This paper has been mechanically scanned. Some errors may have been inadvertently introduced.
A Simple Detection Scheme for Delay-Inducing Freeway Incidents

Wei-Hua Lin  
Carlos F. Daganzo

April 1996
This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do no necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.
A Simple Detection Scheme for Delay-Inducing Freeway Incidents*

Wei-Hua Lin
and
Carlos F. Daganzo

Institute of Transportation Studies and
Department of Civil Engineering
University of California at Berkeley
Berkeley, CA 94720

*Research supported by PATH MOU 90, Institute of Transportation Studies, Berkeley, CA
Abstract

This paper describes a freeway incident detection scheme that does not rely on complicated theories. The procedure compares the occupancy information recorded by two neighboring loop detectors to determine whether an incident has occurred in the intervening segment, using a recipe that is directly related to an intrinsic property of delay-inducing incidents. The method, which can also signal the termination of a detected incident, assumes that any significant bottlenecks are located outside the segment in question; i.e., that the segment is “homogeneous”.

Independent of complicated theories, the proposed method can be applied with little calibration to any site despite the presence of detector errors and statistical fluctuations. It is also possible to use site-independent (default) parameters, although one should then expect a somewhat inferior performance. Tests with five sets of 30-second data from three different sites in the United States, Canada, and the Netherlands, using the recommended default parameters, were encouraging. The resulting graphical displays included in this paper clearly illustrate the scheme’s effectiveness in distinguishing non-recurrent from recurrent congestion.

Keywords: Incident Detection, Traffic Surveillance, Advanced Traveler Information Systems, Data Collection
Acknowledgement

We wish to express our gratitude to Mr. Hisham Noeimi, Mr. Karl Petty, and Mr. Daniel Rydzewski from the University of California at Berkeley for their help in providing us with the California data. We also wish to thank Professor Michael Cassidy from the University of California at Berkeley and Dr. J. M. del Castillo from the University of Seville, Spain, for generously sharing their data with us. The comments of two anonymous reviewers are also gratefully acknowledged.
1 Introduction

The literature on automatic freeway incident detection on loop detectors has grown tremendously during the past two decades and its complexity seems to be increasing. Recent proposals tend to be based on methods and ideas that have little to do with traffic such as neural networks, Bayesian analysis, Kalman filtering, fuzzy logic, and catastrophe theory. They require many parameters that are either very difficult to calibrate in the field or sensitive to traffic composition and roadway geometries.

Older, more established algorithms can be roughly divided in two categories: “threshold-based”, such as the California algorithm (Kahn, 1972) and its extensions (see a review given by Payne et al., 1975), and “dynamic-flow-prediction-based”, such as the double exponential smoothing algorithm (Cook and Cleveland, 1974) and the standard normal deviation (SND) algorithm (Dudek et al., 1974).

Threshold-based algorithms look for incidents by examining the pattern formed by the most recent groups of observations from a detector or group of detectors at each decision time. By disregarding the past, algorithms of this type cannot filter random fluctuations in the data effectively. They also tend to be sensitive to roadway geometries and traffic conditions. In practice, calibrating threshold values so as to balance the false alarm rate and the incident detection rate is tedious and strongly site-dependent. In some cases, a computer software package is required (Payne and Tignor, 1978).

Dynamic-flow-prediction-based algorithms should be more robust because they base their decisions on more past observations. Unfortunately, these algorithms are based on a questionable premise: that under normal traffic conditions the observed series of traffic counts (or occupancies) can be represented as a linear function of the present and past traffic counts. An incident alarm is usually triggered whenever the observed traffic deviates significantly from the linearly predicted one. In actuality though, flows passing a specific point of the road are governed either by the downstream or upstream traffic conditions, depending on the level of congestion; traffic counts are non-linear when congestion is growing or dissipating. This means that this type of incident detection algorithm might be able to detect transitions between congested and uncongested traffic but not distinguish between recurrent and non-recurrent congestion.

The incident detection scheme presented in this paper takes a different approach. Independent of complex theories, it is based on a comparison of cumulative occupancy data for the two detectors on both sides of a hypothetical incident. (By not relying on complicated traffic flow models, the chance that a flawed theory may corrupt the results is
eliminated.) Cumulative sums allow the past to be automatically remembered and robust results to be obtained despite random fluctuations in the data. Occupancy is preferred to flow as a system status indicator because (unlike flow) occupancy is unambiguously related to the level of congestion. Average speed could serve a similar purpose but we found it exhibited more variability in our data. The closest relative of the proposed algorithm appears to be the exponential smoothing scheme proposed in Stephanedes and Chassiakos (1993), which uses (non-cumulative) occupancy data. This approach, however, is not efficient for incidents that induce a small but lasting change in occupancy because, like threshold-based algorithms, exponential smoothing emphasizes recent observations over past ones.

2 The Proposed Scheme Under Ideal Conditions

An incident can be detected with loop detectors only if it generates a significant traffic disturbance. Thus, we shall focus our attention on incidents whose disturbance generates congestion. Ignoring incidents that do not fall in this category is not a serious omission if our main goal is to reduce congestion because the excluded incidents do not generate significant delays.

The proposed algorithm is based on an effect of delay-inducing incidents that does not arise otherwise on homogeneous road sections; i.e. the generation of a traffic pattern where the road is more crowded upstream than downstream for an extended period of time. Although this effect may also be observed in incident-free road sections containing a recurrently congested bottleneck, the effect should not arise in the homogeneous sections directly upstream or downstream of the bottleneck. Thus, the proposed scheme should be relevant for a good part of most freeways.

Before starting our discussion we recall that the occupancy recorded at a detector site over a time interval is defined as the number of vehicle-hours spent over the detector divided by the length of the interval, and that occupancy is a measure of crowdedness. [In this paper we take occupancy to be the cumulative sum of the occupancies for individual lanes; as a result, it can be greater than 1.] Since an incident causes occupancies downstream of the incident site to be significantly lower than those upstream, it is natural to use their difference as an incident indicator. More specifically, we propose to use the time series, \( \{ Z(t); t = t_1, t_2, \ldots \} \), formed by the difference in the sums of the occupancies

\footnote{We note that incident detection on inhomogeneous road segments is still possible, although somewhat less efficient. A full discussion of this can be found in Lin (1995).}
recorded at the two detectors in all the sampling intervals between the time at which incident detection is activated, $t_0$, and $t_j$. We assume that all the sampling intervals are of equal length; i.e.: $t_1 - t_0 = t_2 - t_1 = \cdots = t_j - t_{j-1} = At$. The expression for $Z(t_j)$ is then:

$$Z(t_j) = \sum_{i=1}^{j} O_u(t_i) - \sum_{i=1}^{j} O_d(t_i), \quad (j = 1, 2, \ldots, )$$

where $O_u(t_i)$ and $O_d(t_i)$ are the occupancies during time interval $(t_{i-1}, t_i)$ at the upstream and downstream detectors, respectively; i.e.,

$$O_n(t_i) = \frac{\sum_{k=1}^{k_{n,i}} t_{n,i}^k}{At},$$

where $t_{n,i}^k$ is the time the $k$th vehicle spends over detector $n$ within time interval $(t_{i-1}, t_i)$.

Note that $Z(t)\Delta t$ can be interpreted as the difference in the total number of vehicle-hours that have been spent over the upstream and downstream detectors since $t = t_0$; i.e. as the difference in the cumulative times that the upstream and downstream traffic streams have used to cover a typical vehicle length. Clearly, $Z(t)$ must increase with time until the incident is removed. Once the incident and its effects have dissipated, $Z(t)$ should stop varying with $t_i$, stabilizing at a value greater than zero which should depend on the severity and duration of the incident. By choosing a high threshold for triggering an alarm we can avoid false alarms due to random fluctuations in $Z(t)$, while remaining able to detect any incident that lasts long enough. This is the advantage of basing detection on combined information from all previous observations rather than a group of the most recent observations.

Because random fluctuations, e.g. caused by the selection of vehicle speeds/spacings by individual drivers and/or by detector errors, may cause $Z(t)$ to drift away from zero in the absence of an incident as time passes (see Figure 1) one should choose a trigger that increases with time so as to remain beyond the range of normal (incident-free) fluctuations. Although we could theorize that $Z(t)$ should vary as a random walk, and that therefore the threshold should be increased with the $\frac{1}{2}$ power of $(t - t_0)$, we choose to be more conservative and use instead a linear threshold: $\tau_0 + r \left(\frac{t - t_0}{At}\right)$. (We are more conservative in the sense that for suitable choices of $\tau_0$ and $r$, the linear threshold is higher than any $\frac{1}{2}$ power threshold for all positive $t$.) This choice prevents small incidents that would cause $Z(t)$ to drift at a rate smaller than $r$ to be detected reliably\(^2\), but also prevents

\(^2\) $\tau$ is a critical occupancy difference beyond which incidents can be reliably detected. The significance of $\tau_0$ will become more apparent later.
Figure 1: Occupancy and the cumulative sum of the difference of the occupancy vs. time for a single lane.
non-random walk fluctuations (if they were to exist in significant amounts) to generate false alarms.

In order for an incident alarm not to be triggered at time $t_j$, we require all past fluctuations of $Z(t)$ to remain within bounds; i.e.:

$$Z(t_j) - Z(t_i) \leq \tau_0 + \tau \left( \frac{t_j - t_i}{\Delta t} \right) \quad (\forall t_i \leq t_j). \quad (2)$$

Because this test requires a large number of paired comparisons we now develop an equivalent condition that avoids them.

If we let $S(t)$ be the following transformation of $Z(t)$:

$$S(t_j) = Z(t_j) - \left( t_j - t_0 \right) / \Delta t \quad (j = 1, 2, \ldots), \quad (2a)$$

condition (2) may be rewritten as follows:

$$S(t_j) - S(t_i) \leq \tau_0 \quad (\forall t_i \leq t_j; \quad j = 1, 2, \ldots); \quad (2b)$$

which in turn is equivalent to:

$$S(t_j) - \min_{t_i \leq j} S(t_i) \leq \tau_0 \quad (j = 1, 2, \ldots). \quad (2c)$$

Letting $Y(t_j)$ denote the left hand side of (2c), we note that it satisfies the recursion:

$$Y(t_{j+1}) = \begin{cases} 0 & (j = 0) \\ S(t_{j+1}) - \min\{S(t_{j+1}), S(t_j) - Y(t_j)\} & (j = 1, 2, \ldots), \end{cases}$$

since the second argument of the minimum function is $\min_{t_i \leq j} \{S(t_i)\}$. If we collect terms, the right side of this last expression becomes $\max\{0, Y(t_j) + S(t_{j+1}) - S(t_j)\}$; and if we now eliminate $S$ from the formula with (2a) the final result is:

$$Y(t_{j+1}) = \begin{cases} 0 & (j = 0) \\ \max\{0, Y(t_j) + Z(t_{j+1}) - Z(t_j) - \tau \} & (j = 1, 2, \ldots). \end{cases} \quad (3)$$

Note from (1) that $Z(t_{j+1}) - Z(t_j)$ is simply the occupancy difference recorded from the current interval $\Delta O(t_{j+1}) = O_u(t_{j+1}) - O_d(t_{j+1})$. Without any paired comparisons, an incident alarm is simply triggered whenever (3) exceeds $\tau_0$.

It should also be clear that $Y(t)$ is non-negative, and that it will be a random walk if $Z(t)$ is a random walk. The $-\tau$ term in (3) essentially introduces a negative drift relative
to $Z(t)$, and the maximum operation introduces a reflecting barrier at $Y = 0$. Thus, whenever appropriate, the theory of random walks may be used to select values of $\tau$ and $\tau_0$ that will reduce the false alarm rate to a prescribed amount. In any case, we have succeeded in constructing a time series that increases due to significant incidents (i.e. those causing $Z(t)$ to increase more than $\tau$ units per step) and that remains near zero in the absence of incidents (due to the negative drift); see Figure 2.

3 Practical Considerations

Here we concern ourselves with practical issues pertaining to the selection of $\tau$ for a freeway with $n$ lanes. These include the effects of incident magnitude and detector errors. The default values about to be proposed are based on the hypothesis that the stationary traffic states that may arise in our freeway section lie (approximately) on a triangular flow-occupancy curve such as the one shown in Figure 3.

3.1 Incident Magnitude Considerations

We show first that an incident induces a positive drift in $Z(t)$, i.e. a positive average change in $Z(t)$ per time interval, $\Delta \tau$, that is roughly equal to two thirds of the equivalent number of lanes, $n_b$, blocked by the incident; i.e.: $\Delta \tau \approx \frac{2}{3} n_b$. Note that this formula is independent of the prevailing flow.

The basis for this approximation is contained in Figure 3. It depicts a flow-occupancy curve of triangular shape with maximum flow $q_{\text{max}}$ (vph) and maximum occupancy $O_j$ where the variables are aggregated across all lanes. To derive our estimates we assume that such a curve is a good approximation for the traffic states observed over long intervals of time without an incident. We also assume that when $n_b$ lanes are blocked ($n_b < n$), the maximum allowable passing flow drops approximately from $q_{\text{max}}$ to $(1 - \frac{n_b}{n}) q_{\text{max}}$. Then, if the initial traffic is in a state with flow less than $(1 - \frac{n_b}{n}) q_{\text{max}}$, the blockage does not create a significant disruption, and the incident is not in the class of delay-reducing incidents that are the focus of this paper. On the other hand, if the initial flow is greater than $(1 - \frac{n_b}{n}) q_{\text{max}}$, then the incident will have an impact: as we said earlier, occupancy will increase upstream and decrease downstream. In as much as conservation of vehicles between the detectors ensures that the average flows over the detectors are equal to the restricted flow, we know that the average upstream and downstream occupancies must be those shown in Figure 3. We see from the geometry of the figure that the gap in


to $Z(t)$, and the maximum operation introduces a reflecting barrier at $Y = 0$. Thus, whenever appropriate, the theory of random walks may be used to select values of $\tau$ and $\tau_0$ that will reduce the false alarm rate to a prescribed amount. In any case, we have succeeded in constructing a time series that increases due to significant incidents (i.e. those causing $Z(t)$ to increase more than $\tau$ units per step) and that remains near zero in the absence of incidents (due to the negative drift); see Figure 2.

3 Practical Considerations

Here we concern ourselves with practical issues pertaining to the selection of $\tau$ for a freeway with $n$ lanes. These include the effects of incident magnitude and detector errors. The default values about to be proposed are based on the hypothesis that the stationary traffic states that may arise in our freeway section lie (approximately) on a triangular flow-occupancy curve such as the one shown in Figure 3.

3.1 Incident Magnitude Considerations

We show first that an incident induces a positive drift in $Z(t)$, i.e. a positive average change in $Z(t)$ per time interval, $\Delta \tau$, that is roughly equal to two thirds of the equivalent number of lanes, $n_b$, blocked by the incident; i.e.: $\Delta \tau \approx \frac{2}{3} n_b$. Note that this formula is independent of the prevailing flow.

The basis for this approximation is contained in Figure 3. It depicts a flow-occupancy curve of triangular shape with maximum flow $q_{\text{max}}$ (vph) and maximum occupancy $O_j$ where the variables are aggregated across all lanes. To derive our estimates we assume that such a curve is a good approximation for the traffic states observed over long intervals of time without an incident. We also assume that when $n_b$ lanes are blocked ($n_b < n$), the maximum allowable passing flow drops approximately from $q_{\text{max}}$ to $(1 - \frac{n_b}{n}) q_{\text{max}}$. Then, if the initial traffic is in a state with flow less than $(1 - \frac{n_b}{n}) q_{\text{max}}$, the blockage does not create a significant disruption, and the incident is not in the class of delay-reducing incidents that are the focus of this paper. On the other hand, if the initial flow is greater than $(1 - \frac{n_b}{n}) q_{\text{max}}$, then the incident will have an impact: as we said earlier, occupancy will increase upstream and decrease downstream. In as much as conservation of vehicles between the detectors ensures that the average flows over the detectors are equal to the restricted flow, we know that the average upstream and downstream occupancies must be those shown in Figure 3. We see from the geometry of the figure that the gap in


to $Z(t)$, and the maximum operation introduces a reflecting barrier at $Y = 0$. Thus, whenever appropriate, the theory of random walks may be used to select values of $\tau$ and $\tau_0$ that will reduce the false alarm rate to a prescribed amount. In any case, we have succeeded in constructing a time series that increases due to significant incidents (i.e. those causing $Z(t)$ to increase more than $\tau$ units per step) and that remains near zero in the absence of incidents (due to the negative drift); see Figure 2.

3 Practical Considerations

Here we concern ourselves with practical issues pertaining to the selection of $\tau$ for a freeway with $n$ lanes. These include the effects of incident magnitude and detector errors. The default values about to be proposed are based on the hypothesis that the stationary traffic states that may arise in our freeway section lie (approximately) on a triangular flow-occupancy curve such as the one shown in Figure 3.

3.1 Incident Magnitude Considerations

We show first that an incident induces a positive drift in $Z(t)$, i.e. a positive average change in $Z(t)$ per time interval, $\Delta \tau$, that is roughly equal to two thirds of the equivalent number of lanes, $n_b$, blocked by the incident; i.e.: $\Delta \tau \approx \frac{2}{3} n_b$. Note that this formula is independent of the prevailing flow.

The basis for this approximation is contained in Figure 3. It depicts a flow-occupancy curve of triangular shape with maximum flow $q_{\text{max}}$ (vph) and maximum occupancy $O_j$ where the variables are aggregated across all lanes. To derive our estimates we assume that such a curve is a good approximation for the traffic states observed over long intervals of time without an incident. We also assume that when $n_b$ lanes are blocked ($n_b < n$), the maximum allowable passing flow drops approximately from $q_{\text{max}}$ to $(1 - \frac{n_b}{n}) q_{\text{max}}$. Then, if the initial traffic is in a state with flow less than $(1 - \frac{n_b}{n}) q_{\text{max}}$, the blockage does not create a significant disruption, and the incident is not in the class of delay-reducing incidents that are the focus of this paper. On the other hand, if the initial flow is greater than $(1 - \frac{n_b}{n}) q_{\text{max}}$, then the incident will have an impact: as we said earlier, occupancy will increase upstream and decrease downstream. In as much as conservation of vehicles between the detectors ensures that the average flows over the detectors are equal to the restricted flow, we know that the average upstream and downstream occupancies must be those shown in Figure 3. We see from the geometry of the figure that the gap in
(a) Without incident

![Graph showing plots of Y(t), S(t), Z(t), and min S(i) without incident.](image)

(b) With incident

![Graph showing plots of Y(t), S(t), Z(t), and min S(i) with incident.](image)

Figure 2: Plots of $Y(t), S(t), Z(t)$, and $\min S(i)$
occupancy $AO$, is related to the maximum occupancy (which we estimate to be about $\frac{2}{3}n$ on most facilities since mean vehicle length over jam spacing is close to two thirds) and to the fractional reduction in capacity by: $\Delta O = \frac{n_n}{n}$; i.e., $\Delta O = \frac{2}{3}n_b$. This gap is the average increase in $Z(t)$ per time interval in the long run. Therefore, it increases the drift of the incident-free $Y$ by an amount $A_T = \Delta O = \frac{2}{3}n_b$, as claimed.

This means that if $T$ is set to about $0.1n$ (our recommended value) an incident that closes one lane will yield a positive drift in $Y(t)$ (which is $A_T = 0.1n$ as per Eq. (3)) if $n \leq 6$. Therefore, our choice should lead to detection; even simple “rubbernecking” may be detected.

The recipe also allows us to estimate the magnitude of an incident. In Figure 2(b) for example, which corresponds to a case with $n = 5$ and $T = 0.5$, $Y(t_j)$ increases with $j$ at a 1.6 rate after the onset of the incident. This means that $Z(t)$ increases at the rate $A_T = \frac{2}{3}n_b = 1.6 + T = 2.1$, which is approximately equivalent to a three lane blockage.

### 3.2 Detector Reading Considerations

We now turn our attention to problems with detector readings. In practice, the average occupancy measured by a loop detector can be consistently higher or lower than the occupancy recorded by another nearby detector. This causes $Z(t)$ to drift towards one side, even in the absence of incidents. Figure 4 shows a typical plot of incident-free data.
Figure 4: Difference of the cumulative occupancy vs. time (with detector error drifts) (with \( n = 5 \)) with a typical small drift (\( \sim 0.05 \)). There are two main reasons for this kind of drift. First it can be due to changes in roadway geometry, such as grades and curves, which may result in different travel speeds (and therefore occupancies) at the two locations despite the same (stationary) flows\(^3\). Second, it can be caused by detector errors. Although, this is a common occurrence\(^4\), it is often neglected. In fact, none of the existing algorithms seem to have treated this problem explicitly.

The drift caused by detector errors can be eliminated by choosing \( \tau \) in Eq. (3) so

\(^3\)It was pointed out by Persaud and Hall (1989) that this kind of occupancy discrepancy has made the conventional occupancy-based incident detection logic difficult to apply. In our case this possibility has been ruled out by stipulating that the freeway should be homogeneous. Section 6 and Lin (1995) discuss how this requirement may be relaxed.

\(^4\)An experimental study conducted by Chan and May showed with field data that the average detector pulse on-times for two longitudinally closely spaced stations, could vary by five to ten percent, or even higher (1986).
as to leave a target residual drift when there is no incident and traffic is uncongested. For example, for the single-lane data in Figure 4, $\tau$ would have to be increased by 0.05. Because this is an inconsequential amount in our procedure compared with $0.1n$, it can be ignored. This is also typical of other cases we have seen. Note, however, that small error drifts could disrupt conventional algorithms.

Since the time series drift has been studied and perhaps adjusted under uncongested conditions, what assurances do we have that it will remain negative under congested conditions? Although the answer to this question must be experimental we note that for it to be affirmative it suffices that the percentage errors in detector readings be independent of occupancy (which seems plausible) and that the occupancy difference at the two locations for any given traffic state remains significantly below $\tau$ (which also seems plausible for “homogeneous” freeway sections). Our limited experience was also encouraging in the respect: as we are about to see, no false alarms were issued with any of the incident-free recurrent congestion scenarios that were examined.

4 Graphical Examples

Five 30-second data sets exhibiting either recurrent or non-recurrent congestion were selected from three different sites. These included a 2194 m long segment of the 1-880 freeway in Oakland, California between Hersperian boulevard and A street (California data), a 2437 m long segment of the A2 freeway from Amsterdam to Utrecht in the Netherlands (Netherlands data), and a 1940 m long segment of the Queen Elizabeth Way near Toronto, Canada (Toronto data). All three sites were thought to be suitable for testing since none had any traffic entrances, exits, or lane drops. A brief description of the five scenarios is given in Table 1; details can be found in the indicated Figures.

<table>
<thead>
<tr>
<th>Scenario number</th>
<th>Data source</th>
<th>Congestion type</th>
<th>Closest detectors straddling the incident</th>
<th>Number of lanes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Fig. 5)</td>
<td>California data; 9/8/93</td>
<td>Non-recurrent</td>
<td>3 and 1</td>
<td>5</td>
</tr>
<tr>
<td>2 (Fig. 6)</td>
<td>California data; 9/23/93</td>
<td>Non-recurrent</td>
<td>1 and 7</td>
<td>5</td>
</tr>
<tr>
<td>3 (Fig. 7)</td>
<td>California data; 10/27/93</td>
<td>Recurrent</td>
<td>N/A</td>
<td>5</td>
</tr>
<tr>
<td>4 (Fig. 8)</td>
<td>Netherlands data; 9/9/91</td>
<td>Recurrent</td>
<td>N/A</td>
<td>3</td>
</tr>
<tr>
<td>5 (Fig. 9)</td>
<td>Toronto data; 3/17/94</td>
<td>Recurrent</td>
<td>N/A</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: Summary of the data sets
Figure 5: (a) Scenario 1 (California data): Flow and occupancy vs. time and space plot
Figure 5: (b) Scenario 1 (California data): The cumulative difference of occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 5: (c) Scenario 1 (California data): The cumulative difference of transformed occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Occancy vs. time and space

Flow vs. time and space

Figure 6: (a) Scenario 2 (California data): Flow and occupancy vs. time and space plot
Figure 6: (b) Scenario 2 (California data): The cumulative difference of occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 6: (c) Scenario 2 (California data): The cumulative difference of transformed occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 7: (a) Scenario 3 (California data): Flow and occupancy vs. time and space plot
Figure 7: (b) Scenario 3 (California data): The cumulative difference of occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 7: (c) Scenario 3 (California data): The cumulative difference of transformed occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 8: (a) Scenario 4 (Netherlands data): Flow and occupancy vs. time and space plot
Figure 8: (b) Scenario 4 (Netherlands data): The cumulative difference of occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 8: (c) Scenario 4 (Netherlands data): The cumulative difference of transformed occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 9: (a) Scenario 5 (Toronto data): Flow and occupancy vs. time and space plot
Figure 9: (b) Scenario 5 (Toronto data): The cumulative difference of occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Figure 9: (c) Scenario 5 (Toronto data): The cumulative difference of transformed occupancy between two loop detectors (Ordinates have been normalized by the number of lanes.)
Each figure has three separate parts. The first part (a) is a pair of time-space diagrams shaded with an intensity that is proportional to the (30-second) occupancy and to the (30-second) flows estimated from the detector data. (The distance axis always points in the direction of flow, as is customary.) These diagrams give at a glance an overall idea of the evolution of traffic conditions for each scenario. Comparisons, however, should not be made across scenarios because the shading intensity was not standardized across sites.

The second part (b) shows the normalized $Z(t)$ at the site. To facilitate comparisons across scenarios, occupancies (in this section only) have been normalized by the number of lanes. The reader can verify that a significant drift exists for almost all pairs but the maximum value is only 0.08 points per lane. This arises for detector pair (23,25) for more than an hour in Figure 9(b).

The third part (c) shows the normalized $Y(t)$. In all five scenarios, we chose $\xi_n = 0.1$ as recommended in section 3.1. The diagrams are self explanatory and quite encouraging. They show that the $Y(t)$'s corresponding to every pair of detectors straddling an incident (but not those of other pairs) increase steadily with time after the onset of the incident.

It is interesting to note from the oscillating changes in shading intensity over time seen in all the diagrams that the interval data are quite noisy (many things could cause this) and that such noise should create problems for any method that bases detection on just a few time intervals. Remarkably, the noise is greatly reduced by the cumulative transformations leading to $Y(t)$, as we can see.

For a more detailed look at the proposed incident detection scheme, we use the plots for scenario 2. As shown in Figure 6(a), initially traffic is fairly light and smooth. At about 7:50 a.m., the occupancy downstream of loop detector station 1 suddenly becomes lighter and the occupancy upstream of loop detector station 7 becomes heavier, suggesting the occurrence of an incident between the loop detector stations 1 and 7. This was confirmed by “floating driver” reports included in the data. The incident appears to last for about 40 minutes and then traffic returns to normal operation.

In Figure 6(b), the normalized $Z(t)$ is given for every loop detector pair. Note the existence of a systematic drift in $Z(t)$ for the time period preceding the incident; i.e. from $i = 0$ (6:30 a.m.) to $i = 160$ (7:50 a.m.). During that period, occupancies from detector 3 are consistently higher than those measured from detector 1, and those from detector 7 are consistently lower than the ones measured from detector 20.

In Figure 6(c), the normalized $Y(t)$ is given for every loop detector pair. Note the absence of detector error drift and the clear signal that is issued when the incident occurs.
The incident location according to Figure 6(a) is also correctly identified since nothing happens to the $Y(t)$'s between detector pairs $(3,1)$ and $(7,20)$.

## 5 Triggers and Diagnosis

To complete our description of the incident detection algorithm we need a rule for selecting $\tau_0$. This should be done so as to ensure that $Y(t_j)$ has a negligible probability of exceeding $\tau_0$ if there is no incident but $\tau_0$ should not be so large that major incidents are missed.

A “quick and dirty” choice can be made without looking at data if we can accept that $Z(t_j)$ behaves like a random walk (i.e. with independent increments $\Delta O(t_j)$). Then $Y(t_j)$ can be treated like a random walk with a negative drift $-\tau$ and a reflecting barrier. The long term average of such a process is $E(Y) \approx \frac{1}{2} \frac{\text{Var}(\Delta O)}{\tau}$, and the fraction of time, $f$, that it spends above a given threshold $\tau_0$ declines exponentially with $E\left(\frac{\text{Var}(\Delta O)}{\text{Var}(\Delta O)}\right)$.

We propose here to estimate $\text{Var}(\Delta O)$ by $2 \text{Var}(O)$. This choice is very conservative on a scale of minutes because of the positive correlation in upstream and downstream occupancies. The choice is also valid for time intervals less than the trip time between the two detectors because then the two occupancies contributed by different groups of vehicles should be independent. A crude estimate of $\text{Var}(O)$ for all traffic conditions can be made by recognizing from our experience that the standard deviation of the occupancy with $At = 30$ secs rarely exceeds 0.15. Thus a reasonable choice for the variance of the occupancy for multiple lanes would be about $0.03n$, and $\text{Var}(\Delta O)$ would be $0.06n$ if $At = 30$ secs. Then, if we wish to have fewer than one false alarm per year (a logical criterion if a system includes many detectors that could be issuing false alarms all over a city) we would want $f \ll 10^{-6}$ since a year has on the order of $10^6$ 30 second intervals. Thus, triggering an alarm when the exponential of $f$ is about $-15$ should do the job. This yields $\tau_0 \approx 5$ if $\tau = 0.1n$. Note that this is sufficient to identify both incidents in our data since both generated $Y$’s in excess of $15n$. The value can be adjusted if one is willing to live with a different false alarm rate or if the sampling method is such that $At \neq 30$ secs.

If $Z(t)$ doesn’t vary like a random walk or one wishes to refine the choice of $\tau_0$ (always a wise decision since our choice for $\text{Var}(\Delta O)$ was rather speculative) one can simply observe the maximal excursions of $Y(t_j)$ for a couple of days and then double or treble

---

*For large time intervals, most of the vehicles cross both detectors within the same time interval. As a result, the two occupancies should exhibit a positive correlation. Even if traffic is congested, vehicles crossing both detectors are likely to move in the same queue, again resulting in a positive correlation in the two occupancies.*
the result. The largest excursion in our 15 hours of data appears in Figure 8(c). The normalized displacement can be seen to be about 0.8 which (for \( n = 3 \)) corresponds to \( Y_{\text{max}} \sim 2.5 \). This value is consistent with our recommendation for \( \tau_0 \).

When an incident is cleared (including the elimination of rubbernecking effects), traffic flows upstream and downstream of its location gradually return to the capacity level. In as much as occupancies at both locations also approach each other, this property can be used to issue an all-clear signal. This can be done by looking for a negative drift in \( Y(t) \) after reversing the direction of the reflecting barrier; i.e. changing “\( \text{max} \)” for “\( \text{min} \)” in Eq. (3) and redefining \( t_0 \) as some time after the incident. If desired, the parameters \( \tau_0 \) and \( \tau \) can be changed but a discussion of this possibility is not particularly insightful and is not given here. Figure 10 displays the resulting \( Y(t) \) for scenario 2; the starting and

![Diagram](image)

Figure 10: Time for the occurrence and termination of an incident (30 second data) ending times identified in the figure are consistent with the intensity plots of Figure 6(a)
and match the on-the-scene tach vehicle report.

It should be clear from Figure 10 that choosing \( \tau_0 = 5 \) delays the alarm by about 3 minutes for statistical reasons; clearly, if the on-site traffic characteristics allow a lower value of \( \tau \) than the all-purpose value recommended in this paper, the statistical delay would be reduced. On the other hand, we don’t want to make the response time smaller than the time that it takes for decreases in occupancy to propagate from the downstream detector to the upstream detector (in congested traffic) since otherwise there is a chance that the removal of a downstream bottleneck could be mistaken by an incident\(^6\). We believe that the pace of such signals is on the order of 5 min per mile, and this is yet another reason for spacing detectors closely.

### 6 Suggestions for More Detailed Diagnosis

We note that the incident starting time as identified by this algorithm is the time at which incident signals are received by the detectors and not the true incident starting time. The difference between these two times is an additional delay that depends on the detector spacing, incident location and magnitude, and traffic condition. [An upper bound for this delay is given by the formula, \( \frac{\delta k_j}{\Delta q} \), where \( \delta \) is the detector spacing, \( k_j \) is the jam density and \( \Delta q \) is the reduction in flow induced by the incident; see Lin (1995) for details.\] Clearly, the delay can be reduced by spacing detectors closely. We now show that our data contains information that may be used to estimate the true incident starting time and its location.

Close inspection of Figure 6(c) reveals that \( Y(t) \) dips down towards the end in two stages, and that these dips are related to the waves and shocks seen in Figure 6(a). The first dip corresponds to the onset of near-capacity occupancies (and flows) shortly after the incident is cleared and the second dip to the passage of the recovery shock over the upstream detector. [As expected, this second dip occurs first for detector pair (1,20) then for (1,7).\] If the shock passes the downstream detector before \( Y(t) = 0 \), as occurs in our example, the slope of \( Y(t) \) then returns to its equilibrium value. This can also be seen in our data as a slight concave bend in \( Y(t) \) around \( t \approx 300 \). Since the timing and magnitude of these slope changes are a function of the incident location, magnitude, and duration, they may be used as the basis for a more detailed diagnosis. The details of such a procedure, however, would depend on an underlying traffic model, which we do not\(^6\)

\(^6\)A similar problem could be caused by increases in occupancy that propagate forward (in light traffic) but these tend to move much faster.
want to introduce in this paper. Lin (1995) discusses a method based on the kinematic wave model.

We mentioned earlier that the equivalent number of lanes closed can be obtained from the slope of $Y(t_j)$ during the incident, and estimated with this method that $n_b \approx 3$ for the scenario of Figure 6. A more accurate estimate for $n_b$ can be obtained from direct measurement of the average flows during the incident, which in our case were 5250 vph. For a freeway with a lane capacity of 2000 vphpl, this through flow represents $\frac{5250}{2000} = 2.6$ operating lanes. Since $n = 5$, there must be 2.4 blocked lanes, which coarsely matches the other estimate. The rough equivalence of both estimates, as in this example, can be used as a sign that the incident has indeed occurred, and that the detectors are functioning properly.

7 Conclusion

This paper has presented a simple incident detection algorithm that is based on an intrinsic property of delay-inducing incidents without relying on complex and possibly questionable theories of traffic behavior. In this spirit, and in order to avoid introducing noise into our benchmarks, the algorithm was tested against real data rather than simulated data.

Unfortunately (for researchers only) delay-inducing incidents are rare events that do not always occur where a freeway happens to be properly instrumented. As a result, an exhaustive search only resulted in a limited number of incidents, and this obviously prevented us from establishing detection rates and false alarm rates for comparison with other algorithms. Perhaps other researchers can continue the tests. Both of our incidents happened when traffic was relatively free-flowing but we would have liked to have conducted a test when the incident happened during or after the onset of recurrent congestion. We believe that in this latter scenario our strategy is more susceptible to failure induced by geometric effects but we remain ignorant of the magnitude of the geometric inhomogeneities that will cause it. This, of course, is a question that cannot be answered by simulation; only direct experiments and practical experience may.

The problems caused by geometric inhomogeneities (including drops and rises in capacity) may in principle be avoided by an algorithm that would look for contemporary changes in the slopes of the cumulative occupancies upstream and downstream (positive upstream and negative downstream, of course) instead of positive changes in their difference as we have proposed in this paper. [Note that our current strategy will fail if a positive trend in the difference, $Z$, arises because of increases or decreases of different
magnitude in the trend of both cumulative occupancies.] It should be noted, however, that proper detector placement (e.g. closely spaced) would obviate the need for such an algorithm, and that close spacing is also needed for a quick response. The latter should in fact be true of any algorithm since traffic signals travel with a finite speed.

It is often said that the quickest incident detectors are the calls made by affected cellular-phone users on the road, and the proposal of this paper is unlikely to change that. The advantage of automated detection is that (with closely spaced detectors) the location and magnitude of the incident are automatically generated in a way that can be used by traffic forecasting models. This, of course, is a prerequisite of a detection algorithm that is to be part of an information distribution system able to provide anticipatory route guidance advice to drivers.
References


