O.R. Applications

A sequential detection approach to real-time freeway incident detection and characterization

Jiuh-Biing Sheu [*]

Institute of Traffic and Transportation, National Chiao Tung University, 4F, 114 Chung Hsiao W. Rd., Sec. 1, Taipei, 10012 Taiwan, ROC

Abstract

In this paper, a new methodology is presented for real-time detection and characterization of freeway incidents. The proposed technology is capable of detecting freeway incidents in real time as well as characterizing incidents in terms of time-varying lane-changing fractions and queue lengths in blocked lanes, the lanes blocked due to incidents, and duration of incident, etc. The architecture of the proposed incident detection approach consists of three sequential procedures: (1) symptom identification for identification of anomalous changes in traffic characteristics probably caused by incidents, (2) signal processing for stochastic estimation of incident-related lane traffic characteristics, and (3) pattern recognition for incident detection. Lane traffic count and occupancy are two major types of input data, which can be readily collected from point detectors. The primary techniques utilized to develop the proposed method include: (1) discrete-time, nonlinear, stochastic system modeling used in the signal processing procedure, and (2) modified sequential probability ratio tests employed in the pattern recognition procedure. Off-line tests were conducted to substantiate the performance of the proposed incident detection algorithm based on simulated data generated employing the calibrated INTRAS simulation model and on real incident data collected on the I-880 freeway in Oakland, California. The test results indicate the feasibility of achieving real-time incident detection and characterization utilizing the proposed method.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Transportation; Markov–Gaussian processes; Discrete-time nonlinear stochastic system; Real-time incident detection

1. Introduction

Real-time freeway incident detection and characterization is an important function for freeway traffic management in urban areas. Studies have shown that 60% of the urban freeway delay may be caused by freeway incidents (Lindley, 1987). This may increase approximately to 70% by year 2005. Incidents on freeways interrupt traffic flows unexpectedly, and thus, they can be the major cause of such unusual events as bottlenecks and secondary accidents. It has been suggested that the risk of secondary accidents can be significantly reduced by earlier detection and warning (Busch, 1991). Clearly, earlier detection and warning are two important factors in decreasing the impact of

[*] Tel.: +886-2-2349-4963; fax: +886-2-2349-4953.
E-mail address: jbsheu@mail.nctu.edu.tw (J.-B. Sheu).

incidents on freeways. Therefore, functions of real-time automatic incident detection (AID) and incident characterization should be taken into account simultaneously in developing advanced techniques for freeway incident traffic management.

Studies dealing specifically with freeway incident detection started in the mid-1960s in the USA. The principles of published AID methods can be broadly grouped into four categories: (1) direct comparison (West, 1969; Payne, 1976; Payne and Knobel, 1976; Tignor and Payne, 1977; Payne and Tignor, 1978), (2) pattern recognition (Levin and Krause, 1978; Tsai and Case, 1979; Madanat and Cassidy, 1996), (3) data processing (Persaud and Hall, 1989; Stephanedes and Chassiakos, 1993; Cheu and Ritchie, 1995; Abdulhai and Ritchie, 1997), and (4) temporal/spatial forecast (Ahmed and Cook, 1982; Willsky et al., 1980; Cremer, 1981; Balke et al., 1996).

Early AID algorithms on freeways were developed primarily on the basis of simple comparison approaches using inductive loop data as input. California algorithms, freeway AID techniques developed for the use of real-time surveillance systems in Los Angeles, are best known (Payne, 1976; Payne and Knobel, 1976; Tignor and Payne, 1977; Payne and Tignor, 1978). A common problem with comparison approaches is the determination of thresholds employed for identification of incident occurrence. The determination of proper threshold values requires a large amount of data which limits the practical utility of comparison technologies.

Pattern recognition is another typical technique used in developing freeway AID algorithms. Levin and Krause (1978) proposed the use of Bayesian techniques for incident detection. A feature of their algorithm is that conditional probabilities of incident occurrence given various incident-occurrence signals are allowed to be computed in advance based on field data. These conditional probabilities are then provided to operators to reduce the false alarm rate. To improve false-alarm rate, Tsai and Case (1979) attempted to utilize Bayes' optimum decision rule in developing a freeway lane-blocking incident detection algorithm. Madanat and Cassidy (1996) developed a freeway incident response decision-making system

employing the technique of sequential probability ratio tests (SPRT). In their algorithm, the decision making for freeway incident detection is conducted under a condition of minimum decision cost which is determined by a dynamic cost function. Since the techniques above are probabilistic approaches, the issue associated with the determination of probability density functions in these algorithms is critical.

Recently, researchers interested in freeway incident detection have focused greater attention on techniques related to raw data processing than to AID methods. Some researchers, for example, have proposed new approaches using pre-processed data for freeway incident detection rather than raw data used in early methods. According to a study by Persaud and Hall (1989), freeway incidents could be inferred from specific patterns exhibited by 30-second data sets on a catastrophe theory surface. However, some critical issues, such as incident detection during congestion, noisy data patterns on shoulder lanes, model testing under conditions complicated by weather or geometry, were not addressed in Persaud's study. They therefore suggested that the catastrophe-theory-based method be used in parallel with other existing AID algorithms. Because short-term random fluctuations which may hamper incident detection, exist in traffic measurements, Stephanedes and Chassiakos (1993) proposed a filtering approach to smooth traffic occupancy data used in freeway incident detection. Their test results indicated that smoothing data can substantially reduce the false-alarm risk. The advantage of utilizing the data trained by artificial neural network technologies for freeway incident detection has also been suggested in several studies (Cheu and Ritchie, 1995; Abdulhai and Ritchie, 1997). However, the model retraining problem which requires a large amount of incident data for model transferability remains in any artificial neural networks based algorithms.

In terms of temporal and spatial forecast approaches, time series modeling (Ahmed and Cook, 1982) and dynamic traffic modeling (Willsky et al., 1980; Cremer, 1981) are two typical types of techniques respectively used in developing freeway incident algorithms. The performance of the above

techniques depends primarily on the validity of traffic modeling. Furthermore, these dynamic models are complicated. Any invalid traffic modeling may lead to inaccuracy in incident detection where either temporal or spatial forecast approaches are used.

This study aims to investigate a naval approach to real-time AID and characterization on freeways. To accomplish the aforementioned goal, the segment-wide inter-lane and intra-lane traffic dynamics are formulated as a discrete-time nonlinear stochastic system. The time-varying system states including lane-changing fractions and queue lengths are then estimated in real-time, and are used further for detecting incidents utilizing the extended Kalman filtering and modified sequential probability ratio test (MSPRT) techniques, respectively. In contrast to conventional incident detection methods, which may aim to detect incident occurrence (Busch, 1991), the new method presented in this paper possesses several distinctive characteristics:

(1) The proposed AID approach is capable of both detecting freeway lane-blocking incidents promptly and characterizing incidents in real time. The characteristics of incidents that are identified in the proposed approach include time-varying lane-changing fractions either from blocked lanes to adjacent lanes (upstream to incident sites) or from adjacent lanes to blocked lanes (downstream to incident sites), queue lengths in blocked lanes, the number of vehicles in each adjacent lane, the lanes blocked due to incidents, and incident duration. These incident characteristics are estimated in real time using a discrete-time nonlinear stochastic model embedded in the proposed real-time AID algorithm.

(2) Compared to most existing AID algorithms which primarily use raw traffic data, e.g., traffic counts and occupancies, as the direct input of algorithms, the proposed method utilizes time-varying lane-changing fractions and queue lengths both which are estimated in real time in the proposed algorithm using collected raw traffic data for real-time incident detection. According to our previous research experiences (Sheu and Ritchie, 2001), the time-varying mandatory lane-changing fractions from blocked lanes to adjacent lanes and

the changes of queue lengths in blocked lanes can be conceptually treated as two major indicators to promptly discriminate between incident and incident-free cases regardless of the effect of shock waves on the measures of raw traffic data collected upstream to the incident site. Therefore, they are used as the major decision variables for real-time incident detection in the proposed method.

(3) The maximum time to detection is controllable using the proposed AID algorithm. According to the principles of MSPRT, the threshold values of final decision-making vary with the collected data samples and are convergent to the same value under the condition that the input data samples reach to the pre-determined maximum sample size. Utilizing the aforementioned attribute, the proposed AID algorithm controls the maximum detection time by means of restricting the maximum sample size of the estimated lane-changing fractions.

(4) The probabilities of a miss and a false alarm in the final decision-making process can be pre-set to determine the minimum decision-making cost which dominates the performance of real-time incident detection.

2. Prerequisite setting

In this study, we utilized raw lane traffic data collected from pairs of point detectors installed on the mainline segments of freeways as the input of the proposed AID algorithm. Conveniently, only lane traffic counts and occupancies are needed in our approach. To collect the raw lane traffic data, the upstream and downstream detector stations are specified as shown in Fig. 1, where the area between the upstream and downstream detector stations is defined as the detection zone corresponding to the "scene window" in image-processing systems. Note that the definition of "detection zone" aforementioned is also applicable to cases of sequential detectors on freeways in which the sequential detection zones can be defined by pairs of sequential detectors. Based on the above detector configurations, lane-by-lane traffic data are readily obtained in each time step.
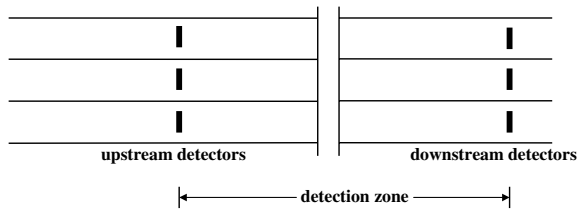
Fig. 1. Detector configurations for AID on freeways.

In any given detection zone, only one of the following two hypotheses may hold true anytime: an incident has occurred and no incident occurs. Thus, real-time incident detection can be conceptually regarded as prompt decision making to identify the true hypothesis between these two in the given detection zone using real-time traffic data samples.

Nevertheless, raw traffic data which can be readily collected from point detectors may not be sufficient for real-time identification of incident occurrence for the following reasons. The speed of shock waves remains as a critical factor in determining the time to detection in the case of direct utilization of raw traffic data. It is perceptible that the shock wave speed also depends on traffic volume conditions and detector spacing, and thus, may lead to a variety of detection performance under diverse lane-blocking incident conditions. In addition, the possibility that such events as the errors of data collection and recurrent traffic jams cause similar patterns of raw traffic data to those caused by incidents does exist in any stochastic systems, which may contribute to the difficulty in distinguishing incident-induced traffic patterns from incident-free traffic patterns based merely on collected raw traffic data.

By contrast, the alternative of utilizing lane-changing fractions and queue lengths for real-time AID is proposed in this study. Incident occurrence may immediately alter inter-lane and intra-lane traffic maneuvers such as lane changing and queuing regardless of the shock wave effect mentioned above. Accordingly, employing their distinct patterns changing in the temporal domain, time-varying lane-changing fractions from blocked lanes to adjacent lanes and the changes of queue lengths in blocked lanes are treated as two types of

significant indicators to discriminate between incident and incident-free cases in this paper. Once an incident has occurred, either lane changing fractions or queue lengths in blocked lanes may unusually increase during the incident. Thus, the technology of pattern recognition can be used to execute real-time incident detection according to the recognition of changing patterns of lane-changing fractions and queue lengths.

In addition, two assumptions are postulated to facilitate model formulation in the following sections. They are summarized as follows:

1. Each vehicle in a given blocked lane has the same lane-changing probability in a given time step. This assumption helps to deduce the unmeasured lane-changing probabilities from the measured lane-changing fractions in blocked lanes for the use of real-time incident detection in the proposed AID algorithm.
2. All state variables in the stochastic system follow homogeneous Gaussian–Markov processes. This assumption favors the setup of the recursive equations of the stochastic model proposed for real-time estimation of lane-changing fractions and queue lengths.

3. Methodology development

The entire architecture of the proposed methodology includes three sequential procedures: (1) symptom identification, (2) signal processing, and, (3) pattern recognition. Fig. 2 graphically illustrates these procedures.

The following is the description of the three procedures:

3.1. Symptom identification

This procedure aims to rapidly recognize incident symptoms with knowledge-based logical rules. Herein, incident symptoms are defined as anomalous changes of raw traffic data whose time-varying change patterns are significantly different from the patterns of incident-free cases. In this procedure, raw traffic data collected in each lane, including lane traffic counts and occupancies, are
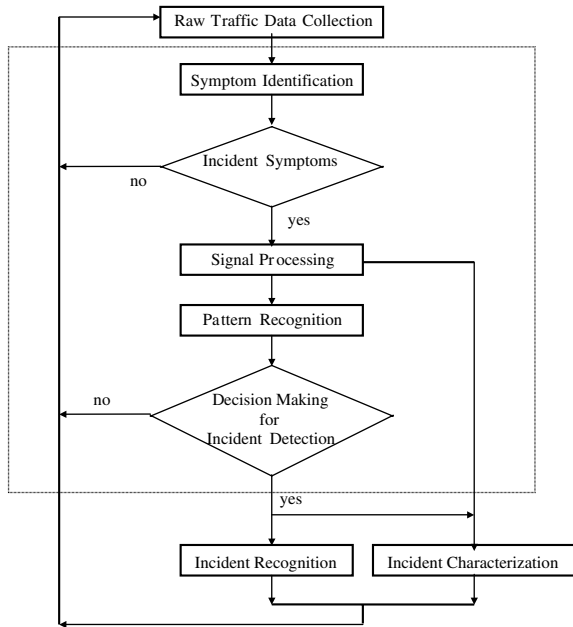
Fig. 2. Architecture of methodology.

examined in each given time step via these logistical rules. Once any pre-specified incident symptoms in a given lane are identified in a given time step, the case of incident occurrence is hypothesized, and the lane with incident symptoms is coded as a blocked lane. Meanwhile, the next two procedures, Signal processing and pattern recognition, are triggered for further identification and characterization.

Note that the aforementioned logical rules are constructed primarily on the basis of direct comparison techniques as extensively used in classical AID approaches. They are broadly classified into three groups: (1) direct comparisons of current-time-step raw traffic data to predetermined thresholds, (2) spatial comparisons of raw traffic data (e.g., comparison of upstream lane occupancy to downstream lane occupancy in a given time step), and (3) temporal comparisons of raw traffic data (e.g., comparison of current-time-step raw traffic data to previous-time-step raw traffic data at a given detector station). To construct the logical rules, the temporal and spatial relationships of lane traffic data including traffic counts and occupancies were investigated and categorized based

on the nature of lane-blocking incidents on freeways. The thresholds associated with differing incident symptoms were predetermined, and then used in the logical rules for identification of incident symptoms.

3.2. Signal processing

The procedure of signal processing performs the function of estimating time-varying lane traffic states for the use in real-time incident detection (conducted in the next procedure) and characterization. As mentioned previously, lane changing and queuing are regarded as two significant traffic characteristics for detecting and characterizing incidents in the proposed AID method. For this reason, a discrete-time, nonlinear stochastic model and a recursive estimation algorithm are proposed to estimate time-varying lane-changing fractions and queue lengths in blocked lanes in real time. The following describes the development of the stochastic model and algorithm.

The system specified for modeling lane-changing behavior during freeway incidents can be treated as an extended system of lane changing during arterial incidents (Sheu and Ritchie, 2001). The mainline segments on freeways, unaffected by either merging or diverging movements, are similar to arterials on surface streets. Without considering the weaving areas of freeways, vehicular lane-changing behavior during freeway incidents are almost the same as arterial incidents except that vehicles often return to blocked lanes from adjacent lanes after passing incident sites. When a lane-blocking incident occurs at the upstream section of the detection zone on a freeway, vehicles present in the blocked lane upstream to the incident site need to change lanes to avoid the incident. A portion of these vehicles may change lanes back to the blocked lane after passing the incident, all within the detection zone. Such abnormal lane-changing behavior, termed "return-lane-changing" in this paper, makes the model previously developed for lane-changing prediction on arterials (Sheu and Ritchie, 2001) inappropriate for freeway incident cases.

To appropriately formulate the aforementioned lane-changing maneuvers, the system within the

detection zone is divided into two subsystems: Subsystem 1, which is upstream to an incident, and Subsystem 2, which is downstream to the incident. Fig. 3 illustrates these two subsystems. Note that the above definition of subsystems also applies to the case of sequential detector stations on a freeway as mentioned in the previous section. When an incident occurs in any detection zone on a freeway, the potential lane-changing maneuvers within the detection zone are depicted using the state variables in the subsystems of the detection zone. In addition, one postulation is set to simplify the stochastic model: Drivers execute only one lane change during a given time step. This means that drivers attempting to maneuver around an obstacle blocking their lane may change lanes during a given time step, but may not return to their original lane within the same time step.

Herein, six types of state variables are specified to formulate the stochastic lane traffic system. The following formally describes these state variables:

(1) $p_{i,j}(k)$ is the lane-changing fraction from blocked lane i to adjacent lane j in Subsystem 1 in time step k, and given by

$$p_{i,j}(k) = \frac{c_{ij}(k)}{[a_i(k) + q_i(k|k-1)]}, \tag{1}$$

where $a_i(k)$ corresponds to the upstream traffic count in blocked lane i in time step k; $q_i(k|k-1)$

Fig. 3. Illustration of subsystems for freeway incident detection.

represents the number of vehicles queuing in blocked lane i in Subsystem 1 at the beginning of time step k; $c_{ij}(k)$ represents the number of lane-changing vehicles moving from blocked lane i to adjacent lane j in Subsystem 1 in time step k.

(2) $r_j(k)$ represents the proportion of the vehicles present in adjacent lane j of Subsystem 1 which can arrive at Subsystem 2 in time step k, and is given by

$$r_j(k) = \frac{a_{jj}(k)}{(a_j(k) + q_j(k|k-1))}, \tag{2}$$

where $a_{jj}(k)$ represents the traffic arrivals from adjacent lane j of Subsystem 1 to adjacent lane j of Subsystem 2 in time step k (exclusive of lane-changing vehicles from blocked lane i to adjacent lane j in time step k); the denominator on the right-hand side of Eq. (2) represents the number of vehicles in adjacent lane j of Subsystem 1 in time step k, including: (1) $a_j(k)$ which corresponds to the upstream traffic count in adjacent lane j in time step k, and (2) $q_j(k|k-1)$ which represents the number of vehicles present in adjacent lane j of Subsystem 1 at the beginning of time step k.

(3) $r_{i,j}(k)$ is the proportion of the vehicles conducting lane-changing maneuvers from blocked lane i to adjacent lane j in Subsystem 1 in time step k which can arrive at Subsystem 2 in time step k:

$$r_{i,j}(k) = \frac{a_{ij}(k)}{c_{ij}(k)}, \tag{3}$$

where $a_{ij}(k)$ is a proportion of $c_{ij}(k)$, and corresponds to the number of those lane-changing vehicles which can arrive in Subsystem 2 in time step k.

(4) $p_{j,i}(k)$ is the return-lane-changing fraction from adjacent lane j to blocked lane i in Subsystem 2 in time step k, and is given by

$$p_{j,i}(k) = \frac{a_{ji}(k)}{[a_{jj}(k) + q_{jj}(k|k-1)]}, \tag{4}$$

where $q_{jj}(k|k-1)$ represents the number of vehicles present in adjacent lane j of Subsystem 2 at the beginning of time step k; $a_{ji}(k)$ represents the number of return-lane-changing vehicles of
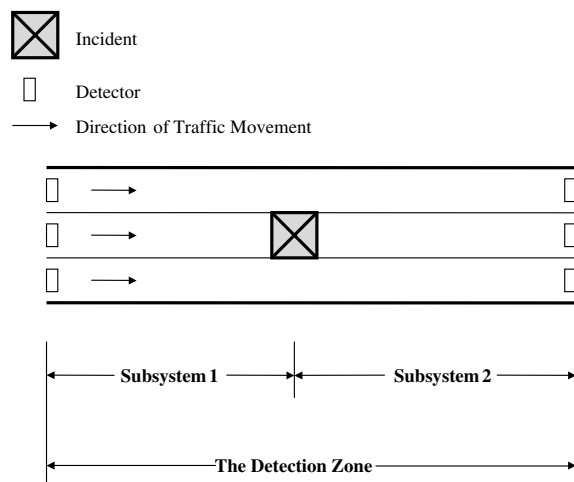
Subsystem 2 from adjacent lane j to blocked lane i in time step k.

(5) $r_{j,j}(k)$ is the proportion of the vehicles present in adjacent lane j of Subsystem 2 which can pass the downstream detector in adjacent lane j in time step k, and given by

$$r_{j,j}(k) = \frac{D_j(k)}{[a_{ij}(k) + (a_{jj}(k) + q_{jj}(k|k-1))(1 - p_{j,i}(k))]},$$
(5)

where $D_j(k)$ represents the traffic count collected from the downstream detector in adjacent lane j in time step k; the denominator on the right-hand side of Eq. (5) represents the total number of vehicles in adjacent lane j within Subsystem 2 in time step k.

(6) $r_{j,i}(k)$ represents the proportion of the vehicles in adjacent lane j of Subsystem 2 which pass the downstream detector in blocked lane i in time step k, and is given by

$$r_{j,i}(k) = \frac{D_i(k)}{[a_{jj}(k) + q_{jj}(k|k-1)]p_{j,i}(k) + q_{ii}(k|k-1)},$$
(6)

where the denominator of Eq. (6) represents the sum of the return-lane-changing vehicles from adjacent lane j to blocked lane i in Subsystem 2 in time step k and the number of vehicles present in blocked lane i in Subsystem 2 at the beginning of time step k ($q_{ii}(k|k-1)$); the numerator, $D_i(k)$, represents the number of downstream traffic count in blocked lane i in time step k.

It is noted that among the variables shown in Eqs. (1)–(6), only the upstream traffic counts ($a_i(k)$ and $a_j(k)$) and the downstream traffic counts ($D_i(k)$ and $D_j(k)$) are measurable directly from point detectors; the other variables should be derived in the estimation algorithm which is described later in this section.

Using the state variables defined above, a discrete-time nonlinear stochastic model is formulated to characterize the time-varying relationships of state variables and raw traffic data exhibited under freeway lane-blocking incident conditions. The following presents the generalized form of the proposed model, which is composed of (1) recursive equations (Eq. (7)), (2) measurement equations (Eq. (8)), and (3) boundary constraints (Eq. (9)).

$$\mathbf{X}(k+1) = \mathbf{f}[x(k),k] + \mathbf{L}[x(k),k]\mathbf{w}(k),$$
(7)

$$\mathbf{Z}(k+1) = \mathbf{h}[x(k+1),k+1] + \mathbf{v}(k+1),$$
(8)

$$0 \leqslant \forall x(k+1) \leqslant 1.$$
(9)

Eq. (7) represents a group of recursive equations indicating the relationships between the next-time-step and current-time-step state variables in the stochastic model. Each recursive equation involves a deterministic vector $\mathbf{f}[x(k),k]$ and a noise term which is decomposed into two elements as shown in matrices $\mathbf{L}[x(k),k]$ and $\mathbf{w}(k)$. In Eq. (7), the deterministic vector $\mathbf{f}[x(k),k]$ is constructed on the basis of the second assumption described in the previous section. If the noise term does not exist in the model, the state variables will follow Markov property only, and thus, the next-step state variables will depend only on the current-step state variables. This special case mimics incident-free lane-changing conditions since the model without noise terms may generate the homogeneous lane-changing fractions which are similar to the constant transitional probabilities in Worrall's lane-changing model (Worrall et al., 1970) for freeway lane-changing prediction. However, in case of incidents, the state variables are unstable and affected sharply by traffic characteristics, and thus, the noise terms of the recursive equations are developed to clarify the traffic effects in the prediction of the state variables. Herein, $\mathbf{X}(k+1)$, $\mathbf{f}[x(k),k]$, $\mathbf{L}[x(k),k]$ and $\mathbf{w}(k)$ can be further expressed respectively as

$$\mathbf{X}(k+1) = \begin{bmatrix} p_{i,j}(k+1) \\ r_j(k+1) \\ r_{i,j}(k+1) \\ p_{j,i}(k+1) \\ r_{j,j}(k+1) \\ r_{j,i}(k+1) \\ \vdots \end{bmatrix},$$
(10)

where $\mathbf{X}(k+1)$ is a $6n \times 1$ state variable vector in time step $k+1$; n is the number of the adjacent lanes in the system.

$$\mathbf{f}[x(k),k] = \begin{bmatrix} p_{i,j}(k) \\ r_j(k) \\ r_{i,j}(k) \\ p_{j,i}(k) \\ r_{j,j}(k) \\ r_{j,i}(k) \\ \vdots \end{bmatrix}, \tag{11}$$

where $\mathbf{f}[x(k),k]$ is a $6n \times 1$ deterministic vector which depends on system states.

$\mathbf{L}[x(k),k]$

$$= \begin{bmatrix} l_{11}(k) & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & l_{22}(k) & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & l_{33}(k) & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & l_{44}(k) & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & l_{55}(k) & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & l_{66}(k) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{12}$$

where $\mathbf{L}[x(k),k]$ is a $6n \times n$ noise matrix which also depends on system states; in $\mathbf{L}[x(k),k]$,

$$l_{11}(k) = \left[1 - \sum_{\forall j \in J} p_{i,j}(k) \right] r_{i,j}(k), \tag{13}$$

$$l_{22}(k) = \left[1 - \sum_{\forall j \in J} p_{i,j}(k) \right] p_{i,j}(k) + [1 - r_j(k)], \tag{14}$$

$$l_{33}(k) = \left[1 - \sum_{\forall j \in J} p_{i,j}(k) \right] p_{i,j}(k), \tag{15}$$

$$l_{44}(k) = [1 - p_{j,i}(k)] r_{j,i}(k), \tag{16}$$

$$l_{55}(k) = [1 - p_{j,i}(k)] + [1 - r_{j,j}(k)], \tag{17}$$

$$l_{66}(k) = \sum_{\forall j \in J} [1 - p_{j,i}(k)] p_{j,i}(k). \tag{18}$$

The other noise term (i.e., $\mathbf{w}(k)$) in the recursive equations is

$$\mathbf{w}(k) = \begin{bmatrix} w_{p_{i,j}}(k) \\ w_{r_j}(k) \\ w_{r_{i,j}}(k) \\ w_{p_{j,i}}(k) \\ w_{r_{j,j}}(k) \\ w_{r_{j,i}}(k) \\ \vdots \end{bmatrix}, \tag{19}$$

where $\mathbf{w}(k)$ is a $6n \times 1$ Gaussian noise vector which is independent of system states.

The measurement equations (see Eq. (8)) indicate the relationships among measurable traffic counts and state variables. Vector $\mathbf{Z}(k+1)$ represents a group of the downstream lane traffic counts measured at the end of time step $k+1$. For each given lane, the downstream lane traffic count is composed of (1) a proportion of the vehicles present in the given lane, and (2) a proportion of the lane-changing vehicles from the adjacent lane. Accordingly, these relationships are specified in the elements of vector $\mathbf{h}[x(k+1),k+1]$. In addition, a white noise vector $\mathbf{v}(k+1)$ is added in Eq. (8) to take account of the errors of collected data due to malfunction of detectors, or inaccuracy of input data. Herein, $\mathbf{Z}(k+1)$, $\mathbf{h}[x(k+1),k+1]$, and $\mathbf{v}(k+1)$ are given as follows

$$\mathbf{Z}(k+1) = \begin{bmatrix} D_i(k+1) \\ ----- \\ D_j(k+1) \\ \vdots \end{bmatrix}, \tag{20}$$

where $\mathbf{Z}(k+1)$ is a $(n+1) \times 1$ measurement vector; $D_i(k+1)$ and $D_j(k+1)$ represent the downstream lane traffic counts in blocked lane i and adjacent lane j in time step $k+1$, respectively.

$$\mathbf{h}[x(k+1),k+1] = \begin{bmatrix} h_i(k+1) \\ ----- \\ h_j(k+1) \\ \vdots \end{bmatrix}, \tag{21}$$

where $\mathbf{h}[x(k+1),k+1]$ is a $(n+1) \times 1$ deterministic vector which depends on state variables. In vector $\mathbf{h}[x(k+1),k+1]$,

$$h_i(k+1) = \sum_{\forall j \in J} \left[\left(u_j(k+1)r_j(k+1) \right. \right.$$
$$+ q_{j,j}(k+1|k) \big) p_{j,i}(k+1)$$
$$\left. + q_{i,i}(k+1|k) \right] r_{i,i}(k+1), \qquad (22)$$

$$h_j(k+1) = \left[u_i(k+1)p_{i,j}(k+1)r_{i,j}(k+1) \right.$$
$$+ \left(u_j(k+1)r_j(k+1) + q_{j,j}(k+1|k) \right)$$
$$\left. \times \left(1 - p_{j,i}(k+1) \right) \right] r_{j,j}(k+1), \quad (23)$$

where $h_i(k+1)$ and $h_j(k+1)$ respectively represent the components of the downstream traffic counts in blocked lane i and adjacent lane j in time step $k+1$; $u_i(k+1)$ is the sum of the upstream lane traffic count in blocked lane i in time step $k+1$ $(a_i(k+1))$ and the queue length in blocked lane i in Subsystem 1 $(q_i(k+1|k))$ at the beginning of time step $k+1$; $u_j(k+1)$ is the sum of the upstream lane traffic count in adjacent lane j in time step $k+1$ $(a_j(k+1))$ and the number of vehicles present in adjacent lane j in Subsystem 1 at the beginning of time step $k+1$ $(q_j(k+1|k))$.

$$\mathbf{v}(k+1) = \begin{bmatrix} v_i(k+1) \\ -\,-\,-\,-\,-\,- \\ v_j(k+1) \\ \vdots \end{bmatrix}, \qquad (24)$$

where $\mathbf{v}(k+1)$ is a $(n+1) \times 1$ Gaussian vector; the elements $v_i(k+1)$ and $v_j(k+1)$ correspond to the noise terms associated respectively with the downstream traffic counts measured in blocked lane i and adjacent lane j in time step $k+1$.

The boundary constraints of state variables (Eq. (9)) are used to restrict the state variables to the range between lower and upper bounds. In the proposed model, state variables such as lane-changing fractions and proportion variables are bounded within the range 0 and 1.

To dynamically estimate the state variables of the model, a recursive estimation algorithm is proposed. The primary computational steps involved in the estimation algorithm include: (1) an extended Kalman filter, (2) truncation and normalization, and (3) queue-length prediction. The following steps primarily summarize the proposed recursive estimation logic, and the sequence of major computational steps is shown in Fig. 4.
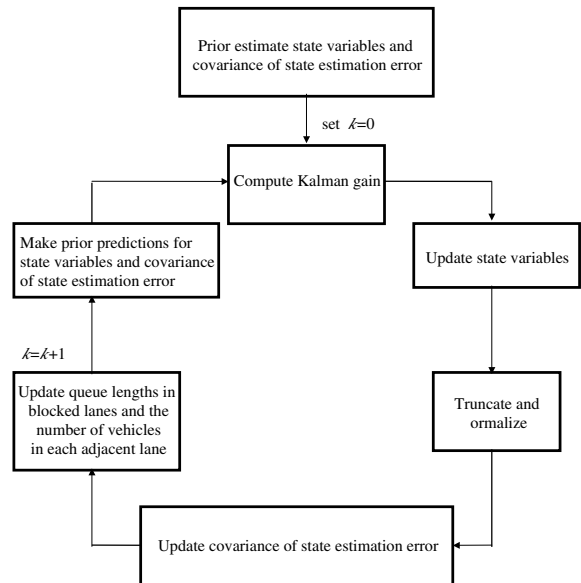

Fig. 4. Computational loop of the state estimation algorithm.

Step 0. Initialize state variables and the covariance matrix of the state estimation error. Note that the covariance matrix of the state estimation error is a significant element used to compute the Kalman gain. Through Step 1, the prior estimate of this matrix is then utilized in Step 2 for the calculation of the Kalman gain.

Step 1. Compute prior estimates of lane traffic state variables and the covariance matrix of the state estimation error. The prior estimates of state variables are herein predicted employing the recursive equations shown in Eq. (7).

Step 2. Calculate the Kalman gain. According to Kalman filtering theories, the Kalman gain is the critical element in the procedure of updating the state variables. In the proposed algorithm, the Kalman gain is updated in each time step for the real-time state estimation.

Step 3. Update the prior estimates of state variables using the current-time-step raw traffic data together with the Kalman gain calculated in the previous step.

Step 4. Truncate and normalize the estimates of state variables to meet the requirements of boundary constraints.

Step 5. Update the covariance matrix of the state estimation error.

Step 6. Predict the number of vehicles in each given adjacent lane and the queue length in the blocked lane using the estimates of lane traffic state variables.

Step 7. Input the next-time-step raw traffic data and go back to Step 1 to continue the recursive estimation procedure.

The signal processing procedure characterizes incidents, and generates decision variables for use in incident detection once incident symptoms are identified in the previous procedure (symptom identification). Among the estimates of state variables generated in the signal processing procedure, the time-varying lane-changing fraction in blocked lane i of Subsystem 1 ($p_{i,j}(k+1)$) and the queue length in blocked lane i of Subsystem 1 ($q_i(k+1)$) are used as decision variables in the next procedure (pattern recognition) for real-time incident detection. In addition, other incident-related temporal and spatial characteristics such as initialization and termination of incidents, incident duration, and the numbers of vehicles present either in blocked lanes or in adjacent lanes are recorded in the signal processing procedure for incident characterization. The above incident characteristics are updated every time step until incident symptoms do not exist.

3.3. Pattern recognition

The pattern recognition procedure conducts the decision-making process for real-time incident detection based on the estimated lane-changing fractions and queue lengths which are generated dynamically from the previous procedure (signal processing). In this procedure, we developed an AID algorithm utilizing the technique of the MSPRT which is extended from the technique of the SPRT proposed by Wald (1973). The major distinction between MSPRT and SPRT is that the maximum sample size for terminal decision-making is controllable in MSPRT and uncontrollable in SPRT. Therefore this distinct property of MSPRT is employed to control the maximum detection time in the AID algorithm. The following presents the primary steps in the AID algorithm:

Step 1: Initialize (if the iteration $k = 0$)
- pre-specify the probabilities of a false alarm P_f, a miss P_m and a detection P_d such that

$$P_f = \alpha, \quad P_m = \beta, \quad P_d = 1 - \beta, \qquad (25)$$

where

$$0 < \alpha, \quad \beta < 1;$$

- pre-determine the maximum time to detection N;
- specify time-varying thresholds $e^{g_0(k)}$ and $e^{g_1(k)}$ such that

$$g_1(k) = a\left(1 - \frac{k}{N}\right)^{r_1}, \quad a > \log(\eta_1), \qquad (26)$$

$$g_0(k) = -b\left(1 - \frac{k}{N}\right)^{r_0}, \quad b > -\log(\eta_0), \qquad (27)$$

where

$$\eta_1 = \frac{1 - \beta}{\alpha}, \quad \eta_0 = \frac{\beta}{1 - \alpha}, \qquad (28)$$

$$0 < r_0, \quad r_1 \leqslant 1, \quad a > 0, \quad b > 0; \qquad (29)$$

- set the hypotheses: $H_1 =$ an incident has occurred, and $H_0 =$ no incident occurs;
- let Z_k be a binary variable in a given time step k, indicating whether or not vehicular lane-changing behavior occurring in the given time step k, where Z_k is equal to 1 if vehicular lane changing occurs in time step k; otherwise, is set to be 0;
- set $k = 1$.

Step 2: Estimate the current-time-step lane-changing probability, $P_i(Z_k, H_1)$, and queue length, $q_i(k)$, in blocked lane i from the signal processing procedure, where

$$P_i(Z_k, H_1) = \sum_{\forall j \in J} p_{i,j}(k). \qquad (30)$$

Step 3: Compute the likelihood function of lane-changing probability, $\Lambda_i(\underline{Z_k})$, as

$$\begin{aligned} \Lambda_i(\underline{Z_k}) &= \frac{P_i(Z_k, Z_{k-1}, \ldots, Z_1 | Z_0, H_1)}{P_i(Z_k, Z_{k-1}, \ldots, Z_1 | Z_0, H_0)} \\ &= \prod_{m=1}^{k} \frac{P_i(Z_m | Z_{m-1}, H_1)}{P_i(Z_m | Z_{m-1}, H_0)}. \end{aligned} \qquad (31)$$

Step 4: Make a decision for incident detection:

IF $\quad \Lambda_i(\underline{Z_k}) > e^{g_1(k)}$,

then *say H_1 is true, and stop the algorithm*; (32)

ELSE IF $\quad \Lambda_i(\underline{Z_k}) < e^{g_0(k)}$,

then *say H_0 is true, and stop the algorithm*; (33)

ELSE IF $e^{g_0(k)} \leqslant \Lambda_i(\underline{Z_k}) e^{g_1(k)}$: (34)

IF $q_i(k) - q_i(k-1) > c$, then say H_1 is true and stop the algorithm;
ELSE consider the estimates of the next-time-step lane-changing probability and queue length for incident detection;

where c is a predetermined threshold associated with the change of queue length in a given time step.

Step 5: Set $k = k + 1$, and go to the symptom identification procedure.

It is noted that to address the issue of unknown joint probability density functions remaining in either SPRT or MSPRT, the use of time-varying joint probability density functions of lane-changing probabilities (rather than the functions of lane-changing fractions) is proposed in developing the above MSPRT-based AID algorithm. According to the first assumption postulated in the previous section, the value of the estimated lane-changing fraction in blocked lane i in a given time step ($\sum_{\forall j \in J} p_{i,j}(k)$) is consistent with the time-varying lane-changing probability ($P_i(Z_k, H_1)$), which makes Eq. (30) hold true. In addition, following the Markov property (see the assumptions mentioned previously), $P_i(Z_k, H_1)$ can be further expressed as

$$P_i(Z_k, H_1) = P_i(Z_k | Z_{k-1}, \ldots, Z_0, H_1) \qquad (35)$$
$$= P_i(Z_k | Z_{k-1}, H_1). \qquad (36)$$

Therefore, Eq. (31) shown above is derived.

For the incident-free hypothesis, the incident-free lane-changing probabilities can also be notated by replacing H_1 with H_0 in Eqs. (35) and (36). In the algorithm, these incident-free lane-changing probabilities are assumed to be identical during a given time period such as morning peak hours, afternoon peak hours, and off-peak hours. The incident-free lane-changing probabilities are pre-determined according to observations from field data. It is recommended that previous to model tests, the incident-free lane-changing probabilities associated with differing study sites should be calibrated.

4. Off-line tests and results

In this paper, the off-line tests were focused on testing the performance of the proposed method for real-time incident detection. Two data sources were utilized in the tests: (1) simulated data generated using INTRAS, a microscopic simulation model, and (2) real I-880 data. Test results with different data sources are described respectively in the following.

In the tests using simulated data, the study site was located at the westbound direction of SR-91 Riverside Freeway in Orange County, California, between SR-57 and I-5 freeways. The average daily volume at the study site is about 200,000 vehicles/day. This study site is 5 miles in length, and divided into seven sections by loop detector stations. Detection zones range from 0.34 to 1.02 miles. Fig. 5 graphically depicts the geometry of the study site.

Thirty-six simulated data sets generated from INTRAS were used for testing the freeway AID algorithm. Calibration of INTRAS is needed for simulating lane-blocking incidents on freeways. In
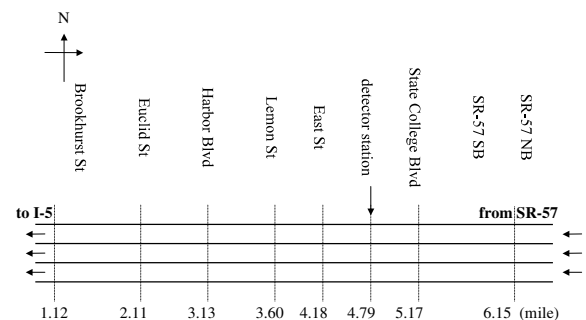

Fig. 5. The study site for freeway incident simulation.

this paper, an input file with calibrated parameters generated by Cheu et al. (1993) was directly used for freeway incident simulation. Thirty-six lane-blocking incidents associated with different incident durations, blocked lanes, and locations were simulated at the study site.

Three performance measures, detection rate (DR), false alarm rate (FAR), and time to detection (TTD) were used in these tests. In addition, four different persistence tests were conducted associated with each simulation, including (1) none-interval persistence tests (i.e., an incident is detected once the decision associated with incident-occurrence is made), (2) one-interval persistence tests (i.e., an incident is not detected until two consecutive decisions associated with incident-occurrence are made), (3) two-interval persistence tests, and (4) three-interval persistence tests. The results using four persistence tests are summarized in Table 1.

The results shown in Table 1 indicate that the proposed sequential AID algorithm is suitable for real-time incident detection on freeways. Using change patterns of lane-changing probabilities and queue lengths improves the AID performance in terms of TTD very well when compared to conventional methods. A similar study conducted earlier by Cheu and Ritchie (1995) indicated that the time to detect one-lane-blocking incidents on freeways employing neural network approaches takes 206 seconds using INTRAS simulated data at the same study site. The average TTD associated with persistence 0 tests in Table 1 is only 20.83 seconds which shows significant improvement using the proposed approach.

The following testing scenario describes the off-line AID tests using I-880 real data sets. The I-880 data were collected on a fully instrumented section of the I-880 freeway (Nimitz Freeway) in Oakland, California. The selected segment on the I-880 freeway used for data collection is 49,700 feet in length, and is located between the Marina exit ramp and Wipple exit ramp. There are a total of 18 inductive loop stations at this study site. Each detector station covers all lanes on the freeway.

In this testing scenario, 32 incident data sets were generated from the I-880 database. Out of the 32 incident data sets, 11 data sets were randomly selected for the use in modification of the symptom identification procedure of the AID algorithm. Efforts in calibrating threshold values of the logic rules used in the procedure of symptom identification are necessary to identify the specific incident symptoms on the I-880 freeway. The rest of data sets were used in off-line tests. The following describes the off-line tests and results using I-880 data.

Based on the same performance measures used above, four persistence tests were conducted using the I-880 database, which involves 21 lane-blocking incident cases. Wherein, any secondary accidents induced by a given incident are regarded as the same case as the given incident. The test results are summarized in Table 2, along with results from an earlier study by Abdulhai and Ritchie (1997) which developed AID algorithms using the Bayesian-based probabilistic neural network (PNN) approaches. The first PNN shown in Table 2 is a Bayesian-based PNN algorithm without a retraining process. The second PNN is a calibrated PNN algorithm with a retraining process using 45 I-880 data sets.

The test results shown in Table 2 suggest two advantages of using the proposed AID algorithm. First, the test performance of the proposed AID algorithm demonstrates its competitiveness with the advanced neural network incident detection algorithm developed previously, namely the PNN algorithm. Out of the 21 real incidents, only one incident was missed in the persistence-3 test for the reason that the symptoms of the incident were too weak to identify. Second, in terms of TTD, the overall performance is improved much using the proposed algorithm. The TTD values associated with the proposed AID algorithm are relatively

Table 1
Results of off-line tests for AID on freeways (based on simulated data)

Sequential AID algorithm			
Persistence	TTD (seconds)	DR (%)	FAR (%)
0	20.83	100	0
1	57.92	100	0
2	87.08	100	0
3	110.12	100	0

Table 2
Results of off-line tests for AID on freeways (based on I-880 data)

Approach	Persistence	TTD (seconds)	DR (%)	FAR (%)
Sequential AID algorithm	0	9	100	0
	1	47	100	0
	2	71	100	0
	3	90	95	0
PNN (before patterns update)	0	1080	33	0.4
	1	1180	29	0.02
	2	1218	29	0
	3	902	24	0
PNN (after patterns update)	0	15	100	0.5
	1	79	95	0
	2	112	95	0
	3	142	95	0

short, compared to either the calibrated PNN algorithm or the uncalibrated one. In the off-line tests, once the proposed AID algorithm identified incident symptoms the optimum decisions for incident detection could for the most part be made within 20 seconds.

As a whole, our preliminary test results, based either on simulated data or on field data, show the feasibility of achieving real-time incident detection and characterization on freeways using the proposed methodology which is never found in early studies. Another interesting finding in our off-line testing is that incidents are detected primarily on the basis of the change of queue-length in high-volume cases. Conversely, in low-volume cases, detection is primarily the result of lane-changing behavior. This is understandable since vehicles change lanes to pass by incidents in low-volume cases more easily than in high-volume cases. More importantly, this finding supports our insistence on using the change of queue-length and lane-changing probabilities as two decision variables for real-time incident detection.

5. Conclusions and recommendations

Despite the early technologies which have been proposed for incident detection on freeways, two crucial issues remain: (1) real-time incident detection, and (2) automatic incident characterization.

Without a clear understanding of abnormal traffic behavior, conventional methods are restricted to incident identification using raw traffic data. The result is that problems such as long detection time, unsatisfying detection performance caused by the errors of data collection, etc. continue to be a source of frustration in incident management.

This paper has presents a new AID methodology to address the above two issues. The proposed method consists of three sequential procedures including: (1) symptom identification, (2) signal processing, and (3) pattern recognition. In view of the disadvantages of directly using raw traffic data for real-time incident detection, we suggested the use of real-time lane-changing probabilities and the change of queue-length as two decision variables which are estimated based on lane traffic counts collected from point detectors. Time-varying lane-changing probabilities and queue lengths are estimated in each time step in the signal processing procedure which is conducted by the proposed stochastic lane-changing model. The estimated lane-changing probabilities and queue lengths are then used as the decision variables in the Pattern Recognition procedure which is conducted by a MSPRT-based algorithm for real-time incident detection as well as incident characterization.

The off-line tests for evaluating the performance of the proposed real-time AID algorithm were conducted on the basis of simulated and real incident data. The results not only indicate the feasibility of achieving real-time incident detection using the proposed method but also support our suggestion of utilizing lane-changing probabilities and the change of queue-length as decision variables for incident detection.

In contrast to conventional incident detection technologies, which may aim merely at identification of incident occurrence, the proposed method possesses several distinctive features summarized below:

1. In addition to incident detection, the proposed AID algorithm is capable of providing real-time incident-related information including: (1) the lanes blocked, (2) start and end time points of the incidents, (3) lane-changing fractions and queue lengths in blocked lanes, (4) the number of vehicles

in each adjacent lane, and (5) the nature of incident symptoms in terms of the relationships among volume and occupancy values. The above information provided by the algorithm is important to incident management because it may help us better understand incident characteristics and enhance the functionality of advanced incident-responsive traffic control and management systems.

2. The sequential detection approach proposed in this paper is applicable to both freeway and surface street systems. The critical distinction between freeway and surface street incident detection utilizing the proposed AID approach is the use of different lane-changing estimation models since traffic characteristics on freeways are apparently different from that on surface streets. In addition, modification of the logic rules in the symptom identification procedure is suggested.

3. The threshold values of decision-making for incident detection are time-varying and convergent. This feature controls the maximum detection time by means of restricting the sample size of lane-changing fractions needed in the pattern recognition procedure. Note that the probabilities of a miss and a false alarm for decision-making can be predetermined in the proposed AID algorithm to control the performance of real-time incident detection.

We really hope the method proposed in this paper has not only addressed the critical issues on freeway incident detection but also provided the link between incident detection and incident management. The potential for developing incident-responsive traffic control and management systems is promising. The use of more real data in either on-line or off-line tests is also necessary in our future studies.

Acknowledgements

References

Abdulhai, B., Ritchie, S.G., 1997. Development of a universally transferable freeway incident detection framework. In: Proceedings of Transportation Research Board, 76th Annual Meeting, Washington, DC.

Ahmed, S.A., Cook, A.R., 1982. Discrete dynamic models for freeway incident detection systems. Transportation Planning and Technology 7, 231–242.

Balke, K.N., Dudek, C.L., Mountain, C., 1996. Using probe-measured travel times to detect major freeway incidents in Houston, Texas. In: Proceedings of Transportation Research Board, 75th Annual Meeting, Washington, DC.

Busch, F., 1991. Incident detection. In: Papageorgiou, M. (Ed.), Concise Encyclopedia of Traffic and Transportation Systems. pp. 219–225.

Cheu, R.L., Recker, W.W., Ritchie, S.G., 1993. Calibration of INTRAS for simulation of 30-second loop detector output. In: Proceedings of the 72nd Transportation Research Board Meeting, Washington, DC.

Cheu, R.L., Ritchie, S.G., 1995. Automatic detection of lane-blocking freeway incidents using artificial neural networks. Transportation Research 3C (6), 371–388.

Cremer, M., 1981. Incident detection on freeways by filtering techniques. Preprints of 8th International Federation of Automatic Control (IFAC) World Congress, Kyoto, XVII, 96–101.

Levin, M., Krause, G.M., 1978. Incident detection: A Bayesian approach. Transportation Research Record 682, 52–58.

Lindley, J.A., 1987. Urban freeway congestion: Quantification of the problem and effectiveness of potential solutions. Journal of Institute of Traffic Engineering 57, 27–32.

Madanat, S., Cassidy, M.J., 1996. Sequential hypothesis testing based freeway incident response. In: Proceedings of Transportation Research Board, 75th Annual Meeting, Washington, DC.

Payne, H., 1976. Federal Highway Administration Work Report: Development and testing of incident-detection algorithms: vol. 1, Summary of results. Federal Highway Administration.

Payne, H., Knobel, H., 1976. Federal Highway Administration Work Report: Development and testing of incident-detection algorithms: vol. 3, Users guidelines. Federal Highway Administration.

Payne, H., Tignor, S., 1978. Freeway incident detection algorithms based on decision trees with states. Transportation Research Record 682, 30–37.

Persaud, B.N., Hall, F.L., 1989. Catastrophe theory and patterns in 30-second freeway traffic data-implications for incident detection. Transportation Research 23A (2), 103–113.

Sheu, J.-B., Ritchie, S.G., 2001. Stochastic modeling and real-time prediction of vehicular lane-changing behavior. Transportation Research—Part B 35B (7), 695–716.

Stephanedes, Y.J., Chassiakos, A.P., 1993. Freeway incident detection through filtering. Transportation Research 1C (1), 219–233.

Tsai, J., Case, R., 1979. Development of freeway incident-detection algorithms by using pattern-recognition techniques. Transportation Research Record 722, 113–116.

Tignor, S., Payne, H., 1977. Improved freeway incident-detection algorithms. Public Roads, 32–40.

Wald, A., 1973. The sequential probability ratio test for testing a simple hypothesis H_0 against a single alternative H_1. In: Wald, A. (Ed.), Sequential Analysis. pp. 37–70.

Worrall, R.D., Bullen, A.G., Gur, Y., 1970. An elementary stochastic model of lane-changing on a multilane highway. Highway Research Records 308, 1–12.

West, J., 1969. Real-time surveillance and control systems for Los Angeles. In: Proceedings of HRB Committee on Freeway Operations.

Willsky, A.S., Chow, E.Y., Gershwin, S.B., Greene, C.S., Houpt, P.K., Kurkjan, A.L., 1980. Dynamic model-based techniques for the detection of incidents on freeways. IEEE Transactions on Automatic Control 25, 347–360.