On-Line Testing of the McMaster Incident Detection Algorithm Under Recurrent Congestion

FRED L. HALL, YONG SHI, AND GEORGE ATALA

The work reported here represents an elaboration of the logic for incident detection identified in previous work conducted at McMaster University. The improved incident detection logic has gone through three levels of testing; data from the Freeway Traffic Management System on the Queen Elizabeth Way in Ontario were used. An improved logic that could recognize and then ignore recurrent congestion and that could identify incidents that occurred within recurrent congestion was developed and tested off-line. The data used for this stage of the work consisted of 39 days from early summer 1990. The results were sufficiently promising that the algorithm was then installed on-line, and its results were reported to a file instead of to the system operator. Following a period of initial testing and revision to the algorithm and parameters, a major on-line test was conducted during 64 normal weekdays from March 12 to June 18, 1992. The algorithm detected 19 of 28 incidents, a 68 percent success rate. For the 19 incidents, the algorithm time to detection averaged 2.1 min after the time recorded in the operator's log; the median time to detection was 1 min later than for the operator. The false alarm experience was 20 in the 64 days of test, or one in every 6.4 operator shifts.

The purpose of this paper is to present the results of extensive testing of an idea for incident detection developed by Gall (*1*), a feasibility test of which was reported by Gall and Hall (*2*). Gall used a method to identify the cause of congestion by identifying the nature of flows downstream of a traffic queue as suggested by Wattleworth and Berry (*3*). This idea also formed the foundation for the California comparative algorithm (*4*). Gall's contribution was to frame the idea in terms of the congestion-detection logic suggested by Persaud and Hall (*5*). However, Gall was not able to test her idea for distinguishing between recurrent and incident-caused congestion on more than a few days of data. To properly show that it is a feasible incident-detection method, a more extensive test is needed.

There has been considerable previous work on incident detection on freeways, such as that by Payne (*6*), Dudek and Messer (*7*), Cook and Cleveland (*8*), and Levin and Krause (*9*). Stephanedes et al. (*10*) documented some of the difficulties of these existing algorithms in terms of the trade-offs between false alarms and detection rates. It is tempting to use the results from their work for a direct comparison with the results of the study reported in this paper, but because differences in the data bases may be important, caution is needed in the comparison. These differences will be discussed subsequently.

Three criteria are used to evaluate the algorithm: detection rate, false alarm rate, and time to detection. The detection rate is defined as the percentage of operator-identified incidents with an effect on traffic that were detected by the algorithm. False alarms are defined to be alarms recorded by the algorithm that do not correspond to an incident in the operator's log. The false alarm rate is the number of false alarms divided by the number of decisions made by the algorithm. That rate is calculated as the number of stations involved in the test multiplied by the number of time intervals per day in the test multiplied by the number of days in the test. This definition is consistent with earlier work (*4,9*), and with that used by Stephanedes et al. (*10*).

The first section of this paper describes the logic of the algorithm with reference to Gall's ideas. The second section describes the nature of the data used for the testing. The third section presents the results of off-line testing of the new version, on 39 days of data from the Queen Elizabeth Way (QEW) Freeway Traffic Management System (FTMS). Following successful off-line testing, the algorithm was implemented on-line, in the background. (Results were written to a file instead of being sent to the operators.) The final section of the paper reports the results of 64 days of on-line background testing, following some modifications to the algorithm and parameters.

THE LOGIC FOR INCIDENT DETECTION

Gall's idea was expressed in the form of a template drawn on a flow-occupancy diagram, defining four different states for traffic. Her initial template has been modified to create two templates, depending on the location of the detector station with respect to recurring bottlenecks such as those caused by heavily used entrance ramps (*11*). The template for a normal station, away from ramps, is shown in Figure 1; that for a station affected by recurrent congestion is in Figure 2.

For the stations not affected by recurrent congestion (Figure 1) the template is composed of 4 areas, which are divided by the lower bound of uncongested data (LUD), the critical occupancy (Ocrit), and the critical volume (Vcrit). Area 1, above the LUD, is uncongested data. The area below LUD and Vcrit and to the left of Ocrit is Area 2, one type of congested traffic operation. The area to the right of Ocrit and below Vcrit is Area 3, more heavily-congested traffic operation compared with the data in Area 4, which is below LUD

Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

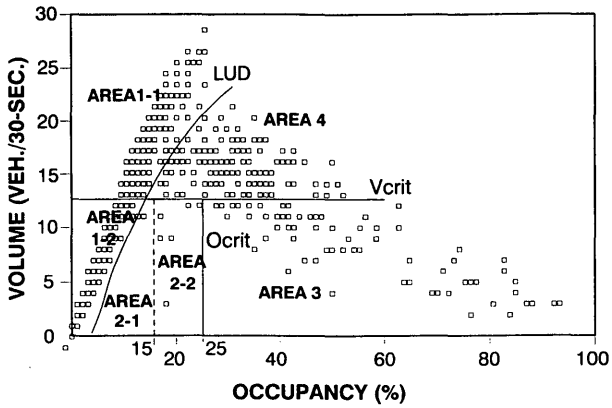

FIGURE 1 Template showing typical parameters for a normal station.

but above Vcrit. The division between Areas 3 and 4 is used for detecting incidents within congestion. Area 1 is further divided into Areas 1-1 and 1-2 by Vcrit, and Area 2 is divided into Areas 2-1 and 2-2 by Ocrit-10. These sub-areas are also used for detecting incidents within congestion.

For the template at the stations affected by recurrent congestion (Figure 2), the only difference from the template just described is that Area 4 represents queue discharge flow (QDF), the flow generated by recurrent bottlenecks. QDF is divided from other congested data by the lower bound of queue discharge flow (LQDF) and by a constant volume (labeled Qconst). Although Qconst is shown in Figure 2 as equal to Vcrit, they may take on different values.

Calibration procedures to establish the parameters displayed in these two figures have undergone considerable development since the methods described by Persaud et al. (*12*). The most important change is in the procedure for identifying the LUD line. That paper described a procedure based on subtracting a constant value from a quadratic function fit through the uncongested data by means of regression. Experience with more data has shown that the uncongested data do not display constant variance for volume as a function of occupancy. This means first that one of the fundamental assumptions of regression analysis is not met, and second that subtracting a constant

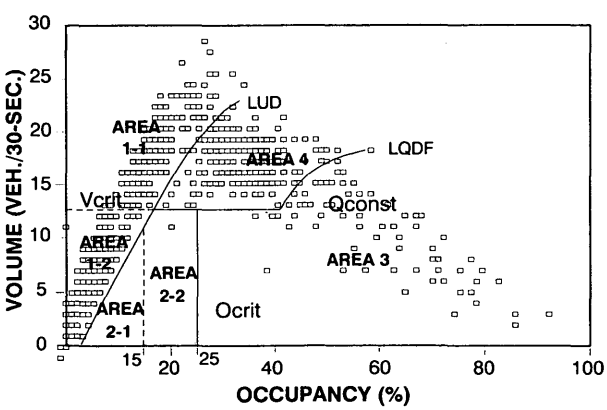

FIGURE 2 Template for a typical station affected by recurrent congestion.

value from the regression function will not reflect the correct location of the boundary of the data.

As a result, LUD is now fit directly to the boundary. The procedure is to start with a standard quadratic function, and then to plot the function against the data. Visual inspection shows if any aspect of the curve needs to be modified (intercept, slope, curvature), and the relevant coefficient for the equation is then adjusted. Although this procedure is not easily automated, experience with it can be gained quite quickly, and then complete calibration of a station may be accomplished in less than 2 hr, including acquiring the necessary data from tape, and when necessary separating the uncongested from the congested data.

The other parameters are also manually identified. Ocrit is simply the occupancy at which the highest observed volume occurs. Vcrit and Qconst may be harder to establish. To properly identify them requires some congested data (which not all stations have), and it must be possible to identify the volume that is normal within recurrent congestion, as opposed to that which occurs only during capacity reductions.

The raw data received from the loop detectors are compared with the appropriate template values. Instead of smoothing the data, a persistence check is used, such that the same state needs to be maintained for a certain number of intervals for a change of condition to be identified. For most of the testing, this persistence check has been set at three (30-sec) intervals.

Any data falling below LUD for longer than three intervals are considered to be congested. The algorithm then attempts to identify the cause. If the cause can be identified as one of two categories, then the congestion is not considered to be from an incident. If the cause cannot be placed in one of these categories, then the congestion is deemed to be caused by an incident. The simplest of the two nonincident categories is secondary congestion, representing the extension of primary congestion to the next station in sequence, either further upstream as a consequence of queue growth, or further downstream when queue discharge effects carry on further than expected.

The other main type of nonincident congestion is recurrent congestion in the vicinity of an entrance ramp. On the basis of experience with the QEW system and data, three stations in the vicinity of each entrance ramp have been defined as stations where recurrent congestion is a possibility: the first station upstream of an entrance ramp and the first two stations downstream of the ramp in the bottleneck. The downstream stations are included because at the time that a queue forms upstream of the entrance ramp, speeds in the queue discharge downstream from the ramp decrease considerably, with the result that the data fall below LUD (although they remain above the volume Qconst shown in Figure 1). Table 1 presents the combination of template states that may occur at the station being checked and at the downstream station and the resulting decision by the algorithm about incident presence. This table also identifies the method for distinguishing incidents from recurrent congestion at stations where recurrent congestion might first be seen, immediately upstream or downstream of an entrance ramp. For the first station immediately upstream of a bottleneck section, if the volume-occupancy data are in Area 2-2, Area 3, or Area 4 of the template (Figure 1), and the data at the downstream station

TABLE 1 Assessment Procedure for Stations Where Recurrent Congestion May Occur

DOWNSTREAM STATION	STATION BEING CHECKED					
VOL-OCC AREA*	1-1	1-2	2-1	2-2	3	4
1-1	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION
1-2	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
2-1	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
2-2	NO CONGESTION	NO CONGESTION	CONGESTION	INCIDENT	INCIDENT	INCIDENT
3	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION
4	NO CONGESTION	NO CONGESTION	CONGESTION	CONGESTION	CONGESTION	CONGESTION

* See Figures 1 and 2.

are in Area 1-2 or Area 2 (Figure 2), it is likely that an incident happened between the two stations. When there is no incident, the downstream data can be expected to be above QDF. If entrance ramp volume is high enough, however, the downstream data may be above QDF even after an incident has occurred, in which case the incident may not be detected.

It was also expected that the algorithm would be able to detect incidents that occurred within congestion. The logic for detecting an incident within congestion is that if the volume-occupancy data at Station i are in Area 2-2 or 3 while the data at Station $i+1$ are in Area 1-2 or 2, the algorithm will declare an incident at Station i. There are two main differences between detecting incidents from recurrent congestion and incidents that occurred within congestion. The first is that Area 3 in the template may be quite different, as seen in a comparison of Figures 1 and 2. The second is that data in Area 4 at the station being checked will not lead to an incident declaration. Thus Table 1 would be modified for this situation such that all categories under Area 4 would read "congestion," instead of some being "incident."

DESCRIPTION OF THE DATA FOR THE TESTS

Both the off-line and on-line testing of the algorithm were done with data from the FTMS on the QEW west of Toronto. The relevant parts of the FTMS are shown in Figure 3. The road is three lanes in each direction from the western limit until just after Station 25, where it becomes four lanes. Queues regularly form at the entrance ramps from Mississauga Road, Highway 10, and Cawthra Road. Ramp metering is used at these three interchanges, and at the next two west as well, although these other two do not often form congestion on the expressway.

The section from Erin Mills Parkway to the east side of Dixie Road was chosen for the off-line testing because of the daily recurrent congestion there. This section covers 15 eastbound detector stations (from 11 to 25) and is 8.8 km long. The vehicle detectors are installed in each lane at roughly 800-m intervals. Traffic volumes, occupancies, and speeds are recorded for these stations every 30 sec, 24 hr per day. For

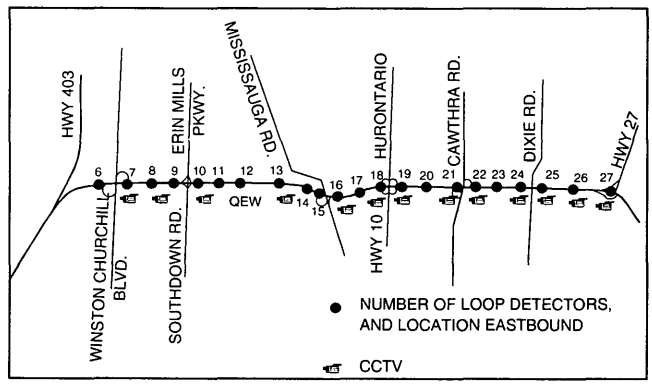

FIGURE 3 Configuration of the QEW FTMS (modified from map by Ministry of Transportation, Ontario).

the on-line testing, the application was extended to cover Stations 6 through 27.

The results from the algorithm are evaluated against the operator's log for both the off-line and on-line tests. During both test periods the FTMS was operated 16 hr per day, from 6:00 a.m. to 10:00 p.m. Hence the algorithm output can be evaluated for those hours only because there is no record against which to compare the algorithm between 10:00 p.m. and 6:00 a.m.

During these tests no automatic incident detection occurred at the QEW. The on-line test was carried on in the background (i.e., the results were written to a file, to be matched subsequently against the operator's log) instead of being reported directly to the operator. Operators relied on closed-circuit television (CCTV) and routine patrols of the Ontario Provincial Police for incident detection. As on most FTMSs that use CCTV, the remote-controlled cameras swivel almost 360 degrees, and tilt and zoom as well. With these features, and the placement of the 18 cameras (as shown in Figure 3), there is complete coverage of this section of the QEW by CCTV.

Many of the incidents noted on the operator's log are vehicles on the shoulder of the roadway. A critical question in the evaluation of the algorithm results is whether vehicles on the shoulder should be included in the data base of "inci-

dents." For example, for the 39 days of the off-line test, there were 152 events noted in the operator's log, of which only 28 occurred on the traveled lanes of the QEW. If the instances of vehicles on the shoulder affected traffic, then they should be included in the base for calculating the detection rate. If the vehicles on the shoulder did not affect traffic, they should be excluded because any algorithm can operate only on the basis of "observable changes in the traffic flow" [as Stephanedes et al. (*10*) wrote in excluding two incidents that caused no impact on traffic].

The operators report that it is unusual for a vehicle on the shoulder to cause sufficient disruption to traffic flow that there would be any noticeable effect on the data. In early testing, the authors examined the raw data for all missed events, including the vehicles on the shoulder. This inspection supported the operators' impressions: few of the vehicles on the shoulder affected traffic operations. Consequently almost all of the "vehicle on shoulder" events have been excluded from the evaluation of the algorithm.

OFF-LINE TESTING

Thirty-nine days of data from the Mississauga FTMS were used to test the algorithm off-line, from the period May 15 to July 15, 1990. The 39 days cover almost all weekdays during this 2-month period. Five weekdays were not included because the operator's log was not available at the time when the off-line testing was conducted. Some inclement weather conditions, such as a heavy storm, are included in the test. For the 39 days of off-line testing, the operator's log shows 28 incidents on the traveled lanes of the QEW. Twenty-nine incidents were declared by the algorithm during the times the operators were on duty. Fifteen incidents were detected by both the operator and the algorithm, 14 detected only by the algorithm, and 13 on the operator's log that were missed by the algorithm. Complete details of this testing appear in work by Shi (*11*).

Of the 15 incidents that appear in both the operator's log and the output of the algorithm, 2 were detected at the same time by both the operator and the algorithm; 3 incidents were detected 1 or 2 min earlier by the algorithm; and the remaining 10 incidents were detected 1 to 12 min later by the algorithm. For the 15 matched incidents, the mean time to detection was 2.2 min later for the algorithm (including the 12-min delay, or 1.5 min without it). The off-line tests were run with a persistence check of three (30-sec) intervals. Because the first interval is always needed for the congestion to appear, this three-interval persistence check has the potential to add 1 min of delay to the detection time. Any remaining delay in detection must be systemic (i.e., it takes that amount of time for the congestion to move upstream from the point of the incident to the next closest detector station).

The 14 incidents detected only by the algorithm are deemed to be false alarms. Dividing 14 by the 1,123,200 decisions made by the algorithm (16 hr per day, 60 min/hr, 2 intervals per min, or 1,920 intervals per day, per station, for 39 days at 15 stations) gives a false alarm rate of 0.0012 percent. There were problems with bad or missing data in the off-line test set, so one could estimate more conservatively that only 70 percent of those potential decisions actually were made by the algorithm, which would yield a false alarm rate of 0.0018 percent.

Of the 13 incidents recorded only by the operators, 3 were missed because the data were not collected around the incident location during the time the incidents happened. One incident identified by the operators as occurring east of Dixie Road probably happened beyond the test limits because there is no evidence for it in the data at Station 25 (which is the only station in this test east of Dixie Road). These four incidents should not be considered incidents for which the algorithm made an error because the data were not present for the test. For the remaining nine incidents recorded by the operator but missed by the algorithm, seven have relatively high volumes, from 8 to 21 vehicles per 30 sec and speeds from 75 to 120 km/hr at those stations with speed data. Hence these seven incidents had only a slight impact on traffic, if any. The remaining two incidents happened during the peak period. Congested data existed at the stations both upstream and downstream of the incidents before and during the incident times. These congested data do not show any difference compared with the data at the same time and location on previous incident-free days. Of these nine incidents that appear to have had no effect on traffic, the lane of occurrence is specified for only one (which occurred in the shoulder lane). It is possible that a number of these occurred on the shoulder. Conservatively, however, all nine of these incidents are considered to be incidents that should have been detected by the algorithm.

Fifteen incidents were detected successfully. The last nine incidents are considered to be the incidents missed by the algorithm. Hence at worst the detection rate is 15/24, or 62.5 percent. However if the seven incidents that had no effect on traffic are also omitted, then the detection rate is 88 percent (15/17).

The effectiveness at distinguishing between incident-caused and recurrent congestion may be evaluated by considering the number of wrongly classified occurrences of congestion. For the off-line test, each occurrence of recurrent congestion was also printed out; in total there were nearly 800 such occurrences in the 39 days. Three of the nine missed incidents can be matched by recurrent congestion declarations by the algorithm at a similar place and time, although the match is not exact. Five of the 14 false alarms occurred in locations that allow for the possibility that the algorithm identified as an incident something that might have been caused by recurrent congestion. Compared with the nearly 800 correct identifications of recurrent congestion and the 15 correctly identified incidents, these numbers confirm the acceptable performance of the algorithm in distinguishing the two types of congestion.

The ability of the algorithm to detect incidents within congestion is compared with its overall ability, during both congested and uncongested periods, in Table 2. (The congested period was taken to be 6:30 to 9:30 a.m. daily at all stations.) Nine of the 15 successfully detected incidents were detected within the congested period. With regard to false alarms, the percentage obtained by dividing the number of false alarms by the total number of incident declarations is a useful indicator in this instance. This percentage for the time within the congested period is considerably lower than that during the full off-line testing period, indicating that the algorithm is efficient at avoiding false alarms during recurrent

TABLE 2 Comparison of Incident Detection During Congestion and During Entire Period

	within congested period	during the entire period
(1) successful incident detection	9	15
(2) false alarm	5	14
(3) total declaration	14	29
(2)/(3)	35.7%	48.2%
(4) total incidents that might have been detected	14	24
(5) detection rate (1)/(4)	64.3%	62.5%
(6) total incidents that affected traffic	11	17
(7) detection rate based on incidents that affected traffic (1)/(6)	81.8%	88.2%

congestion. The detection rate within congestion is measured as defined earlier: the number of successful detections divided by the number that should have been found. Two values are reported: 64.3 percent, based on counting all nine missed incidents, and 81.8 percent, counting only those two that had a noticeable effect on traffic. Considering all three of these indicators, the ability of the algorithm to find incidents within congestion was judged to be more than satisfactory in the off-line tests.

A sensitivity analysis using four values for the persistence check was conducted with the same 39 days of data following the main off-line testing. The results (Figure 4) show that a value of three intervals is the most effective in terms of the trade-off between detection rate and false alarm rate and between false alarm rate and time to detection. It is interesting to compare this figure with Figures 3 through 7 in the paper by Stephanedes et al. (*10*). Their data cover 14 stations, spanning 5.5 mi (8.8 km); the QEW data come from 15 stations, also spanning 8.8 km. In making such a comparison, however, it is important to keep in mind the differences between the two data sets. Their data were "confined to the afternoon peak period (4:00–6:00 p.m.), since incident detection under moderate-to-heavy traffic conditions is of greatest importance for Advanced Freeway Management," whereas the data in

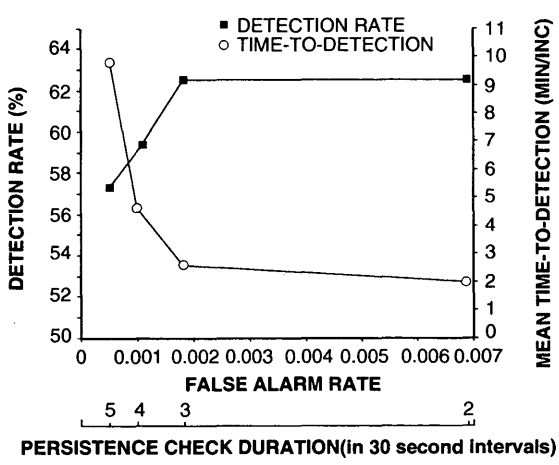

FIGURE 4 Effects of different persistence check durations on false alarm rate, detection rate, and mean time to detection.

the study reported in this paper were collected from 6:00 a.m. to 10:00 p.m.

With regard to mean time to detection, the McMaster algorithm results of roughly 2 min, shown in Figure 4, are two to four times longer than the 0.5 to 1.0 min results at a 60 percent detection rate in Figure 7 of the paper by Stephanedes et al. (*10*). This is perhaps attributable to the different time periods in the two studies; congestion takes longer to reach an upstream detector station in lighter traffic than in moderate-to-heavy traffic. The more striking feature of the comparison is the magnitude of the false alarm rates. At detection rates near 60 percent, Figure 4 shows false alarm rates in the vicinity of 0.001 percent. The results in the paper by Stephanedes et al. (*10*) at comparable detection rates range from 0.2 percent to 0.7 percent for comparative algorithms and are in the vicinity of 1 percent for time series algorithms. At first glance, this difference may also be attributed to the longer daily duration of the McMaster test, in that false alarms may be less likely during light traffic than during moderate-to-heavy traffic. However the rates in the Stephanedes test for the comparative algorithms are similar to those reported in the original literature, so not all of the difference can be due to the daily duration of the test. Hence there is some indication that the McMaster algorithm has an improved false alarm rate and perhaps a slightly increased time to detection (for a 70 percent detection rate).

ON-LINE TEST

Implementation of the algorithm for use on-line had the usual difficulties, but these were overcome in the space of a few months of preliminary on-line testing. One difficulty was that several stations needed to be recalibrated. Another was that two stations in the extended set for the on-line tests should have been defined as recurrent congestion stations and had not been. The final difficulty was that it proved necessary to look further downstream for those cases in which a particular station was missing data at the time congestion had been detected upstream of it. The problem of missing data remained an important one in the testing, with perhaps one-quarter of the data missing or suspect in general.

Testing of version 2.3 began March 12, 1992. Results from that time until June 18, 1992, a total of 67 days, are presented in this paper. Friday and Monday of the Easter holiday weekend are excluded, as is Monday, April 27, because there are no operator's logs for those days. Hence the test consists of 64 days, 16 hr per day, or 1,024 hr. Version 2.3 covers Stations 6 through 27, for a total of 22 stations. Because of treatment of the system end points, only 21 sections are covered. With data coming in every 30 sec, the test includes 2,580,480 decisions (21 sections 1,024 hr 120 data transmissions per hour per station).

During the 64 days of the test, the operators reported 230 incidents, of which 191 were identified as being vehicles on the shoulder. All but four of these have been ignored in evaluating algorithm performance. In one case there was a disturbance visible in the data at approximately the same location, 5 min before the operator reported the vehicle on the shoulder. This has been counted as a matched incident. For another three, the algorithm declared incidents at the

correct locations, but the time is later than that on the operator's log. Checking the stored data for these three confirms that there is a clear effect on the data consistent with an incident pattern. Because the only entry in the operator's log to equate these incidents to is the vehicle on the shoulder, these have also been counted as matched incidents.

An additional 10 incidents were identified as "debris on road," or as a truck losing "part" of its load. The operator's practice is to log any occurrence of material on the roadway, whether or not it affects traffic. Consequently the stored data for these 10 incidents were checked to identify the effect on traffic. In 7 of the 10, no visible effect on traffic occurred. In another two, the system was not reporting good data at the relevant stations at the time that the debris was noted. Hence only one of the "debris on road" incidents was of a type that might affect traffic and therefore be visible to an incident-detection algorithm.

For an additional four incidents, the operator's log unfortunately provides insufficient information about the incident to know whether the disabled vehicle is on the shoulder or in one of the traveled lanes. In one of the four, the vehicle might even be on an entrance ramp. The stored data for all four were checked. Three showed no effect on traffic; one had bad data at both stations in the vicinity of the logged location. Hence none of these four logged events would be found by an algorithm.

Thus of the 230 reported events in 64 days, 187 have been removed from the test because they occurred on the shoulder and had no effect on traffic that was noticed by the algorithm, another 9 have been removed as occurrences of debris on the road that did not affect traffic (or for which there was no recorded data), and 4 have been removed because of insufficient information about them (together with no effect on the available data). That leaves a total of 30 incidents on which the algorithm is to be tested. The algorithm identified 19 of these. The stored data were scanned for the 11 missed incidents, and it was found that for an additional 2 incidents, the system was not recording good data at the necessary stations. Those incidents should also be removed from the test set. Hence the algorithm correctly identified 19 of 28 possible incidents, for a detection rate of 68 percent.

With regard to the difference in time to detection between the operator's log and the algorithm, 3 of the 19 matched detections occurred at the same time, 11 were found later by the algorithm, and 5 were found earlier. It is useful to separate the incidents detected on the shoulder from those that occurred on the traveled roadway. Four are identified by the operators as being solely on the shoulder; another two are partially (in time or space) on the shoulder. Of the four completely on the shoulder, the algorithm was 5 min earlier on one, and 1, 2.5, and 11.5 min later on the others. For the two partly on the shoulder, the algorithm was 12 min later on one, and at the same time as the operator on the other. Thirteen incidents occurred solely in the traveled lanes of the roadway, and for these incidents detection time differences were 3, 1.5, 1, and 1 min earlier; two at the same time; and 0.5, 1, 2, 3, 3.5, 4, and 11 min later for the algorithm. The mean time to detection for all incidents is 2.1 min, and for only those on the traveled lanes is 1.4 min. The median difference in detection times for all incidents is 1 min later for the algorithm; for those on the traveled lanes only, it is 0.5 min.

The false alarm rate for the on-line test is similar to the off-line result. There were only 20 false alarms during the 64 days of the test. That is an average of one every 6.4 operator shifts. On the basis of the 2,580,480 decisions identified earlier, this is a false alarm rate of 0.00078 percent. In future work, it might be appropriate to perform sensitivity testing of some other parameters, given the possibility of allowing the false alarm rate to increase slightly if more of the incidents could be captured.

These results may also be compared with those obtained by Stephanedes et al. (*10*), recognizing the differences in the data discussed earlier, as well as with earlier operating characteristic curves such as those in work by Payne and Tignor (*4*). The first point is that the false alarm rate remains two to three orders of magnitude lower in these on-line results. For a detection rate of roughly 68 percent Stephanedes et al. (*10*) reported false alarm rates of 0.1 percent to 0.3 percent. Mean times to detection remain marginally better (by 1 min or less) for most of the algorithms reported by Stephanedes et al. (*10*) than for the McMaster algorithm. Payne and Tignor (*4*) do not report detection times, but the false alarm rates they report are a similar order of magnitude to those in the paper by Stephanedes et al. (*10*). At the closest reported detection rates to that in the McMaster algorithm, the false alarm rates reported by Payne and Tignor (*4*) ranged from 0.13 percent to 0.8 percent.

In summary, the on-line test results suggest that the McMaster algorithm has an excellent false-alarm performance and acceptable detection rates, but that the false alarm performance may have been achieved at the expense of the mean time to detection. For small systems such as that on the QEW and the one used in testing done by Stephanedes et al. (*10*), it may well seem that the false alarm rate for the McMaster algorithm is excessively low because it results in only one false alarm in several operator shifts. However there are larger FTMSs in operation in North America. The system under development on Highway 401 in Toronto, for example, has 136 detector stations, as opposed to the 22 in the QEW on-line test. The system being designed for the Boston Central Artery and Third Harbor Tunnel may have as many as 150, and the Chicago area systems already have 1,800 mainline detectors (*13*), which implies 600 stations, if there are an average of three lanes per station with detectors. Other systems in the design stage with even more detectors (3,000 in Phoenix and 7,000 in Fort Worth) are reported in *Transportation Research Circular 378* (*13*). On a system the size of Chicago's, with 30 times as many stations as on the QEW, the false alarm rate for the McMaster algorithm would mean an average of roughly 4.5 false alarms per shift. Taking the best of the results from the Payne and Tignor report (*4*), a false alarm rate of roughly 0.1 percent, the number of false alarms in a Chicago-size system would be well over 600 per shift, which means more than 1 per minute. False alarm rates need to be as low as those found with the McMaster algorithm for automatic incident detection to be feasible on large systems.

The next step in development of the algorithm for the QEW is to develop version 2.4, which will report to the operators. That will allow for a different on-line test because it will show how many incidents are noticed by the operator before they are detected by the algorithm. However it will no longer be possible in such a test to show the algorithm detecting an

incident before the operator because when the algorithm identifies it the operator will know of it, too. It will also allow for a better test of false alarm rates because it may well be that some of the alarms identified as such in this paper were in fact real events that the operators did not notice on the CCTV. In addition to this last step for the algorithm on the QEW system, work has begun on adapting the algorithm for the Highway 401 system, with its more complex geometry and larger number of detector stations.

ACKNOWLEDGMENTS

The work described in this paper has been supported by contracts from the Ministry of Transportation in Ontario. The ideas expressed in the paper represent the opinions of the authors and not of the sponsor. The assistance of Mark Fox, David Tsui, and Phil Masters, of the Ministry's Freeway Traffic Management Section, has been most appreciated. Thanks are also due to Colin Rayman, former head of the section, who had the foresight to take a chance on an idea, and to Peter Korpal, the current section head, who has displayed the confidence to continue that work. The financial support of the Natural Sciences and Engineering Research Council of Canada in developing the original idea should also be acknowledged, as should their support for George Atala through an undergraduate student research stipend. Special mention is due Lisa Aultman-Hall, who worked on several aspects of the algorithm as an undergraduate student before moving to Queen's University for her graduate work.

REFERENCES

1. A. I. Gall. *Distinguishing Between Incident and Recurrent Congestion.* M. Eng. thesis. Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada, 1990.
2. A. I. Gall and F. L. Hall. Distinguishing Between Incident Congestion and Recurrent Congestion: A Proposed Logic. In *Transportation Research Record 1232,* TRB, National Research Council, Washington, D.C., 1990, pp. 1–8.
3. J. A. Wattleworth and D. S. Berry. Peak Period Control of a Freeway System—Some Theoretical Investigations. In *Highway Research Record 89,* HRB, National Research Council, Washington, D.C., 1965, pp. 1–25.
4. H. J. Payne and S. C. Tignor. Freeway Incident Detection Algorithms Based on Decision Trees with States. In *Transportation Research Record 682,* TRB, National Research Council, Washington, D.C., 1978, pp. 30–37.
5. B. N. Persaud and F. L. Hall. Catastrophe Theory and Patterns in 30-Second Freeway Data—Implications for Incident Detection. *Transportation Research A,* Vol. 23A, 1989, pp. 102–113.
6. H. J. Payne. *Traffic Flow Theory with Application to Freeway Ramp Control and Incident Detection.* Technology Service Corporation, 1974.
7. C. L. Dudek and C. J. Messer. Incident Detection on Urban Freeways. In *Transportation Research Record 495,* TRB, National Research Council, Washington, D.C., 1974, pp. 12–24.
8. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp. 1–11.
9. M. Levin and G. M. Krause. Incident Detection Algorithms—Part 1: Off-Line Evaluation, Part 2: On-Line Evaluation. In *Transportation Research Record 722,* TRB, National Research Council, Washington, D.C., 1979, pp. 49–64.
10. Y. J. Stephanedes, A. P. Chassiakos, and P. G. Michalopoulos. Comparative Performance Evaluation of Incident Detection Algorithms. In *Transportation Research Record 1360,* TRB, National Research Council, Washington, D.C., 1992.
11. Y. Shi. *Improvements to the McMaster Incident Detection Algorithm.* M. Eng. thesis. Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada, 1991.
12. B. N. Persaud, F. L. Hall, and L. M. Hall. Congestion Identification Aspects of the McMaster Incident Detection Algorithm. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 167–175.
13. *Transportation Research Circular 378: Freeway Operations Projects Summary.* TRB, National Research Council, Washington, D.C., Sept. 1991.

Publication of this paper sponsored by Committee on Freeway Operations.