DISTRIBUTED SURVEILLANCE ON FREEWAYS EMPHASIZING INCIDENT DETECTION AND VERIFICATION

Benjamin A. Coifman
corresponding author,
Associate Professor
The Ohio State University,
Joint appointment with the Department of Civil and Environmental Engineering and Geodetic Science and the Department of Electrical and Computer Engineering,
Hitchcock Hall 470
2070 Neil Ave
Columbus, OH 43210, USA,
phone: 614-292-4282;
fax: 614-292-3780;
e-mail: Coifman.1@OSU.edu

Ramachandran Mallika
Graduate Research Assistant
Department of Civil and Environmental Engineering and Geodetic Science,
The Ohio State University,
Columbus, OH 43210, USA,
e-mail: Mallika.1@OSU.edu

ABSTRACT

Efficient freeway management requires continuous decision-making based on conditions on the network and an understanding of the impacts of the decisions made. These conditions are usually measured with fixed-point surveillance systems, most of which are deployed in such a manner as to require communication links that are always connected and are polled at regular intervals. All of the sensor data are typically sent to a Traffic Management Center (TMC) for assessment, yet most of the time no action is taken in response to the data, leading to unnecessarily high communication costs.

To reduce communication costs without a significant loss in the quality of information received at the TMC this paper lays the foundation for an event driven communication system by examining the sequence of events at the detector stations in the context of incident detection. Where, following the broadest convention on freeways, an incident is any non-recurring event that causes a temporary bottleneck and restricts flow. Although the focus is incident detection, the proposed communication system could easily support many other applications that use aggregate data, e.g., measuring average annual daily travel (AADT). The methodology is generalizable to most common freeway geometries and care is taken in the paper to specifically address the situation where an incident interacts with a recurring bottleneck.

To address the normally high communications costs, a portion of the decision-making process is transferred from the TMC to the field controllers, which would make the initial evaluation of conditions and only send data that might elicit a control response or benefit comparative decisions between detector stations. In other words, rather than relying on the conventional, centrally polled communication system, these events could be used to initiate communication from the field when the potential value outweighs the cost per communication. The process could also lead to better data handling for decision-making or archiving in a conventional, polled communication system as well. We develop the methodology by deconstructing several incidents on a freeway and identify the observable events at a pair of detector stations that may be upstream, downstream or straddle the incident. This analytical process could be repeated for any other condition of interest.

INTRODUCTION

Urban roadways are becoming more congested, the *Urban Mobility Report* (Schrank and Lomax, 2005) estimated the net cost of urban congestion in the United States was over $63 billion in 2003, almost double the cost of ten years earlier. They also found that the average traveler lost 47 hours annually to traffic delays. Of these costs, 52 percent were attributed to incidents.

In response to increasing congestion, greater emphasis is being placed on active monitoring of the freeway network. Many cities around the world have deployed freeway traffic monitoring systems (FTMS).[1] Typically these surveillance systems use loop detectors or similar sensors to monitor traffic flow at discrete points in the network. The data from these sensors are normally transmitted to a Traffic Management Center (TMC) at regular intervals for assessment. Each detector station has a controller (i.e., a specialized microprocessor) or equivalent to monitor the detectors, measure aggregate traffic data over regular sampling periods, and handle communication with the TMC. Sampling periods commonly range between 20 sec and 5 min. Although the specific means of transmission varies[2], most FTMS around the world send the current conditions from every detector station to the TMC every sample, e.g., as employed by Smith et al (1996). Yet most of the time no action is taken in response to the data, leading to unnecessarily high communication costs. FTMS are typically used for three broad applications: traffic control such as ramp metering (e.g., Haj-Salem and Papageorgiou, 1995; Levinson and Zhang, 2004); traffic information such as compiling performance measures to understand operations (e.g., Petty, et al, 1998; Sheu, 1999; Norland, 2002; Shankar and Mannering, 1998) or providing traveler information (e.g., Emmerink et al, 1996; Al-Deek et al, 1998); and maintaining flow such as incident response and management (e.g., Lin and Daganzo, 1997; Hall, 2002).

Deviating from the centrally polled communications architecture normally used in FTMS, we consider the problem from the standpoint of distributed decision making, i.e., having the remote detector stations initiate communications when local conditions indicate the potential need for a response from the TMC. Specifically, the focus of this paper is on incident detection under such distributed decision-making to reduce the normally high communication costs, where, following the broadest convention on freeways, an incident is any non-recurring event that causes a temporary bottleneck and restricts flow (excluding planned road closures, e.g., for maintenance). Detection and monitoring are the first steps for incident response and management aimed at mitigating the impacts on the rest of the traffic stream. Responses could range from dispatching a wrecker to clear damaged vehicles to deploying controls to reduce demand past the incident via stricter metering at ramp upstream or providing information about the incident to travelers.[3]

---

[1] For example, most real time traffic data currently available on the Internet originate from such FTMS.

[2] Conventional phone lines are common in older systems, while fiber optic links or wireless Internet modems are common in more recent systems.

[3] Schrank and Lomax (2005) estimated that such countermeasures saved almost $3 billion in the US in 2003.

This paper develops a distributed decision-making process to greatly reduce the amount of data transfer needed to operate an FTMS effectively and reduce the communications costs without sacrificing measurement fidelity. Such an approach has been employed in many other disciplines, but has seen little use in freeway traffic monitoring. Even if distributed surveillance is not deployed in the field, the approach presented herein could allow for more efficient data handling for decision-making processes and archiving in a conventional, centrally polled communication system. Though incident detection on freeways using loop detectors is emphasized in this paper, the principles can be applied to other monitoring applications. The main motivation of this research is to reduce the frequency of communications (and thus, potentially cost) by transferring a portion of the decision-making process from the TMC to the field controllers without significant impacts to the performance of the overall decision-making process.

The analytical process employed is at least as important as the results. This paper presents a methodology to identify observable events arising from an incident at an unknown location on a homogeneous freeway with two loop detector stations. After accommodating for noise and measurement uncertainty, e.g., using Jain and Coifman (2005), these events can be used to filter the data at the field instead of the TMC, and preempt the transmission of non-informative data. Naturally, this 'filtering' of data to determine the benefit requires intelligence in the field to understand the decision-making process at the TMC. The methodology is generalizable to most common freeway geometries and the basic process can \ be extended to any other conditions of interest, e.g., this paper also considers incident detection in the presence of inhomogeneous segments such as recurring bottlenecks.

Implicit in the distributed communication approach is that "no news is good news," so periodic transmissions will be necessary to confirm the operational status of a station that does not see any critical events. Such transmissions can also be used to transmit summary data at a coarse aggregation, thereby allowing the station to provide a comprehensive overview of conditions at the location over time without a continuous communication link. This latter point is significant, since it will enable planning applications that require continuous monitoring, e.g., average annual daily travel (AADT).

There are a few existing examples of distributed surveillance applied to traffic monitoring, the most common one being the use of cellular phones for incident detection. Many operating agencies have developed call-in programs to accept cellular phone calls by motorists to report incidents. These wireless equipped drivers act as probes in the field, communicating information to a central decision-making system on an event driven basis. But coverage can be uneven since this information is voluntarily provided by the motorists, a given incident may be reported multiple times or not at all. It may also be difficult to establish precisely the location or magnitude of the event.

There have been a few applications of conventional detectors in a distributed surveillance system, e.g., Weil et al (1998) deals with anomaly detection (as opposed to incident detection) to lower the communication costs. The authors proposed that loop detectors in the field communicate to a TMC only when an anomaly is detected and then, transfer data at a higher frequency to aid the center in making a decision about the traffic states. Their theoretical model includes a modified McMaster incident detection algorithm (Persaud et al, 1990) with an online calibration technique based on spatial-based averages. There has also been at least one field operational test but the results were poor due to hardware integration problems (Banks and

Powell, 1997a and 1997b). Unlike these earlier efforts, this paper focuses on identifying the key events that would facilitate decision-making.

DISTRIBUTED SURVEILLANCE

This section both provides motivation for the distributed surveillance concepts and a description of the analytical methodology in the context of detecting incidents. It begins with a discussion on incident detection and observable events both before and after an incident occurs. Some simplifications are employed to highlight the critical events in the process, e.g., assuming that traffic only assumes a finite number of traffic states. These simplifications do not preclude practical application, rather, they bypass otherwise prohibitive nuances.[4] The entire set of observable traffic states at a pair of detectors adjacent to an incident on a homogeneous segment is then examined, along with the temporal progression. This information is used to develop distributed decision-making to determine what information is important and should be transmitted. The work is then extended to inhomogeneous segments with a recurring bottleneck. To control for the additional source of queuing, the section includes a discussion on identifying and locating recurring bottlenecks, including guidelines on measuring recurring bottleneck capacity. After completing this development, the impacts of the initial assumptions are considered before presenting the conclusions of this work.

Incident detection and observable events

Two signals emanate from an incident, a backward moving shock wave and a forward moving drop in flow (e.g., see Lin and Daganzo, 1997). Reliable detection of an incident using point measurements can happen only when both of the signals have been received at detector stations straddling the incident. The first stage of a distributed surveillance algorithm must nonetheless operate in the field using data from a single detector station. No communication would be made until some anomaly is observed, at which point communication would be initiated with the TMC to allow it to integrate information from adjacent stations and localize problems, e.g., distinguishing between an incident and recurring congestion from a known bottleneck. In the absence of such events, the TMC and field controller would share a common history of transmitted data and the field controller may mimic subsequent processing conducted at the TMC.

This section develops a representation of detectable events arising from an incident to aid in understanding the incident detection decision-making process. The underlying concept is a simple, yet effective model of representing the progression of the traffic state over time. Using this concept, we represent the progression of an incident on a homogeneous freeway segment on a 'time progression chart', as defined below. This chart gives insights to develop a set of feasible controls to aid the decision-making process at the TMC, representing graphically how the traffic states evolve over time. This initial development is limited to isolated incidents on a freeway segment without ramps, but the procedure could easily be expanded to include such features, as

[4] It is envisioned that in practice the discrete states would be replaced by a range of measurements, perhaps using a fuzzy logic approach that would allow a given observation to belong to more than one group.

will be illustrated later in this paper in the context of incident detection near recurring bottlenecks.

First consider a traditional time-space diagram depicting the evolution of traffic states on a homogeneous freeway segment, showing both what is observable at the detector stations and what is not. The presentation is made with the simplifying assumption that the observed traffic states on the freeway segment are discrete, stable and can be distinguished from one another. In this fashion, the problem is rarefied down to the fundamentally observable events. This presentation is followed by the approach to construct time progression charts to track the evolution of the traffic states at adjacent stations in the presence or absence of an incident.

Fig. 1 shows the time-space diagram corresponding to an isolated incident on the homogeneous segment and the corresponding flow-density diagram. The interface between two traffic states in the time-space diagram is modeled using the theories of Lighthill and Whitham (1955), and Richards (1956). For convenience in the following discussion, the flow-density relationship is assumed triangular (e.g., as employed by Newell, 1993). The results still apply if the relationship takes another shape, but the signals will propagate at different velocities. It is assumed that there are five discrete states of flow: free flow (F0, F1, and F2), congested (X1), and capacity (C). The flows at states F0, F1, and F2 increase in that order, with F0 is taken as the initial demand of the freeway segment. The flow at states F1 and X1 are equal to the capacity of the incident, corresponding respectively to conditions immediately upstream and downstream of it. Finally, flow at state F2 represents an arbitrary increase in demand from flow at F0 in the homogeneous freeway segment prior to the incident. The focus is on delay-inducing incidents, so the capacity of the incident is chosen to be below the demand in the freeway segment at the time and location of its occurrence. If there is no flow restriction due to an incident, then it is unlikely that it will disrupt traffic enough to be easily discernable, but similarly, if it does not restrict flow, it will cause little traveler delay.

The time-space diagram of Fig. 1 shows an incident at (t_1, d_I). After being in state F2, the service is restricted by the incident capacity at F1 and this state is subsequently observed downstream of the incident. A queue forms upstream of the resulting bottleneck in state X1, the rate the queue grows depends on the severity of the incident, i.e., capacity of the incident relative to the demand on the segment. For simplicity of presenting the relationships, the incident is assumed to have a constant capacity (flow at state F1) and the demand (flow at state F2) does not change after the incident occurs.

To capture all possible relative relationships between a pair of detectors and an incident, as will be discussed shortly, four possible detector locations are considered and they are represented in Fig. 1A as Detectors A-D. Far upstream of the queue vehicles remain free flowing at F2, and in this case demand exceeds incident capacity, X1. The backward moving shockwave traces the tail of the queue, in which flow drops and density increases. The queue grows upstream with an interface velocity corresponding to the slope of the line segment F2-X1 in the flow-density diagram, represented as dashed lines connecting the two traffic states. The incident starves demand downstream, causing both flow and density to drop, and this forward moving interface travels with a velocity equal to the slope of the line segment F2-F1 in the flow-density diagram. Once the incident is cleared, the queue discharges at the maximum service rate, the capacity of the freeway segment. This interface velocity corresponds to the slope of the line segment X1-C. Similarly, the interface velocity representing the raise in flow downstream of the incident corresponds to the slope of the line segment F1-C. Once the entire queue is discharged, traffic

returns to the pre-incident states. It should be noted that as with the assumption of a triangular flow-density relationship, the discretization may impact the exact velocity of signals traveling through the traffic stream, it will not, however, impact the sequence of events that the following discussion is based on.

Time progression chart for an isolated incident on a freeway

Since the detectors on the freeway segment are located at discrete points, the entire time-space diagram cannot be observed or inferred from a single detector station's data alone. Furthermore, an incident's location relative to any given pair of detectors cannot be known a priori. Instead, we consider the progression of traffic states for any pair of detectors, A-D, from Fig. 1A to facilitate this understanding, i.e., both stations may be upstream or downstream of the incident, or the pair of stations may straddle it. We develop a time progression chart to describe the traffic state evolution for all six combinations of detector and incident locations and aid the decision-making process. The time progression chart further condenses the time-space diagram to the states observed by *any* pair of fixed-point detectors in the field. Consider the time progression chart shown in Fig. 2, which illustrates the sequence of possible states that any pair of detectors will observe if an incident occurs on the homogeneous freeway segment. Each unique combination of the states observed by any pair of detectors is shown in a separate box, using the following notation: (upstream detector traffic state/downstream detector traffic state). Arrows indicate the transition from one pair of states to another with all feasible combinations considered. So in this case, the bottom left-most box contains (F0/F0), which indicates traffic state F0 is observed initially at both the upstream and downstream detector for any pair of detector present in the homogeneous freeway segment. Then due to raising demand, the upstream detector rises first, as shown in the adjacent box (F2/F0), and the signal propagates with the flow of traffic, impacting the downstream detector in the next box, (F2/F2). For any pair of detectors in Fig. 1A, these transitions occur prior to time t1.

There are four arrows emanating from (F2/F2). If there is no incident, the state remains (F2/F2) and when the demand drops, the state changes to (F0/F2) as shown with the dashed arrow. If an incident occurs upstream of both the detectors, then the state progresses from (F2/F2) to (F1/F2) as the drop in demand propagates downstream from the incident, i.e., detector locations C and D in Fig. 1. But again, the incident location or even its occurrence is not known a priori. So next we consider the possibility that the incident occurs downstream of both detectors, then the state progresses from (F2/F2) to (F2/X1), as the queue grows upstream from the incident, i.e., detector locations A and B in Fig. 1A. If the incident is severe enough that the end of queue reaches the further upstream detector, the state changes from (F2/X1) to (X1/X1). On the other hand, until the shockwave reaches the downstream detector location, no state change from (F2/F2) can be observed from this pair. Finally, if the detectors straddle the incident, e.g., detector locations B and C in Fig. 1A, then the state is shown as progressing from (F2/F2) to (F2/F1) since the slow moving shockwave will usually reach the upstream station after the fast moving drop in flow reaches the downstream station. This exercise can be continued to obtain the recovery states on the right hand side of Fig. 2.

The chart captures the history leading to a given pair of traffic states, rather than just the current traffic state. In addition, some paths are highlighted with solid squares indicating where historic information about the freeway segment and traffic states (e.g., characteristics of the particular freeway segment, temporal variations of traffic like time-of-day, day-of-week etc.) is particularly

useful to distinguish between alternatives. Similarly, stars highlight paths where historic information is particularly helpful to identify the fact that the state has progressed from one box to the next, e.g., to differentiate flow downstream of an incident from the normal drop in flow at the end of the peak period. Long-lived traffic state combinations that may be sustained for an extended period are shown with bold outlines while short-lived states, such as F2/F0 during the transition from F0/F0 to F2/F2, are shown with thinner outlines. Naturally, the precise time for the signal to reach a detector depends on the relative location of an incident to the detector location as well as the severity of an incident.

Decision-making under distributed surveillance

The time progression chart captures the observable traffic states. After this calibration one would use the chart to determine the freeway conditions from observations at each consecutive pair of detector stations. Under distributed surveillance, intelligence is required in the field to make decisions on communicating the observed data to the TMC, including what data to send and when to initiate communication. The time progression chart can be used to help answer these questions by constructing a feasible set of control points, as will be defined below, under various cost scenarios. The set of feasible controls acts as a proxy for the value of observed data and any control has a set of decision states, i.e., the response by the traffic control system to observed events. The initial decisions must occur in the field in a distributed surveillance system. To this end, each box in Fig. 2 is grouped into one of five decision states: No Action, Suspect Incident, Confirm Incident, Recovery, and Post-recovery, as noted at the bottom of each box. Consider the progression from (F0/F0) to (F2/F0) on the left hand side of the figure, since this change simply shows a rise in demand, No Action is taken. On the other hand, the progression from (F2/F2) to (F2/X1) indicates that the downstream detector observes queued traffic from an obstruction further downstream, and hence the decision state indicates Conform Incident. Similar reasoning is applied to the remaining boxes and reflects the typical actions appropriate for the states that are observed in the field. The assignment of decision states includes the history of feasible paths, i.e., the value of the current state depends on the previous states.

After establishing the decision states, control points are the set of measurements observed by the field that are chosen to trigger a shift of decision-making from the field to the TMC. They are characterized by the deviation from the expected set of measurements, indicating an anomaly in the field (though not necessarily an incident). The control points can be established by tracing the feasible paths in time progression chart while considering the traffic state observed at only *one* detector at a time. In this work, control points are denoted by a set of traffic states at a given detector, [expected state, measured state], e.g., if the historical trends or the last reported traffic state is F2 and the observed traffic state in field becomes X1, the control point is denoted by [F2, X1]. Provided care is taken to reduce noise in the field data, these transition points would occur relatively infrequently since successive observations of the traffic state will normally remain at the same state.

Table 1 lists the control points when the time progression chart is reduced to the observations from a single station. Depending on the cost of a single transmission, control actions can be chosen selectively from the feasible set of controls. A sub-set of control points could be chosen depending on the cost per transmission incurred. When the cost of initiating individual connections dominate the cost per unit of data sent, any intermediate states from Table 1 that are not transmitted could be sent to the center along with the next state when the next transmission

eventually does occur. In practice, one would likely want to extend the progression chart and include control points at [not X1, X1] to capture queues arising from incidents when demand is below F2 and perhaps [X1, not X1] to capture an upstream drop in demand.

Incidents in the Presence of Recurring Bottlenecks

Thus far, this paper has considered a homogeneous freeway segment, with no other capacity constraints beyond an incident. The basic process can easily be generalized to other situations. For example, although queuing is a key indicator of an incident, obviously queuing can also arise upstream of a recurring bottleneck on inhomogeneous segments. Any practical solution must be able to differentiate between recurring congestion and queuing arising from an incident.

To illustrate the process of generalizing the methodology, this section examines the interaction between a recurring bottleneck and an incident. Building off of the simplified model used above, it is assumed that an inhomogeneous segment can be modeled with a single recurring, point bottleneck limiting flow in the absence of any incident. Now it is assumed that there are seven discrete states of flow: free flow (F0, F1, F3, and F4), congested (X1 and X3), and capacity (C). Once more, flow increases with increasing number, and the flow is the same for a given number whether free flow (F) or congested (X), e.g., the states F3 and X3 correspond to the recurring bottleneck capacity. An incident must restrict flow below this level to impact the segment throughput and we arbitrarily denote such a restriction with X1 and F1, corresponding respectively to conditions immediately upstream and downstream of the incident. Flow at state F4 represents an arbitrary increase in demand above the capacity of the recurring bottleneck, prior to the incident. Finally, note that state F2 is omitted to avoid any confusion with the arbitrary increase in flow used in the homogeneous example.

In the absence of an incident, when demand climbs to state F4, a queue would build up behind the recurring bottleneck until demand drops below that of state F3, at which point, the queue would recede back towards the bottleneck and eventually dissipate. In the presence of queuing, conditions downstream of the recurring bottleneck would be expected to be at F3. The relationship becomes more complicated when an incident limits flow below that of state F3. Fig. 3 shows the time-space diagram corresponding to an incident at t_1, upstream of a recurring bottleneck. The construction of the diagram follows the same process used in Fig. 1. Notice how the incident starves demand at the recurring bottleneck when it limits downstream flow to that of state F1. Given the two restrictions, an arbitrary pair of detector stations could both be downstream, between, or upstream of both restrictions; or, the detector stations might straddle one or both of the restrictions. To capture all of these combinations, six possible detector locations are shown in the figure, denoted E-J. Alternatively, an incident may occur downstream of the recurring bottleneck, as shown at t_1 in the time-space diagram of Fig. 4. Once more, one needs to consider six detector station placements to capture all possible relationships between an arbitrary pair of detectors and the two restrictions. Highlighting the fact that a given recurring bottleneck may be downstream of one incident and upstream of another, this figure repeats the detector station labels G-J with the same relative positions to the recurring bottleneck from Fig. 3, while adding stations K and L further downstream of the incident.

As with the homogeneous example, this section develops time progression charts for an incident at an arbitrary location relative to any pair of detector stations in Fig. 3-4. Presumably, the recurring bottleneck location would be known to an operating agency (as briefly discussed below), so the analysis is split across three different time progression charts based on the detector

locations relative to the recurring bottleneck. First, consider the case where both of the detector stations are upstream of the recurring bottleneck, i.e., any two stations from E-H in Fig. 3-4. This progression chart is shown in Fig. 5 and uses information from the two time-space diagrams. For example, if it turned out that both detectors were downstream of the incident (stations G and H in Fig. 3), the combined traffic state could progress as follows: (F0/F0) as the initial state, (F4/F0) and (F4/F4) as demand rises, then (F1/F4) as the incident starves supply, and finally settling in (F1/F1) until the incident is cleared. The drop in flow suggests that an incident may have occurred, but given the fact that traffic remained free flowing at both stations, the possibility that demand simply waned cannot be eliminated. Thus, the decision state is denoted as Suspect Incident. On the other hand, if the incident occurred downstream of the recurring bottleneck, the combined traffic state for the same two stations (this time as shown in Fig. 4) could progress as follows: (F0/F0) as the initial state, (F4/F0) and (F4/F4) as demand rises, then (F4/X3) as the queue from the recurring bottleneck overruns station H, and finally passing to (F4/X1) as the queue from the incident overruns station H. Provided the capacity of the recurring bottleneck is known, traffic state X1 with a lower flow would clearly indicate the incident, hence, this decision state is denoted as Confirm Incident. This process is repeated for each of the detector station pairs upstream of the recurring bottleneck to complete the progression chart in Fig. 5. The procedure is then repeated for the case where the detectors straddle the recurring bottleneck, Fig. 6, and when both are downstream of it, Fig. 7, in each case combining information from Fig. 3-4.

Once more, after establishing the decision states, the control points can be chosen to trigger a shift of decision-making from the field to the TMC. As before, they are established by tracing the feasible paths in time progression charts while considering the traffic state observed at only *one* detector at a time and depending on the cost of a single transmission, control actions could be chosen selectively from the feasible set of controls. Combining information from all three of the progression charts for the inhomogeneous segment, Table 2 lists the control points when the time progression chart is reduced to observations from an individual station upstream of the recurring bottleneck, and Table 3 repeats the process for an individual station downstream of the bottleneck.

Identifying Recurring Bottlenecks

Distinguishing recurring congestion from an incident as well as the location of the incident using detector data can be difficult. To this end it is important to know the location of the bottlenecks and their capacity before applying the time progression charts to an inhomogeneous segment. Historical traffic data can be used to identify and locate the existing, recurring bottlenecks, e.g., a contour plot of traffic speed over time and space. Time-series plots of occupancy and speed can also be used to localize the bottlenecks. When identifying and characterizing recurring bottlenecks it is important to analyze the traffic conditions over several days to establish a trend and confirm the existence (Varaiya, 2002) that would otherwise confound detection of the non-recurring incidents. If there are no existing loop detectors present on the given freeway segment, one can employ probe vehicles to identify and locate the bottlenecks in the system.

Once the presence of recurring bottlenecks and their locations are established, the capacity of a given bottleneck can be approximated with historic cumulative arrival curves from the adjacent detector stations. The slope of this curve is the instantaneous flow and an extended sampling period can be used to reduce the impact of transient fluctuations. When the recurring bottleneck

is active, the maximum observed flow at the downstream station is its capacity. Care should be taken to use the station downstream of the bottleneck or restrict analysis to queued conditions at a station upstream of the bottleneck, since a station upstream can observe sustained flows in excess of the bottleneck capacity prior to the onset of queuing, e.g., state F4 at detector E in Fig. 3 has a flow in excess of the bottleneck capacity. Naturally, multiple days should be studied to reduce the chance that incident conditions are accidentally included in this calibration.

Practical Issues

There are several practical issues that must be considered for a distributed surveillance system, including: the benefit of integrating information across adjacent stations, relaxing the assumptions used in this work, accommodating noisy measurements, and deployment issues. Stepping through these issues one at a time, information from adjacent detectors is useful in the decision-making process at the TMC, it complements the temporal information from a single station with the spatial knowledge along the freeway. This information is useful to confirm or dispel an anomaly observed at a single station, to predict a trend or anticipate the state of the larger system. In the proposed distributed surveillance system, a detector communicates with the TMC when it observes a control point and thereby, transfers the decision-making process from the field to the TMC. The TMC must integrate the information from several detector stations before reaching a conclusive decision on the course of action. Here, information from adjacent detectors (upstream and downstream) is useful to get a better understanding of what is happening, e.g., the information is typically valuable in confirming incidents, localizing their location, and distinguishing them from recurring congestion or a drop in demand. In a centrally polled communication system, this information from adjacent detector stations is readily available, allowing for easy comparison of input-output (i/o) on a link, as well as speed differences and occupancy differences between stations. In a distributed surveillance system, the initial decisions must be made without such comparisons between stations. However, the performance of the distributed surveillance can approach that of centralized surveillance if after one detector station initiates communication with the TMC, the TMC can poll the adjacent stations to collect additional information. For example, a controller could store data from the most recent 10 minutes, then, if i/o diagrams are needed for diagnosing an anomaly, the TMC could request the recent history from any station.

In the construction of a time progression charts, it was assumed that the traffic states observed were discrete and stationary in nature. However, traffic states are a continuum. The characteristics that a given traffic state exhibit in this analysis, (e.g., F1 in Fig. 1) may be representative of a continuous range of traffic states. For decision-making, however, it remains important to distinguish those traffic states that can be differentiated from other traffic states by the sensors in the field. This concept plays a vital role, since the field measurements are noisy due to transient events, caused by either inherent traffic characteristics or detector errors. Discernable traffic states play a key role in assessing the current status of the freeway without ambiguity, thereby leading to decisions to return conditions to the desired status. The discrete traffic states used in a time progression chart can potentially be a sub-set of the discernable traffic states. It is possible to empirically establish the bounds between some of the discrete traffic states, e.g., F0 and C in Fig. 1, while others may be infeasible given the noise and transient errors in the field measurements, e.g., F1 and F2 in Fig. 1 may be too close on the continuum. However, the key problem of distinguishing between free flow and queued states is generally feasible with most traffic detectors. Although this task is complicated by the typically

noisy measurements, several approaches have been developed for reducing the impact in aggregate measurements, including: diagnosing detectors for malfunctions (e.g., Chen et al, 2003; Coifman, 1999), validating speed estimates using flow-occupancy relationships, exploiting the redundant information from adjacent lanes (Jain and Coifman, 2005). To capture the real measurement uncertainty in general, one can define an error bound that is a measurable function that maps the uncertainty for every feasible measurement relative to the last reported state. So a measurement outside of the error bound would indicate that conditions on the roadway have changed since the last report to the TMC. It is envisioned that in practice the discrete states would be replaced by a range of measurements, perhaps using a fuzzy logic approach that would allow a given observation to belong to more than one group.

Fig. 8 presents the time series speed and flow measured from detector stations on I71 in Columbus, Ohio, USA, during four incidents. Each row is a different incident, with the left column showing median speed sampled at 30 sec and the right column showing the flow sampled at 5 min. The flow time series are shown over a longer time window than the corresponding speed time series because deviation from "normal" flow requires a large window to establish the non-incident baseline flow while deviation from free flow speed is easier to identify because the free flow speed is known a priori. On the far right of each row the legends show the stations from top to bottom in the order a vehicle would pass them. Each plot includes two stations downstream of the incident (equivalent to Detectors C and D in Fig. 1), these stations do not become congested, i.e., speed remains free flowing, but the flow drops due to the incident starving flow upstream. The plots include several stations within the queue (equivalent to Detector B in Fig. 1), here both speed and flow drop when the queue reaches the given station. Finally, the plots include two stations upstream of the queue (equivalent to Detector A in Fig. 1), neither speed nor flow drop in this case because the stations are upstream of the furthest reach of the queue.[5] Since the present work considers an arbitrary pair of detector stations, the first row of Fig. 8 is expanded in Fig. 9, highlighting the observations from four different pairs of detector stations for the same incident. Now the flow time series is plotted over the same time window used for the speed series. The first three rows of this figure correspond respectively to the three solid arrows emanating from box (F2/F2) on the left hand side of Fig. 2. In the first row the two stations are downstream of the incident and the state transitions to (F1/F1), with the free flow speed and drop in flow after the incident evident from the two plots. In the second row the stations straddle the incident and the state transitions to (X1/F1), similar to the first row, except for the fact that speeds drop in the queue at the upstream station. In the third row both stations are upstream of the incident and the state transitions to (F2/X1), but the queue does not reach the furthest station before recovery. While in the final row of Fig. 9, again both stations are upstream of the incident but now the queue also reaches the upstream station, so after reaching (F2/X1) the state progresses to (X1/X1) before recovery. Much as the capacity of recurring bottlenecks is site specific, as one generalizes the methodology from the discrete traffic states to some collection of bins or regions, it may require site specific fine tuning, e.g., in the third row of Fig. 8 the

[5] Note that the queue in Fig 1 shows that once the bottleneck is cleared the queue dissipates upstream, it is also possible for the queue to recede back to towards the bottleneck location, as is the case in the four example incidents. In each of these incidents the first station to become congested is also the last station to return to free flow. Both dissipation patterns are common.

background flow at station 106 is about a third of that at stations 105 and 107.[6] The precise definition of such bins is left to future research.

From a deployment standpoint, one would have to weigh the value of retrofitting an existing system to the proposed methodology. Such an analysis needs to balance the sunk costs of building the existing communication links against the future costs of maintaining service. In many cases where the operating agency owns dedicated communication links it would likely prove more cost effective to continue sending all of the data all of the time. In other cases where the agency leases phone lines or other services, it could prove cost effective to migrate to a system as proposed herein. Such a migration would require major revisions to the software running at the TMC and on the field controllers, as well as possibly replacing some of the communications links to ones that charge per unit of data sent, e.g., wireless modems.

Where this methodology would likely make the most economic sense would be for new deployments, either in a city without any traffic monitoring system or expanding to new corridors that are not currently monitored. In the latter case, presumably to the extent conventional traffic monitoring systems have already been deployed within a region, they are most likely to have been placed on the most congested corridors. In other words, new deployments would be on corridors that have lower congestion, i.e., fewer incidents in the new deployment that would translate into a greater possible communications savings. For example the California State Department of Transportation (Caltrans) spends several million dollars per year on communications with detector stations and the proposed approach could allow monitoring traffic conditions in normally uncongested areas at a much smaller incremental on-going communication cost per station.

The biggest costs of deployment are likely in developing the mundane details, e.g., defining thresholds. Even if the methodology is not initially used as the basis for a communications system, it could still prove beneficial for data storage, effectively forming the basis of a compression algorithm to archive data. In fact such an archiving application in a conventional system that collects all of the data all of the time would allow an opportunity to work through the sometimes-difficult tasks of implementation without having to also deal with additional complications arising from the missed observations if a controller erroneously decided not to send data when it otherwise should have. Focusing first on archiving would thereby reduce the development costs for a first adopter.

CONCLUSIONS

Most traffic surveillance systems have been deployed with simple communication network architecture, detector stations are polled regularly from a TMC. The authors believe that agencies should either transmit (and archive) traffic flow data at the highest feasible resolution, or transmit (and archive) as little data as possible. Most FTMS today fall in between these two extremes. We have previously examined new applications that would be enabled by transmitting the data at the highest feasible resolution, e.g., the individual vehicle actuations at the detectors allow for vehicle reidentification, travel time measurement, density estimation, and many more

[6] Stations 105 and 107 are upstream and downstream, respectively, of the interchange between I70 and I71, while station 106 is on a connector within the interchange.

metrics that are not presently available from conventional FTMS (Coifman and Cassidy, 2002, Coifman, 2003a and 2003b).

Considering the other extreme -- sending as little data as possible -- this paper proposes an alternative approach that would start evaluating data in the field, before making a decision to transmit data. In the short term, such a process could allow communication using limited bandwidth, wireless connections. Because communications costs often dominate other operating costs for traffic surveillance systems, this approach could be used to inexpensively extend monitoring to freeway links that do not normally see congestion and are not cost effective to monitor using conventional strategies.

Given the rapidly improving wireless communication networks, in the long run perhaps the cost benefit will disappear and the biggest impact of this research will be the process of identifying key events in traffic data using a finite number of states to reduce the impact of noisy data on the analysis, aiding the study of the temporal progression in a meaningful manner that facilitates an understanding of the system. This fact, in turn, could be used to streamline many control processes even in a conventional, centrally polled surveillance system. The fundamental elements presented in this paper can be broken down into three steps, first establish the space of traffic states on a time space diagram using only enough discrete traffic states as necessary while adding enough detectors to capture all feasible sets of states observed from any pair of detector stations. In more complicated scenarios, such as an incident in the presence of a recurring bottleneck, one may have to develop several time space diagrams to capture the entire set of observable states, e.g., Fig. 3-4. Second, trace each pair of detectors to identify the evolution of the traffic states and record this information on a time progression chart. The chart should be constructed to clearly delineate events by information value, e.g., 'No Action' versus 'Confirm Incident'. Finally, control points can be identified from the time progression chart and the responsiveness can be determined as a function of communication cost. If these costs change, the degree of responsiveness can be adjusted accordingly. Of course this final step requires accommodating the continuous spectrum of traffic states and noisy measurements. The impact of the latter point can be reduced through data cleaning techniques, e.g., Jain and Coifman (2005), while both points can be addressed with reasonable error bounds, perhaps by employing fuzzy logic or a similar technique. Obviously, if there were more than two detector stations on the freeway segment, one would rely on whichever successive pair provided the fastest response. One could then either suppress a subsequent response from the other detector stations if communications costs are high, or anticipate subsequent notification from them for verification.

By assuming discrete traffic states, it is possible to simplify the otherwise difficult problem and gain a better fundamental understanding of how the traffic state evolves over time in the presence of an incident. In closing, although the methodology was developed by deconstructing an incident on a freeway, this analytical process could be repeated for any other condition of interest on the freeway, e.g., where to locate detectors on a freeway.

ACKNOWLEDGEMENTS

The authors would like to thank Ho Lee for his help preparing Figs. 8-9, we also appreciate the constructive input from the reviewers that has improved the quality of this paper.

This material is based upon work supported in part by the National Science Foundation under Grant No. 0127944 and by the California PATH (Partners for Advanced Highways and Transit) Program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation. The Contents of this report reflect the views of the author who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification or regulation.

REFERENCES

Al-Deek, H., Khattak, A., Thananjeyan, P., (1998). "A combined traveler behavior and system performance model with advanced traveler information systems," *Transportation Research. Part A: Policy and Practice*, 32(7), pp. 479-493.

Banks, J., Powell, P., (1997a) "San Diego Field Operational Test of Smart Call Boxes: Institutional Issues," TRB, *Transportation Research Record 1603*, pp 20-26.

Banks, J., Powell, P., (1997b) "San Diego Field Operational Test of Smart Call Boxes: Technical Aspects," TRB, *Transportation Research Record 1603*, pp 27-33.

Chen, C., Kwon, J., Skabardonis, A., Varaiya, P., (2003). "Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems," TRB, *Transportation Research Record 1855*, pp 160-167.

Coifman, B. (1999) "Using Dual Loop Speed Traps to Identify Detector Errors," TRB, *Transportation Research Record 1683*, pp 47-58.

Coifman, B., Cassidy, M. (2002). "Vehicle Reidentification and Travel Time Measurement on Congested Freeways", *Transportation Research. Part A: Policy and Practice*, 36(10), pp. 899-917.

Coifman, B. (2003a). "Identifying the Onset of Congestion Rapidly with Existing Traffic Detectors", *Transportation Research. Part A: Policy and Practice*, 37(3), pp. 277-291.

Coifman, B. (2003b). "Estimating Density and Lane Inflow on a Freeway Segment", *Transportation Research. Part A: Policy and Practice*, 37(8), pp 689-701.

Emmerink, R., Nijkamp, P., Rietveld, P., Van Ommeren, J., (1996). "Variable message signs and radio traffic information: an integrated empirical analysis of drivers' route choice behaviour," *Transportation Research. Part A: Policy and Practice*, 30(2), pp. 135-153.

Haj-Salem, H., Papageorgiou, M., (1995). "Ramp Metering Impact on Urban Corridor Traffic: Field Results," *Transportation Research. Part A: Policy and Practice*, 29(4), pp. 303-319

Hall, R., (2002). "Incident dispatching, clearance and delay," *Transportation Research. Part A: Policy and Practice*, 36(1), pp. 1-16.

Jain, M., Coifman, B., (2005) "Improved Speed Estimates from Freeway Traffic Detectors," *ASCE Journal of Transportation Engineering*, Vol 131, No 7, pp 483-495.

Levinson, D. and Zhang L., (2006). " Ramp meters on trial: Evidence from the Twin Cities metering holiday," *Transportation Research. Part A: Policy and Practice*, 40(10), pp. 810-828.

Lighthill, M. and Whitham, G. (1955) "On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads," *Proc. Royal Society of London, Part A*, vol 229, no 1178, pp 317-345.

Lin, W., Daganzo, C., (1997) "A Simple Detection Scheme for Delay-Inducing Freeway Incidents," *Transportation Research. Part A: Policy and Practice*, 31(2), pp. 141-155.

Newell, G.F., (1993). "A simplified theory of kinematic waves in highway traffic i: General theory. ii: Queuing at freeway bottlenecks. iii: Multi-destination flows", *Transportation Research Part B* 27, no 4, 281-313.

Noland, R., (2002). "Relationships between highway capacity and induced vehicle travel," *Transportation Research. Part A: Policy and Practice*, 35(1), pp. 47-72.

Persaud, B., Hall, F., and Hall, L., (1990) "Congestion Identification Aspects of the McMaster Incident Detection Algorithm," TRB, *Transportation Research Record 1287*, pp 167-175.

Petty, K., Bickel, P., Ostland, M., Rice, J., Schoenberg, F., Jiang, J., Ritov, Y., (1998). "Accurate estimation of travel times from single-loop detectors," *Transportation Research. Part A: Policy and Practice*, 32(1), pp. 1-17.

Richards, P., (1956). "Shock Waves on the Highway", *Operations Research*, 4 (1), pp 42-51.

Schrank and Lomax, (2005). *Urban Mobility Report*, Texas Transportation Institute, College Station, TX, USA.

Shankar, V., Mannering, F., (1998). "Modeling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach," *Transportation Research. Part A: Policy and Practice*, 32(5), pp. 311-322.

Sheu, J.B., (1999). "A stochastic modeling approach to dynamic prediction of section-wide inter-lane and intra-lane traffic variables using point detector data," *Transportation Research. Part A: Policy and Practice*, 33(2), pp. 79-100.

Smith, W. S., Hall, F.; Montgomery, F., (1996). "Comparing the speed-flow relationship for motorways with new data from the M6," *Transportation Research. Part A: Policy and Practice*, 30(2), pp. 89-101.

Varaiya, P. (2002). "Freeway Data Collection, Storage, Processing, and Use", Presented at the Workshop on the National Transportation Information Infrastructure, Transportation Research Board.

Weil, R., Wootton, J., Garcia-Ortiz, A., (1998) "Traffic Incident Detection: Sensors and Algorithms," *Mathematical. Computer Modeling*, Vol 27, no 9-11, pp 257-291.

LIST OF FIGURES

LIST OF TABLES

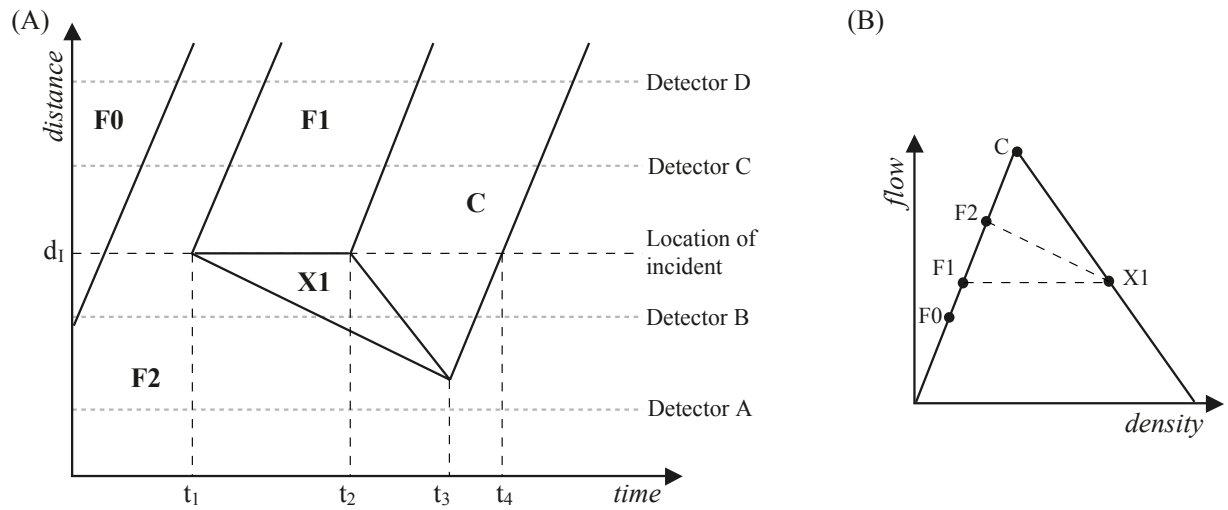

Fig. 1. (A) Time-space diagram for an isolated incident on a homogeneous freeway segment, (B) the corresponding flow-density relationship.

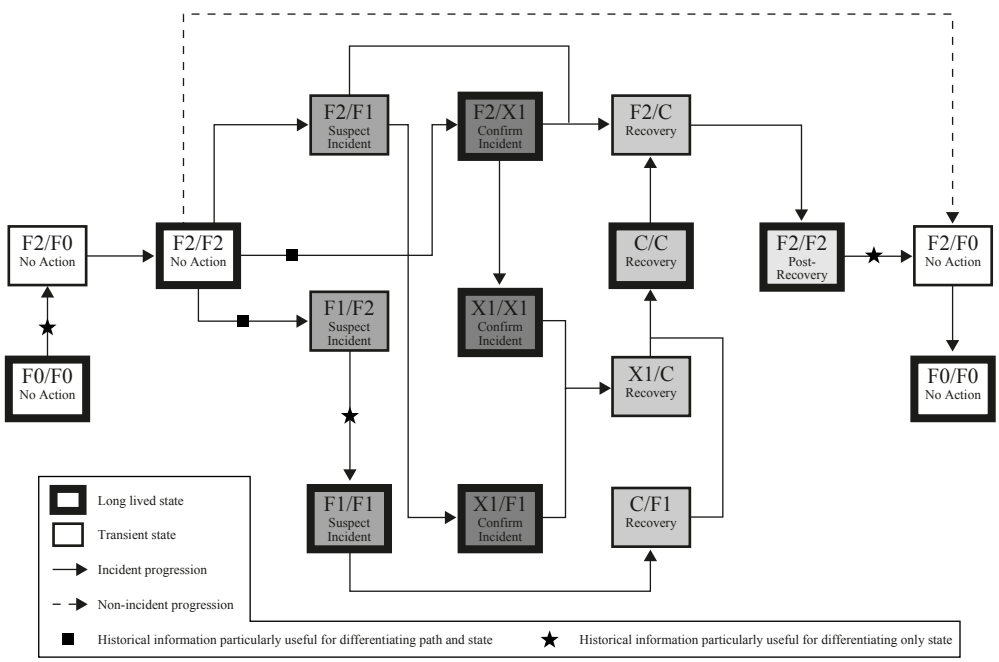

Fig. 2. Time progression chart (us/ds) for an isolated incident on a homogeneous section

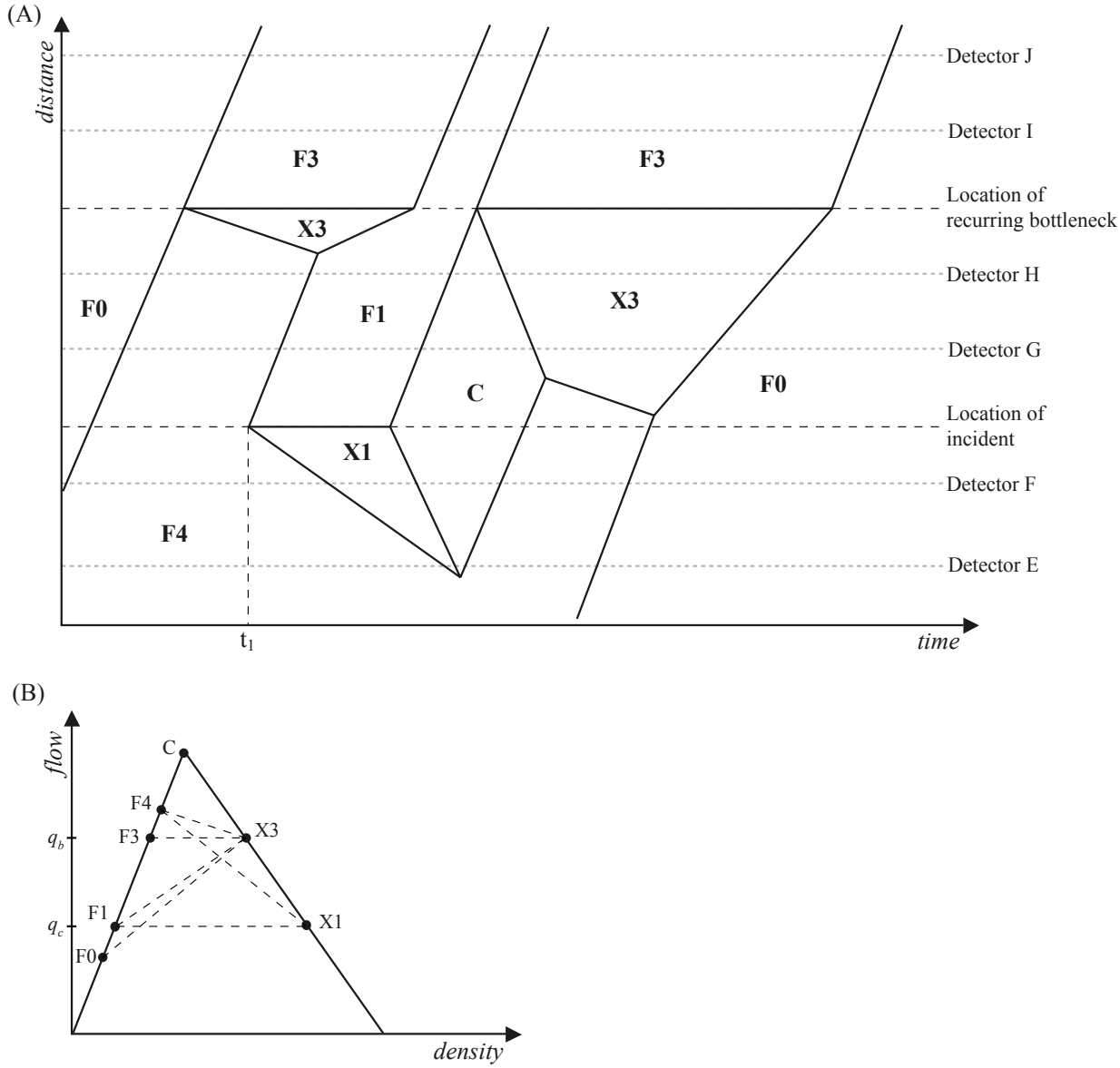

Fig. 3. (A) Time-space diagram for an incident upstream of a recurring bottleneck, (B) the corresponding flow-density relationship.

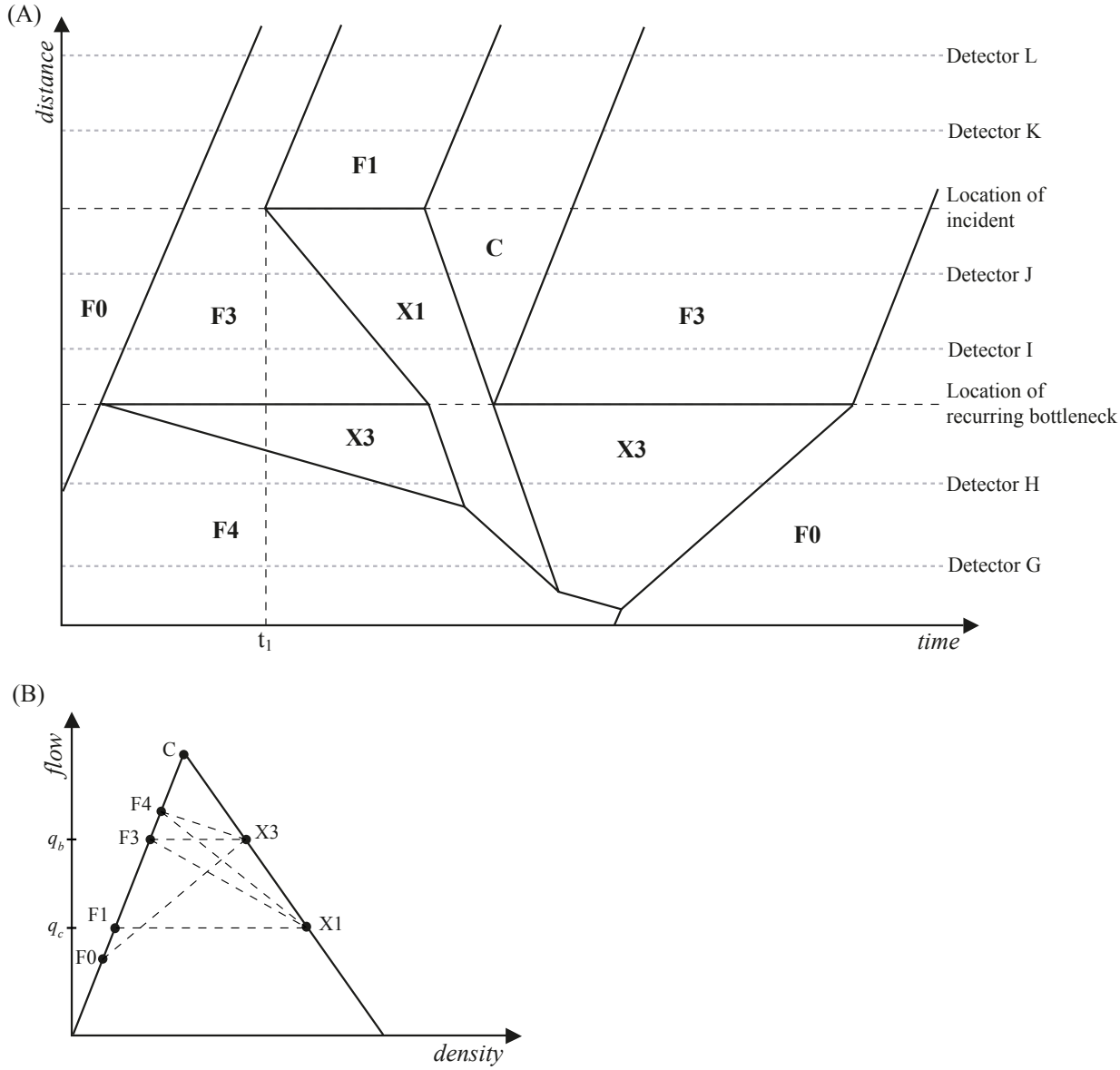

Fig. 4. (A) Time-space diagram for an incident downstream of a recurring bottleneck, (B) the corresponding flow-density relationship.

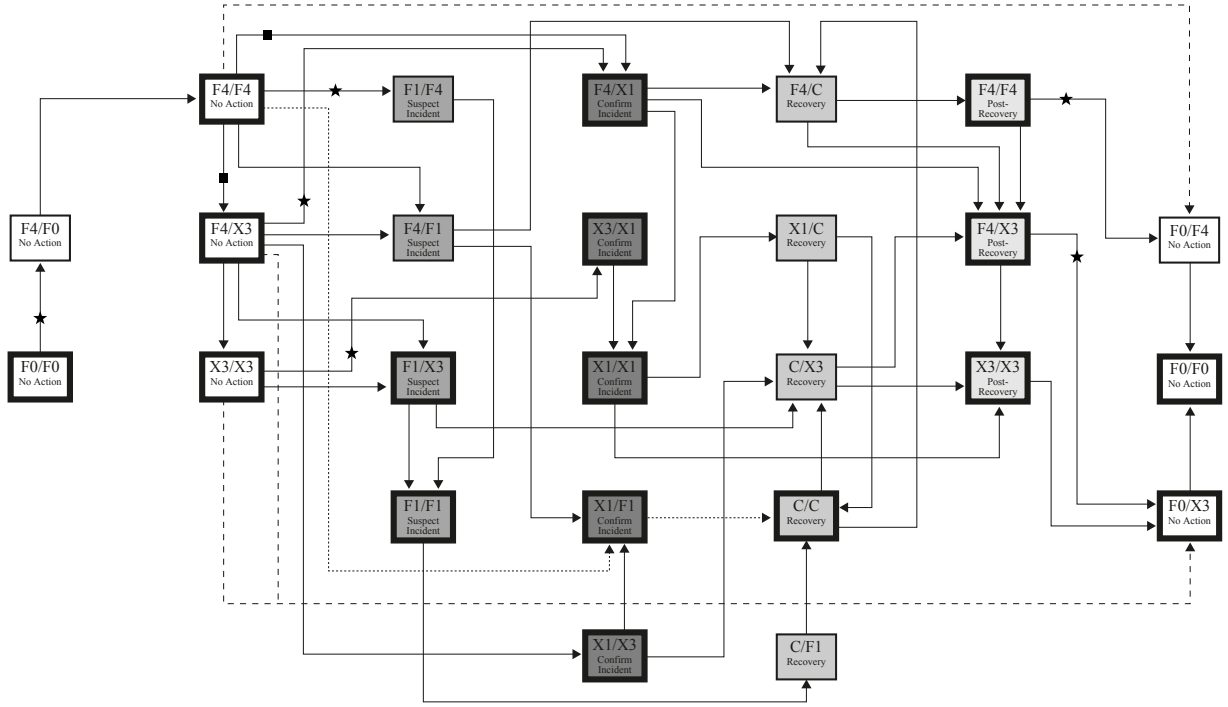

Fig. 5. Time progression chart (us/ds) for an incident as observed by a pair of detectors upstream of a recurring bottleneck. The legend is given in Fig. 2. Note that the incident may be downstream of the bottleneck.

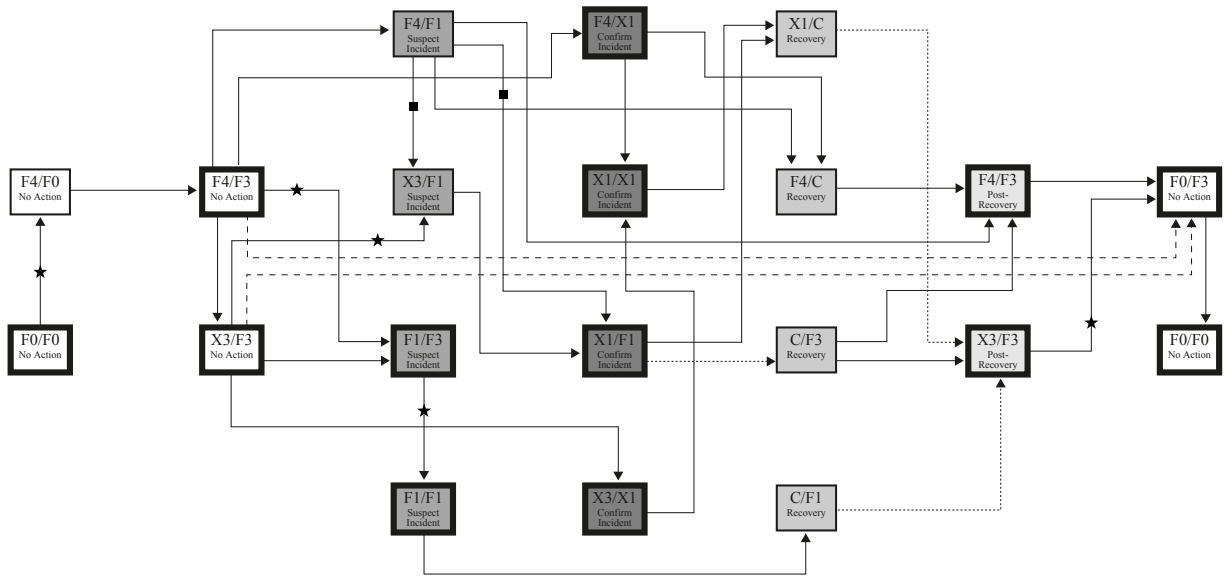

Fig. 6. Time progression chart (us/ds) for an incident as observed by a pair of detectors straddling a recurring bottleneck. The legend is given in Fig. 2.

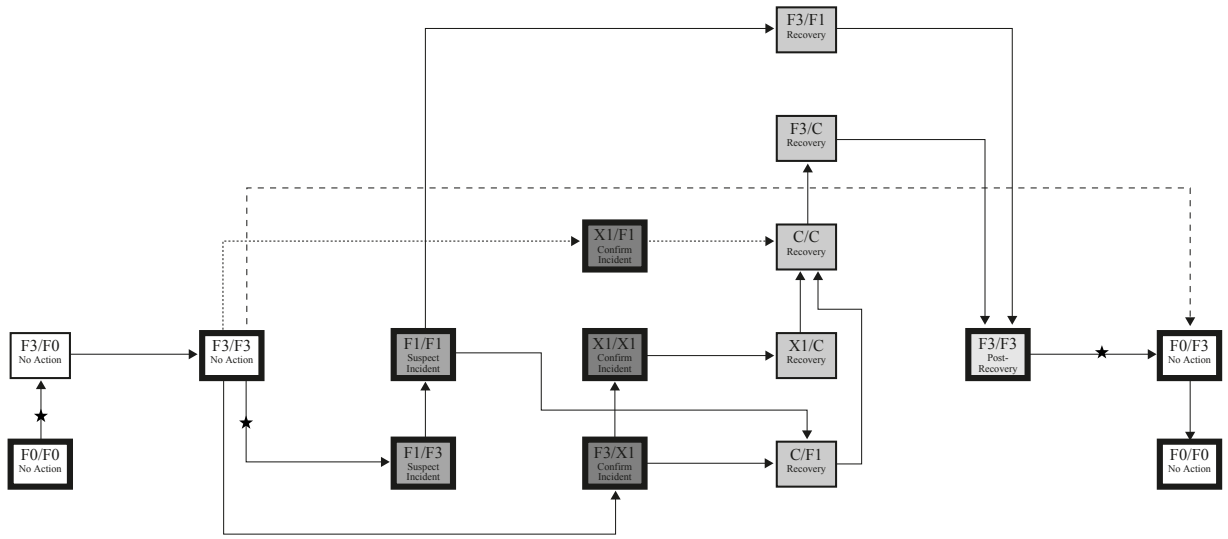

Fig. 7. Time progression chart (us/ds) for an incident as observed by a pair of detectors downstream of a recurring bottleneck. The legend is given in Fig. 2. Note that the incident may be upstream of the bottleneck.

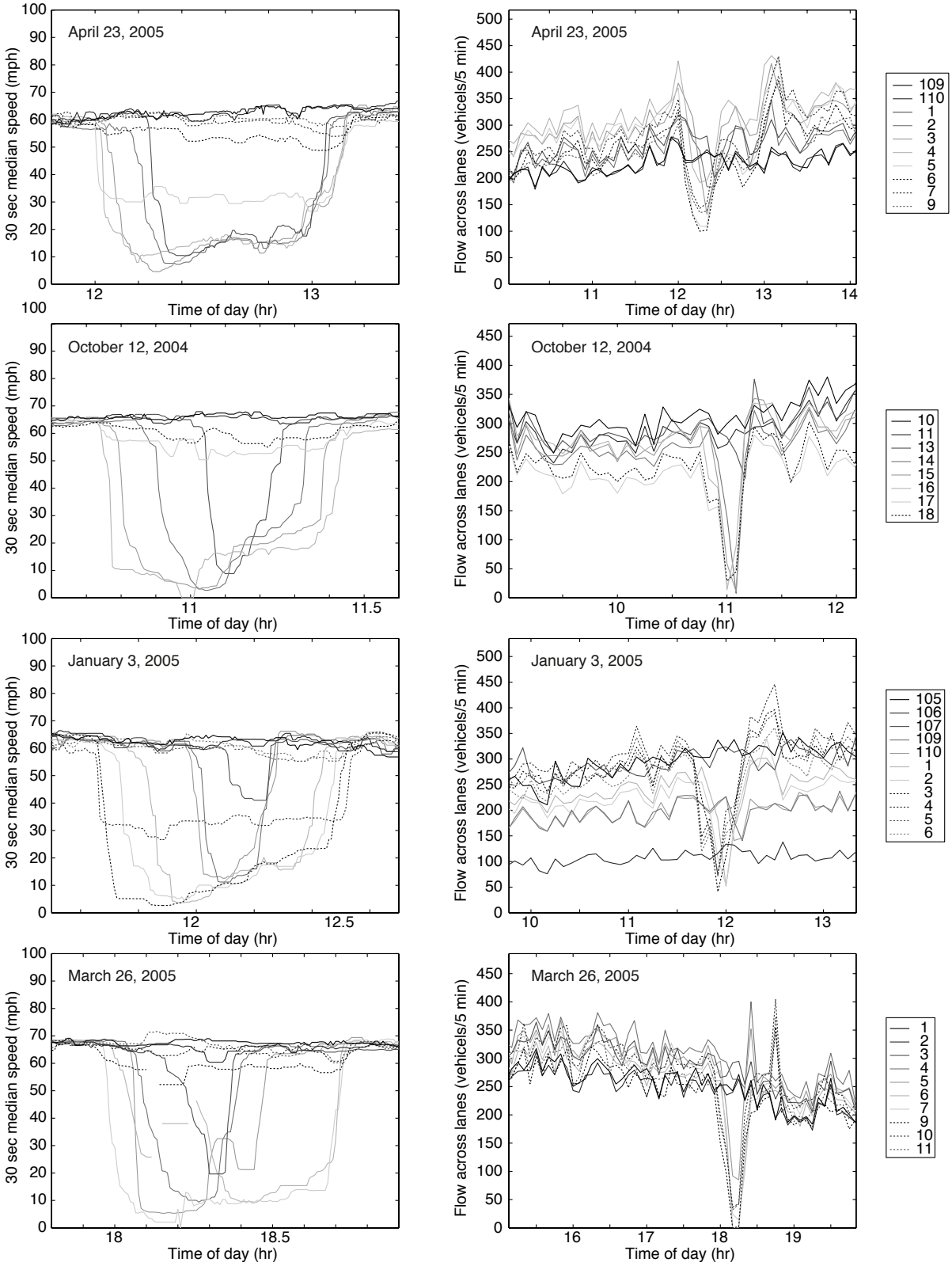

Fig. 8. Detector measurements during four incidents (one per row) illustrating the observed states, the left column shows speed and the right column flow.

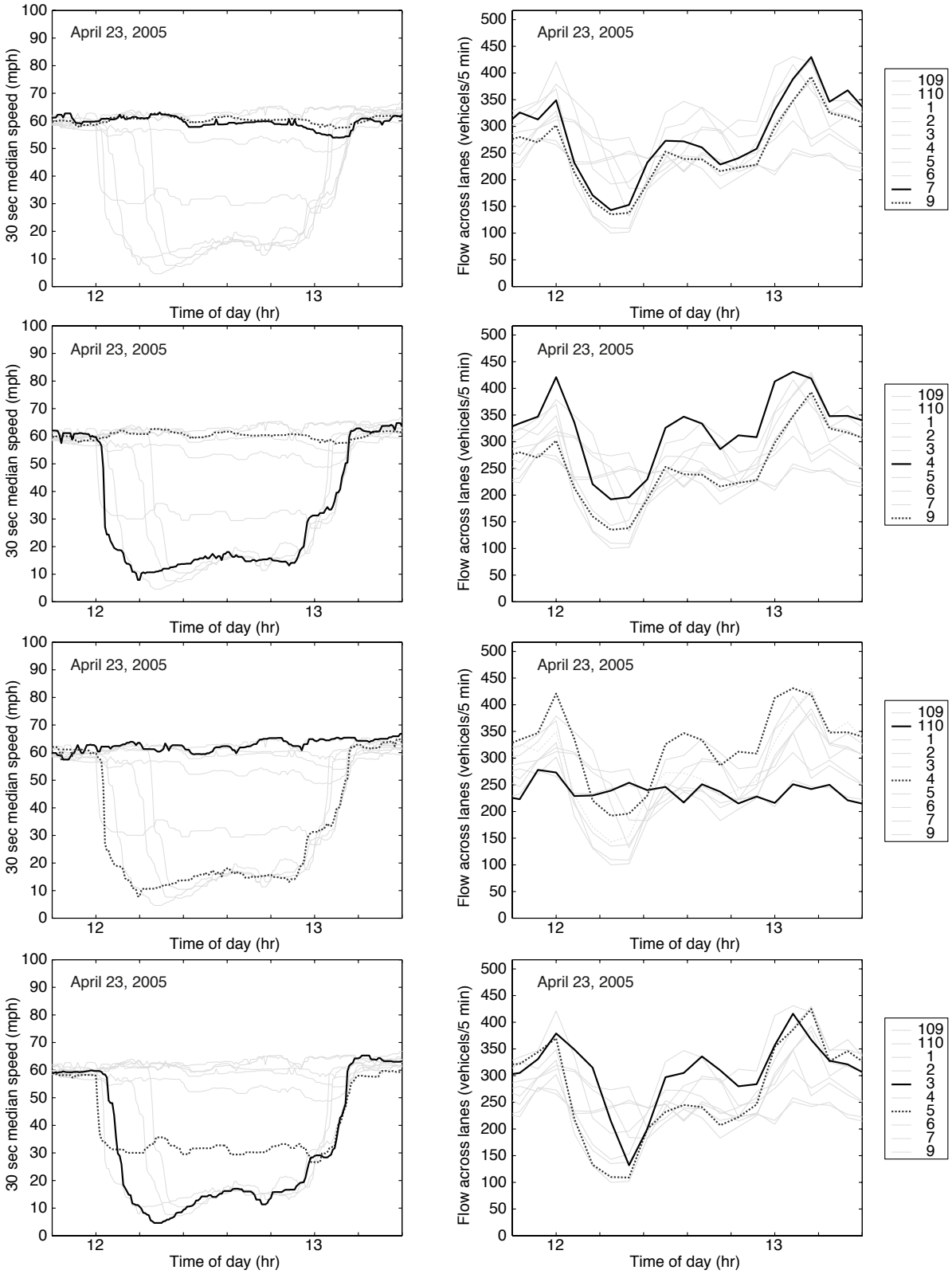

Fig. 9. Examining one incident in detail, highlighting observations from various pairs of stations (one pair per row) during the same incident illustrating the observed states. The left column shows speed and the right column flow.

Table 1. Control Points for an individual detector station on a homogeneous freeway based on time progression chart.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F2], [C,F2]	No Action	X				Update trends
[X1,C], [F1,C]	Detect Recovery	X	X			Update trends
[F2,F1], [F2,F0]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F2,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.

Table 2. Control Points for an individual detector station upstream of a recurring bottleneck based on the time progression charts.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F4], [F4,F3], [C,F4], [F4,X3], [C,X3], [X3,F0]	No Action	X				Update trends
[X1,X3], [X1,C], [F1,C]	Detect Recovery	X	X			Update trends
[F4,F1], [F4,F0], [X3,F1]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F4,X1], [X3,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.

Table 3. Control Points for an individual detector station downstream of a recurring bottleneck based on the time progression charts.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F3], [C,F3]	No Action	X				Update trends
[F1,F3], [F1,C], [X1,C]	Detect Recovery	X	X			Update trends
[F3,F1], [F3,F0]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F3,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.