A network enhancement model with integrated lane reorganization and traffic control strategies

Peng Li, Ph.D. Student
Department of Civil Engineering and Mechanics, University of Wisconsin at Milwaukee
P.O. Box 784, Milwaukee, WI
Tel.: 414-229-3857
Email: pengli@uwm.edu

Yue Liu, Ph.D. (Faculty Advisor)
Assistant Professor
Department of Civil Engineering and Mechanics, University of Wisconsin at Milwaukee
P.O. Box 784, Milwaukee, WI
Tel.: 414-229-3857
Email: liu28@uwm.edu
ABSTRACT

Lane reorganization strategies such as lane reversal, one-way street, turning restriction and cross elimination have demonstrated their effectiveness in enhancing transportation network capacity. However, how to select the most appropriate combination of those strategies in a network remains challenging to transportation professionals considering the complex interactions among those strategies and their impacts on conventional traffic control components. This article contributes to developing a mathematical model for a traffic equilibrium network, in which optimization of lane reorganization and traffic control strategies are integrated in a unified framework. The model features a bi-level structure with the upper-level model describing the decision of the transportation authorities for maximizing the network capacity. A variational inequality (VI) formulation of the user equilibrium (UE) behavior in choosing routes in response to various strategies is developed in the lower level. A GA-based heuristic is used to yield meta-optimal solutions to the model. Results from extensive numerical analyses reveal the promising property of the proposed model in enhancing network capacity and reducing congestion.

Keywords: Lane reorganization, network optimization; capacity enhancement; traffic management, bi-level programming
1. INTRODUCTION

How to efficiently utilize the existing transportation network has been one of the most important issues faced by transportation professionals as congestion on roadways in many cities across the world continues getting worse due to increasing traffic volumes. Reorganization of current roadway lane configurations through traffic management strategies can significantly reduce the construction cost of building new roads and has proved to be effective to increase the capacity of transportation network. In review of literature, commonly adopted lane reorganization strategies for network enhancement include turning restriction, lane reversal, cross elimination, and one-way street operation [1].

Turning restriction is one of the most commonly used strategies to improve the capacity of signalized intersections in an urban network [2]. The resulting capacity increase is due to the reduced number of signal phases and less loss time. However, vehicles in prohibited movements are forced to detour, which may induce extra driving distances in the network. In the NCHRP Report 457 [3], proposed guidelines to identify the conditions for left-turn restriction at existing intersections; however where to implement turning restrictions was not discussed in the report. Some researchers in recent years have formulated discrete network design problems with a bi-level structure to optimize turning restriction settings in the transportation network [4,5]. Other studies [6] attempt to eliminate conflicts between movements at an intersection and convert signalized intersections into uninterrupted flow ones (also called crossing elimination). Liu and Luo [7,8] optimized the distribution of signal control and uninterrupted flow intersections with resource constraints in static and dynamic network settings. Integrated models [9,10] were also proposed to combine crossing elimination and lane reversal strategies during emergency evacuation.
Lane reversal (also called tidal flow) has long been used across the world to accommodate frequent and predictable unbalanced traffic demand during peak periods [11-13], special event management [14-16], and emergency evacuation [17-19]. The key idea is to configure the lanes of a roadway to match available capacity with traffic demand. A handful of bi-level network models have been formulated to optimize the setting and selection of lane reversals while accounting for various types of route choice behaviors of network users [20-24] further developed time-varying reversibility with different reversibility durations for various candidate link pairs in a bi-level program model such that the optimal starting times and the optimal reversibility durations for candidate link pairs can be determined for peak-period traffic management on a daily basis.

An extreme case of the lane reversal is the one-way street strategy in which the conversion takes place in the entire roadway. Experimental studies have quantified the trip-serving capacity of the one-way street strategy [25-27]; however those results tend to be site-specific and generalization to other networks cannot be made. Gayah and Daganzo [28] compared the trip-serving capacities of one-way and two-way networks based on macroscopic analyses. It was found that two-way networks can serve more trips per unit time than one-way networks when average trip lengths are short. Similar to the turning restriction strategy, one-way street is not always beneficial due to the resulting extra vehicle detour distance. Realizing this, some researchers have developed optimization models to select the most appropriate segments for one-way traffic organization [29-31].

As above stated, lane reorganization strategies have been used for several decades and much is known about their effectiveness, feasibility and safety. However, few efforts have been made to investigate the interactions among those strategies as well as their combinational
impacts on the transportation network. Furthermore, implementation of lane reorganization strategies may also affect other conventional traffic management and control components in the network, such as lane channelization and signal timings at intersections. Neglect of such interactions may result in non-optimal design results and unsatisfactory operational performances in the network.

To remedy this deficiency, it is necessary to develop integrated models for design and operation of those strategies. Several integrated model [32-34] have been established to combine the design of lane markings and signal timings for isolated signalized intersections. It was shown that substantial improvement in the intersection performance can be achieved by using the integrated model. In the network level, the benefit of integrated design and operation should be even more significant as there are more combinations of decisions and flexibility to accommodate different traffic flow patterns. However, only limited studies have been done regarding network enhancement with lane reorganization and traffic control strategies optimized in a unified framework. Some studies have developed models for the integrated design of signal settings with network routing decisions [35,36]. However, the optimization was not done in a concurrent manner and lane reorganization strategies were not considered. Xie and Turnquist [9,10] proposed an integrated model combining crossing elimination and lane reversal strategies during emergency evacuation; but lane channelization and signal timings were neglected in their models.

In this paper, a lane-based optimization model is proposed, integrating signal timings and reorganization of lane configurations in a unified framework. The proposed model aims to address the following critical question that has long challenged transportation authorities during traffic management, namely: given the target transportation network and demand distribution
how to select the most appropriate combination of traffic management strategies to enhance the network capacity? The rest of the paper is organized as follows. In Section 2, network representation and notations adopted in this paper are described. The optimization model is proposed in Section 3. Section 4 develops the heuristic to solve the model. Performance of the integrated model is evaluated through numerical analysis in Section 5. Findings and concluding remarks are made in the end.

2. NETWORK REPRESENTATION

As illustrated in Figure 1, the target network $\mathcal{G} = (\mathcal{N}, \mathcal{S})$ consists of a set of intersections denoted by $\mathcal{N}, r \in \mathcal{N}$ and a set of links joining intersections denoted by $\mathcal{S}, a = (r, r') \in \mathcal{S}$. Each intersection $r$ consists of a set of arms denoted by $\mathcal{A}_r, i \in \mathcal{A}_r$. And each arm $i$ consists of a set of turning arcs denoted by $\mathcal{T}_i, w \in \mathcal{T}_i$, and a set of lanes denoted by $\mathcal{L}_i, k \in \mathcal{L}_i$. A pair of two directional links between two intersections $r$ and $r'$ can be defined as a segment, denoted by $(a, a')$ with $a = (r, r')$ and $a' = (r', r)$.

For any lane $k \in \mathcal{L}_i$, denote $x_{riwk}$ be the permission of arc $w$ on lane $k$ on arm $i$ at intersection $r$. Then, turning restriction can be easily realized by setting the prohibited movement not permitted in any lanes in the arm. For any link $a \in \mathcal{S}$, denote $n_a$ be the number of lanes on the link, $n_{a'}$ be the number of lanes on the link of opposing direction, and $n_{aa'}$ be the total number of lanes in both directions. Therefore, lane reversal and one-way street strategy can be realized by deciding the values of $n_a$ and $n_{a'}$. A set of auxiliary binary variables $\delta_{riw}$ and $\delta_a$ are used to represent the permission of arc $w$ on arm $i$ at intersection $r$ and on segment $a$, respectively (1- Yes, 0- No). Figure 2 illustrates an example network operated with different type of traffic management strategies.
Figure 1 Example network representation.

Figure 2 Traffic management strategies in a network.
Let $\mathcal{O} \subseteq \mathcal{N}$ represent the set of demand origins, $\mathcal{D} \subseteq \mathcal{N}$ represent the set of demand destinations, and $(o, d)$ represent an OD pair with $o \in \mathcal{O}, d \in \mathcal{D}$. Let $Q_{o,d}$ be the traffic demand between $(o, d)$, $q_{o,d}$ be the scaled traffic demand between $(o, d)$, and $Z_{o,d}$ represents the set of routes between $(o, d)$. For any route $z \in Z_{o,d}$, it may include a sequence of links and turning arcs. The binary variables $\delta_{o,d}^{az}$ and $\delta_{o,d}^{wz}$ are used to represent whether a route $z$ between $(o, d)$ traverses link $a$ or turning arc $w$, respectively (1- Yes, 0- No). The scaled traffic demand on route $z$ between $(o, d)$ is denoted by $q_{o,d}^z$.

3. The Optimization Model

3.1 Decision Variables

The optimization model aims to simultaneously determine the best set of lane reorganization and traffic control strategies, expressed by the following set of decision variables:

- $n_a = \text{the number of lanes on link } a, \forall a \in S$
- $x_{r i w k} = \text{the permission of turning } w \text{ on lane } k \text{ on arm } i \text{ at intersection } r \text{ (1- Yes, 0- No),}$
- $\xi_r = \text{reciprocal of signal cycle length at intersection } r \text{ (1/s), } \forall r \in \mathcal{N}$
- $Y_{r i k} = \text{start of green for lane } k \text{ on arm } i \text{ at intersection } r, \forall k \in L_i; i \in \mathcal{A}_r; r \in \mathcal{N}$
- $\lambda_{r i k} = \text{green split for lane } k \text{ on arm } i \text{ at intersection } r, \forall k \in L_i; i \in \mathcal{A}_r; r \in \mathcal{N}$
- $P_{r(iw,jw')} = \text{order of signal phase for a pair of conflicting traffic movements at}$
  - intersection $r$ (1- movement $(i, w)$ follows movement $(j, w')$), 0- movement $(j, w')$ follows movement $(i, w)$), $\forall w \in T_i; w' \in T_j; i, j \in \mathcal{A}_r; r \in \mathcal{N}$

These decision variables can further be grouped into the solution vector $\mathbf{\eta} = (n, x, \xi, Y, \Lambda, P)$ with $n = (n_a | a \in S)$, $x = (x_{r i w k} | k \in L_i; w \in T_i; i \in \mathcal{A}_r; r \in \mathcal{N})$, $\xi =$
\( (\xi_r | r \in \mathcal{N}) \), \( Y = (Y_{rik} | k \in \mathcal{L}_i; i \in \mathcal{A}_r; r \in \mathcal{N}) \), \( \Lambda = (A_{rik} | k \in \mathcal{L}_i; i \in \mathcal{A}_r; r \in \mathcal{N}) \), and \( P = (P_{r(iw,jw')} | w \in \mathcal{T}_i; w' \in \mathcal{T}_j; i, j \in \mathcal{A}_r; r \in \mathcal{N}) \).

### 3.2 The Upper-Level Problem

The proposed model aims to maximize the reserved capacity for the given network. By adopting the commonly used assumption that the proportions of the traffic demand remain constant, maximizing the reserve capacity is equivalent to maximizing the common flow multiplier [33,34,37], \( \mu \), given by:

\[
\max \quad \mu \tag{1}
\]

The optimization problem should include the flow conservation constraints,

\[
\mu Q_{o,d} = q_{o,d}, \quad \forall o \in \mathcal{O}; \; d \in \mathcal{D} \tag{2}
\]

\[
q_{riw} = \sum_{k=1}^{n_{ri}} q_{riwk}, \quad \forall w \in \mathcal{T}_i; \; i \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{3}
\]

the lane assignment constraints,

\[
n_{aa'} = n_a + n_{a'}, \quad \forall a = (r, r') \in \mathcal{S} \tag{4}
\]

\[
n_{ri}^s = n_{rl} + n_a, \quad \forall i \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{5}
\]

\[
\sum_{w \in \mathcal{T}_i} x_{riwk} \geq 1, \quad \forall k \in \mathcal{L}_i; \; i \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{6}
\]

\[
M \delta_{riw} \geq \sum_{k=1}^{n_{ri}} x_{riwk} \geq \delta_{riw}, \quad \forall w \in \mathcal{T}_i; \; i \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{7}
\]

\[
M \delta_a \geq n_{rl} \geq \delta_a, \quad \forall a = (r, r') \in \mathcal{S}; \; i \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{8}
\]

\[
M \delta_a \geq \sum_{w \in \mathcal{T}_i} \delta_{riw'} \geq \delta_a, \quad \forall a = (r, r') \in \mathcal{S}; \; i' \in \mathcal{A}_r; \; r' \in \mathcal{N} \tag{9}
\]

\[
M \delta_a \geq \sum_{(rw, a) \in \mathcal{C}_r} \delta_{jw} \geq \delta_a, \quad \forall a = (r, r') \in \mathcal{S}; \; w \in \mathcal{T}_j; \; j \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{10}
\]

\[
n_a \geq \sum_{k=1}^{n_{ri}} x_{rjwk}, \quad \forall (rjw, a) \in \mathcal{C}_r; \; a = (r, r') \in \mathcal{S}; \; w \in \mathcal{T}_j; \; j \in \mathcal{A}_r; \; r \in \mathcal{N} \tag{11}
\]

\[
1 - x_{riw(k+1)} \geq x_{riw'}k, \quad \forall k \in \{1, \ldots, n_{ri} - 1\}, w \in \{1,2,3\}, w' \in \{w + 1, \ldots, 4\}, i \in \mathcal{A}_r, r \in \mathcal{N} \tag{12}
\]

the signal timing constraints,
\[
\frac{1}{c_{\min}} \geq \xi_r \geq \frac{1}{c_{\max}}, \forall r \in \mathcal{N}
\]  
\[1 \geq y_{riw} \geq 0, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[1 \geq \lambda_{riw} \geq 0, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M\delta_{riw} \geq \lambda_{riw} \geq -M\delta_{riw}, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(1 - x_{riwk}) \geq Y_{rik} - y_{riw} \geq -M(1 - x_{riwk}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(1 - x_{riwk}) \geq \Lambda_{rik} - \lambda_{riw} \geq -M(1 - x_{riwk}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[\gamma_{rijk} = \frac{\sum_{w \in \mathcal{T}_i} q_{riwk}}{s_{rik}}, \forall k \in \mathcal{L}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(2 - x_{riwk} - x_{riw(k+1)}) \geq Y_{ri(k+1)} - y_{riw} \geq -M(2 - x_{riwk} - x_{riw(k+1)}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[d_{\max}\Lambda_{rik} \geq \gamma_{rijk}, \forall k \in \mathcal{L}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[d_{\max}S_a n_a \geq q_a, \forall a \in \mathcal{S}\]  
\[\frac{1}{c_{\min}} \geq \xi_r \geq \frac{1}{c_{\max}}, \forall r \in \mathcal{N}\]  
\[1 \geq y_{riw} \geq 0, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[1 \geq \lambda_{riw} \geq 0, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M\delta_{riw} \geq \lambda_{riw} \geq -M\delta_{riw}, \forall w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(1 - x_{riwk}) \geq Y_{rik} - y_{riw} \geq -M(1 - x_{riwk}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(1 - x_{riwk}) \geq \Lambda_{rik} - \lambda_{riw} \geq -M(1 - x_{riwk}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[\gamma_{rijk} = \frac{\sum_{w \in \mathcal{T}_i} q_{riwk}}{s_{rik}}, \forall k \in \mathcal{L}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[M(2 - x_{riwk} - x_{riw(k+1)}) \geq Y_{ri(k+1)} - y_{riw} \geq -M(2 - x_{riwk} - x_{riw(k+1)}), \forall k \in \mathcal{L}_i; \ w \in \mathcal{T}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[d_{\max}\Lambda_{rik} \geq \gamma_{rijk}, \forall k \in \mathcal{L}_i; \ i \in \mathcal{A}_r; \ r \in \mathcal{N}\]  
\[d_{\max}S_a n_a \geq q_a, \forall a \in \mathcal{S}\]  
the acceptable level-of-service constraints,

Constraint (2) obtains the scaled traffic demand matrix given the set of demand origins and destinations as an exogenous input. Constraint (3) indicates that the sum of the flows of a movement on different lanes should be equal to the total assigned flow of that movement, where \(q_{riw}\) represents the flow of movement \(w\) on arm \(i\) at intersection \(r\), and \(q_{riwk}\) presents the flow of movement \(w\) on lane \(k\) on arm \(i\) at intersection \(r\).

Constraint (4) sets the total number of lanes in a segment, which adds the number of lanes on links of two directions. Constraint (5) sets the total number of lanes in an arm, which is
equal to the sum of the number of approach and receiving lanes, where \( n_{rl}^a \) represents the total number of lanes in an arm, \( n_{rl} \) represents the number of approach lanes, and the number of receiving lane is assumed to be equal to the number of lanes on the link. Constraint (6) allows each lane to carry at least one turning or through movement. Constraint (7) sets the turning restriction strategy: if a movement at the intersection is prohibited, the number of lanes permitted for the prohibited movements should be equal to 0; otherwise, the movement should be permitted in at least one lane, where \( M \) is an arbitrary large positive constant number. Constraint (8)-(10) sets the one-way street strategy: if the right-of-way of a link is prohibited, the number of lanes in the link should be set to 0, as illustrated by Constraint (8); all movements in the approach connecting to the link should be prohibited, as indicated by Constraint (9); and all the movements entering the link should also be prohibited, as indicated by Constraint (10), where \( C_r \) represents the set of turning movements and their receiving links at intersection \( r \). Constraint (11) sets that the number of lanes in a movement’s corresponding receiving link should be at least as many as the number of lanes assigned to that movement to ensure safety and operational efficiency. Constraint (12) prevents internal conflicts among lanes in an arm.

Constraint (13) limits the common cycle length for the intersections in the network to be within \( C_{\text{min}} \) and \( C_{\text{max}} \), which represent the minimum and maximum cycle lengths. Instead of defining the cycle length directly as the control variable, its reciprocal \( \xi_r = 1/C_r \), is used to preserve the linearity in the mathematical formulation [33,34]. Constraint (14) confines the start of the green to be within a fraction between 0 and 1 of the cycle length. Constraint (15) indicates that the green split of a movement is confined between 0 and 1 of the cycle length. Constraint (16) sets that the green split of a movement should be equal to 0 if the movement is prohibited. Constraints (17)-(18) define the lane signal timings. Constraint (19) sets the order of signal phase
display for a pair of conflicting traffic movements at intersection \( r \), which is governed by a successor function [38]. Constraint (20) limits the start of greens for any pair of conflicting traffic movements considering the minimum clearance time and movement prohibition, where \( I_{r(iw,jw)} \) represents the clearance time for a pair of conflicting traffic movements.

Constraint (21) obtains the flow ratio, where \( \gamma_{rik} \) is the flow ratio of lane \( k \) on arm \( i \) at intersection \( r \), and \( s_{rik} \) represents the saturation flow rate of lane \( k \) on arm \( i \) at intersection \( r \). Constraint (22) sets the flow ratios on a pair of adjacent approach lanes to be identical if they share a common lane marking. Constraints (23) and (24) limits the degree of saturation for each approach lane and each link to be no more than the maximum limit to ensure acceptable level of service, where \( ds_{\text{max}} \) is the maximum acceptable degree of saturation.

3.3 The Lower-Level Problem

The lower-level problem specifies the destination distribution and routing assignment of the traffic demand. Given a feasible solution \( \eta = (n, x, \xi, Y, A, P) \), drivers will route in the network without violating the turning restrictions and signal control constraints. The network flow distribution will therefore reflect their route choice behaviors. Since network enhancement is usually not a short-term event, we adopt the user equilibrium (UE) principle to capture the resulting network flow pattern in this model. Based on the UE principle, no driver could unilaterally decrease his/her transportation disutility by changing routes between a certain OD pair.

The disutility along a route could be expressed as the sum of disutilities along its comprising links and turning arcs, which is expressed by the travel times. As show in Equation (25) and (26), the BPR-form function is adopted to estimate the disutility on the link and turning arcs.
\[ u_a(q, \eta) = t^0_a [1 + k_a (d_{s_a})^{b_a}], \; \forall a \in S \]  
\[ u_w(q, \eta) = t^0_{riw} [1 + k_w (d_{s_{riw}})^{b_w}], \; \forall w \in T_i; \; i \in A_r; \; r \in N \]

where \( t^0_a \) and \( t^0_w \) are the free flow travel time at road section link \( a \) and turning arc \( w \) on arm \( i \) at intersection \( r \), respectively; \( d_{s_a} \) and \( d_{s_{riw}} \) are the degree of saturation on link \( a \) and turning arc \( w \) on arm \( i \) at intersection \( r \), respectively; \( k_a, b_a, k_w, b_w \) are function parameters.

Then, the UE flow pattern in the network can be captured by solving the following parametric variational inequality (VI) problem, where \( T \) represents the set of all turning arcs in the network; \( q = (q_a, q_w) = \{(q_a, \; \forall a \in S) \cup (q_w, \; \forall w \in T)\} \) represents the flow pattern under the solution \( \eta \); \( q_a \) and \( q_w \) are the flow vectors of link and turning arcs, respectively.

\[ \sum_{a \in S} u_a^T (q, \eta) (q_a^* - q_a) + \sum_{w \in T} u_w^T (q, \eta) (q_w^* - q_w) \geq 0 \]  
\[ \forall v^* \in \Omega(\eta) = \{ \eta | q_a = \sum_{o \in O} \sum_{d \in D} \sum_{z \in Z_{o, d}} q_{o, d}^z \delta_{o, d}^{a^z}, \; \forall a \in S \} \]

4. SOLUTION

The proposed optimization model has a bi-level structure with a mix-integer-non-linear-programming problem at the upper-level and a parametric variational inequality at the lower-level. It is therefore NP-hard and difficult to solve due to its non-convexity and non-differential characteristics. In this section, we developed a genetic algorithm (GA) based heuristic method to yield viable and approximate optimal solutions to the model in a reasonable time period. The procedure of the algorithm flow is illustrated in Figure 3.
Initialization
- common flow multiplier, $\mu^l$
- population of chromosome, $\mathcal{H}^n = \{ \eta^m_n | m = 1,2,\ldots, \alpha \}$
- $l = 0, n = 0, m = 1$

Crossover and mutation
$\hat{\mathcal{H}}^n = \{ \hat{\eta}^m_n | m = 1,2,\ldots, \alpha \}$

Network flow assignment
- obtaining flow pattern, $q^m_n$

Degree of saturation verification
$(\eta^m_n, q^m_n)$ satisfying the constraints (24)-(28)

Fitness evaluation
$h^*_{mn} = h(\eta^m_n, q^m_n)$

Breed a new population
$\mathcal{H}^{n+1} = \{ \eta^m_{n+1} | m = 1,2,\ldots, \alpha \}$

Update common flow multiplier
$\mu^{l+1}$

Stoping criteria

Output
- common flow multiplier, $\mu^*$
- network optimization design scheme, $\eta^*$

Figure 3 Procedure of the heuristic algorithm.

Specifics on each module of the algorithm are illustrated as follows:

Initialization

An initial common flow multiplier $\mu^l$ ($l = 0$) and a population consisting of $\alpha$ distinct chromosomes, denoted by $\mathcal{H}^n = \{ \eta^m_n | m = 1,2,\ldots, \alpha \}$, are generated satisfying constraints (2)-(20), where $l$ represents the index of iteration; $n$ represents the index of GA generation; $m$ represents the index of chromosome; $\eta^m_n$ is a binary string representing chromosome $m$ at generation $n$.  

Crossover and mutation
One-point crossover and mutation operations are performed on the chromosomes selected from $\mathcal{H}^n$ to generate new solution populations, and $\tilde{\mathcal{H}}^n$ is the set of all resulting offspring chromosomes that satisfy constraint (2)-(20).

**Network flow assignment**

The diagonalization algorithm is used to solve the parametric variational inequality (27) to obtain the UE flow pattern $\mathbf{q}^n_m$, corresponding to the design decision $\mathbf{n}^n_m$. At each sub-iteration of diagonalization algorithm, the vector disutility function is diagonalized at the current solution, yielding a normal UE problem, which can be solved by the Frank-Wolf method [39].

**Degree of saturation verification**

Though each chromosome in the population is generated satisfying constraints (2)-(20), the degree of saturation constraints may still be violated since the traffic flows on links and turning arcs cannot be obtained before finishing the network flow assignment step. To deal with this problem, the feasibility verification step is designed. For the solution $(\mathbf{n}^n_m, \mathbf{q}^n_m)$, if the degree of saturation constraints (21)-(24) are satisfied, the algorithm will go to the step of stopping criteria check directly; otherwise, the algorithm will continue searching the solutions.

**Fitness evaluation**

Given $\mathbf{n}^n_m$ and $\mathbf{q}^n_m$ ($m = 1, 2, ..., \alpha$), the evaluation function of chromosome $m$ at generation $n$, $h^n_m = h(\mathbf{n}^n_m, \mathbf{q}^n_m)$, can be calculated by Equation (29). Then the fitness of each chromosome can be then computed by normalizing its evaluation value with Equation (30):

$$h(\mathbf{v}, \mathbf{n}) = \sum_{r \in \mathcal{N}} \sum_{l \in \mathcal{A}_r} \sum_{w \in \mathcal{T}_l} \left( \max(ds_{r\ell w} - ds_{\max}, 0) \right) + \sum_{a \in \mathcal{S}} \left( \max(ds_a - ds_{\max}, 0) \right)$$  \hspace{1cm} (29)

$$\bar{h}^n_m(\mathbf{v}, \mathbf{n}) = \frac{h^n_m - h(\mathbf{n}^n_m, \mathbf{q}^n_m) + \varepsilon}{h^n_{\max} - h^n_{\min} + \varepsilon}$$  \hspace{1cm} (30)

where $h(\mathbf{v}, \mathbf{n})$ is the evaluation function; $h^n_{\max}$ and $h^n_{\min}$ denote the maximum and minimum evaluation function values at generation $n$, respectively; $\varepsilon$ is a positive value between
0 and 1 which functions to prevent (30) from zero division and adjust the selection behavior between fitness proportional selection and pure random selection [40].

Breed a new population

Generate the new population $\mathcal{H}^{n+1}$ of size $\alpha$ by using a binary tournament selection method [40] according to the fitness of each chromosome calculated with Equation (30).

Update the common flow multiplier

The common flow multiplier is adjusted according to the dichotomy principle, given by:

$$\mu^{l+1} = \mu^l + \rho \prod_{i=0}^l \beta^l + \frac{\mu^l - \mu^{l-1}}{2} (2\beta^l - 1)(2\beta^{l-1} - 1)(1 - \prod_{i=0}^l \beta^l)$$

where $\beta^l$ is a binary indicator showing whether a solution is found at iteration $l$ (1- Yes, 0-No).

Stopping criteria

The algorithm would not stop unless the difference between the common flow multipliers of two adjacent iterations $l$ and $l-1$ is less than a threshold $\epsilon$:

$$|\mu^l - \mu^{l-1}| \leq \epsilon$$

5. A NUMERICAL EXAMPLE

In this section, a network (see Figure 4) with 40 segments (80 links) and 32 nodes (16 nodes are demand origins and destinations) is employed to test the proposed model. Each link in the network has 3 lanes and all intersections are four-arm signalized intersections.
Traffic demand of all OD pairs is set to be 100vph; length of each link is 600m; $d_{max}$ is set to be 0.9; $c_{min}$ and $c_{max}$ are set to be 60s and 120s; $s_{ik}$ is set to 1800veh/h/ln for all lanes; clearance time for any pair of conflicting traffic movements is set to be 4s; the arbitrary large positive constant number, $M$, is set to be $10^9$. In the BPR function, $k_a$, $b_a$, $k_w$ and $b_w$ are set to be 0.15, 4.0, 20 and 3.5, respectively. In the GA, the crossover probability is set to be 0.25, mutation probability is set to be 0.01, the maximum number of GA generations $n_{max}$ is set to be 100, and the population size $\alpha$ is set to be 50.

Performance of the proposed optimization model (Strategy 1) is compared with other three strategies:

Strategy 2: Conventional design (lane marking and signal timing only, no reversible lanes or turning restrictions)

Strategy 3: Left-turn restriction system only

Strategy 4: One-way street system only

Table 1 illustrates the optimized signal timings for Strategy 1 and detailed lane configuration plans for four strategies are illustrated in Figures. 5-8. Table 2 summarizes the
performance comparison results. Measures of effectives used for comparison include the
multiplier value, the average travel distance of vehicles, and the maximum, average, and
standard deviation of degrees of saturation for intersections and links.

One can observe from Table 2 that the proposed model outperforms other three strategies
in terms of enhancing the network capacity (i.e. significantly larger multiplier) due to its
advantage of properly selecting and utilizing different types of traffic management strategies.
Under the conventional design (Strategy 2), the average degree of saturation for intersections is
almost the same as in Strategy 1, while the average degree of saturation for links is much lower,
indicating that intersections comprise the major bottlenecks under Strategy 2 while the capacity
of links is not fully utilized. The proposed model can reduce the number of bottleneck
intersections and increase the capacity of intersections by reducing the number of signal phases
and using links with low degrees of saturation as detour paths. Strategy 3 can reduce the average
degree of saturation for intersections but shift worse congestion to less number of critical
intersections that carry the detoured left-turn traffic (i.e. 6, 9, 24, and 27). Strategy 4 successfully
reduces the degree of saturation at intersections and increases the network capacity; however the
degree of saturation for links and the average travel distance of vehicles increase compared with
the conventional design. Compared with Strategy 3 and Strategy 4, the proposed model yields
significantly lower standard deviation of degrees of saturation for intersections, showing its
capability of properly balancing the network load to prevent over-congestion at specific
intersections.
<table>
<thead>
<tr>
<th>Intersection</th>
<th>Phase 1 Movements</th>
<th>Duration of green (s)</th>
<th>Phase 2 Movements</th>
<th>Duration of green (s)</th>
<th>Phase 3 Movements</th>
<th>Duration of green (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>EB-T, SB-L+R</td>
<td>39.55</td>
<td>NB-L+T+R</td>
<td>72.45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>EB-L+T+R</td>
<td>72.44</td>
<td>SB-L+T</td>
<td>39.56</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>EB-L+T</td>
<td>22.81</td>
<td>NB-T+R</td>
<td>47.07</td>
<td>SB-L</td>
<td>38.12</td>
</tr>
<tr>
<td>9</td>
<td>EB-L+T+R</td>
<td>72.45</td>
<td>WB-L+R, SB-T</td>
<td>39.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>WB-T+R</td>
<td>47.07</td>
<td>EB-L</td>
<td>38.12</td>
<td>NB-L+T</td>
<td>22.81</td>
</tr>
<tr>
<td>13</td>
<td>WB-L+T</td>
<td>62.13</td>
<td>SB-T+R</td>
<td>49.87</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>WB-T+R</td>
<td>49.87</td>
<td>NB-L+T</td>
<td>62.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>WB-L+T</td>
<td>39.56</td>
<td>SB-L+T+R</td>
<td>72.44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>EB-L+T</td>
<td>39.56</td>
<td>NB-L+T+R</td>
<td>72.44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>EB-T+R</td>
<td>49.87</td>
<td>SB-L+T</td>
<td>62.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>EB-L+T</td>
<td>62.13</td>
<td>NB-T+R</td>
<td>49.87</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>WB-L</td>
<td>38.12</td>
<td>EB-T+R</td>
<td>47.07</td>
<td>SB-L+T</td>
<td>22.81</td>
</tr>
<tr>
<td>24</td>
<td>WB-L+T+R</td>
<td>72.45</td>
<td>EB-L+R, NB-T</td>
<td>39.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>WB-L+T</td>
<td>22.81</td>
<td>NB-L</td>
<td>38.12</td>
<td>SB-T+R</td>
<td>47.07</td>
</tr>
<tr>
<td>26</td>
<td>WB-L+T+R</td>
<td>72.44</td>
<td>NB-L+T</td>
<td>39.56</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>WB-T, NB-L+R</td>
<td>39.55</td>
<td>SB-L+T+R</td>
<td>72.45</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: L, T, and R represent left-turn, through movement, and right-turn, respectively.
Table 2 Performance comparison of different strategies

<table>
<thead>
<tr>
<th>Performance indices</th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum $\mu$</td>
<td>1.029</td>
<td>0.690</td>
<td>0.562</td>
<td>0.745</td>
</tr>
<tr>
<td>Maximum degree of saturation for intersections</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>6, 7, 8, 9,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 13, 14,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical intersections</td>
<td>8, 12, 21, 25</td>
<td>15, 18, 19,</td>
<td>6, 9, 24, 27</td>
<td>6, 9, 24, 27</td>
</tr>
<tr>
<td></td>
<td>20, 21, 24,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25, 26, 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average degree of saturation for intersections</td>
<td>0.879</td>
<td>0.899</td>
<td>0.606</td>
<td>0.724</td>
</tr>
<tr>
<td>Standard deviation of degree of saturation for intersections</td>
<td>0.027</td>
<td>0.000</td>
<td>0.175</td>
<td>0.105</td>
</tr>
<tr>
<td>Maximum degree of saturation for links</td>
<td>0.505</td>
<td>0.215</td>
<td>0.258</td>
<td>0.366</td>
</tr>
<tr>
<td>Average degree of saturation for links</td>
<td>0.320</td>
<td>0.194</td>
<td>0.195</td>
<td>0.239</td>
</tr>
<tr>
<td>Standard deviation of degree of saturation for links</td>
<td>0.093</td>
<td>0.009</td>
<td>0.039</td>
<td>0.069</td>
</tr>
<tr>
<td>Average travel distance (m)</td>
<td>3680</td>
<td>3040</td>
<td>3760</td>
<td>3640</td>
</tr>
</tbody>
</table>
Figure 5 Optimized plans (Strategy 1).
Figure 6 Optimized plans (Strategy 2).
Figure 7 Optimized plans (Strategy 3).
Figure 8 Optimized plans (Strategy 4).

6. FINDINGS

A lane-based optimization model integrating signal timings and reorganization of lane configurations is developed in this paper with the objective to enhance network capacity and relieve congestion. The model features a bi-level structure with a mix-integer-non-linear-
programming problem at the upper-level and a parametric variational inequality at the lower-level. A GA-based heuristic is used to yield meta-optimal solutions to the model.

Numerical analyses are conducted to evaluate the performance the proposed model. Comparisons are made between the proposed model and other candidate strategies such as conventional design, left-turn restriction system, and one-way street system. The following finds can be made:

1) The proposed model can optimize the lane reorganization and traffic control strategies in a unified framework. Compared with using them separately, integrated traffic management strategies may further expand the network capacity and improve the operational efficiency;

2) Compared with the conventional design, the proposed model can significantly reduce the number of bottleneck intersections and increase the capacity of the network by reducing the number of signal phases at intersections and using underutilized links as detour paths;

3) Compared with left-turn restriction and one-way street strategies, the advantage of the proposed model lies in its capability of properly balancing the network load to prevent over-congestion at specific intersections.

Future work along the line will be extending the model into a dynamic setting and introduction of stochastic elements into the network flow patterns to accommodate the dynamic process of traffic control and management. Application and evaluation of the proposed model with a real-world transportation network will also be performed in the next step.

ACKNOWLEDGEMENT

The research is supported by the National Natural Science Foundation of China under Grant No.51238008 and No.51178345.

REFERENCES


33. Wong, C.K., and S.C. Wong. Lane-Based Optimization of Signal Timings for Isolated


