Refining Lane-Based Traffic Signal Settings to Satisfy Spatial Lane Length Requirements

Liu, Yanping; Wong, C. K.

Published in:
Journal of Advanced Transportation

Published: 17/07/2017

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1155/2017/8167530

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
Research Article

Refining Lane-Based Traffic Signal Settings to Satisfy Spatial Lane Length Requirements

Yanping Liu and C. K. Wong

Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong

Correspondence should be addressed to C. K. Wong; wongck@cityu.edu.hk

Received 26 January 2017; Revised 16 May 2017; Accepted 31 May 2017; Published 17 July 2017

Academic Editor: David F. Llorca

In conventional lane-based signal optimization models, lane markings guiding road users in making turns are optimized with traffic signal settings in a unified framework to maximize the overall intersection capacity or minimize the total delay. The spatial queue requirements of road lanes should be considered to avoid overdesigns of green durations. Point queue system adopted in the conventional lane-based framework causes overflow in practice. Based on the optimization results from the original lane-based design, a refinement is proposed to enhance the lane-based settings to ensure that spatial holding limits of the approaching traffic lanes are not exceeded. A solution heuristic is developed to modify the green start times, green durations, and cycle length by considering the vehicle queuing patterns and physical holding capacities along the approaching traffic lanes. To show the effectiveness of this traffic signal refinement, a case study of one of the busiest and most complicated intersections in Hong Kong is given for demonstration. A site survey was conducted to collect existing traffic demand patterns and existing traffic signal settings in peak periods. Results show that the proposed refinement method is effective to ensure that all vehicle queue lengths satisfy spatial lane capacity limits, including short lanes, for daily operation.

1. Introduction

In the past few decades, traffic signal settings at signal-controlled intersections have been optimized using different optimization frameworks, including the stage-based method [1–4], the group-based method [5–12], and the lane-based method [13–16]. Traffic signal settings are optimized by maximizing the overall intersection capacity or by minimizing the total intersection delay.

Aggregated lane-use patterns were first considered as design variables to optimize intersection performance [17]. In the latest lane-based optimization method, individual lane-marking arrows for left turn, straight-ahead, and right turn movements were defined as discrete binary variables in a mixed-integer linear programming optimization framework [13–16, 18, 19]. As a direct extension of the conventional group-based (phase-based) design approach, the lane-based approach is an offline design method, and peak hourly vehicular flow rates from traffic surveys are usually set as demand flow inputs to optimize a set of cooperative lane markings for operation. The use of surveyed demand flow rates as inputs is still reasonable because, once painted on the ground, lane markings are not expected to vary in different signal cycles. Other real-time or online methods may still be applied to fine-tune the traffic signal timings when all lane-marking arrows are optimized and given. The lane-based method has also been extended and applied for traffic signal optimization in signalized road networks by using the peak hourly flow rates as demand flow inputs. Platoon dispersion is modeled to realize the effects of coordinated traffic signal settings across upstream and downstream intersections [15]. Han and Gayah [20] even developed a continuum model for dynamic traffic networks that considers offset, spillback, and multiple signal phases. A more innovative idea of adopting a displaced left-turn (DLT) lane was reported by Sun et al. [21] to facilitate a continuous flow design to improve oversaturated bottleneck intersections. However, shifting and reserving a DLT lane may sometimes be impractical because of insufficient available road lanes in urban street networks.

D’Acierno et al. [22] studied two-way coordinated arterials...
and minimized the total delay by adopting a microsimulation model. Stage sequence optimization was reported by Memoli et al. [23] for network traffic signal settings. Other studies relating to coordinated signalized systems were performed by Gayah and Daganzo [24] for two-way arterials and by Zhou and Zhuang [25] for tandem intersections. Route choices influenced by traffic signal controls were studied by Liu and Smith [26].

In this study, we aim to refine the existing lane-based method for isolated intersections connected by short-approach lanes with limited spatial holding capacities. Liu and Chang [27] reported that adjacent lane groups could easily be blocked if lane markings and shared lane markings are not designed in a cooperative manner, especially for approach lanes with (short) flare lanes under oversaturated conditions. Lu and Yang [28] recently estimated the dynamic queue distribution in a signalized network via a probability-generating model in which queue formation and dissipation, platoon dispersion, queue merging and diverging, queue spillover, and downstream blockage were all modeled as stochastic events. In our case study, the intersection is one of the busiest and most complicated intersections located in Hong Kong Island (details can be found in the case study). Major legs are relatively long and minor legs are relatively short and traffic signals are operated without coordination. Connecting to the short lanes, there is another signalized intersection of two phases controlling two conflicting movements. Both of these conflicting movements will be discharged in turn to enter to the short lanes continuously. Uniform vehicle arrival pattern entering the short lanes is assumed. Referring to the street map, similar intersection settings could be found near the case study area. The proposed formulation is to deal with signal-controlled intersections with these kinds of geometric settings. In our proposed formulation, the lane-based approach is used initially to optimize the lane markings (single-arrow marking or shared lane markings, with two or three arrows on each approach lane) to effectively avoid potential internal blocking. However, the existing defect of the lane-based design method is that it adopts a point queue system that neglects the spatial requirements and limits along the approaching traffic road links (i.e., the maximum queue length should not be longer than the actual physical road length), which are critical for short-approach traffic lanes. “Short lane” means that the physical length of the lane is relatively short and can be occupied by only a few vehicles (or pcu). Signal-controlled intersections with short physical road lengths with limited holding capacities are common in urban areas. Ignoring the limited spatial requirements of such approaching traffic lanes may lead to unrealistic designs for traffic signal settings. Short lanes will overflow. Existing lane-based traffic signal optimization tends to design longer green times and longer cycle lengths to maximize the overall intersection capacity [13–16]. To overcome this weakness, we propose a new set of refinements to modify the traffic signal settings at isolated intersections to prevent overflow.

In this study, we compute vehicle queue lengths under various traffic and control conditions. Comparing the vehicle queue lengths and the spatial lane-holding capacities, overflowing and nonoverflowing approaching traffic lanes are identified and grouped. Excess green durations are extracted from the nonoverflowing lane group and used to control the overflowing lane group. Based on this new control logic, a set of refinement procedures to fine-tune lane-based-optimized traffic signal settings is proposed. A search heuristic is also developed to modify the green start times, green durations, and cycle length until the developed vehicle queues satisfy the spatial requirements in which the optimized lane-marking patterns are implemented. To evaluate the refinement results, a VISSIM simulation model is built to simulate the maximum queue lengths obtained from implementing the original lane-based optimized settings and the refined traffic signal settings. This paper is organized as follows. In Section 3, evaluations of traffic queue lengths and their effects on traffic signal settings are discussed. In Section 4, a mathematical formulation of the traffic signal refinement model and the corresponding solution algorithms are presented. Finally, a case study of a typical signal-controlled intersection in Hong Kong is analyzed, and new traffic signal settings are produced following the proposed refinements to eliminate overflowing vehicle queues.

2. Notations and Symbols

To formulate the proposed traffic signal refinement procedures, a signal-controlled intersection with $N_T$ traffic legs and $M_i$ approaching traffic lanes from each leg $i$ is considered. For each leg $i$, traffic lanes are numbered consecutively from 1 to $M_i$, starting from the curbside lane (near the pavement). Other symbols used are shown in Notations section.

3. Queue Lengths under Different Control Patterns

Traffic signal timings at an intersection operate and repeat to form a signal cycle with a cycle time $c$. Conflicting traffic movements are not allowed upon entering the common area of an intersection and receiving the right-of-way during the same green period. Their respective signal phases become “incompatible,” implying that their green durations must be well separated within a signal cycle. When a signal displays red, the arriving vehicles must not be discharged, and they are held up in the form of vehicle queues along the approaching traffic lanes. For approaching traffic lane $k$ from leg $i$, $\ell_{i,k}$ is defined as the initial queue length (just before a signal cycle starts). The vehicle queue length along approach lane $k$ from leg $i$ within a signal cycle $[0, c]$ is denoted by $L_{i,k}(t)$ in

$$L_{i,k}(t) = \ell_{i,k} + \int_0^1 r_{i,k}(\tau) d\tau - \int_0^1 s_{i,k}(\tau) d\tau,$$

where $r_{i,k}(t) = \sum_{j=1}^{N_T} q_{i,j,k}(t)$ is the total traffic demand that enters approach lane $k$ from leg $i$ at a particular time $t$, $q_{i,j,k}(t)$ represents the traffic flow from leg $i$ to leg $j$ on lane $k$ at time $t$, and $s_{i,k}(t)$ expresses the saturation flow (i.e., discharge rate) of lane $k$ from leg $i$ at time $t$. Depending on the width of the
approach lane, we use

\[
\text{saturation flow (pcu/hour)} = \frac{1940 + (\text{lane width} - 3.25) \times 100 - 42 \times (\text{uphill gradient})}{1 + 1.5(\text{lane turning proportion/turning radius})}
\]  

(2)

as specified in the Transportation Planning and Design Manual (for the nearside lane). The constant "1940" is replaced by "2080" if the traffic lane is a nonnearside lane [29]. In this study, we measure the lane width to estimate the lane saturation flow assuming a constant turning radius across different approach lanes and zero uphill gradient. The turning proportion depends on the optimized lane flows, which are assigned to different approach lanes in the optimization process.

Mathematically, \(\int_0^t r_{i,k}(\tau) d\tau\) in (1) represents the total flow arrivals within time period \([0, t]\), whereas \(\int_0^t s_{i,k}(\tau) d\tau\) calculates the total number of vehicles discharged within time period \([0, t]\). These total arrivals and total discharges within time period \([0, t]\) are represented by the two shaded areas in Figures 1 and 2, respectively. Figure 1 shows the typical shape of traffic arrivals, and the (shaded) area under the curve represents the cumulative arrivals in lane \(k\) from leg \(i\). Figure 2 plots a typical saturation flow (discharge rate), and the (shaded) area under the saturation flow curve represents the cumulative discharges in lane \(k\) from leg \(i\).

To model the vehicle queue development and to realize the spatial requirements along an approaching traffic lane under different traffic conditions, the curve in Figure 3 shows how the vehicle queue length varies within a signal cycle and how it can be deduced from the cumulative arrivals and cumulative discharges. For unsaturated traffic conditions, all traffic arrivals can be fully discharged during a signal cycle, assuming that the approaching traffic lanes are sufficiently long to hold all arriving vehicles without overflowing. The traffic arrival flow must reach an ideal traffic state (i.e., \(E^*\) in Figure 3) within the green duration (period) of a signal cycle. The green start time and green duration of lane-based signal settings are given by \(\theta_{i,k}\) and \(\phi_{i,k}\), respectively. The vehicle queue length, \(L_{i,k}(t)\), in signal cycle \([0, c]\) can be evaluated by

\[
L_{i,k}(t) = \begin{cases} 
\varepsilon_{i,k} + \int_0^t r_{i,k}(\tau) d\tau, & t \in [0, \theta_{i,k}], \\
\int_0^t r_{i,k}(\tau) d\tau - \int_{\theta_{i,k}}^{\theta_{i,k}+\phi^*_{i,k}} s_{i,k}(\tau) d\tau, & t \in (\theta_{i,k}, \theta_{i,k}+\phi^*_{i,k}], \\
0, & t \in (\theta_{i,k}+\phi^*_{i,k}, c].
\end{cases}
\]  

(3)

\(\forall i = 1, 2, \ldots, N_i, k = 1, 2, \ldots, M_i\), where the arrival rate \(r_{i,k}(t) = \sum_{j \neq i} q_{i,j}(t)\) represents the total number of arriving vehicles entering lane \(k\) from leg \(i\) to all other leg \(j\). In (3), \(\phi^*_{i,k}\) specifies the green duration required to reach an ideal traffic state \(E^*\) when all arrival vehicle queues are just dissolved, as given in Figure 3. This particular moment is further illustrated in Figures 5 and 6 under different traffic conditions.

Figure 3 shows graphically the general vehicle queue development along approach lane \(k\) from leg \(i\). At the end of the red signal, \(\theta_{i,k}\), a maximum queue length \(F_{i,k}\) should
have developed. When the green time starts, vehicles begin to discharge and the vehicle queue length should decrease. $E^*$ represents an ideal traffic state in which the cumulative traffic arrivals just meet the cumulative traffic discharges, inducing no residual queue at the end of the green time. $\theta_{i,k}$ is the green start time (equivalent to the end of the red time), and $\phi_{i,k}$ is the green duration. $\phi^*_{i,k}$ is the green duration required to ensure that an ideal traffic state $E^*$ exists to discharge the arrival vehicle queue. Mathematically, the duration $\phi^*_{i,k}$ could be determined by satisfying the equality constraint

$$\int_{0}^{\theta_{i,k} + \phi^*_{i,k}} r_{i,k}(\tau) d\tau = \int_{\theta_{i,k}}^{c} r_{i,k}(\tau) d\tau.$$  

In general, Figure 4 lists four possible traffic signal patterns that display green and red signals within a signal cycle. In case (a), the signal cycle pattern is the same as that presented in Figure 3, in which the red signal starts when the signal cycle starts, $\theta_{i,k} > 0$. The green signal ends at the end of signal cycle $c$, $\theta_{i,k} + \phi_{i,k} = c$. The maximum queue length would then be found at the end of the red signal (or the start of the green signal), as given in

$$F_{i,k} = \int_{0}^{\theta_{i,k}} r_{i,k}(\tau) d\tau. \quad (4a)$$

In case (b), the green signal starts at the beginning of the signal cycle, $\theta_{i,k} = 0$. The end of the red signal must then be found at the end of the signal cycle. The maximum queue length would thus be found at the end of the red signal or at the end of the signal cycle, as given in

$$F_{i,k} = \int_{\theta_{i,k} + \phi_{i,k}}^{c} r_{i,k}(\tau) d\tau = \int_{0}^{\phi^*_{i,k}} r_{i,k}(\tau) d\tau. \quad (4b)$$

In case (c), the green signal starts and ends within the signal cycle (i.e., $0 < \theta_{i,k} + \phi_{i,k} < c$). The maximum queue length would still be found at the end of the red signal (just before the green signal starts in the signal cycle), and the queue would form during the whole red signal period from the end of the green signal, $\theta_{i,k} + \phi_{i,k}$, until the end of the signal cycle, $c$, and from the start of the red signal (equivalent to the start of signal cycle) until the end of the red signal (equivalent to the start of the green signal, $\theta_{i,k}$). The following gives the maximum queue length accordingly:

$$F_{i,k} = \int_{\theta_{i,k} + \phi_{i,k}}^{c} r_{i,k}(\tau) d\tau + \int_{0}^{\theta_{i,k}} r_{i,k}(\tau) d\tau. \quad (4c)$$

Finally, in case (d), the start of the green time period is greater than the end of the green time period within a signal cycle, $\theta_{i,k} + \phi_{i,k} - c < \theta_{i,k}$. This inequality holds only when $c < \theta_{i,k} + \phi_{i,k}$, implying that the green signal starts later in one signal cycle; the green period then lasts until the end of the signal cycle.
cycle and further extends to the next signal cycle. The end of the green signal is thus found inside the next signal cycle. The maximum queue length could be evaluated by (4d), with the start of the red signal (equivalent to the end of the green signal, $\theta_{i,k} + \phi^*_i$) at $\theta_{i,k} + \phi^*_i - c$ and the end of the red signal (equivalent to the start of the green signal) at $\theta_{i,k}$:

$$F_{i,k} = \int_{\theta_{i,k} + \phi^*_i - c}^{\theta_{i,k} + \phi^*_i} r_{i,k}(\tau) \, d\tau. \quad (4d)$$

The goal of this study is to enhance the traffic signal settings optimized by the lane-based model to fulfill the spatial queue requirements. We must therefore estimate the physical holding capacities of the approach road lanes for further analysis. Let $l_{i,k}$ be the actual length of approach road lane $k$ from leg $i$ (in meters), and let $\zeta$ be the physical length of standard vehicle $\nu$ plus a gap length between front and rear bumpers of two consecutive vehicles occupying the road space (in meters/pccu). The lane-holding capacity can then be defined by $l_{i,k}/\zeta$, which represents the maximum number of standard vehicles that could be held up and occupy the approaching traffic lane (in pccu). Comparing the maximum queue length $F_{i,k}$, evaluated from (4a)–(4d) and the physical holding capacity $l_{i,k}/\zeta$, approaching traffic lane $k$ from leg $i$ can always be classified as one of the following two lane types: (1) a nonoverflowing traffic lane (i.e., $F_{i,k} < l_{i,k}/\zeta$) or (2) an overflowing traffic lane (i.e., $F_{i,k} > l_{i,k}/\zeta$). In case (1), the maximum queue length is lower than the actual holding capacity, as shown in Figure 5, and thus approach lane $k$ from leg $i$ is categorized into lane set $B_i$. Lane-based traffic signal settings for approaching traffic lanes in set $B_i$ will not induce overflow in practical operations. More importantly, an excess green duration, $\phi^*_i - \phi_i^*$, may exist. In case (2), the maximum queue length exceeds the actual physical holding capacity, as shown in Figure 6. Approach lane $k$ from leg $i$ is thus grouped into lane set $A_i$. The lane-based traffic signal settings for approaching traffic lanes in set $A_i$ would induce overflow. Because of the spatial requirements, the maximum queue lengths must be reduced for these overflowing traffic lanes by refining the traffic signal settings. There is a special case in which the maximum queue length equals the physical holding capacity (i.e., $F_{i,k} = l_{i,k}/\zeta$). The lane-based traffic signal timings under this condition do not induce overflow, and the green durations are just enough to clear the incoming traffic. If lane set $A_i$ is nonempty, some approaching traffic lanes overflow, and the lane-based traffic signal settings should be refined to satisfy the spatial queue requirements.

The two sets of vertical double solid lines in Figures 5 and 6 indicate the maximum physical holding capacity, $l_{i,k}/\zeta$, of approaching traffic lane $k$ from leg $i$. In Figure 5, the maximum queue length is smaller than the spatial lane-holding capacity. In Figure 6, the maximum queue length is larger than the spatial lane-holding capacity. Thus, $\phi^*_i$ is the green duration required to keep the maximum queue length in lane $k$ from leg $i$ from exceeding its physical holding capacity. $\phi^*_i$ (in Figure 5) is the required green duration to reach the ideal traffic state $E^*$. By using the maximum queue length evaluated by (4a)–(4d), the minimum green duration, $\phi^*_i$, can be estimated by

$$F_{i,k} = \int_{\theta_{i,k} + \phi^*_i}^{\theta_{i,k} + \phi^*_i} r_{i,k}(\tau) \, d\tau = \frac{l_{i,k}}{\zeta}. \quad (5a)$$

The minimum green duration, $\phi^*_i$, can be derived from

$$F_{i,k} = \int_{\theta_{i,k}}^{\theta_{i,k} + \phi^*_i} (s_{i,k}(\tau) - r_{i,k}(\tau)) \, d\tau. \quad (5b)$$

Both (5a) and (5b) can be numerically evaluated by applying a line search technique until the equality constraints hold. The feasible range for searching the time length $\phi^*_i$ should be set between zero (as the lower bound) and the optimized green duration obtained from the lane-based optimization model (as the upper bound).

For a given lane-based traffic signal setting, as shown in Figure 6, if the optimized green duration, $\phi_{i,k}$, from the lane-based method in approach lane $k$ from leg $i$ is shorter than the required green duration, $\phi^*_i$ (ensuring the maximum queue length would not exceed the spatial capacity), overflow is likely to occur. It is expected that such overflow could be prevented if a longer green duration is provided. In Figure 5, in contrast, approaching traffic lane $k$ from leg $i$ is not overflowing and there is excess green duration. It is considered that reducing the green duration could still maintain an ideal nonoverflowing traffic state $E^*$. It is found that the green duration could be reduced by no more than $(\phi_{i,k} - \max(\phi^*_i, h_{i,k}))$ time units.

4. Refinement Model for Lane-Based Optimization Signal Settings

The lane-based optimization method is the latest approach for optimizing lane markings and traffic signal settings to maximize the overall intersection capacity (i.e., to maximize the reserve capacity) or to minimize the total delay. One critical issue is to examine its effectiveness in satisfying the spatial queue requirements (not to induce overflow) for all approaching traffic lanes. In this section, based on the optimization results from the original lane-based design framework, a new refining optimization process is proposed to modify the lane-based traffic signal settings to satisfy the physical spatial limits for approaching traffic lanes. A search heuristic is developed to fine-tune the lane-based signal settings, including the green start times, green durations, and cycle times, by considering the vehicle queuing patterns and physical holding capacities along all approaching traffic lanes.

4.1. Problem Formulation. In urban areas, intersections are closely spaced and connected by short traffic lanes. For signal-controlled intersections in such high-density street networks, the actual capacity of the individual intersections is
restricted by the holding capacity of the approach lanes. Long cycle times and long red durations (or long green durations in conflicting phases) lead to long vehicle queues before discharge. Overflow may easily occur, and vehicle queues may block upstream intersections, inducing huge delays. Using a conventional point queue modeling system may result in overdesign of the green durations for practical operations. The capacity of a signalized intersection is mainly controlled by its geometric layout and the operation of the signal settings. Geometric layout refers to the available number of traffic lanes and individual lane usage in the form of lane-marking patterns. Once fixed, these parameters cannot be changed frequently because of safety concerns. Daily traffic fluctuations and variations can then be well controlled and managed by fine-tuning the traffic signal settings. Practically, it is comparatively easy to refine (shorten or extend) green durations and revise the cycle times obtained from the proposed refinement method, which should be compatible with the optimized lane-marking patterns.

**Optimized Results from the Lane-Based Model as the Basic Inputs**

**Cycle length of lane-based signal optimization:** $c$

**Green start times of turning movements from lane-based optimization model:** $\Theta = \{\theta_{i,j}; i, j = 1, 2, \ldots, N_f \ (i \neq j)\}$

**Green durations of turning movements from lane-based optimization model:** $\Phi = \{\phi_{i,j}; i, j = 1, 2, \ldots, N_f \ (j \neq i)\}$

**Assigned traffic lane flows:** $q = \{q_{i,j,k}; i, j = 1, 2, \ldots, N_f \ (i \neq j), k = 1, 2, \ldots, M_j\}$

**Saturation flows of traffic lanes:** $s = \{s_{i,k}; i = 1, 2, \ldots, N_f, k = 1, 2, \ldots, M_j\}$

**Lane-marking patterns:** $\delta = \{\delta_{i,j,k}; i, j = 1, 2, \ldots, N_f \ (i \neq j), k = 1, 2, \ldots, M_j\}$

**Successor function from lane-based optimization model:** $\Pi = \{\Pi_{i,j,m,n}; \forall (i, j), \forall (m, n)\}$

---

**Input Data**

- Actual physical road length of approaching traffic lanes: $l = \{l_{i,k}; i = 1, 2, \ldots, N_f, k = 1, 2, \ldots, M_j\}$
- Weighting factor of the randomness of vehicle arrivals: $\eta = \{\eta_i; i = 1, 2, \ldots, N_f\}$
- Weighting factor of the importance of physical road space: $\Omega = \{\Omega_i; i = 1, 2, \ldots, N_f\}$
- Required green durations for nonoverflowing lanes: $\phi^* = \{\phi_{i,k}^*; i = 1, 2, \ldots, N_f, k \in A_i\}$
- Required green durations of nonoverflowing lanes: $\phi^* = \{\phi_{i,k}^*; i = 1, 2, \ldots, N_f, k \in B_i\}$
- Intergreen time matrix between turning movements: $\Gamma = \{\gamma_{i,j,k}; u = (i, j, k), \nu = (m, n, o)\}$
- Minimum green durations of turning movements: $G = \{g_{i,j}; i, j = 1, 2, \ldots, N_f \ (j \neq i)\}$

---

**Control Variables**

- Cycle length: $c$
- Green start times on approaching traffic lanes from different legs: $\Theta = \{\theta_{i,k}; i = 1, 2, \ldots, N_f, k = 1, 2, \ldots, M_j\}$
- Green durations on approaching traffic lanes from different legs: $\Phi = \{\phi_{i,k}; i = 1, 2, \ldots, N_f, k = 1, 2, \ldots, M_j\}$

**Objective Function.** In the present study, we aim to examine and verify the lane-based traffic signal optimization model results, including the lane markings and traffic signal settings, to satisfy spatial queue requirements on existing road lanes and to modify them if necessary to ensure that the refined traffic signal settings are capable of managing the incoming traffic without leading to overflow. When operating the refined traffic signal settings, it is expected that the maximum queue lengths due to vehicle arrivals should not exceed the actual physical length of the approach road lanes.

To form an effective mathematical objective function to prevent overflow, we first defined a parameter $E_{i,k}$ to depend on $\eta_i E_{i,k} = (l_{i,k}/\xi_v)/\Omega_i$ in (6), where $\eta_i$ and $\Omega_i$ are user-defined inputs for weighting the randomness of traffic arrivals (which may lead to increased maximum queue lengths) and weighting the importance of traffic legs in avoiding traffic overflow, respectively. $E_{i,k}$ and $l_{i,k}/\xi_v$ are the maximum queue length and lane-holding capacity of lane $k$ from leg $i$, respectively. $E_{i,k}$ then denotes numerically the “space shortage” to hold up all incoming vehicles without overflowing:

$$E_{i,k} = \eta_i F_{i,k} - \frac{l_{i,k}/\xi_v}{\Omega_i}.$$ (6)

Based on the numerical value of $E_{i,k}$ for an individual approaching traffic lane, if $E_{i,k} > 0$, approach lane $k$ from leg $i$ will be classified as “overflowing” and lane $k$ will be collected in the set $A_i$. If $E_{i,k} < 0$, then lane $k$ from leg $i$ will be classified as an approaching traffic lane with a spatial buffer and lane $k$ will be collected in the set $B_i$. Clearly, for every lane $k \in A_i$, $E_{i,k}$ can be regarded as the overflow queue length (to be minimized). For each approaching traffic lane $k \in B_i$, negative $E_{i,k}$ indicates that excess green durations may exist on lane $k$ from leg $i$. These excess green durations can be extracted to control the overflowing road lanes in set $A_i$.

With the parameter $E_{i,k}$ introduced above, the objective now is to reduce the overflow by minimizing $E_{i,k}$ whenever its numerical values are positive (overflow occurs on lane $k$ from leg $i$ because the actual lane length is shorter than the maximum queue length). To define the objective function, it is thus reasonable to include the overflowing vehicle queues from all overflowing approach lanes. Mathematically, the following equation is adopted as the objective function for refining the lane-based traffic signal optimization results:

$$J (\Theta, \Phi, c) = \sum_{i=1}^{N_f} \sum_{k \in A_i} E_{i,k},$$ (7)
where $A_j$ is the set of overflowing approach lanes from leg $i$ in which the physical road length is shorter than the maximum queue length $F_{i,k}$.

In the original lane-based model, vehicle queues are not considered explicitly and a point queue modeling system is applied. Although this maximizes the overall intersection capacity (throughputs), the optimized cycle length is always binding at its maximum allowable limit (usually 120 s in Hong Kong). If there are short lanes with very limited capacities (lane lengths) to hold incoming vehicles, overflow and spillback of vehicle queues would easily occur. This is a defect of the existing lane-based design approach. In the proposed design framework, cycle length (time) is also considered as a control variable to be adjusted (reduced) to allow faster changes of rights-of-way among conflicting turning movements. It is expected that incoming traffic could be discharged in time with better use of all green durations.

**Identifying Overflowing Approaching Traffic Lanes Depending on Spatial Holding Capacity.** The maximum queue length along an approaching traffic lane is determined by its incoming traffic demands and the traffic signal settings, including the green start times, green durations, and cycle times. These could be varied to influence the maximum vehicle queue length along approach road lanes. Actual lane lengths are fixed parameters and cannot be changed during daily operations. Overflow may depend heavily on the traffic signal settings. In the proposed refinement process, it is necessary to identify which approaching traffic lanes are overflowing or not overflowing.

1. **Approach Road Lanes Exceeding Spatial Capacity.** Operating improper traffic signal timings may cause the maximum queue length of an approach lane to exceed its spatial holding capacity, as described in Figure 6, resulting in overflow. The objective function is used to minimize the overflow on all approaching traffic lanes. Mathematically, the maximum queue lengths on overflowing approaching traffic lanes should be reduced. Overflowing approaching traffic lanes can be identified using

$$F_{i,k} > \frac{l_{i,k}}{c}, \quad (8)$$

for $i = 1, 2, \ldots, N_T, k \in A_j$. $F_{i,k}$ represents the maximum queue length, and $l_{i,k}/c$ is the spatial holding capacity of lane $k$ from leg $i$. Equation (8) identifies the overflowing approaching traffic lanes, and the respective $E_{i,k}$ values are positive. Thus, the weighted sum of all $F_{i,k}$ values will form the objective function for the traffic signal refinement model.

2. **Approach Road Lanes without Exceeding Spatial Capacity.** If a spatial lane-holding capacity is large enough to hold all incoming traffic vehicles without overflowing, such an approach road lane is regarded as a nonoverflowing lane, given in Figure 5. In this case, excess green duration may exist that could be extracted to control the overflowing approach lanes. With sufficient spatial holding capacities and green durations to manage the maximum vehicle queues, the assigned green durations on nonoverflowing approaching traffic lanes could be reduced at the expense of increasing the residue queue lengths. There is no harm in doing so if the maximum queue lengths so developed are shorter than the spatial lane-holding capacities. The following equation was developed to identify all these nonoverflowing approaching traffic lanes forming a part of the solution region:

$$F_{i,k} < \frac{l_{i,k}}{c}, \quad (9)$$

for $i = 1, 2, \ldots, N_T, k \in B_i$.

**Lane-Based Traffic Signal Settings for Refinement.** For operating traffic signal settings optimized from the lane-based method, overflow may occur because the signal timings may not be able to manage the incoming vehicles effectively; that is, they may not be able to maintain the maximum queue length below the lane-holding capacity. To maximize the intersection capacity, the lane-based traffic signal optimization could assign excess green time to nonoverflowing approaching traffic lanes. Also, the respective long red durations assigned to the incompatible phases may lead to long vehicle queues. If the maximum queue lengths exceed the lane-holding capacities, overflow occurs. If the spatial queue requirements along short-approach road lanes are ignored, the design of traffic signal settings becomes unrealistic and unreliable. To eliminate or reduce the degree of overflow, the green durations assigned to nonoverflowing and overflowing approaching traffic lanes should be refined. The proposed refinement procedures for the traffic signal settings are enhanced by restructuring the lane-based optimized signal timings with consideration of spatial lane-holding capacities, safety requirements for practical implementation, and the ideal state of vehicle queue development for all approach road lanes.

1. **Overlapping Approaching Traffic Lanes under Lane-Based Traffic Signal Settings.** Considering the spatial requirements of approach road lanes, it is important to examine the ability of the optimized signal settings from the lane-based optimization framework to control all incoming traffic so that the maximum vehicle queue lengths are all less than the respective lane-holding capacities. If the cycle time is fixed at $c$, the green duration should be longer than a threshold value $\phi_{i,k}^*$ (i.e., the maximum queue length would at most be equal to the spatial holding capacity). To confirm that an approach lane $k$ from leg $i$ is overflowing, the green duration optimized by the lane-based method should be less than this minimum requirement, as given in

$$\phi_{i,k} < \phi_{i,k}^*, \quad (10)$$

for $i = 1, 2, \ldots, N_T, k \in A_j$. $\phi_{i,k}$ represents the green duration used to control approach lane $k$ from leg $i$ optimized by the lane-based method. $\phi_{i,k}^*$ is the required green duration to avoid overflow that is determined numerically by solving (5a) using a line search technique.

2. **Nonoverlapping Approach Lanes under Lane-Based Traffic Signal Settings.** When operating lane-based optimized traffic
signal settings, vehicle queues develop. If the spatial lane-holding capacity is large enough to hold the incoming traffic that accumulates during red signals, overflow can be avoided. Numerically, this condition is verified if the maximum queue length is less than the spatial lane-holding capacity. However, in the proposed refinement framework, it is considered that the lane-based settings must also be refined if the assigned green duration is longer than that needed to dissolve the waiting vehicle queue. When the cycle length \( \epsilon \) is fixed, it is possible to reassign the excess green durations by shifting them from nonoverflowing traffic lanes to overflowing traffic lanes. This can effectively alleviate the degree of overflow along those overflowing approach road lanes. The feasible amount of green duration to be reduced and shifted depends on the new minimum green duration, \( \phi_{i,k}^* \) (i.e., the green duration for the traffic on lane \( k \) from leg \( i \) to reach an ideal traffic state) and on the minimum green duration for safety concerns. Thus, the lane-based traffic signal settings for the nonoverflowing approach road lanes should satisfy the constraint given in

\[
\phi_{i,k} > \max \left\{ \phi_{i,k}^*, h_{i,k} \right\}, \quad (11)
\]

for \( i = 1, 2, \ldots, N_T, k \in B_i \). \( \phi_{i,k}^* \) in (11) is the green duration required for the traffic on approach lane \( k \) from leg \( i \) to reach an ideal traffic state, which is numerically evaluated by solving (5b). \( h_{i,k} \) is the minimum green duration for the traffic movement(s) on approach lane \( k \) from leg \( i \) for safety operations to prevent frequent stop-and-go motion. Equation (11) is a governing constraint to ensure that the refined green duration is able to discharge the maximum traffic queue after shifting the excess green duration to overflowing approach lanes.

(3) Size of Time Step for Refining the Traffic Signal Timings. It has been revealed that unbalanced green durations allocated by the lane-based optimization method may lead to unnecessary overflow. Excess green durations may exist, which could be extracted from nonoverflowing approach lanes and added to the green durations assigned to overflowing approach lanes. During this refinement process, a proper time step size, \( \Delta t \), should be used to extract the green durations to balance the green durations of the overflowing and nonoverflowing approach traffic lanes in the heuristic solution process. From (10), the upper bound for choosing \( \Delta t \) should be bound by

\[
(\phi_{i,k}^* - \phi_{i,k}).
\]

Any increase of the green duration may reduce the degree of overflow of an overflowing approach lane. Time step sizes that satisfy (12) are all permitted. The finer the \( \Delta t \) value, the better the solution resolution:

\[
0 < \Delta t \leq (\phi_{i,k}^* - \phi_{i,k}), \quad (12)
\]

for \( i = 1, 2, \ldots, N_T, k \in A_i \).

For nonoverflowing approach traffic lanes, their spatial lane capacities and the lane-based optimized traffic signal settings enable them to hold and discharge their maximum vehicle queues. For nonoverflowing approaching traffic lane \( k \) from leg \( i \), the refined green durations should be good enough to control the incoming traffic without overflow and without violating the basic specifications for safety. Thus, the respective feasible \( \Delta t \) value for refining the green durations should be within the range specified by

\[
0 < \Delta t \leq (\phi_{i,k} - \max \{\phi_{i,k}^*, h_{i,k}\}), \quad (13)
\]

for \( i = 1, 2, \ldots, N_T, k \in B_i \).

Compatible Traffic Signal Settings with the Lane-Based Design Requirements. With the vehicle queue lengths along the approach road lanes computed from (3) and (4a)–(4d), whether an approach road lane is overflowing or not can be detected by comparing its maximum queue length and its spatial lane-holding capacity. Short lanes are usually the problematic source of overflowing because they have very tight spatial capacities to hold incoming vehicle queues. To formulate the proposed traffic signal refinement procedures, governing constraint sets for the traffic signal settings that are compatible with the lane-based design requirements are given as follows.

(1) Feasible Cycle Length. For an intersection under different traffic demand conditions, different signal cycle lengths for practical operations are required. If all green durations are set at their minimum allowable limits (usually 5 s for the display green times), then a minimum cycle time, \( c_{min} \), should be used. Based on the lane-based optimization results, an optimized cycle length \( c \) should always be available, which would be regarded as the maximum bound for the refined cycle length \( c \). The feasible range of \( c \) is governed by

\[
c_{min} \leq c \leq c. \quad (14)
\]

We set the cycle length optimized by the lane-based model as the maximum bound because the point queue system adopted in the lane-based method always overdesigns the cycle length. In the proposed formulation, we refine the traffic signal settings until the maximum queue lengths are all below the spatial lane capacities for all approach road lanes.

(2) Constraints for the Start and Duration of Green Times. Traffic signal settings are operated and repeated in cycles. Green signals could be started arbitrarily along the time axis. The duration of green times should be longer than a required minimum duration for safety operations. In signal cycles, the start and duration of green times are bounded by the length of the signal cycle. Constraints for the start and duration of green times for approach lane \( k \) from leg \( i \) are given, respectively, by the following:

\[
0 \leq \theta_{i,k} \leq c, \quad (15a)
\]

\[
h_{i,k} \leq \phi_{i,k} < c, \quad (15b)
\]

for \( i = 1, 2, \ldots, N_T, k = 1, 2, \ldots, M_i \).

(3) Traffic Signal Settings on Approaching Traffic Lanes. In practical operations, some turning movements from different legs cannot receive the right-of-way simultaneously because they involve incompatible movement turns in which their
destination legs are identical. Turning movements from the same leg are always compatible without clashing in the common intersection area. Shared-lane markings that allow two or more turning movements at the same time along an approaching traffic lane could be designed. Once shared-lane designs are optimized from the lane-based method, all involved turning movements should be controlled by the same traffic signal phase such that the start and duration of green times should be identical. For approach lane \( k \) from leg \( i \) permitting a movement turn to destination \( j \), the following constraint sets are required:

\[
-100 \left(1 - \delta_{i,ijk}\right) \leq \Theta_{ij} - \theta_{ik} \leq -100 \left(1 - \delta_{i,ijk}\right) 100,
\]

\[
-100 \left(1 - \delta_{i,ijk}\right) \leq \Phi_{ij} - \phi_{ik} \leq -100 \left(1 - \delta_{i,ijk}\right) 100, \quad (16)
\]

\[
-100 \left(1 - \delta_{i,ijk}\right) \leq g_{ij} - h_{ik} \leq -100 \left(1 - \delta_{i,ijk}\right) 100,
\]

for \( i, j = 1, 2, \ldots, N_T \) \((j \neq i)\), \( k = 1, 2, \ldots, M \). \( \delta_{i,ijk} \) is the lane-marking output from the lane-based optimization framework. Numerically, \( \delta_{i,ijk} = 1 \) means that a traffic movement turn from leg \( i \) to leg \( j \) is permitted on approach lane \( k \). Equation (16) will then force \( \Theta_{ijk} = \Theta_{ij} \), \( \Phi_{ijk} = \Phi_{ij} \), and \( h_{ik} = g_{ij} \) to ensure identical signal settings on all approaching traffic lanes involving the same turning movements to destination leg \( j \). If a turning movement from leg \( i \) to leg \( j \) is not permitted on approach lane \( k \), then \( \delta_{i,ijk} = 0 \), implying that the binding effects of the constraints given in (16) are relaxed and the traffic signal settings could all be different (i.e., controlled by other constraint sets).

(4) Clearance Time of Movements. Two traffic turning movements are incompatible if they cannot receive the right-of-way at the same time, and the set of all incompatible movements is denoted as \( \Psi \). For any pair of incompatible movements, \( u = (i, j, k) \) and \( v = (m, n, o) \). If lane markings on approaching traffic lanes exist, then, for safety concerns, an intergreen (clearance) time is given to separate their rights-of-way. For approaching traffic lane \( k \) from leg \( i \) and another approach lane \( o \) from leg \( m \), if their turning movements to destination legs \( j \) and \( n \), respectively, are incompatible, then the clearance time constraint can be set as follows:

\[
\Theta_{ij} + \Phi_{ij} + y_{uv} \leq \Theta_{mn} + \Pi_{i,j,m,n}\]

\[
+ (2 - \delta_{i,ijk} - \delta_{m,n,o}) \times 100, \quad (17)
\]

\[
\forall (u, v) \in \Psi,
\]

where \( y_{uv} \) is a user-specified minimum intergreen time to separate turning movements \( u \) and \( v \) and \( \Pi_{i,j,m,n} \) is the successor function [9] regulating the order of signal displays for the two incompatible signal groups \((i, j) \) and \((m, n)\), where \( \Pi_{i,j,m,n} = 0 \) if the green start of signal group \((m, n)\) follows that of signal group \((i, j)\), and vice versa when \( \Pi_{i,j,m,n} = 1 \).

The proposed traffic signal refinement problem can now be formulated by minimizing the objective function in (7) subject to the constraint sets in (8)–(17).

4.2. Solution Heuristics. In the present formulation, the spatial queue requirements of lane-holding capacities are considered. The details of vehicle queue developments in a signal cycle and maximum queue lengths along approach road lanes are evaluated using (3) and (4a)–(4d), and different lane conditions can be modeled. By comparing the maximum queue lengths and the physical lane-holding capacities, all approaching traffic lanes can be classified into one of two groups: (i) overflowing or (ii) not overflowing, as discussed in Section 3. Mathematically, overflowing approach lane \( k \) is grouped in set \( A_k \), and the set of all overflowing approach lanes is defined as \( A \). If approaching traffic lane \( k \) is not overflowing, it is grouped in set \( B_k \) and the set of all nonoverflowing approaching traffic lanes is denoted as \( B \). The green durations are considered insufficient for the overflowing approaching traffic lanes in set \( A \), as illustrated in Figure 6. The weighted sum of \( E_{ijk} \) for all approach lanes in set \( A \) constitutes the objective function for the refinement process. Green durations are, however, considered to be more than sufficient for approach lanes without overflow in set \( B \), as presented in Figure 5. Equation (9) of all approach lanes in set \( B \) provides the necessary constraints in the design framework.

In the proposed solution heuristics for each fixed cycle time \( c \), the green durations are extracted from the nonoverflowing lane set \( B \) and added to the green durations assigned to overflowing lane set \( A \) until all excess green durations are transferred.

When implementing the lane-based traffic signal optimization results, no overflow implies that the traffic signal settings optimized by the lane-based framework satisfy the spatial queue requirements for all approach road lanes and should satisfy the requirements in (8)–(17) mathematically. However, if overflow occurs, the lane-based settings no longer fulfill the spatial requirements for approaching traffic lanes. In general, short approaching traffic lanes with limited lane-holding capacities may lack sufficient road space to hold the maximum number of vehicles. To reduce the degree of overflow, the green durations could be shifted from nonoverflowing approaching traffic lanes to overflowing approach traffic lanes, and thus the overall performance and reliability of the intersection could be improved. If the cycle time \( \zeta \) optimized by the lane-based framework is fixed, the green durations could be refined until the objective function value is minimized. Reducing the objective function value to zero may also be possible. The cycle time should be considered a decision variable in the refinement process. To find practical traffic signal settings, a search heuristic is proposed to solve this traffic signal refinement process. The refining process for the lane-based optimization results is summarized by the flow chart in Figure 7.

Preparations. Assume that traffic signal settings \((\Theta, \Phi)\) are initialized by taking the lane-based model outputs as inputs to start the proposed refinement process. Based on the users’ input demand flow patterns and the optimized traffic signal settings, lane-marking patterns and assigned lane flows are obtained from the lane-based optimization model. We first check whether all approaching traffic lanes are not overflowing. If this is the case (i.e., set \( A \) is empty), then the refinement process terminates, taking the lane-based optimized settings for implementation. If some approach lanes are found to be overflowing (i.e., set \( A \) is not empty),
Results of lane-based optimization model $q, s, \delta, \Pi, \zeta, \Theta, \Phi$ as inputs for refinement.

Preparing data inputs: $l, \phi^a, \phi^b, \eta, \Omega, \Gamma, G$; initialization: set $c = \zeta, (\theta, \varphi) = (\Theta, \Phi)$.

With the signal settings $(\theta, \varphi)$, evaluate $F_{i,k}$ and $E_{i,k}$ by equations (4a)–(4d) and equation (6).

Group approaching traffic lanes into sets $A$ and $B$ by equation (8) and equation (9).

Evaluate the objective value $J(\theta, \varphi, c)$ in equation (7) and set $(\theta^*, \varphi^*) = (\theta, \varphi)$; $J^*$ is the respective objective function value.

Accept the traffic signal settings $(\theta^*, \varphi^*)$ as the solution. Refinement process terminates.

Yes

A empty ?

No

B empty ?

Yes

Shorten the cycle time by reducing it by one time unit $\Delta c: c - \Delta c$ as the new maximum allowable limit to repeat the solution process.

No

From the overflowing lane set $A$, identify the worst lane $(i^*, k^*)$ with the largest $E_{i^*,k^*}$ value; from the nonoverflowing lane set $B$, search an incompatible lane $(i^#, k^#)$ with the shortest excess green time; record $(\theta_{i^*,k^*}, \phi_{i^*,k^*}), (\theta_{i^#,k^#}, \phi_{i^#,k^#})$ and produce a suitable time step $\Delta t$ from equations (12)–(13).

Refine signal settings on incompatible lanes $(i^*, k^*)$ and $(i^#, k^#)$ by updating $\phi_{i^*,k^*}$ by $\phi_{i^*,k^*} + \Delta t$ and $\phi_{i^#,k^#}$ by $\phi_{i^#,k^#} - \Delta t$, revise traffic signal settings for other traffic lanes by equation (10)–(17), and set the revised traffic signal settings as a new solution $(\theta^*, \varphi^*)$ and $J^*$ as the respective objective function value.

Based on the new traffic signal settings $(\theta^*, \varphi^*)$, update the lane sets $A$ and $B$ by reevaluating the values of $F_{i,k}$ and $E_{i,k}$ by equations (4a)–(4d) and equation (6) and group approach traffic lanes using equations (8)–(9) again.

B empty ?

Yes

No

A empty ?

Yes

Terminate the traffic signal refinement procedures and output the refined traffic signal settings.

No

A empty ?

Figure 7: Proposed search heuristics for refining the traffic signal settings.
Reassignment of Green Durations. The reassignment process for the green duration involves extracting the excess green durations from nonoverflowing approaching traffic lanes (i.e., set B) and adding those times to the green durations of the approaching traffic lanes (i.e., set A). During this process, extreme cases are selected first by picking up the most severely overflowing approach lane (i.e., the largest value of $E_{i,k}$ in lane set A). An overflowing approach lane $(i^*, k^*)$ from set A and a nonoverflowing approach lane $(i^*, k^*)$ from set B are identified. A suitable $\Delta t$ for adjusting the green durations between the approach lanes $(i^*, k^*)$ and $(i^*, k^*)$ is selected by satisfying (12)–(13). The $\Delta t$ value should be kept as small as possible to ensure that the excess green durations can be redistributed effectively. The refinement process is composed of two parts: (1) the green duration is reassigned from approaching traffic lane $(i^*, k^*)$ to approaching traffic lane $(i^*, k^*)$ by updating $\phi_{r,k^*}$ to $\phi_{r,k^*} + \Delta t$ and $\phi_{r,k^*}$ by $\phi_{r,k^*} - \Delta t$ and (2) the traffic signal settings for other approach traffic lanes should be subsequently revised to satisfy the design constraints in (10)–(17). As a result, the refined trial settings $(\theta, \varphi)$ are found. It is expected that the degree of overflow would thus be reduced for the overflowing approach traffic lane.

Update Rules. The maximum queue lengths along the approaching traffic lanes could vary and change instantly after the refinement of the traffic signal settings. Upon adjusting one time step $\Delta t$ over the green durations $(\theta^*, \varphi^*)$, approach lane sets A and B should be updated for further refinement. The objective function value $J^*$ is revised. If the termination condition cannot be met, the refinement process continues by updating $F_{i,k}$ and $E_{i,k}$ numerically by (4a)–(4d) and (6). Approach lane sets A and B are updated as well.

Termination Condition. In the proposed refinement process, we aim to eliminate overflow or reduce the degree of overflow of approaching traffic lanes. Whenever approach lane set A is found to be empty (i.e., no overflow), the whole traffic signal refinement and the heuristic search process are terminated. If neither approach lane set A nor B is empty, the refinement of the green durations continues by repeating the reassignment of green durations as described above. A case could occur in which approach lane set B is empty (i.e., no nonoverflowing approach lane exists) but approach lane set A is not empty. To deal with this special case, the lane-based optimized cycle length is adjusted to generate more solution candidates for examination. Again, the refinement for cycle length is also conducted by reducing it by one $\Delta t$ time unit at a time. New trial settings for green durations are then repeated as discussed above until the termination condition is met.

5. A Case Study

To demonstrate the proposed lane-based refinement method for practical applications, a case study is given for illustration. A site survey at one of the busiest and most complex signal-controlled intersections in Hong Kong (Wanchai) was conducted to collect turning flow patterns covering the morning, off, and evening peak periods. The connecting road links have different lengths and widths with different spatial holding capacities and discharge rates. The optimization results from the conventional lane-based approach served as inputs, including the demand turning flows, lane-marking patterns, and traffic signal settings. The geometric details of the intersection are given in Figure 15. The intersection is an isolated crossroad intersection. Hennessy Road is a major road consisting of four approaches and three exit lanes. Fleming Road is a minor road with two approaches and two exit lanes. To account for the spatial lane-holding capacities, the physical road lengths were measured and converted to the number of standard vehicles that could be held along the approach road lanes. It was assumed that a standard vehicle length is 5 m plus 1 m gap between front and rear bumpers of two consecutive vehicles (given as 1 pcu); that is, $\xi_v = 6$ m. The actual measured lane lengths along Hennessy Road were $l_{i,k} = 90$ m for Legs 2 and 4; those along Fleming Road were $l_{i,k} = 30$ m for Legs 1 and 3. The spatial holding capacity for each approach lane from Leg 1 and Leg 3 was 5 pcu and that for each lane from Leg 2 and Leg 4 was 15 pcu. All approaching traffic lanes were of identical importance in terms of overflow, and we used the weighting factors $\eta_l = 1$ and $\Omega_l = 1$ for all leg $i$. Traffic turning demands and the observed traffic signal settings are given in Table 7. Lane-based traffic signal and lane-marking optimization results are shown in Figure 15 (i.e., optimized patterns of lane markings) and in Table 8 (i.e., optimized lane flows and optimized signal settings). These served as refinement model inputs in the proposed refining process to eliminate overflow. To conform to the common code of practice used in Hong Kong, we further assumed that minimum green durations for all vehicular movements were 6 s and that the required intergreen times for pairs of conflicting vehicular movements were also 6 s.

5.1. Spatial Queue Lengths. Directly implementing the optimized traffic signal settings obtained from the lane-based model, it was found that some critical approach lanes could overflow because the spatial lane-holding capacities were practically inadequate to hold all approach demand flow. From the conventional lane-based optimization framework, capacity maximization and delay minimization settings were optimized to generate various traffic signal settings, including various green durations and cycle times, as found in Table 8. Based on these results, the maximum queue lengths $F_{i,k}$ in a signal cycle along the approach traffic lanes were calculated using (4a)–(4d); the results are given in Table 1. The actual lane lengths, $l_{i,k}$, were measured, and the spatial lane-holding capacities were estimated (in pcu), which are the lane lengths for practical operations and the overflow would be either reduced or eliminated (i.e., objective function value is zero).
divided by the standard vehicle length plus a gap between front and rear bumpers of two consecutive vehicles, $\zeta$. In the case study, the approaching traffic lanes from Legs 1 and 3 were found to overflow because the maximum queue lengths were generally greater than the lane-holding capacities in all study periods. This is strong evidence of the weakness of operating according to the lane-based optimization results without taking into consideration the spatial capacities of the approaching traffic lanes.

### 5.2. Overflow due to Improper Green Duration Optimization from Lane-Based Model

To realize the potential problem of the lane-based traffic signal optimization results, we borrowed some previous findings, which are tabulated in Table 8. Total lane flows, $r_{i,k}$, and green durations, $\phi_{i,k}$, in the morning, off, and evening peak hours were optimized using the lane-based optimization model adopting a point queue system. Taking into consideration the spatial lane capacities, the incoming traffic exhibited overflow. It was found that the maximum queue lengths were longer than the lane-holding capacities, as explained in Section 5.1 (Table 1), because the red durations were too long and traffic accumulated in front of stop lines. Under the fixed cycle length, extending the green durations would somewhat shorten the red durations and overflow could be reduced or even eliminated. For this, a new minimum required green duration, $\phi_{i,k}^*$, is calculated numerically from (5a). Table 2 presents the new minimum green times required to avoid overflow. When comparing them to the lane-based optimized green durations, $\phi_{i,k}$, a significant numerical difference was found from $\phi_{i,k}^* - \phi_{i,k}$. If $\phi_{i,k} - \phi_{i,k}^*$ is negative, satisfying inequality (10), the approaching traffic lane $k$ from leg $i$ overflows. If the condition of (10) is violated, the approaching traffic lane should be nonoverflowing. For any nonoverflowing approaching traffic lane, the minimum green duration, $\phi_{i,k}^*$, is evaluated numerically in (5b) to reach an ideal traffic state $E^*$ without developing residue queues.

The approaching traffic lanes from Leg 1 and Leg 3 were found to overflow in all study periods using either the capacity maximization or delayed minimization settings. To prevent such overflow, longer green durations, $\phi_{i,k}^*$, should be used. As for Leg 2 and Leg 4, none of the approaching traffic lanes overflowed when using the optimized green duration, $\phi_{i,k}$, from the lane-based optimization framework. Indeed, the $\phi_{i,k}^*$ values are the minimum green durations for the approaching traffic lanes to operate without overflowing. Whenever $\phi_{i,k} - \phi_{i,k}^*$ is positive, excess green durations are

---

### Table 1: Maximum queue length under the lane-based signal optimization settings.

<table>
<thead>
<tr>
<th>Leg $i$</th>
<th>Lane $k$</th>
<th>Morning peak</th>
<th>Off peak</th>
<th>Evening peak</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max RC</td>
<td>Min total delay</td>
<td>Max RC</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>8.043 (7.900)</td>
<td>6.471 (6.943)</td>
<td>7.821 (7.972)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>8.582 (8.988)</td>
<td>6.905 (7.063)</td>
<td>8.039 (7.902)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5.713 (5.872)</td>
<td>4.271 (4.125)</td>
<td>6.734 (7.007)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6.510 (6.797)</td>
<td>4.866 (5.085)</td>
<td>7.771 (7.808)</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6.395 (6.539)</td>
<td>4.780 (4.954)</td>
<td>7.661 (7.652)</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5.787 (5.967)</td>
<td>4.326 (4.656)</td>
<td>6.908 (6.936)</td>
</tr>
</tbody>
</table>

Note: Cumulative numbers of arrival vehicles that require lane $k$ from leg $i$ to hold up in a signal cycle.
Optimized from the lane-based optimization model. Obviously, approaching traffic lanes from Leg 1 and Leg 3 require longer green durations, and those from Leg 2 and Leg 4 supply excess green duration. Refining the signal settings could certainly minimize the degree of overflow.

Overflow along approaching traffic lanes from Leg 1 and Leg 3 could be avoided if the traffic signal settings obtained from lane-based optimization model are refined by the proposed formulation satisfying (8)–(17).

5.3. Refining Traffic Signal Settings under Fixed Cycle Length. Adopting a point queue system, traffic signal settings from the lane-based optimization model are considered to be optimal for practical operations to maximize the overall intersection capacity or minimize the total delay. In Section 5.2, we mentioned that the conventional lane-based method overdesigns the required green durations once the concepts of spatial queues and spatial lane capacities are introduced. In this study, we developed new design equations, (1)–(17), to refine the traffic signal settings to prevent overflow of approaching traffic lanes. Under a fixed cycle time, \( c \), our strategy for refining the lane-based traffic signal optimization results is to extract excess green duration from nonoverflowing traffic lanes and add those times to overflowing approaching traffic lanes. During the solution process, all traffic signal timings are designed to be safe and compatible with the general lane-based specifications. The solution process is terminated when all excess green duration from the nonoverflowing traffic lanes is shifted to extend the green durations of the overflowing approaching traffic lanes.

Referring to Table 1, we find that approaching traffic lanes from Legs 1 and 3 are overflowing, although the level of overflow differs between legs. According to the maximum spatial queue length, \( F_{i,k} \), the approaching traffic lanes from Leg 1 are more critical than those from Leg 3 in the morning peak period, with longer maximum vehicle queue lengths. Such a difference between Legs 1 and 3 in the maximum spatial queue length is reduced during off and evening peak periods.

To reduce the level of overflow, we first deal with the most critical approaching traffic lanes. In this case study, the approaching traffic lanes from Legs 1 and 3 overflow in all study periods under capacity-maximizing or delay-minimizing settings, and the approaching traffic lanes from Leg 1 overflow more seriously during the morning and off peak periods. Excess green durations along the approaching traffic lanes from Legs 2 and 4 could be extracted and reassigned to overflowing approaching traffic lanes from Legs 1 and 3. To facilitate this refinement process, \( \Delta t = 1 \) s has been adopted to satisfy the constraints of (12) and (13). To refine the signal settings from the maximizing capacity in the morning peak period, the excess green duration from Leg 2 is extracted first (8 s), and then the excess green duration from Leg 4 (9 s) is extracted. These 17 s of excess green duration are added to the green durations of the approach lanes from Leg 1. Figure 8(a) shows the refinement results. With the green duration added to Leg 1, the maximum queue lengths

---

Table 2: Differences of green durations between traffic signal settings with spatial queue requirements and those without spatial queue requirements.

<table>
<thead>
<tr>
<th>Traffic lanes from leg ( i = )</th>
<th>Overflowing legs</th>
<th>Nonoverflowing legs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Max RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_{i,k} )</td>
<td>32.5</td>
<td>21.4</td>
</tr>
<tr>
<td>( \varphi_{i,k} )</td>
<td>69.0</td>
<td>12.9</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-36.5</td>
<td>+8.5</td>
</tr>
<tr>
<td>Min total delay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \varphi_{i,k}^* )</td>
<td>39.4</td>
<td>9.6</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-19.4</td>
<td>+7.1</td>
</tr>
<tr>
<td>Max RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_{i,k} )</td>
<td>26.4</td>
<td>23.8</td>
</tr>
<tr>
<td>( \varphi_{i,k} )</td>
<td>61.8</td>
<td>15.9</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-35.4</td>
<td>+7.9</td>
</tr>
<tr>
<td>Min total delay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \varphi_{i,k}^* )</td>
<td>29.8</td>
<td>11.7</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-12.8</td>
<td>+5.3</td>
</tr>
<tr>
<td>Max RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_{i,k} )</td>
<td>27.1</td>
<td>20.2</td>
</tr>
<tr>
<td>( \varphi_{i,k} )</td>
<td>56.2</td>
<td>11.9</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-29.1</td>
<td>+8.3</td>
</tr>
<tr>
<td>Min total delay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \varphi_{i,k}^* )</td>
<td>17.0</td>
<td>17.0</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>24.2</td>
<td>8.5</td>
</tr>
<tr>
<td>( \phi_{i,k} - \varphi_{i,k}^* )</td>
<td>-7.2</td>
<td>+8.5</td>
</tr>
</tbody>
</table>

Note. \( \phi_{i,k} \) is the green durations (in seconds) optimized by the lane-based optimization model; \( \varphi_{i,k}^* \) and \( \varphi_{i,k}^* \) are the minimum required green durations (in seconds) evaluated by the proposed traffic signal refinement method for not inducing overflowing and reaching ideal traffic state along approach traffic lanes.
along approach lanes 1 and 2 from Leg 1 are shortened from 8.043 pcu and 8.582 pcu (shown in Table 1) to 6.572 pcu and 7.013 pcu (shown in Table 3), respectively. Correspondingly, the maximum queue lengths along the approaching traffic lanes from Legs 2 and 4 increase because of the shortened green durations. However, the shortened queue lengths from Leg 1 are still longer than those from Leg 3 (as shown in Table 3). Figure 8(b) shows very similar results for refining the delay-minimizing settings. For other cases during the off and evening peak periods, Figures 8(c)–8(f) plot the maximum queue lengths along the approaching traffic lanes from Legs 1 and 3 to realize balanced improvements during the solution process.

We have demonstrated that the proposed refining process is able to shorten the maximum queue lengths by reassigning the available green durations among overflowing and nonoverflowing traffic lanes from different legs. These improvements could also be reflected by the numerical objective function values. Figure 9 plots the performance of the objective function in (7) for difference cases in different study periods by the refinement process governed by (8)–(17). The traffic signal settings optimized by the conventional lane-based model serve as the initial inputs to the proposed refining process. With respect to the objective function in (7) encapsulating the spatial queues and lane-holding capacities, the curves in Figure 9 show significant improvements, which are reflected in further reduction in the objective function values in all study cases. The total excess green duration from nonoverflowing approaching traffic lanes is extracted, and every Δt is reassigned iteratively to the overflowing approach traffic lanes. During this refinement process, the objective function value is observed to improve (drop) iteratively. Under this fixed cycle time refinement, the objective function may not be fully optimized because our study aim was to eliminate the overflow until $F_{i,k}$ is reassigned iteratively to the overflowing approach traffic lanes.

Table 3 shows that the maximum queue lengths along the approaching traffic lanes from Legs 1 and 3 could still exceed the spatial holding capacity, $l_{i,k}/\xi_\iota = 5$ pcu, after refinement of the green durations has finished meeting the respective stopping criteria. It is expected that the objective function value could be further reduced if the cycle time was relaxed as one of the fine-tuning parameters in the proposed refinement algorithm. Further discussions are given in Section 5.4.

<table>
<thead>
<tr>
<th>Leg $i$</th>
<th>Lane $k$</th>
<th>Maximum spatial queue length $F_{i,k}$ after the signal refinement (in pcu)</th>
<th>Lane holding capacity (in pcu) $l_{i,k}/\xi_\iota$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6.572, 5.368, 7.069, 5.431, 6.542, 5.044</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.013, 5.728, 7.266, 5.583, 6.807, 5.248</td>
<td>15.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6.177, 4.618, 7.224, 5.250, 5.579, 4.037</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.039, 5.262, 8.336, 6.058, 6.549, 4.739</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.914, 5.169, 8.219, 5.973, 6.519, 4.717</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.257, 4.678, 7.410, 5.385, 5.579, 4.212</td>
<td>15.0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6.061, 4.600, 6.000, 5.077, 6.096, 4.690</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.859, 5.206, 7.253, 5.579, 6.844, 5.265</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6.131, 4.645, 7.610, 5.563, 6.378, 4.691</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.371, 5.584, 8.846, 6.467, 7.515, 5.527</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7.371, 5.584, 8.826, 6.452, 7.515, 5.527</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.367, 3.308, 7.863, 5.748, 5.468, 4.021</td>
<td>15.0</td>
</tr>
</tbody>
</table>

5.4. Refining Traffic Signal Settings with Adjustable Cycle Length. In Section 5.3, it is demonstrated that the objective function...
Figure 8: Variations of maximum queue lengths (in pcu) of approach road lanes in the refinement process of green durations.
function values in (7) could be iteratively reduced by refining the traffic signal settings to consider the spatial queues and spatial lane capacities. Without refining the cycle time, $c$, the objective function may not be fully optimized, implying that the shortened maximum queue lengths along the approaching traffic lanes could still be longer than the spatial lane-holding capacities (i.e., still overflowing). In this section, we further relax the cycle time (length) $c$ as a model variable for refinement to continue the improvement.

Maximum queue lengths are developed according to the lengths of the red times. If full green is given, there must be no (residual) vehicle queues. The longer the red duration is, the more the incoming vehicles are held up and the longer the vehicle queue lengths become. If we keep the cycle time constant, extending the green durations would mean that the red durations are relatively shorter and thus the maximum queue lengths should be shorter. However, there is a limit. Within a signal-controlled intersection, there are conflicting phases in which green durations must not be given during the same timespan, which means that reducing red durations is effective for reducing the maximum queue lengths of the same phase. The resultant long green durations would lead to very long red durations for other conflicting phases. For these, maximum queue lengths could grow longer than the physical approach lane lengths, and overflow could occur again in the other legs. These conflicts are the reason why all green duration and red durations, together with the cycle time, must be well designed to eliminate overflow.

For demonstration, we took the minimum cycle time $c_{\text{min}}$ to be 48 s. The lane-based optimization cycle length $\zeta$ is set to be 120 s in the case of maximizing RC and 90.4 s and 88.0 s in the case of minimizing total delay as the maximum cycle times. A reduced time step size of $\Delta t = 0.1$ s was adopted in performing the refinement process. A feasible range for searching the refined cycle time $c$ was set within the interval $[c_{\text{min}}, \zeta]$, with a step size of $\Delta c = 1.0$ s. A search heuristic was applied for solving the signal refinement model subject to (8)–(17), as described in Figure 7.

In Figure 10, the objective function values in (7) subject to the various model constraints given in (8)–(17) are plotted for all study cases. By adjusting (reducing) the cycle times, all overflowing approaching traffic lanes from Legs 1 and 3 were repaired, in which all maximum queue lengths were controlled to be below the spatial lane-holding capacities (5 pcu). This was achieved as all the objective function values were found to be zero. The resultant cycle times after the refining process were as follows. For the case of maximizing RC, the refined cycle times were 71.0 s during the morning peak period, 74.0 s in the off peak period, and 80.0 s in the evening peak period. For the case of minimizing total delay, the cycle times were 71.4 s, 75.0 s, and 82.0 s for the morning, off, and evening peak periods, respectively.

Tables 4, 5, and 6 tabulate all details of the refinement results for the traffic signal settings in various cases and study periods, including lane traffic flows, turning proportions, flow factors, start and duration of green times, and maximum spatial queue lengths. The maximum queue lengths along the approaching traffic lanes from Legs 1 and 3 were all below their spatial holding capacities (i.e., the maximum queue lengths fulfilled the space requirements of all physical road lengths). The results prove that the refined traffic signal settings are better and more effective for practical operations. For the given patterns of traffic demand turning

---

**Figure 9:** Optimization of the objective function values during the refinement process under a fixed cycle time, $\zeta$.

**Figure 10:** Optimization of objective function values of signal refinement model with adjustable cycle length.
<table>
<thead>
<tr>
<th>From leg</th>
<th>Lane k</th>
<th>To leg j (allocated lane flow, pcu/h)</th>
<th>Total lane flow (pcu/h)</th>
<th>Turning proportion</th>
<th>Saturation flow (pcu/h)</th>
<th>Flow factor</th>
<th>Start of green/duration of green</th>
<th>Maximum queue length (pcu)</th>
<th>Simulated maximum queue length from VISSIM (pcu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>180.0</td>
<td>330.9</td>
<td>0.544</td>
<td>1886.7</td>
<td>0.1754</td>
<td>0.0/20.4</td>
<td>2.3/20.5</td>
<td>4.683</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>154.1</td>
<td>353.1</td>
<td>0.564</td>
<td>2013.2</td>
<td>0.1754</td>
<td></td>
<td></td>
<td>4.997</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>103.0</td>
<td>208.6</td>
<td>0.494</td>
<td>1803.7</td>
<td>0.1157</td>
<td>40.7/7.6</td>
<td>42.8/7.7</td>
<td>3.685</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>237.7</td>
<td>237.7</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1157</td>
<td></td>
<td></td>
<td>4.125</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>33.7</td>
<td>233.5</td>
<td>0.144</td>
<td>2018.6</td>
<td>0.1157</td>
<td></td>
<td></td>
<td>3.733</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>211.3</td>
<td>211.3</td>
<td>1.000</td>
<td>1826.7</td>
<td>0.1157</td>
<td></td>
<td></td>
<td>3.779</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>53.7</td>
<td>224.7</td>
<td>0.761</td>
<td>1812.6</td>
<td>0.1240</td>
<td>54.3/10.7</td>
<td>56.5/11.2</td>
<td>4.277</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>181.3</td>
<td>254.3</td>
<td>0.287</td>
<td>2051.4</td>
<td>0.1240</td>
<td></td>
<td></td>
<td>3.680</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>203.0</td>
<td>210.6</td>
<td>0.964</td>
<td>1709.1</td>
<td>0.1232</td>
<td></td>
<td></td>
<td>4.424</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>253.2</td>
<td>253.2</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1232</td>
<td></td>
<td></td>
<td>4.424</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>253.2</td>
<td>253.2</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1232</td>
<td></td>
<td></td>
<td>2.621</td>
</tr>
</tbody>
</table>

[1]: the optimal cycle length refined from Max RC result is 71.0 s; [2]: the optimal cycle length refined from Min total delay result is 71.4 s.
Table 5: Detailed traffic signal timing results for the off peak period.

<table>
<thead>
<tr>
<th>From leg</th>
<th>Lane</th>
<th>To leg j (allocated lane flow, pcu/h)</th>
<th>Total lane flow (pcu/h)</th>
<th>Turning proportion</th>
<th>Saturation flow (pcu/h)</th>
<th>Flow factor</th>
<th>Start of green/duration of green</th>
<th>Maximum queue length (pcu)</th>
<th>Simulated maximum queue length from VISSIM (pcu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Refined from Max RC [1]</td>
<td>Refined from Min total delay [2]</td>
<td>Refined from Max RC</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>118.0</td>
<td>300.8</td>
<td>0.392</td>
<td>1920.8</td>
<td>0.1566</td>
<td>0.0/16.1</td>
<td>3.3/16.7</td>
<td>4.861</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>82.2</td>
<td>309.2</td>
<td>0.734</td>
<td>1973.9</td>
<td>0.1566</td>
<td></td>
<td></td>
<td>4.997</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>152.0</td>
<td>252.0</td>
<td>0.603</td>
<td>1780.7</td>
<td>0.1415</td>
<td>40.6/9.8/9.8</td>
<td>45.1/10.7</td>
<td>5.050</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>290.8</td>
<td>290.8</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1415</td>
<td></td>
<td></td>
<td>5.199</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>32.5</td>
<td>286.7</td>
<td>0.113</td>
<td>2026.3</td>
<td>0.1415</td>
<td></td>
<td></td>
<td>5.125</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>258.5</td>
<td>258.5</td>
<td>1.000</td>
<td>1826.7</td>
<td>0.1415</td>
<td></td>
<td></td>
<td>4.621</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>89.9</td>
<td>264.9</td>
<td>0.661</td>
<td>1833.6</td>
<td>0.1445</td>
<td>22.1/12.5</td>
<td>26.0/13.1</td>
<td>4.547</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>163.1</td>
<td>291.1</td>
<td>0.440</td>
<td>2014.3</td>
<td>0.1445</td>
<td></td>
<td></td>
<td>4.996</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>178.0</td>
<td>267.0</td>
<td>0.667</td>
<td>1767.7</td>
<td>0.1510</td>
<td></td>
<td></td>
<td>4.643</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>310.4</td>
<td>310.4</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1310</td>
<td></td>
<td></td>
<td>5.398</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>303.6</td>
<td>309.7</td>
<td>0.020</td>
<td>2050.0</td>
<td>0.1510</td>
<td></td>
<td></td>
<td>5.386</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>275.9</td>
<td>275.9</td>
<td>1.000</td>
<td>1826.7</td>
<td>0.1510</td>
<td></td>
<td></td>
<td>4.798</td>
</tr>
</tbody>
</table>

[1]: the optimal cycle length refined from Max RC result is 74.0 s; [2]: the optimal cycle length refined from Min total delay result is 75.0 s.
Table 6: Detailed traffic signal timing results for the evening peak period.

<table>
<thead>
<tr>
<th>From leg ( i )</th>
<th>Lane ( k )</th>
<th>To leg ( j ) (allocated lane flow, pcu/h)</th>
<th>Total lane flow (pcu/h)</th>
<th>Turning proportion</th>
<th>Saturation flow (pcu/h)</th>
<th>Flow factor</th>
<th>Start of green/duration of green</th>
<th>Maximum queue length (pcu)</th>
<th>Simulated maximum queue length from VISSIM (pcu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>131.0</td>
<td>271.0</td>
<td>0.483</td>
<td>1900.2</td>
<td>0.1426</td>
<td>0.0/16.4</td>
<td>4.803</td>
<td>4.998</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>79.0</td>
<td>282.0</td>
<td>0.720</td>
<td>1977.1</td>
<td>0.1426</td>
<td>1.4/17.8</td>
<td>4.805</td>
<td>5.000</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>140.0</td>
<td>186.3</td>
<td>0.751</td>
<td>1750.6</td>
<td>0.1064</td>
<td>3.742/7.8</td>
<td>3.742</td>
<td>3.933</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>46.3</td>
<td>218.7</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1064</td>
<td>4.393/4.50</td>
<td>4.393</td>
<td>4.008</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7.6</td>
<td>210.1</td>
<td>0.035</td>
<td>2046.1</td>
<td>0.1064</td>
<td>4.373/4.43</td>
<td>4.373</td>
<td>4.037</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>194.4</td>
<td>194.4</td>
<td>1.000</td>
<td>1826.7</td>
<td>0.1064</td>
<td>3.905/4.03</td>
<td>3.905</td>
<td>3.984</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>74.3</td>
<td>272.3</td>
<td>0.727</td>
<td>1819.6</td>
<td>0.1496</td>
<td>52.6/21.4</td>
<td>4.450</td>
<td>4.996</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>198.0</td>
<td>305.7</td>
<td>0.321</td>
<td>2043.1</td>
<td>0.1496</td>
<td>54.6/22.8</td>
<td>4.450</td>
<td>4.995</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>172.0</td>
<td>219.3</td>
<td>0.784</td>
<td>1744.0</td>
<td>0.1257</td>
<td>4.244/4.41</td>
<td>4.244</td>
<td>3.970</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>258.4</td>
<td>258.4</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1257</td>
<td>5.001/5.206</td>
<td>5.001</td>
<td>4.998</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>258.4</td>
<td>258.4</td>
<td>0.000</td>
<td>2055.0</td>
<td>0.1257</td>
<td>5.001/5.206</td>
<td>5.001</td>
<td>4.866</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>188.0</td>
<td>188.0</td>
<td>1.000</td>
<td>1826.7</td>
<td>0.1029</td>
<td>3.638/3.78</td>
<td>3.638</td>
<td>3.986</td>
</tr>
</tbody>
</table>

[1]: the optimal cycle length refined from Max RC result is 80.0 s; [2]: the optimal cycle length refined from Min total delay result is 82.0 s.
flows and lane markings, the refined traffic signal results from the two design objectives, capacity maximization and total delay minimization, are quite similar. Thus, the proposed traffic signal refinement procedure is effective and provides consistent results to eliminate overflow problems.

It is also interesting to observe how the spatial queues developed within a signal cycle across different approaching traffic lanes from different legs. Figure 11 gives the details. The vertical axis represents the maximum queue length in pcu and the horizontal axis is the time axis. When using
Figure 11: Queue length (in pcu) against time (from start to end of a signal cycle in seconds) under the refined traffic signal settings.
the refined traffic signal settings, all approaching traffic lanes from Legs 1 and 3 can hold their maximum queue lengths without overflowing (i.e., the maximum queue lengths are controlled within the lane-holding capacities of approaching traffic lanes from Legs 1 and 3).

5.5. Evaluations of the Refinement Procedures Using VISSIM Simulation Model. To evaluate the proposed refinement procedures, the maximum queue lengths are critical parameters to be examined. Overflowing occurs if the maximum queue lengths exceed the physical holding capacities of approach lanes. A VISSIM simulation model was built to model the case study intersection taking the original lane-based optimization model results and the refined traffic signal settings as inputs. From Google map, the intersection layout was captured and inserted to the VISSIM background image. Proper scale was adjusted so that the holding capacities of major roads (15 pcu/lane) and short lanes (5 pcu/lane) are matched. Intersection geometries including all approach and exit lanes were added as VISSIM links with respect to the background image. Based on the lane-marking patterns, connectors were then introduced to connect all approach and exit lanes. For shared lanes, turning proportions were available from the lane-based model outputs and the refinement results. Ratios of turning were defined as relative flows in VISSIM through the interface of vehicle routes. For each link in VISSIM, a signal head was added near the stop line position. Similarly, vehicle inputs (i.e., allocated lane flows from lane-based model outputs and refinement results) were added at the end of links. Nodes were added to VISSIM links with signal heads so that maximum queue lengths, average queue lengths, total discharges, and other engineering parameters could be displayed as output records. For simulating the intersection operations, traffic signal settings including cycle time, ends of green, ends of red, length of amber, and length of red/amber were all given as VISSIM inputs through the interface of signal controllers. And the total simulation period was set to be one hour. From node results in VISSIM, we could then collect the simulated maximum queue lengths for comparisons. A standard vehicle was assumed to be 5.0 m in length. Gap between front and rear bumpers of two consecutive vehicles was set to be 1.0 m with default standard deviation of 0.3 m. Various simulated maximum queue lengths in pcu are tabulated in Tables 1, 4, 5, and 6 for comparisons and evaluations. In general, the difference between the simulated and calculated maximum queue lengths is within ±10.0%. Simulation results are considered to be consistent with the calculated results in terms of the maximum queue lengths. The differences between the calculated and simulated maximum queue lengths may be due to constant gap (in calculation) and varying gap (in simulation) distances adopted in the two modeling platforms. With default standard deviation of 0.3 m in simulating the varying gaps between front and rear bumpers of two consecutive vehicles in VISSIM, the simulated results of maximum queue lengths could be slightly overflowing as presented in the last two columns of Tables 4, 5, and 6. To prevent overflowing due to traffic dynamics, different weighting factors $\Omega$ and $\eta$, for leg $i$ could be specified to provide safety buffers. Relevant applications are discussed in the next section.

5.6. Applications of the Proposed Traffic Signal Refinement Procedures. The proposed lane-based signal refinement procedures are applicable to enhancing traffic signal settings to deal with different traffic conditions and different intersections by specifying different model parameters. Actual traffic flow and arrival patterns are time-varying in nature. Various queuing patterns and initial traffic conditions may be involved. It is necessary to give different initial conditions as inputs to test the proposed refinement algorithms to further demonstrate their modeling features. For signal-controlled intersections, there are T-junctions or cross-intersections, and approach lanes from different legs may have different physical lengths. Some legs could be more critical. Slight overflow may induce long vehicle queues that block upstream road lanes or intersections, causing large total delays. In this formulation, we introduce two sets of numerical weighting factors, $\Omega_i$ and $\eta_i$, for leg $i$ in the objective function. The weighting factor $\Omega_i$ for leg $i$ specifies the relative importance of the road space from different legs. $\eta_i$ for leg $i$ represents the fluctuation of maximum queue lengths due to the randomness of traffic arrivals, which were not included when calculating the maximum queue lengths. In the original case study, all weighting factors of $\eta_i$ and $\Omega_i$ were set to 1.0, thus assuming that all legs were equally important and there was no random effect for minimizing the objective function. $J = \sum_{i=1}^{N_L} \sum_{k \in A_i} (\eta_i F_{i,k} - (i_{k,3}/C_i)/\Omega_i)$. From Figure 12, if heavier weighting factors $\Omega_i$ (Leg 1) and $\Omega_3$ (Leg 3) are given to legs with short lanes, the operating cycle times are further reduced. Because the spatial holding capacities of short lanes are limited, a larger weighting factor simply serves as a safety factor to enlarge the buffer space between the maximum queue lengths $F_{i,k}$ and the actual holding capacity $l_{i,k}/C_i$. To prevent overflow, the right-of-way for short lanes should be granted more frequently by implementing a shorter cycle length. The weighting factors $\eta_1$ (Leg 1) and $\eta_3$ (Leg 3), shown in Figure 12, have similar results. If the random arrival of incoming vehicles or abnormal traffic fluctuations leads to longer maximum queue lengths from Legs 1 and 3, weighting factors $\eta_1$ and $\eta_3$ should be given larger values because overflow would occur more easily. Also, the operating cycle time should be further reduced to avoid overflow along these approach lanes. Similar trends were observed for all study cases.

Another traffic condition that could be varied in daily traffic signal operations is the initial queue length, $l_{i,k}$. In the original case study, we set all initial queue lengths to zero. In reality, initial vehicle queues may exist, and the signal settings must be adjusted to prevent overflow. From an operational point of view, implementing the refined traffic signal settings could effectively reduce the probability of overflow. Because traffic demand patterns are random in nature, it is possible to have residual vehicle queues on the approaching traffic lanes after the display of green signals. To eliminate these residual vehicle queues, the proposed refinement method could be applied by setting a nonzero initial vehicle queue length as an input.
Figure 12: Effect of the road space importance factors \( \Omega_i \) \((i = 1, 3)\) and random arrival influence factors \( \eta_i \) \((i = 1, 3)\) on the signal settings refined by the proposed model.

For short lanes from Legs 1 and 3 with spatial capacities of 5 pcu in the previous case study, as shown in Figure 13, increasing the initial queue length \( l_{i,k} \) from 0.0 to 3.0 pcu has been tested. Different cases show similar results, in which shorter cycle lengths should be used to ensure that all approaching traffic lanes, especially the short lanes, do not overflow. Until the minimum cycle length \( (48 \text{ s}) \) is reached, no further reduction is possible. As for the long approaching traffic lanes from Legs 2 and 4 with spatial capacities of 15 pcu, increasing the initial queue length \( l_{i,k} \) from 0.0 to 14.0 pcu has also been examined. It is found that the cycle length should be reduced initially to ensure that short lanes from Legs 1 and 3 do not overflow. Because the approaching traffic lanes from Legs 2 and 4 are longer with larger spatial capacities to hold more incoming vehicles without overflowing, the cycle length could then be maintained at a steady level, even when longer initial queues are added. Until the longer approaching traffic lanes overflow after adding around 10 pcu as an initial queue length, the cycle length is decreased again and further reduced until the minimum cycle length \( (48 \text{ s}) \) is reached. This decreasing trend stops when no further reduction in cycle time is possible.

In general, nonzero and nonequal initial queue lengths from different legs are common in practice. Figure 14 plots the analysis results from the proposed traffic signal refinement process. The combined effects of different initial queue lengths from different legs are illustrated. As in all the above study cases, the operating cycle times are generally reduced if longer initial queues are given irrespective of the legs.

To sum up, for approaching traffic lanes with limited holding capacities such as approaching traffic lanes from Legs 1 and 3 in the case study, overflow occurs as the maximum queue length of an approaching traffic lane is developed that is proportional to the length of the red time within a signal.
Figure 13: Effects of a single initial queue length, $\ell_{i,k}$ ($i = 1, 2, 3, 4$), in pcu on the refined signal settings.

Figure 14: Combined effects of setting different initial queue lengths, $\ell_{i,k}$ ($i = 1, 3$), in pcu from different legs for the refinement process.
cycle. Because of the spatial queue requirements along short lanes, the original lane-based signal optimization results are no longer suitable for practical operations. In different traffic demand patterns (in different survey periods), the proposed traffic signal refinement was found to be effective for refining the red durations, green durations, and cycle times to prevent overflow.

6. Conclusions

In the conventional lane-based optimization framework, which is developed using a point queue approach, signal timing settings are always optimized to maximize the overall capacity, neglecting the spatial queue requirements of short physical road lengths. One critical issue is to examine whether the optimized settings also satisfy the vehicle queuing requirements without causing overflow along all approaching traffic lanes. Based on the lane-based optimization results from the original design framework, a refinement method is proposed to fine-tune the lane-based signal settings to ensure that the physical spatial limits for approaching traffic lanes, including short lanes, are not exceeded. A search heuristic is developed to modify the lane-based signal settings for isolated intersections, including green start times, green durations, and cycle length, considering vehicle queuing patterns and physical holding capacities along all approaching traffic lanes. The proposed refinement algorithm is found to be effective for refining the red durations, green durations, and cycle times to ensure that all vehicular traffic is well controlled such that all spatial queue lengths satisfy the spatial lane capacities governed by the established lane-marking patterns. A VISSIM simulation model is built to simulate and evaluate the maximum queue lengths. Differences between the calculated and simulated maximum queue lengths are found to be within ±10%. Results of the maximum queue lengths are considered to be consistent. Safety factors could be applied in the refinement procedures to provide additional buffers to avoid overflowing due to traffic dynamics. For further study, lane-marking patterns established at (upstream) intersections surely affect the demand (flow turning) patterns at downstream intersections. In a network, individual lane markings at intersection level would serve as “links” to connect upstream and downstream intersections. Ban turn can be implemented by not providing the respective lane marking (arrow). Entire network configurations would be controlled by lane-marking patterns. OD demand would be composed of path flows and path flows would then be distributed to intersections as turning demand flows. As long as OD demand flow exists, a single path or multiple paths should exist depending on the demand intensities. Path flows then control the turning flows at intersections. With demand flows at intersections, lane markings could be optimized. By developing flow conservation constraints (at different levels), proper assigned lane flows, lane markings, and network link connectivity could be designed. This problem could be developed by extending the lane-based optimization framework to deal with overflowing problems for linked signalized network systems. Challenges would be designing green times with proper offsets in which the arrivals and discharges of traffic at intersections could be optimized taking platoon or other arrival patterns into considerations.

Notations

\( i, m \): Leg \( i \) or \( m \) of a signal-controlled intersection

\( j, n \): Destination (exit) leg \( j \) or \( n \) of a signal-controlled intersection

\( k, o \): Approaching traffic lane \( k \) or \( o \)

\( q_{i,j,k}(t) \): Arrival rate of a turning movement from leg \( i \) to leg \( j \) on lane \( k \) at time \( t \)

\( r_{i,k}(t) \): Total lane flow on lane \( k \) from leg \( i \) at time \( t \)

\( s_{i,k}(t) \): Saturation flow (discharge) rate on lane \( k \) from leg \( i \) at time \( t \)

\( c \): Cycle length after the proposed refinement process (spatial queue system)

\( \zeta \): Cycle length from the lane-based traffic signal optimization (point queue system)

\( \theta_{i,k} \): Green start time for traffic movement(s) on approach lane \( k \) from leg \( i \)

\( \phi_{i,k} \): Green duration for traffic movement(s) on approach lane \( k \) from leg \( i \)

\( h_{i,k} \): Minimum green duration for traffic movement(s) on approach lane \( k \) from leg \( i \)

\( \phi'_{i,k} \): Required green duration of approach lane \( k \) from leg \( i \) to prevent the maximum vehicle queue length from exceeding physical holding capacity

\( \phi''_{i,k} \): Green duration required without leaving residue vehicle queues on approach lane \( k \) from leg \( i \)

\( \Theta_{i,j} \): Green start time for traffic movement from leg \( i \) to leg \( j \)
\( \Phi_{i,j} \): Green duration for traffic movement from leg \( i \) to leg \( j \)
\( g_{i,j} \): Minimum green duration for traffic movement from leg \( i \) to leg \( j \)
\( \gamma_{u,v} \): Intergreen (clearance) time separating the right-of-way of two incompatible turning movements \( u = (i, j, k) \) and \( v = (m, n, o) \)
\( \delta_{i,j,k} \): Binary variable to show the existence of a lane-marking arrow on approach lane \( k \) turning from leg \( i \) to leg \( j \) (=1 permitted or =0 not permitted)
\( \Pi_{i,j,m,n} \): Successor function controlling two incompatible signal groups \((i, j) \) and \((m, n)\)
\( l_{i,k} \): Physical road length of approach lane \( k \) from leg \( i \) (in meters)
\( \xi_v \): Physical length of a standard vehicle plus a gap between front and rear bumpers of two consecutive vehicles occupying actual road space (in meters)
\( \ell_{i,k}(t) \): Initial queue length on approach lane \( k \) from leg \( i \) at the start of a signal cycle (pcu)
\( L_{i,k}(t) \): Queue length on approach lane \( k \) from leg \( i \) at time \( t \) (pcu)
\( E_{i,k} \): Maximum queue length on approach lane \( k \) from leg \( i \) (pcu)
\( E_{i,k} \): Difference between the maximum vehicle queue length and the physical holding capacity of approach lane \( k \) from leg \( i \) (pcu)
\( \nu_i \): Numerical weighting factor to specify the randomness level of traffic arrival from leg \( i \)
\( \nu_i \): Numerical weighting factor to measure the importance of road space in avoiding the traffic overflow from leg \( i \)
\( A_i \): Set of overflowing approach traffic lanes from leg \( i \) with positive value of \( E_{i,k} \) and the union of all lane set \( A_i \) is denoted by \( A \)
\( B_i \): Set of nonoverflowing approach traffic lanes from leg \( i \) with negative value of \( E_{i,k} \) and the union of all lane set \( B_i \) is denoted by \( B \)
\( \theta \): Set of all green start times on approach lanes from all different legs
\( \varphi \): Set of all green durations on approach lanes from all different legs
\( \beta(\theta, \varphi) \): Objective function value to implement signal timings \((\theta, \varphi)\) at cycle time \( c \)
\( \beta^* \): Objective function value while operating the trial signal timings \((\theta^*, \varphi^*)\) at cycle time \( c \)
\( (i^*, k^*) \): Approaching traffic lane with the largest value of \( E_{i,k} \) in the overflowing lane set \( A \)
\( (i^*, k^*) \): Approaching traffic lane with the shortest excess green duration (i.e., part of green duration that is longer than \( \phi_{i,k}^* \)) in the nonoverflowing lane set \( B \)
\( \Delta t \): Size of time step for reassigning green durations
\( \Delta c \): Size of time step for updating cycle length.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper was supported by the GRF funding scheme from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CityU 123313).

References


