Research* 7, 1-16.

Ross, D.W., Sandys, R.C., Schlaefli, J.L., 1971. A computer control scheme for critical-intersection control in an urban network. *Transportation Science* 5, 141-160.

Saaty, T.L., 1961. *Elements of Queueing Theory: With Application*s. McGraw-Hill, New York.

Silcock, J.P., 1997. Designing signal-controlled junctions for group-based operation. *Transportation Research Part A: Policy and Practice* 31, 157-173.

Smith, M.J., 1979a. The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological* 13, 295-304.

Smith, M.J., 1979b. Traffic control and route-choice; a simple example. *Transportation Research Part B: Methodological* 13, 289-294.

Smith, M.J., 1981. Properties of a traffic control policy which ensure the existence of a traffic equilibrium consistent with the policy. *Transportation Research Part B: Methodological* 15, 453-462.

Smith, M.J., 2011. Dynamics of route choice and signal control in capacitated networks. *Journal of Choice Modelling* 4, 30-51.

Smith, M.J., Ghali, M., 1990. The dynamics of traffic assignment and traffic control: A theoretical study. *Transportation Research Part B: Methodological* 24, 409-422.

Smith, M.J., Van Vuren, T., 1993. Traffic equilibrium with responsive traffic control. *Transportation Science* 27, 118-132.

Srinivasan, D., Choy, M. C., Cheu, R. L., 2006. Neural networks for real-time traffic signal control. *IEEE Transactions on Intelligent Transportation Systems* 7(3), 261-272.

Tully, I.M., 1976. Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. Ph.D. Thesis, University of Oxford.

Varaiya, P., 2013. Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies* 36, 177-195.

Webster, F.V., 1958. Traffic signal settings. *Road Research Technical Paper No. 39*. Road Research Laboratory.

Webster, F.V., Cobbe, B.M., 1966. Traffic signal. *Road Research Technical Paper No. 56*, HMSO, London

Wong, C.K., Wong, S.C., 2003a. 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., 2003b. Lane-based optimization of signal timings for isolated junctions. *Transportation Research Part B: Methodological* 37, 63-84.

Wong, C.K., Wong, S.C., Tong, C.O., 2006. A lane-based optimization method for the multi-period analysis of isolated signal-controlled junctions. *Transportmetrica* 2, 53-85.

Wong, S.C., 1993. Phase-based optimisation of signal timings for area traffic control. Ph.D. Thesis, University of London.

Wong, S.C., 1995. Derivatives of the performance index for the traffic model from TRANSYT. *Transportation Research Part B: Methodological* 29, 303-327.

Wong, S.C., 1996. Group-based optimisation of signal timings using the TRANSYT traffic model. *Transportation Research Part B: Methodological* 30, 217-244.

Wong, S.C., 1997. Group-based optimisation of signal timings using parallel computing. *Transportation Research Part C: Emerging Technologies* 5, 123-139.

Wong, S.C., Wong, W.T., Leung, C.M., Tong, C.O., 2002. Group-based optimization of a time-dependent TRANSYT traffic model for area traffic control. *Transportation Research Part B: Methodological* 36, 291-312.

Wong, S.C., Yang, H., 1997. Reserve capacity of a signal-controlled road network. *Transportation Research Part B: Methodological* 31, 397-402.

Wong, S.C., Yang, C., 1999. An iterative group-based signal optimization scheme for traffic equilibrium networks. *Journal of Advanced Transportation* 33, 201-217.