A STOCHASTIC OPTIMAL CONTROL APPROACH TO REAL-TIME INCIDENT-RESPONSIVE TRAFFIC SIGNAL CONTROL AT ISOLATED INTERSECTIONS

Jiuh-Biing Sheu
Assistant Professor
Institute of Traffic and Transportation
National Chiao Tung University
4F, 114 Chung Hsiao W. Rd., Sec. 1
Taipei, Taiwan 10012
R.O.C.
TEL: 886-2-2349-4963
FAX: 886-2-2349-4953
jbsheu@cc.nctu.edu.tw

ABSTRACT

Real-time incident-responsive traffic control and management is vital to development of advanced incident management systems in ITS. More importantly, it provides, from an academic point of view, the linkages between incident detection, incident management, and traffic signal control. This study explores the application of a stochastic optimal control approach to real-time incident-responsive traffic control at isolated intersections. In the methodology development, time-varying lane traffic state variables and control variables are specified to characterize section-wide inter-lane and intra-lane traffic states under conditions of lane-blocking incidents. Following specification of system states, we formulated a discrete-time nonlinear stochastic model which comprises four types of equations, namely (1) recursive equations, (2) measurement equations, (3) incident-induced delay equations, and (4) boundary constraints. From the proposed stochastic model, we then developed a stochastic optimal control algorithm to update the time-varying control variables and lane traffic state variables in real time with lane-blocking incidents at isolated intersections. To generate efficiently traffic data used in model tests, we employed an advanced microscopic traffic simulator, Paramics, Version 3.0, which is developed to model and analyze ITS traffic flow conditions. The preliminary test results indicate that the proposed method can accomplish the goal of real-time incident-responsive traffic signal control. In addition to proposing a new methodology, we hope that this study can initiate investigation into real-time incident-responsive traffic control and management to achieve the final goal of network-wide incident-responsive traffic optimal control for incident management.

1. INTRODUCTION

Real-time incident-responsive traffic control and management can be a critical stage in the development of advanced incident management systems for two reasons. First, it has been pointed out in our early related research (Sheu, Chou, and Shen, 2001) that a comprehensive incident management system should involve three primary mechanisms: (1) incident detection, (2) the prediction of incident congestion and (3) incident-responsive traffic management and control. Such an ideal architecture implies that incident-responsive traffic management and control represents a critical functionality for reducing automatically and efficiently incident impacts on traffic flows in incident management systems. Second, the development of real-time incident-responsive traffic control and management systems may provide the linkages between incident detection, incident management, and traffic signal control. From an academic point of view, incident-responsive traffic control and management can be regarded as a specific field that integrates the aforementioned areas. Such an integrated field is worth investigating since it extends substantially the applicability of the aforementioned aspects to addressing the issues related to non-recurrent traffic congestion in urban areas.

To date, many studies have been devoted to exploring advanced traffic control systems in an effort to address diverse urban traffic congestion problems. One common feature exhibited by the contemporary traffic control systems is that compared to conventional fixed time control modes, the advanced control systems seek to accommodate dynamically control strategies in response to a variety of traffic flow conditions. The most notable of these are modes of vehicle actuated control and demand-responsive control which can also serve as references for development of incident-responsive traffic control systems in the study.

Utilizing specific control logic together with special detector configurations, vehicle actuated controllers have been implemented increasingly for critical intersection control (CIC) which addresses, in particular, the control of an isolated intersection using special control strategies as well as algorithms suited to saturated flow levels (McShane and Roess, 1990). Although they perform various operational functions, one distinctive feature of vehicle actuated control is that traffic signals are actuated primarily by traffic arrivals which are measured directly by detectors placed at intersection approaches. Given parameters such as the initial interval, the vehicle interval, and the maximum interval, pre-set in control logic, vehicle actuated controllers execute specific control rules on the basis of detector observations in response to the diversity of traffic congestion conditions at an individual intersection.

There exists a tremendous variety of demand-responsive intersection control systems proposed with the evolution of computerized technologies. The well-known ones include the British SCOOT system (Hunt, Robertson, and Bretherton, 1982; Robertson and Bretherton, 1991; Bretherton, Wood, and Baker, 2000), the Australian SCATS system (Sims, 1979), the OPAC system (Gartner, 1983), and PRODYN (Henry and Farges, 1990), which are also referred to as the 3rd generation of computer-aided traffic signal control systems. The SCOOT system is extensively utilized in numerous cities around the world, and remains to be updated to improve its congestion management facilities. To date, the updated version of SCOOT has been integrated with the INGRID automatic incident detection system and the system of VAMPIRE which is a computer program to operate variable message signs to execute the mechanism of congestion and incident management. Similarly, SCATS has been installed in several major cities in Australia, New Zealand, Asia, and the US, and is increasingly drawing attention for its potential with respect to handling diverse network-wide traffic congestion problems. From a theoretical point of view, both SCOOT and SCATS tend to perform incremental optimization that makes small changes around predetermined signal plans in response to prevailing traffic conditions; while under control strategies of either OPAC or PRODYN, a binary decision in terms of phase switching is made at each short time step. In addition, a great number of researchers have recently proposed various sophisticated methodologies for improving system performance of demand-responsive control for cases of network-wide oversaturated cases (Eddelbuttel and Cremer, 1994; Abu-Lebdeh and Benekihal, 1997; Lo, 1999).

Despite remarkable progress made in formulating and solving diverse urban traffic congestion problems, the applicability of the aforementioned advanced traffic control modes to surface street incident cases remains arguable because of the following three major concerns.

First, incident effects on lane traffic maneuvers may lead to instability of the traffic flow system. Incident-induced intra-lane and inter-lane traffic maneuvers including vehicular queuing and mandatory lane changing in blocked lanes seem to be considerably different from those in incident-free cases. Consequently, such abnormal time-varying patterns of lane traffic states may precipitate more complicated traffic operational problems such as extreme congestion, gridlock, and secondary accidents. Therefore, the steady-state traffic flow condition that is extensively assumed in numerous published signal control technologies does not hold anymore. Details of these related issues can also be found in our previous research (Sheu and Ritchie, 1999 and 2001; Sheu, Chou, and Shen, 2001).

Second, although the above advanced control systems may meet expectations better than traditional fixed-time control systems, their functionality in terms of responding to sudden changes in traffic flow patterns appears incomplete, particularly in case of incident-induced traffic congestion. This issue can be extensively found in those published demand-responsive control systems, including the up-to-date SCOOT system which must rely, to a great extent, on other sophisticated systems such as INGRID, ASTRID and VAMPIRE to achieve the purpose of incident management. Therefore, it does not seem to have enough evidence, at least, to prove that published advanced control technologies perform well in response to diverse incident-induced traffic congestion conditions in real time.

The last concern is that the issue on real-time prediction of incident effects on traffic congestion such as incident-induced delays and queue lengths remains unsolved, which may cast doubt on the applicability of the developed optimization-based control strategies such as SCOOT and SCATS to incident cases. Note that such objective functions as minimizing delays and queue lengths have been extensively used in developing the optimization-based control algorithms. During optimization, these objective functions are assessable only when the time-varying delays and queue lengths are predictable. That is also one of our major reasons for conducting previous research with respect to real-time prediction of incident effects on traffic congestion (Sheu, Chou, and Shen, 2001; Sheu, Chou, and Chang, 2001) in advance of developing real-time incident-responsive traffic control technologies in this study.

Besides, literature in some related fields including queue-length prediction and incident management may be also worth reviewing (Bretherton and Bowen, 1991; Rouke and Bell, 1991; Cremer and Henninger, 1993; Michalopoulos and Jacobson, 1993; Sellam and Boulmakoul, 1994;

Bretherton, Wood, and Baker, 2000; Harwood, 2000). Dedicated discussion on the early literature can be found elsewhere (Sheu and Ritchie, 1999; Sheu, 2001), and herein, is not further detailed in consideration of the scope of this study. For instance, the architecture of incident management may need to involve other elaborate technologies such as automatic incident detection (AID), route guidance, dynamic traffic assignment (DTA), and detector data acquisition, and they all need to be further elucidated.

This study aims to explore new methodology for the use of real-time incident-responsive local signal control. The primary objective of the proposed control strategy is to minimize the incident impacts on the traffic flows at an isolated intersection under conditions of lane-blocking incidents occurring either on the roadway between two successive intersections or within an intersection. To achieve the study purpose, we propose a prototype of a stochastic optimal control approach which involves modeling stochastic optimal control systems and a real-time control algorithm in response to various traffic congestion problems in case of lane-blocking incidents on surface streets. The most distinctive feature exhibited by the proposed control method is that incident-induced inter-lane and intra-lane traffic states as well as incident impacts, either in the temporal domain or in the spatial domain, on traffic congestion can be estimated in real time, and then used as the parameters in the time-varying objective function to serve specific control purposes during lane-blocking incidents at isolated intersections. Note that the applicability of the proposed method is tentatively limited to real-time incident-responsive local signal control in an attempt to prevent over-congestion at an isolated intersection under conditions of lane-blocking incidents. Correspondingly, the philosophy of local optimization is utilized. Issues with respect to network-wide incident-induced congestion events such as queue overflow and gridlock cases are not addressed in the current research scenario due to the necessity of more sophisticated technologies based on system-optimization control principles.

2. SPECIFICATION OF SYSTEM STATES

Incidents investigated in this study are classified into two categories: (1) arterial lane-blocking incidents (i.e., incidents occurring on the roadway between two adjoining intersections), and (2) intersection incidents (i.e., incidents occurring within a given intersection). In order to estimate in real time lane traffic states as well as decision variables for the two typical incident cases, specific detector configurations using pairs of point detectors are proposed as shown in Fig. 1. The area within any given pair of detector stations is defined as a detection zone in this study. Given a lane-blocking incident occurring in a detection zone, raw lane traffic data collected from the upstream and downstream detector stations of the detection zone are employed for traffic state estimation as well as stochastic signal optimal control in the proposed method. Therefore, as shown in Fig. 1, two successive detection zones are set for each link of the intersection in response to arterial incidents and intersection incidents on a given link.

Fig. 1. Illustration of the proposed detector configurations

Time-varying raw lane traffic data collected from any given pair of point detector stations (upstream and downstream detectors) are the input data of the proposed approach. Given the aforementioned detector configurations, lane traffic data collected from any pair of detector stations at each time step are used for system state estimation, where a time step corresponds to a time unit of data sampling. In this study, 10-sec. traffic count data are used in model tests.

Four groups of time-varying traffic variables are then specified in the system to characterize incident-induced inter-lane and intra-lane traffic maneuvers under real-time stochastic optimal control. They are (1) basic lane traffic states which can be estimated directly from raw traffic data collected from point detectors, (2) space-based incident impacts on traffic congestion (e.g., queue lengths in blocked lanes, and the number of moving vehicles in lanes adjacent to blocked lanes), (3) time-based incident impacts on traffic congestion (e.g., stop delays and acceleration/deceleration delays), and (4) decision variables for stochastic optimal control. They are specified in detail as follows.

Basic lane traffic states are the elements used to characterize section-wide inter-lane and intra-lane traffic maneuvers in the presence of a lane-blocking incident, and to derive in particular the other groups of variables. In the study, four types of basic lane traffic states, shown as follows, are specified

1) $p_{ij}^{m}(k)$ which is the mandatory lane-changing fraction from blocked lane i to adjacent lane j in link m at time step k;
2) $r_{j}^{m}(k)$ corresponding to the proportion of vehicles present in adjacent lane j of link m which can pass the detector station downstream to the detection zone at time step k;
3) $r_{l}^{m}(k)$ representing the proportion of vehicles present in independent lane l of link m which can pass the detector station downstream to the detection zone at time step k, where an independent lane is referred to as a lane which is neither the blocked lane nor the lane adjacent to the blocked lane; and

4) $r_{ij}^{m}(k)$ corresponding to the proportion of vehicles conducting lane changes from blocked lane i to adjacent lane j which pass the downstream detector in adjacent lane j at time step k.

In the aforementioned lane traffic state variables, subscriptions i, j and l are positive integers representing lane codes, and their values range from 1 to N, where N corresponds to the total number of lanes within the detection zone; k is a non-negative integer ranging from 0 to infinity; subscription m represents a code for a given link connecting to the targeted intersection.

Two groups of time-varying lane traffic variables are primarily involved in the system to characterize space-based incident impacts on traffic congestion. They are (1) queue lengths in blocked lanes, and (2) lane traffic loads in either adjacent lanes or independent lanes.

Time-varying queue lengths specified in the system can be further classified into three types of variables. They are (1) the number of vehicles queuing in blocked lane i of link m at time step k ($q_i^m(k)$), (2) the number of vehicles queuing in adjacent lane j of link m at time step k during red intervals ($q_j^m(k)$), and (3) the number of vehicles queuing in independent lane l of link m at time step k during red intervals ($q_l^m(k)$). Utilizing the aforementioned basic lane traffic states, we can denote $q_i^m(k)$, $q_j^m(k)$ and $q_l^m(k)$ respectively by:

$$q_i^m(k) = \left[q_i^m(k \mid k-1) + a_i^m(k)\right]\left[1 - \sum_{\forall j \in J} p_{ij}^m(k)\right] \tag{1}$$

$$q_j^m(k) = q_j^m(k \mid k-1) + a_j^m(k) \tag{2}$$

$$q_l^m(k) = q_l^m(k \mid k-1) + a_l^m(k) \tag{3}$$

where $a_i^m(k)$, $a_j^m(k)$, and $a_l^m(k)$ represent the lane traffic counts collected from the upstream detectors in blocked lane *i*, adjacent lane *j*, and independent lane l, respectively at time step *k*; $q_i^m(k|k-1)$, $q_j^m(k|k-1)$, and $q_l^m(k|k-1)$ correspond to the queue length in blocked lane *i*, adjacent lane *j*, and independent lane *l*, respectively at the beginning of time step *k*.

Compared to queue lengths which characterize the intra-lane traffic states in terms of static vehicles, lane traffic loads which are defined as the numbers of vehicles, excluding queuing vehicles, moving in either adjacent lanes or independent lanes within a detection zone characterize the intra-lane traffic states in terms of moving vehicles. In order to distinguish the lane traffic loads of adjacent lanes ($\text{¢}_j^m(k)$) which are influenced considerably by the traffic conditions in blocked lanes from the lane traffic loads in independent lanes ($\text{¢}_l^m(k)$), they are expressed respectively as:

$$\text{¢}_j^m(k)=[\text{¢}_i^m(k|k-1)+a_i^m(k)]\times p_{ij}^m(k)\times[1-r_{ij}^m(k)]+[\text{¢}_j^m(k|k-1)+a_j^m(k)]\times[1-r_j^m(k)] \quad (4)$$

$$\text{¢}_l^m(k)=[\text{¢}_l^m(k|k-1)+a_l^m(k)]\times[1-r_l^m(k)] \quad (5)$$

where $\text{¢}_j^m(k)$ and $\text{¢}_l^m(k)$ represent the lane traffic loads in adjacent lane *j* and in independent lane *l*, respectively on link *m* at time step *k*, and similarly, $\text{¢}_j^m(k|k-1)$ and $\text{¢}_j^m(k|k-1)$ are their early estimates, respectively, at time step *k*-1.

In contrast to the space-based incident impacts on traffic congestion including the aforementioned time-varying traffic states of queue lengths and traffic loads, delays can be regarded as significant variables indicating the incident impacts on traffic congestion in the temporal domain. In the study, three major types of delays are specified. They are (1) the stopped delay caused by either the vehicles queuing in blocked lane *i* or in red intervals, (2) the approaching delay caused by either vehicular lane changing from blocked lane *i* to adjacent lane *j* or queuing in blocked lane *i*, and (3) travel-time delay in any independent lane. The following details the denotations of these delays.

In the presence of a surface street lane-blocking incident, the stopped delay can be the result of either the lane blockage or the red interval, and thus, four variables of time-varying stopped delays are specified:

$$d_i^m(k) = t \tag{6}$$

$$\cdot_i^m(k) = R_i^m(k) \tag{7}$$

$$\cdot_j^m(k) = R_j^m(k) \tag{8}$$

$$\cdot_l^m(k) = R_l^m(k) \tag{9}$$

where $d_i^m(k)$ corresponds to the stopped delay caused by the queuing vehicles in blocked lane *i* of link *m* at time step *k*; $\cdot_i^m(k)$, $\cdot_j^m(k)$, and $\cdot_l^m(k)$ represent the stopped delays in blocked lane *i*, adjacent lane *j*, and independent lane *l*, respectively on link *m* at time step *k* during a given red interval; $R_i^m(k)$, $R_j^m(k)$, and $R_l^m(k)$ correspond to the lengths of the time-varying red intervals associated with blocked lane *i*, adjacent lane *j*, and independent lane *l*, respectively on link *m* at time step *k*; and *t* represents the length of a time step.

The approaching delays are classified into the following three categories:

1) $d_{ii}^{m}(k)$ corresponds to the deceleration delay caused by an unit vehicle approaching from blocked lane *i* to the end of the vehicles queuing in blocked lane *i* of link *m* at time step *k*, and is given by:

$$d_{ii}^{m}(k) = Min.\left\{\left[e_i^m - s\times q_i^m(k\,|\,k-1)\right]\left[\frac{1}{u_i^m(k)} - \frac{1}{u_i^m(0)}\right],\ t\right\} \tag{10}$$

where $e_i^m(k)$ corresponds to the distance between the upstream detector station and the incident site in blocked lane *i* on link *m*; *s* is defined as the average vehicle length; $u_i^m(k)$ represents the speed detected in blocked lane *i* of link *m* at time step *k* via the upstream detector station; and $u_i^m(0)$ corresponds to the highest speed detected in lane *i* in incident-free cases.

2) $d_{ij}^{m}(k)$ represents the approaching delay caused by an unit vehicle which conducts mandatory lane-changing from blocked lane *i* to approach the downstream detector station in adjacent lane *j* of link *m* at time step *k*, and is given by:

$$d_{ij}^{m}(k) = Min.\left\{\frac{Max.\left\{u_j^m(k),\ u_i^m(k)\right\}\times t}{Min.\left\{u_j^m(k),\ u_i^m(k)\right\}} - t + d_{mc}\ ,\quad t\right\} \tag{11}$$

where d_{mc} means the time spent by an unit vehicle on conducting mandatory lane changing behavior, and is predetermined in this study; and $u_j^m(k)$ is the speed estimated by the upstream detector station in adjacent lane *j* of link *m* at time step *k*.

3) $d_{jj}^{m}(k)$ is defined as the approaching delay caused by an unit vehicle which approaches the downstream detector station from adjacent lane *j* at time step *k*, and is given by:

$$d_{jj}^{m}(k) = Min.\left\{ \frac{u_j^m(0) \times t}{u_j^m(k)} - t \text{ , } t \right\} \tag{12}$$

Travel time delay is viewed as one of the factors in determining the stochastic optimal control strategies of the proposed method. In the study, the time-varying delay in a given independent lane l of link m at time step k ($d_l^m(k)$) is given by:

$$d_l^m(k) = Min.\left\{ \left[\frac{1}{u_l^m(k)} - \frac{1}{u_l^m(0)} \right] \times e_l^m \text{ , } \quad t \right\} \tag{13}$$

where $e_l^m(k)$ corresponds to the detector spacing of the detection zone associated with independent lane l on link m; $u_l^m(k)$ represents the speed detected in independent lane l of link m at time step k via the upstream detector station; and $u_l^m(0)$ corresponds to the highest speed detected in lane l in incident-free cases.

Time-varying decision variables are introduced to determine phase switching as well as the length of a green time interval associated with a given phase in the process of stochastic optimal control. In the system, the generalized form of the time-varying decision variable associated with a given phase a at time step k ($\Omega_a(k)$) is given by:

$$\Omega_a(k) = \frac{G_a(k)}{t} \tag{14}$$

where $G_a(k)$ is referred to as the length of green time associated with phase a at time step k. The time-varying decision variable represents, in reality, a time-varying proportion of the green time to the length of a time step, and thus, it has upper and lower bounds, namely 1 and 0, respectively. In the proposed method, the time-varying decision variable is calculated dynamically at each time step in response to incident impacts in real time.

3. SYSTEM MODELING

In order to characterize the time-varying state variables specified previously under conditions of stochastic optimal control, a discrete-time nonlinear stochastic model is proposed. The proposed stochastic model primarily comprises four groups of dynamic equations, namely (1) recursive equations, (2) measurement equations, (3) delay-aggregation equations, and (4) boundary constraints. These equations are presented respectively as follows.

3.1 Recursive Equations

The recursive equations denote the relationships between the next-time-step and current-time-step basic lane traffic states in the stochastic system on the assumption that these time-varying lane traffic states follow Gaussian-Markov processes. The generalized form of the recursive equations is given by:

$$X(k+1) = f\left[x(k-{}^{2}),\, \Omega_a(k),\, k-{}^{2}\right] + L\left[x(k-{}^{2}),\, \Omega_a(k),\, k-{}^{2}\right] w(k) \qquad (15)$$

In Eq. (15), $X(k+1)$ is a $\left[\sum_{m=1}^{M}\left(3n_j^m + n_l^m\right)\right]\times 1$ time-varying vector of basic lane traffic states at time step $k+1$, where n_j^m and n_l^m are defined as the number of adjacent lanes and the number of independent lanes on link m, respectively; M represents the total number of the links connecting to the targeted intersection; $f\left[x(k-{}^{2}), \Omega_a(k), k-{}^{2}\right]$ represents a $\left[\sum_{m=1}^{M}\left(3n_j^m + n_l^m\right)\right]\times 1$ time-varying vector of basic lane traffic states at time step k; ^{2} is referred to as a time-lag index used to estimate basic lane traffic states during the periods of phase switching, and is elucidated later in this subsection; $x(k-{}^{2})$ represents a set of state variables estimated at time step $k-{}^{2}$; $L[x(k-{}^{2}),\Omega_{a}(k),k-{}^{2}]$ is a $\left[\sum_{m=1}^{M}(3n_j^m+n_l^m)\right]\times\left[\sum_{m=1}^{M}(3n_j^m+n_l^m)\right]$ noise matrix which is dependent on basic lane traffic states as well as time-varying decision variables; $w(k)$ corresponds to a $\left[\sum_{m=1}^{M}(3n_j^m+n_l^m)\right]\times 1$ state-independent Gaussian noise vector. In the recursive equations, $X(k+1)$, $f[x(k-{}^{2}),\Omega_{a}(k),k-{}^{2}]$, $L[x(k-{}^{2}),\Omega_{a}(k),k-{}^{2}]$, and $w(k)$ can be further expressed as:

$$X(k+1)=\begin{bmatrix} p_{ij}^m(k+1) \\ r_j^m(k+1) \\ r_{ij}^m(k+1) \\ \cdots\cdots\cdots\cdots \\ r_l^m(k+1) \end{bmatrix}_{m=1,2,\ldots M} \tag{16}$$

$$f[x(k-{}^{2}),\Omega_{a}(k),k-{}^{2}]=\begin{bmatrix} p_{ij}^m(k) \\ \Omega_{a_j}(k)r_j^m(k-{}^{2}) \\ \Omega_{a_j}(k)r_{ij}^m(k-{}^{2}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \Omega_{a_l}(k)r_l^m(k-{}^{2}) \end{bmatrix}_{m=1,2,\ldots M} \tag{17}$$

$$L[x(k-{}^{2}),\Omega_{a}(k),k-{}^{2}]=\begin{bmatrix} \ell_{11}^m(k-{}^{2}) & 0 & 0 & 0 \\ 0 & \ell_{22}^m(k-{}^{2}) & 0 & 0 \\ 0 & 0 & \ell_{33}^m(k-{}^{2}) & 0 \\ 0 & 0 & 0 & \ell_{44}^m(k-{}^{2}) \end{bmatrix}_{m=1,2,\ldots M} \tag{18}$$

$$W(k) = \begin{bmatrix} w_{p_{ij}^m}(k) \\ w_{r_j^m}(k) \\ w_{r_{ij}^m}(k) \\ \cdots\cdots\cdots\cdots \\ w_{r_l^m}(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{19}$$

Note that Ω_{α_j} and Ω_{α_l} shown in Eq. (17) represent the time-varying decision variables associated with phase α under which the traffic movements in adjacent lane j and independent lane l, respectively are permitted at time step k. In $L[x(k-\tau), \Omega_\alpha(k), k-\tau]$, $\ell_{11}^m(k-\tau)$, $\ell_{22}^m(k-\tau)$, $\ell_{33}^m(k-\tau)$, and $\ell_{44}^m(k-\tau)$, take the following forms, respectively.

$$\ell_{11}^m(k-\tau) = \left[1 - \sum_{\forall j \in J} p_{ij}^m(k-\tau)\right] \times \Omega_{\alpha_j}(k) \times r_{ij}^m(k-\tau) \tag{20}$$

$$\ell_{22}^m(k-\tau) = \left[1 - \sum_{\forall j \in J} p_{ij}^m(k-\tau)\right] \times p_{ij}^m(k-\tau) + \left[1 - \Omega_{\alpha_j}(k) \times r_j^m(k-\tau)\right] \tag{21}$$

$$\ell_{33}^m(k-\tau) = \left[1 - \sum_{\forall j \in J} p_{ij}^m(k-\tau)\right] \times p_{ij}^m(k-\tau) \tag{22}$$

$$\ell_{44}^m(k-\tau) = 1 - \Omega_{\alpha_l}(k) \times r_l^m(k-\tau) \tag{23}$$

In order to deal with the issue of discontinuity of system state estimation in any signal transition step, a time-lag index τ is introduced in formulating the above-mentioned recursive equations, where a signal transition step is defined as a time step during which the traffic control signal turns either from GREEN to RED or from RED to GREEN. In $f[x(k-\tau), \Omega_\alpha(k), k-\tau]$, the value of τ is determined by one of the following four conditions.

Condition 1: If time step k+1 is a signal transition step during which the traffic signal turns from GREEN to RED (see Fig. 2), then τ=0.

Fig. 2. Signal condition (1)

Condition 2: If time step k+1 is a signal transition step during which the traffic signal turns from RED to GREEN (see Fig. 3), then $\tau=\tau_1-1$,

Fig. 3 Signal condition (2)

where τ_1 is the time lag corresponding to the number of time steps between the transition step (RED to GREEN) and the full-green time step one step prior to the last transition step (GREEN to RED), counting this previous full-green time step (e.g., τ_1=5 in Fig.3).

Condition 3: If time step k+1 is the first full-green time step of the current green interval (see Fig. 4), then, $\tau=\tau_1$.

Fig. 4. Signal condition (3)

In Fig. 4, τ_1 has the same definition as above, and is equal to 5.

Condition 4: For all other steps, τ=0.

3.2 Measurement Equations

The measurement equations denote the time-varying relationships between the measured lane traffic counts and the basic lane traffic states. The generalized form of the measurement equations is expressed as:

$$Z(k) = h[x(k)\,,\,k] + v(k) \tag{24}$$

where $Z(k)$ is a $\left[\sum_{m=1}^{M}\left(n_j^m + n_l^m\right)\right]\times 1$ time-varying measurement vector; $h[x(k),\ k]$ is a $\left[\sum_{m=1}^{M}\left(n_j^m + n_l^m\right)\right]\times 1$ time-varying vector which expresses the relationships between the measured lane traffic counts and the basic lane traffic states; $v(k)$ is a $\left[\sum_{m=1}^{M}\left(n_j^m + n_l^m\right)\right]\times 1$ Gaussian vector which represents the error terms of the collected traffic counts at time step k. $Z(k)$, $h[x(k),\ k]$ and $v(k)$ are given respectively by:

$$Z(k) = \begin{bmatrix} z_j^m(k) \\ --- \\ z_l^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{25}$$

$$h[x(k)\,,\,k] = \begin{bmatrix} h_j^m(k) \\ --- \\ h_l^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{26}$$

$$v(k) = \begin{bmatrix} v_j^m(k) \\ --- \\ v_l^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{27}$$

where $z_j^m(k)$ and $z_l^m(k)$ correspond to the lane traffic counts collected from the downstream detectors in adjacent lane j and independent lane l, respectively on link m at time step k; $v_j^m(k)$ and

$v_l^m(k)$ are the Gaussian error terms associated with $z_j^m(k)$ and $z_l^m(k)$, respectively; $h_j^m(k)$ and $h_l^m(k)$ denote the components of $z_j^m(k)$ and $z_l^m(k)$, respectively, and can be further expressed as:

$$h_j^m(k) = \left\{\left(a_i^m(k) + \hat{c}_i^m(k \mid k-1)\right) \times p_{ij}^m(k) \times r_{ij}^m(k) + \left[\left(a_j^m(k) + \hat{c}_j^m(k \mid k-1)\right) \times r_j^m(k)\right]\right\} \quad (28)$$

$$h_l^m(k) = \left[a_l^m(k) + \hat{c}_l^m(k \mid k-1)\right] \times r_l^m(k) \quad (29)$$

3.3 Delay-Aggregation Equations

The delay-aggregation equations govern the mechanism of updating time-based incident impacts under the circumstances of stochastic optimal control by estimating aggregated delays in real time. The generalized form of the delay-aggregation equations is given by:

$$D(k) = G[x(k), k]Y(k) \quad (30)$$

where $D(k)$ is a $\left\{\sum_{m=1}^{M}\left[3n_i^m + 3n_j^m + 2n_l^m\right]\right\} \times 1$ time-varying aggregated delay vector in which each element corresponds to the aggregate associated with a given delay variable shown in Vector $Y(k)$; $G[x(k), k]$ is a $\left\{\sum_{m=1}^{M}\left[3n_i^m + 3n_j^m + 2n_l^m\right]\right\} \times \left\{\sum_{m=1}^{M}\left[3n_i^m + 3n_j^m + 2n_l^m\right]\right\}$ time-varying traffic matrix in which each element represents the number of vehicles associated with a given type of delay; $Y(k)$ is a $\left\{\sum_{m=1}^{M}\left[3n_i^m + 3n_j^m + 2n_l^m\right]\right\} \times 1$ time-varying disaggregated delay vector which comprises all types of the specified delay variables. $D(k)$, $G[x(k), k]$, and $Y(k)$ can be expressed respectively as:

$$D(k)=\begin{bmatrix} \tilde{d}_i^m(k) \\ \tilde{\cdot}_i^m(k) \\ \tilde{\cdot}_j^m(k) \\ \tilde{\cdot}_l^m(k) \\ ---- \\ \tilde{d}_{ii}^m(k) \\ \tilde{d}_{ij}^m(k) \\ \tilde{d}_{jj}^m(k) \\ ---- \\ \tilde{d}_l^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{31}$$

$$G[x(k),k]=\begin{bmatrix} g_{11}^m(k) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & g_{22}^m(k) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_{33}^m(k) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & g_{44}^m(k) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_{55}^m(k) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & g_{66}^m(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_{77}^m(k) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{88}^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{32}$$

$$Y(k)=\begin{bmatrix} \Omega_{a_i} d_i^m(k) \\ [1-\Omega_{a_i}(k)]\cdot {}_i^m(k) \\ [1-\Omega_{a_j}(k)]\cdot {}_j^m(k) \\ [1-\Omega_{a_l}(k)]\cdot {}_l^m(k) \\ ---- \\ \Omega_{a_i} d_{ii}^m(k) \\ \Omega_{aj} d_{ij}^m(k) \\ \Omega_{a_j} d_{jj}^m(k) \\ ---- \\ \Omega_{a_l} d_l^m(k) \end{bmatrix}_{m=1,2,\ldots M} \tag{33}$$

where the element in $G[x(k), k]$ can be further expressed as:

$$g_{11}^m(k) = q_i^m(k\,|\,k-1) \times \left[1 - p_{ij}^m(k)\right] \quad (34)$$

$$g_{22}^m(k) = q_i^m(k\,|\,k-1) + a_i^m(k) \quad (35)$$

$$g_{33}^m(k) = q_j^m(k\,|\,k-1) + a_j^m(k) \quad (36)$$

$$g_{44}^m(k) = q_l^m(k\,|\,k-1) + a_l^m(k) \quad (37)$$

$$g_{55}^m(k) = a_i^m(k) \times \left[1 - p_{ij}^m(k)\right] \quad (38)$$

$$g_{66}^m(k) = \left[q_i^m(k\,|\,k-1) + a_i^m(k)\right] \times p_{ij}^m(k) \quad (39)$$

$$g_{77}^m(k) = \left[\cent_j^m(k\,|\,k-1) + a_j^m(k)\right] \times \left[1 - r_j^m(k)\right] \quad (40)$$

$$g_{88}^m(k) = \left[\cent_l^m(k\,|\,k-1) + a_l^m(k)\right] \times \left[1 - r_l^m(k)\right] \quad (41)$$

3.4 Boundary Constraints

In order to yield feasible solutions efficiently in the procedures of real-time estimation of system states, four boundary conditions are incorporated in the proposed model, and denoted by four boundary constraints. The generalized forms of the boundary constraints are shown as follows:

$$\mathbf{0} \le X(k+1) \le \mathbf{1} \quad (42)$$

$$\mathbf{0} \le Y(k) \le t \cdot \mathbf{1} \quad (43)$$

$$\mathbf{0} \le \Omega_a(k) \le 1, \ \forall\, a \quad (44)$$

$$T_{g,\min} \le \sum_{\pounds=0}^{n_a} \left[\Omega_a(k+\pounds) \times t\right] \le T_{g,\max} \quad (45)$$

where $X(k+1)$ and $Y(k)$ are, as indicated by Eqs. (16) and (33), vectors of basic lane traffic states and disaggregated delays, respectively; $\Omega_a(k)$, as defined previously, is a time-varying decision variable associated with a given phase a; n_a is the total number of the sequential time steps which belong to a given phase a; $T_{g,\min}$ and $T_{g,\max}$ represent the minimum and maximum intervals of green time, respectively.

4. REAL-TIME STATE ESTIMATION AND STOCHASTIC CONTROL

The primary purpose of the proposed incident-responsive traffic signal control is to minimize the incident impacts on traffic flows at the isolated intersection, and correspondingly, to minimize the differences between the ideal and the estimated values of the basic lane traffic states in the presence of a lane-blocking incident at the intersection. Therefore, we have the objective function ($\P$):

$$\P = \min E\left\{\sum_{k=0}^{N}\left[X(k)-X^{*}(k)\right]^{T} Q_1(k)\left[X(k)-X^{*}(k)\right]+\left[\Omega(k)-\Omega^{*}(k)\right]^{T} Q_2(k)\left[\Omega(k)-\Omega^{*}(k)\right]\right\} \tag{46}$$

where $Q_1(k)$ and $Q_2(k)$ represent the $\left[\sum_{m=1}^{M}\left(3n_j^m+n_l^m\right)\right]\times\left[\sum_{m=1}^{M}\left(3n_j^m+n_l^m\right)\right]$ time-varying diagonal, positive-definite weighting matrix associated with the estimation vector of the basic lane traffic states ($X(k)$), and that of the decision variables ($\Omega(k)$), respectively; N corresponds to the total number of time steps in terms of incident duration, and is pre-determined in the study; from a theoretical point of view, N can be estimated via other external technologies such as incident prediction models (i.e., methods used to characterize incidents), and practically, the condition that the estimate of N is not less than its real value should hold in the optimization process; $X^{*}(k)$ and $\Omega^{*}(k)$ are the time-varying target vectors associated with $X(k)$ and $\Omega(k)$, respectively. Note that each element in $X^{*}(k)$ represents the ideal value of a given basic lane traffic state that can relieve incident-induced traffic congestion to the greatest extent, and conveniently, it is set to be 1 in this study; the elements of $\Omega^*(k)$ are set to be the same as the elements of $\Omega(k-1)$ to serve the purpose of minimizing the cost caused by signal phase switching.

To perform the functionality of real-time incident-responsive traffic control utilizing the proposed stochastic model, we develop a stochastic optimal control algorithm. The primary computational steps involved in the proposed algorithm include (1) an extended Kalman filter, (2) truncation and normalization, (3) incident impact prediction, and (4) calculation of time-varying decision variables. The following steps summarize the proposed recursive estimation logic.

Step 0. Initialize system states and the input raw traffic data. Given $k = 0$ and ${}^{2} = 0$, system states including (1) the basic lane traffic states $X(0\,|\,0)$, (2) the covariance matrix of the state estimation error $\Phi(0\,|\,0)$, and (3) the weighting matrix $Q_1^{\tilde{a}^*}(0)$ are initialized. In addition, let the phase with green time ($\tilde{a}^*$) be assigned to the lane group which involves blocked lane i by setting $\Omega_{\tilde{a}^*}(0) = t$. Note that the total number of phasing (Θ) together with lane groups associated with the specific phases are also pre-specified in this step.

Step 1. Compute prior estimates of lane traffic state variables ($X(k+1\,|\,k)$) and the covariance matrix of the state estimation error ($\Phi(k+1\,|\,k)$) respectively by:

$$X(k+1\,|\,k) = f\left[x(k-{}^{2}), \Omega_{\tilde{a}^*}(k), k-{}^{2}\right] \tag{47}$$

$$\Phi(k+1\,|\,k) = F(k)\Phi(k\,|\,k)F^T(k) + L\left[x(k-{}^{2}), \Omega_{\tilde{a}^*}(k), k-{}^{2}\right]Q_1^{\tilde{a}^*}(k)L\left[x(k-{}^{2}), \Omega_{\tilde{a}^*}(k), k-{}^{2}\right]^T \tag{48}$$

where matrix $F^T(k)$ is the transpose matrix of $F(k)$; $F(k)$ is given by:

$$F(k) = \frac{\partial f[x(k-{}^{2}), \Omega_{\tilde{a}^*}(k), k-{}^{2}]}{\partial X(k-{}^{2})}\Bigg|_{x(k-{}^{2})=x(k-{}^{2}|k-{}^{2})} \tag{49}$$

Step 2. Calculate the Kalman gain by:

$$K(k+1) = \Phi(k+1 \mid k)H^T(k+1)\left[H(k+1)\Phi(k+1 \mid k)H^T(k+1) + Q_2(k+1)\right]^{-1} \tag{50}$$

where $Q_2(k+1)$ is pre-specified in the algorithm based on the covariance matrix of $v(k+1)$; and $H(k+1)$ is denoted by:

$$H(k+1) = \frac{\partial h\left[x(k+1 \mid k),\, k+1\right]}{\partial X(k+1 \mid k)} \tag{51}$$

Step 3. Update the prior estimates of the basic lane traffic states ($X(k+1 \mid k+1)$) by:

$$X(k+1 \mid k+1) = X(k+1 \mid k) + K(k+1)\Delta Z(k+1 \mid k) \tag{52}$$

where $\Delta Z(k+1|k)$ is given by:

$$\Delta Z(k+1 \mid k) = Z(k+1) - h\left[x(k+1 \mid k)\,,\, k+1\right] \tag{53}$$

Step 4. Truncate the estimates of disaggregated delays and basic lane traffic states variables by employing boundary constraints.

Step 5. Normalize mandatory lane-changing fractions such that:

$$\sum_{\forall j \in J} p_{ij}^m(k+1) \quad \leq \quad 1 \tag{54}$$

Step 6. Update the covariance matrix of the state estimation error ($\Phi(k+1 \mid k+1)$) as:

$$\Phi(k+1 \mid k+1) = \left[I - K(k+1)H(k+1)\right]\Phi(k+1 \mid k) \tag{55}$$

Step 7. Update the time-based and space-based incident impacts using the estimates of basic lane traffic states at the end of time step $k+1$.

Step 8. Determine the phase number a^* with green time, and associate a^* with the time-varying weighting matrix $Q_1(k+1)$ (i.e., $Q_1^{a^*}(k+1)$). In this step, each lane group associated with a given phase a is scanned. $Q_1^{a^*}(k+1)$ and a^* are then determined by the following rule:

$$Q_1^{a^*}(k+1)=\begin{bmatrix} \pm_1^{a^*}(k+1) & 0 & 0 & 0 & \cdots \\ 0 & \pm_1^{a^*}(k+1) & 0 & 0 & \cdots \\ 0 & 0 & \pm_1^{a^*}(k+1) & 0 & \cdots \\ 0 & 0 & 0 & \pm_1^{a^*}(k+1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \qquad (56)$$

where $\pm_1^{a^*}(k+1)$ is given by:

$$\pm_1^{a^*}(k+1)=\max\left\{ \frac{q_i^a(k+1\,|\,k+1)+q_j^a(k+1\,|\,k+1)+q_l^a(k+1\,|\,k+1)}{\sum_{a=1}^{\Theta}\left[q_i^a(k+1\,|\,k+1)+q_j^a(k+1\,|\,k+1)+q_l^a(k+1\,|\,k+1)\right]}\,\right|_{a=1,2,\ldots\Theta} \left. \vphantom{\frac{q}{\sum}} \right\} \qquad (57)$$

Step 9. Calculate the decision-variable vector $\Omega(k+1)$. From the principles of stochastic optimal control, the estimates of the basic lane traffic states ($X(k+1\,|\,k+1)$) produced by the extended Kalman filter are fed back through the optimal gain matrix $E(k+1)$ to minimize the pre-specified cost function (see Eq. (46)) by:

$$\Omega(k+1)=-E(k+1)X(k+1\,|\,k+1)+\text{¦}(k+1) \qquad (58)$$

In Eq. (58), $E(k+1)$ and $\text{¦}(k+1)$ are denoted respectively by:

$$E(k+1)=\left[B^T(k+1)S(k+2)B(k+1)+Q_2(k+1)\right]^{-1}B^T(k+1)S(k+2)F(k+1) \qquad (59)$$

$$\text{¦}(k+1)=\left[B^T(k+1)S(k+2)B(k+1)+Q_2(k+1)\right]^{-1}\left[B(k+1)Q_1^{a^*}(k+1)X^*(k+1)+Q_2(k+1)\Omega^*(k+1)\right] \qquad (60)$$

where matrix $S(k+2)$ should satisfy the Riccati equation shown as follows:

$$S(k+1) = Q_1^{a^*}(k+1) + F^T(k+1)S(k+2)F(k+1) - F^T(k+1)S(k+2)B(k+1)E(k+1) \qquad (61)$$

, and matrix $B(k+1)$ can be further expressed as:

$$B(k+1) = \frac{\partial f[X(k-\tau), \Omega_a^*(k), k-\tau]}{\partial \Omega(k)} \qquad (62)$$

Step 10. Check the current phase associated with green time a^* using the boundaries of the minimum and maximum green time intervals ($T_{g,\min}$ and $T_{g,\max}$) as follows: if the sum of the green time associated with the phase a^* is less than the minimum green time $T_{g,\min}$ (i.e., $\left[\sum_{\tau=0}^{\tau^*} \Omega_{a^*}(k-\tau)\right] \times t < T_{g,\min}$), then let $\left[\sum_{\tau=0}^{\tau^*} \Omega_{a^*}(k-\tau)\right] \times t = T_{g,\min}$, and the rest of the green time at a given time step be assigned to the lane group with the second-worst traffic congestion by using Eq. (57); if $\left[\sum_{\tau=0}^{\tau^*} \Omega_{a^*}(k-\tau)\right] \times t > T_{g,\max}$, then let $\left[\sum_{\tau=0}^{\tau^*} \Omega_{a^*}(k-\tau)\right] \times t = T_{g,\max}$, and similarly assign the rest of the green time at the given time step to the lane group with the second-worst traffic congestion.

Step 11. Check incident status by conducting the following rules:

If the incident is removed, then stop the control algorithm. Otherwise, input the next-time-step raw traffic data; let the time step index $k=k+1$, and then go back to **Step 1** to continue the control algorithm.

5. NUMERICAL STUDY

The purpose of this numerical study is to initialize investigation into the feasibility of the proposed signal control method in terms of responding, in real time, to incident impacts on traffic congestion at isolated intersections. In addition to verifying the performance of the proposed incident-responsive control method, testing the capability of the proposed stochastic model with respect to estimating valid incident-induced traffic states used for incident-responsive traffic control is also important in the study. Studies with respect to calibrating and testing the proposed model in an effort to ensure the model's capability of characterizing incident-induced lane traffic states as well as incident impacts have been previously completed. Details can be found in our previous research (Sheu and Ritchie, 2001). The following focuses on presenting our preliminary evaluation of system performance in incident-responsive traffic control.

Owing to the difficulty in collecting enough real incident-related traffic data for diverse incident cases, simulation data generated from Paramics, Version 3.0 which is a microscopic traffic simulator particularly promising for modeling and analyzing Intelligent Transportation Systems (ITS) under faster-than-real-time conditions were used in the numerical study. The Paramics simulator was calibrated prior to this study. Efforts spent in evaluating, qualitatively and quantitatively, the Paramics simulator can also be found in our early related research (Abdulhai, Sheu, and Recker, 1999). It may also detail our reasons of using the Paramics simulator in the study. Furthermore, the Paramics programmer which is an application programming interface (API) for traffic modeling was used as an assistant tool for generating as well as collecting time-varying lane traffic data during incidents.

To simulate diverse lane-blocking incidents at isolated intersections, a small traffic network comprising five intersections was built via Paramics. Figure 5 illustrates graphically the scheme of the study network, where each intersection represented by a specific node in Fig. 5 was coded with an integer value for its identification. Lane-blocking incidents were mainly generated on the link between nodes 1 and 3. The output data simulated from Paramics, including lane traffic counts, lane-changing fractions, queue lengths, and delays were collected at each 10-sec. time step.

Fig. 5. The scheme of the simulation network

Seventy-two lane-blocking incident cases associated with diverse incident position on the link, the lanes blocked, and traffic flow conditions were simulated in this study. Out of the 72 lane-blocking incidents, 54 simulated incidents occurred on the main segment of the link and the rest were located at the approach. Each simulation event in the study was set to be 30 minutes, including the first 5 minutes for warming up, the next 20 minutes for incident duration, and the rest for incident removing. Table 1 summarizes the characteristics of the simulated incidents designed in the numerical study.

Table 1. Characteristics of simulated incidents designed in the numerical study

To evaluate the system performance with respect to the improvements in reducing incident impacts in the spatial domain and the temporal domain, four types of state-derived incident-impact measures including $TD(k)$, $\overline{AD}$, $SQ(k)$, and $SP(k)$ are utilized, and defined as follows. $TD(k)$ is referred to as the time-varying system delay at a given time step k, and given by the sum of all the elements of $D(k)$ (see Eq. (31)). AD is defined as the average system delay during a given incident, and given by:

$$\overline{AD} = \frac{\sum_{k=1}^{N} TD(k)}{\sum_{k=1}^{N} A(k)} \tag{63}$$

where $A(k)$ represents the sum of lane traffic counts collected at the upstream detector station at time step k; N, as defined previously, is the incident duration. In contrast with $TD(k)$ and $\overline{AD}$, $SQ(k)$ and $SP(k)$ are two space-based incident-impact indexes, and given respectively by:

$$SQ(k)=\sum_{m=1}^{M}\left\{\sum_{j=1}^{J}\left[\hat{c}_j^m(k|k-1)-r_j^m(k)\times[d_j^m(k)+\hat{c}_j^m(k|k-1)]\right]+\sum_{l=1}^{L}\left[\hat{c}_l^m(k|k-1)-r_l^m(k)\times[d_l^m(k)+\hat{c}_l^m(k|k-1)]\right]\right\} \quad (64)$$

$$SP(k)=\frac{SQ(k)}{A(k)}\times 100\% \quad (65)$$

where J and L represent the total number of the adjacent lanes and that of the independent lanes, respectively. It is worth mentioning that the aforementioned four types of incident-impact measures were proposed elsewhere (Sheu et al., 2001) for the use of real-time incident impact prediction, and herein, they were conveniently used to evaluate the performance of the proposed control method.

In addition to using the aforementioned incident-impact indexes to measure directly the system performance under the proposed incident-responsive control, these measures were further compared with simulations under the two-phase vehicle actuated control which mimicked the full-actuated control mode with the following settings on each of two phases:

- detector setback: 100ft
- initial interval: 14 sec
- maximum green: 40 sec
- yellow change: 3 sec
- red clearance: 2 sec

Herein, $\frac{\overline{\Gamma}_{act}-\overline{\Gamma}_{inc}}{\overline{\Gamma}_{inc}}\times 100\%$ is used to indicate the relative improvements in reducing incident impacts, where $\overline{\Gamma}_{act}$ and $\overline{\Gamma}_{inc}$ represent the average values of a given incident-impact index under vehicle actuated control and incident-responsive control, respectively. The results of the comparisons are summarized in Table 2, and graphically illustrated in Fig. 6.

Table 2. Comparisons of system performance
(incident-responsive control vs. vehicle actuated control)

Fig. 6. Summary of the relative improvements in incident-impact reduction

The results indicated in Table 2 and Fig. 6, overall, revealed that the proposed control method performed better than vehicle actuated control strategies under diverse incident-induced congestion cases. This generalization is agreeable for the following reasons. It is noted that the proposed control method accommodates, in real time, signal timing to the time-varying estimates of section-wide inter-lane and intra-lane traffic states; however, the functionality of the actuated control mode relies, to a great extent, on the raw traffic data such as traffic counts and occupancies which may not be able to characterize appropriately incident-induced lane traffic states at the incident site. Moreover, it is hard to find any related logic rules and settings that are involved in the existing actuated control modes in response to a variety of traffic flows under incident conditions. Consequently, the vehicle actuated controllers simulated in the study failed to respond to incident cases.

Besides, some findings from our observations in the numerical tests are provided for further discussion as follows.

First, under low-volume traffic conditions, the proposed method achieves more significant reductions in incident impacts than that for high-volume incident cases. The major reason inferred from the result is that under low-volume traffic conditions, vehicles queuing in blocked lanes may conduct lane changing more easily than that in high-volume cases during the phase with green time, and thus, the proposed control strategy that assigns dynamically green time to the lane group with blocked lanes in response to time-varying changes in incident impacts seems to perform better in low-volume incident cases than in high-volume incident cases. Such a finding may also imply that it seems very promising to well control incident impacts via appropriate incident-responsive control strategies right after the occurrence of an incident, and apparently, the fulfillment of this idea must rely on other related technologies such as real-time automatic incident detection (AID) algorithms.

Second, in contrast with low-volume incident cases, traffic over-congestion caused by incidents may remain a critical issue in the proposed incident-responsive control method as well as any demand-responsive control technologies. We found that the improvements in the system performance shown in Fig. 6 turned out to be insignificant when traffic volume vacillates around the value 750 veh/hr. This may motivate our further interest in finding out the suitable range of traffic volume to which the proposed incident-responsive method can be applied. Furthermore, integrating system optimization control strategies with the instrumentality of either restricting the total entry flow or re-routing vehicles in the specific incident-affected zone warrants more research to address incident-induced over-congestion cases.

Finally, the function of real-time estimation of incident-induced intra-lane and inter-lane traffic states as well as incident impacts provided by the proposed control method is worth mentioning in the evaluation of system performance. Compared to either vehicle actuated controllers or published signal control modes, the proposed control method exhibits its unique capability in terms of elaborately characterizing incident-induced lane traffic states as well as time-varying incident impacts in the process of real-time signal control. Note that the aforementioned functionality can make available the proposed control method with benefits not only for efficiently responding to diverse incident-induced traffic congestion cases, but also for monitoring incident characteristics in real time.

6. CONCLUSIONS AND RECOMMENDATIONS

This paper has presented a stochastic optimal control-based method in response to lane-blocking incidents at isolated intersections. The proposed control approach performs incident-responsive traffic signal control by means of minimizing a time-varying function cost which is measured on the basis of comparing the real-time estimates of inter-lane and intra-lane traffic states with their ideal values. To achieve the greatest reduction of incident impacts on traffic congestion in real-time via stochastic optimal control-based technologies, we specified four groups of time-varying lane traffic variables, and then proposed a discrete-time nonlinear stochastic model as well as a real-time incident-responsive signal control algorithm.

Our preliminary test results revealed the applicability of the proposed method to real-time traffic signal control for the cases of intersection incidents as well as arterial incidents. Results presented in the study also suggested the relative advantages of the proposed control method compared with specified vehicle actuated control. More importantly, the proposed approach may indicate its potential in terms of characterizing incident-induced intra-lane and inter-lane traffic states together with incident impacts in real time in the procedure of real-time incident-responsive control, and by contrast, published advanced traffic control systems appear incomplete in providing such functionality to monitor the control performance in real time when an incident occurs.

Nevertheless, further tests as well as modifications may be necessary to verify the robustness of the proposed incident-responsive control method, and its applicability to diverse incident cases on surface streets. More complicated cases such as multi-lane-blocking incidents, queues spilling back to the upstream detectors, and incidents occurring within intersections seem to be challenging cases for further studies. Further comparison of the output of the proposed control method with that of other advanced signal control algorithms on the same basis of incident-induced traffic congestion can also help to demonstrate the potential advantages of the proposed method. Moreover, efforts on either integrating the proposed control method with automatic incident detection (AID) systems or extending it for network-wide system optimal control seem to be necessary for the development of advanced incident management systems.

Acknowledgements

This research was supported by National Science Council grants, Taiwan, NSC 90-2211-E-009-071. The author would like to thank Professor Yi-Hwa Chou and our master student Mei-Chuan Weng at the National Taiwan University, Taiwan, for their assistance in data collection and model testing. The author is also grateful to the referees for their helpful comments.

References

B. Abdulhai, J.-B. Sheu and W. W. Recker, "Simulation of ITS on the Irvine FOT Area Using the 'Paramics 1.5' Scalable Microscopic Traffic Simulator: Phase I : Model Calibration and Validation," California PATH research report UCB-ITS-PRR-99-12 1055-1425, CA, 1999.

G. Abu-Lebdeh and R. Benekihal, "Development of a Traffic Control and Queue Management Procedure for Oversaturated Arterials," presented at Transportation Research Board Annual Meeting, Washington D.C., January, 1997.

D. Bretherton, K. Wood, and K. Baker, "Congestion and Incident Management Using the SCOOT UTC System," presented at Road Transport Information and Control, IEE London, November 2000.

R. D. Bretherton and G. T. Bowen, "Incident Detection and Traffic Monitoring in Urban Areas," presented at DRIVE, Brussels, Belgium, 1991.

M. Cremer, and T. Henninger, "Estimation of Queue Lengths in Urban Road Networks," presented at the ASCE Third International Conference on Applications of Advanced Technologies in Transportation Engineering, 1993.

J. Eddelbuttel and M. Cremer, "A New Algorithm for Optimal Signal Control in Congested Networks," *Journal of Advanced Transportation* 28(3), 275-297 (1994).

N. H. Gartner, "OPAC: a Demand-Responsive Strategy for Traffic Signal Control," *Transportation Research Record*, 906, 75-81 (1983).

N. H. Gartner, "Road Traffic Control, Demand-Responsive," in *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou (ed), 386-391 (1991).

N. W. Harwood, "Incident Management Using MOLA," presented at Dynamic Control of Strategic Inter-Urban Road Networks, IEE London, November 2000.

J. J. Henry and J. L. Farges, "PRODYN," *Control, Computers, Communications in Transportation*, Oxford; New York: Pergamon Press, 253-255 (1990).

P. B. Hunt, D. I. Robertson and R. D. Bretherton, "The SCOOT On-Line Traffic Signal Optimization Technique," *Traffic Engineering and Control* 23, 190-192 (1982).

H. K. Lo, "A Novel Traffic Signal Control Formulation," *Transportation Research Part A* 33A(6), 433-448 (1999).

W. R. McShane and R. P. Roess, "Actuated Signals and Detectors," in *Traffic Engineering*, W. R. McShane and R. P. Roess (eds), 498-506, Prentice Hall, New Jersey, 1990.

P. G. Michalopoulos and R. D. Jacobson, "Automatic Incident Detection Through Video Image Processing," *Traffic Engineering & Control*, 34 (2), 66-75 (1993).

D. I. Robertson and R. D. Bretherton, "Optimizing Networks of Traffic Signals in Real-Time: the SCOOT Method," *IEEE Transactions on Vehicular Technology*, 40(1), 11-15 (1991).

S. Sellam and A. Boulmakoul, "Intelligent Intersection: Artificial Intelligence and Computer Vision Techniques for Automatic Incident Detection," *Artificial Intelligence Applications to Traffic Engineering*, 189-200 (1994).

A. Rouke and M. G. H. Bell, "Queue Detection and Congestion Monitoring Using Image Processing," *Traffic Engineering and Control* 32 (9), 412-419 (1991).

J.-B. Sheu, "A Stochastic Modeling Approach to Real-Time Prediction of Queue Overflows," Paper accepted by *Transportation Science* (2001).

J.-B. Sheu, Y.-H. Chou and A. Chen, "Stochastic Modeling and Real-Time Prediction of Incident Effects on Surface Street Traffic Congestion," Paper submitted to *Transportation Research Part C* (2001).

J.-B. Sheu, Y.-H. Chou and L.-J. Shen, "A Stochastic Estimation Approach to Real-Time Prediction of Incident Effects on Freeway Traffic Congestion," *Transportation Research Part B* 35B(6), 575-592 (2001).

J.-B. Sheu and S. G. Ritchie, "Stochastic Modeling and Real-Time Prediction of Vehicular Lane-Changing Behavior," *Transportation Research, Part B*, 35B(7), 695-716 (2001).

J.-B. Sheu and S. G. Ritchie, "A New Methodology for Incident Detection and Characterization on Surface Streets," *Transportation Research Part C 6C* (5-6), 315-335 (1999).

A. Sims, "The Sydney Coordinated Adaptive Traffic System," presented at the ASCE Engineering Foundation Conference on Research Priorities in Computer Control of Urban Traffic Systems, New York, 1979.