Congestion Identification Aspects of the McMaster Incident Detection Algorithm

BHAGWANT N. PERSAUD, FRED L. HALL, AND LISA M. HALL

Automatic detection of incidents on a freeway traffic management system (FTMS) can be thought of as two distinct tasks. The first is the detection of congestion and the second is the determination of whether or not the congestion is incident related. Development and testing of the congestion detection task for a new incident detection algorithm for an FTMS are described. A new logic has recently been proposed for the determination of whether congestion is recurrent or incident caused. The congestion detection logic uses flow, occupancy, and speed (if available) from a single station to automatically detect congestion near that station. This logic has been subjected to on- and off-line tests on a system on which congestion is largely incident related. The results show a good false-alarm rate and a high detection rate, with some incidents detected earlier than they were identified by FTMS operators.

Automatic incident detection is one of the most important ingredients of a freeway traffic management system (FTMS). Of the current incident detection approaches used in North America, the most popular appears to be the comparative (California-type) algorithms (*1*) in which specified differences in traffic operations between two adjacent detector stations indicate the presence of an incident. Another approach (*2–4*) also bases detection on traffic operations at a single station.

An ideal logic would detect all incidents immediately on occurrence and would not produce false alarms when there are no incidents. Recent discussions with FTMS managers have emphasized the need for algorithms performing closer to this ideal than those now in existence. This desire for a better algorithm is not surprising in view of the fact that no algorithm has been consistently superior in evaluations reported in the literature. For example, the California-type algorithms developed by Technology Services Corporation (TSC) (*1*) did not perform as well during independent off- and on-line tests (*5,6*) as they did in initial off-line tests (*1*). More recently, Peat, Marwick, Mitchell & Co. (*7*) field-tested three of the better TSC algorithms during a 5-month period and found a low (45 percent) detection rate and what they thought to be a fairly high ratio of false alarms to detections of 11 to 1.

As Dudek et al. (*3*) point out, "the results of incident-detection model capabilities reported in the literature must be placed in the proper perspective," because these capabilities are significantly affected by factors such as detector station spacing, operating conditions, duration of the incident, and the location of the incident relative to the detector stations. These factors appear to be especially critical for comparative algorithms that expect an increase in occupancy upstream of an incident and a drop downstream. For instance, an incident occurring just upstream of a detector station may not cause the anticipated drop in downstream occupancy because vehicles might still be traveling slowly at the downstream station. Also, geometric factors such as grade, lane drops, and ramps between stations might cause uncongested traffic operations to mimic the pattern recognized by these algorithms.

The comparative algorithms remain popular despite the fact that many of their weaknesses seem to be overcome by single-station algorithms. The prevailing wisdom seems to be that a single-station logic tends to generate excessive false alarms. For instance, as part of the project to develop the comparative algorithms, TSC evaluated several single-station algorithms and reported that "the multiplicity of incident indications they generated was deemed to be an operational disadvantage" (*1*). Cook and Cleveland (*4*) found, by contrast, that "algorithms that used traffic data at two adjacent stations were less effective than those that used data from just one station. . . ." However, they reported with respect to their own single-station algorithms (double exponential smoothing) that "neither the variables used nor the sudden change in values over time were exclusive to incident situations as opposed to incident-free operations." On balance, it seems that the two main problems in developing effective single-station algorithms are the complexity of distinguishing incident from nonincident congestion and the difficulty of controlling for non-incident-related changes in traffic operation because of factors such as weather.

Development and testing of the congestion detection aspect of an approach—the McMaster incident detection algorithm—that appears to retain the advantages of single-station logic while overcoming some of the weaknesses are described. The second aspect of the logic, distinguishing between incident and nonincident congestion, has been described by Gall and Hall (*8*). The congestion detection logic can be used without modification when nonincident congestion is rare, as is the case for the Burlington Skyway FTMS used for on-line testing. The intention is to incorporate a facility for distinguishing between recurrent and incident-related congestion that has been described by Gall and Hall (*8*). With that logic, congestion is detected, but the FTMS operator would have to confirm through video surveillance or other means whether the congestion is caused by an incident.

To date, testing of the algorithm has gone through three stages, all on the Burlington Skyway FTMS (Figure 1). At the time of the on-line testing, 12 detector stations were operational. On that part of the Skyway system, there was no

B. N. Persaud, Department of Civil Engineering, Ryerson Polytechnical Institute, 350 Victoria Street, Toronto, Ontario, Canada M5B 2K3. F. L. Hall and L. M. Hall, Traffic Research Group, Department of Civil Engineering and Department of Geography, McMaster University, Hamilton, Ontario, Canada L8S 4L7.

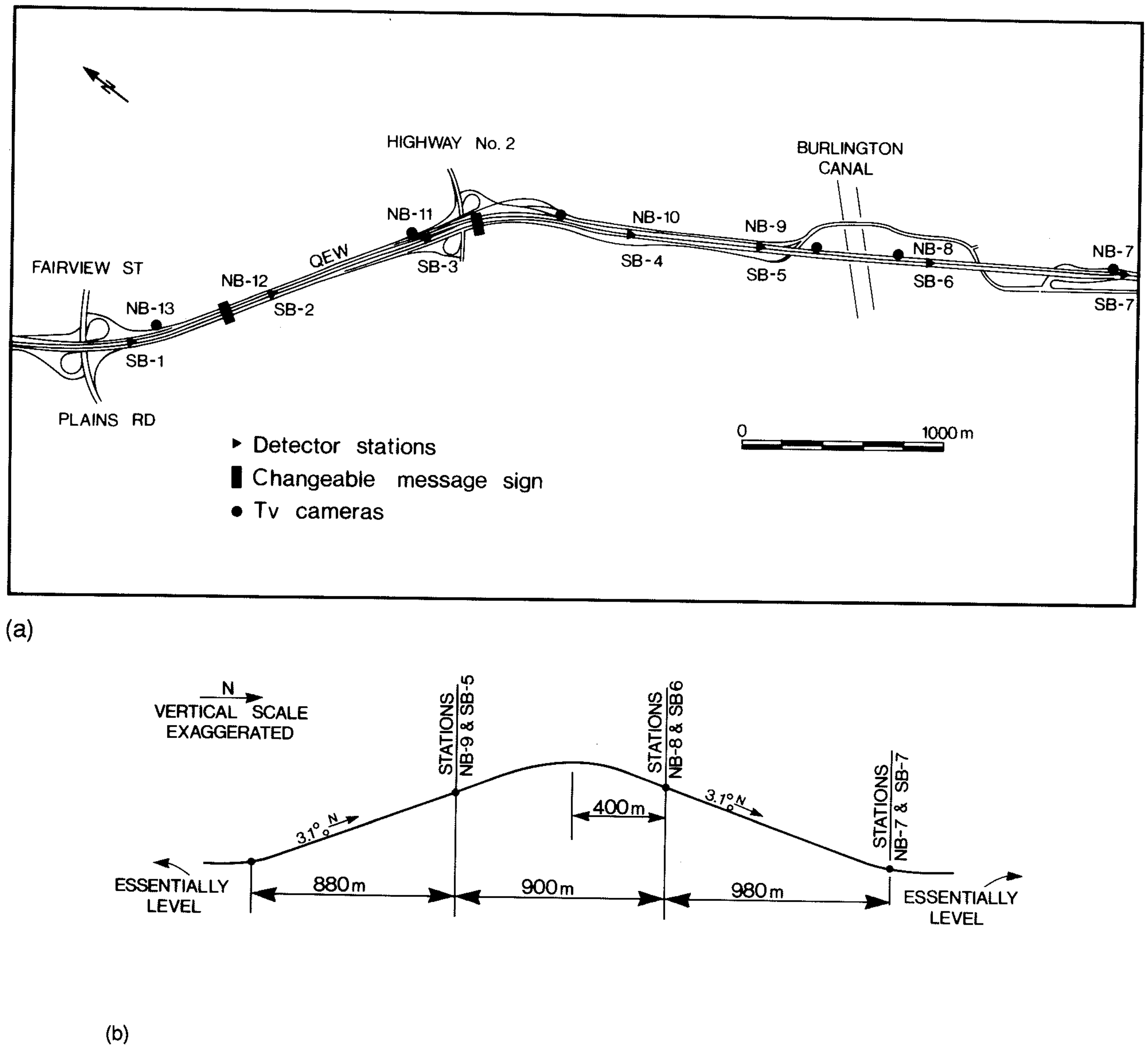

FIGURE 1 Plan (a) and profile (b) views of the Burlington Skyway, showing detector stations and other FTMS elements.

recurrent congestion, because of a doubling of capacity in summer 1988. Any congestion tends to be incident caused. Hence, discussion of the testing refers to incidents, even though it is only the congestion-finding logic that is under test. Each detector station on the Skyway measured and recorded 30-sec averages of speeds, flows, and occupancies in each lane. Speed was measured by paired-loop detectors. The first round consisted of off-line tests on data from the summers of 1986 and 1987, plus some from November 1986. These tests were on the Skyway where there were two lanes in each direction. In August 1988, a parallel bridge was opened to traffic, resulting in a change from two to four lanes in each direction for almost the full system. The second round was an on-line test of the initial algorithm, conducted during May 1989. In that test, the results from the algorithm were not directed to the system operators, but instead were simply written to a file to be inspected later. On the basis of the results of this test, a number of modifications were made to the details of the algorithm, and the third phase of testing was begun in June 1989, again on line, and with results going to a file for later inspection, rather than to the operator.

DESCRIPTION OF NEW LOGIC

Conceptual Framework

The conceptual basis for the proposed logic was suggested over 20 years ago by Athol (*9*) and has been expanded recently by Persaud and Hall (*10*), who elaborated on a catastrophe theory model to describe the relationship between flow, occupancy, and speed. The essence of the logic is shown in Figure 2, which is a plot of 30-sec flow-occupancy data for the median lane at a level station (NB-7) on the Burlington Skyway. The data were observed before, during, and after several incidents downstream of that station. In Figure 2, uncongested flow-

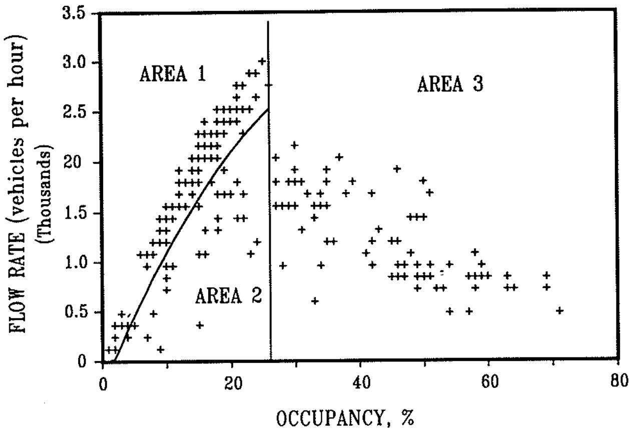

FIGURE 2 Three flow-occupancy areas shown with Skyway Station NB–7 30-sec data.

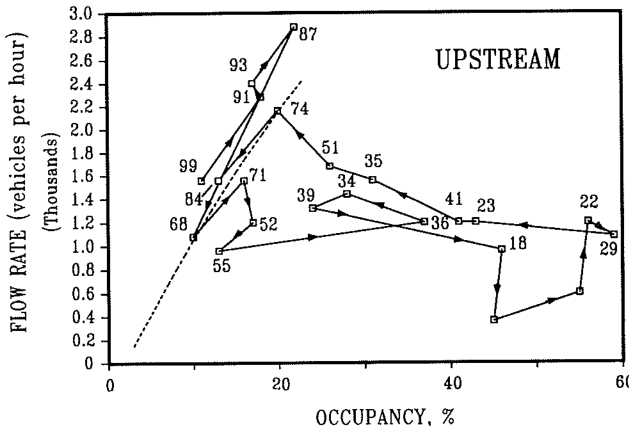

FIGURE 3 Median-lane 30-sec flow-occupancy-speed sequence at Station NB–7 for a downstream incident.

occupancy operation occurs in Area 1. An incident causes movement from Area 1 to congested operation in Areas 2 or 3. The data also suggest that this movement is accompanied by a drop in speed to values below 70 km/hr. The boundary between Areas 1 and 2 was drawn on the basis of the off-line calibration process that is described later, whereas the boundary between Areas 2 and 3 is simply a vertical line at what appears to be the maximum occupancy for uncongested operation (sometimes called critical occupancy).

The principle of the basic version of the congestion detection logic is that a congestion flag can be given by either operations in Area 2 or 3 or a slow speed (or, of course, by both). If a flag is given for *P* consecutive periods, congestion is present, and hence a potential incident is indicated. The logic has been designed and tested with a system that both provides speed and flow-occupancy data. The logic should work with flow-occupancy data alone, but that possibility has not yet been tested. It will be important to test that possibility for systems without paired-loop detectors, because the results of a related analysis (*11*) indicate that speed calculated from single-loop detector data is too unreliable for use in automatic incident detection. The logic is also more efficient if it uses data for a single lane in which there are few or no trucks, as increasing heterogeneity of the vehicle fleet tends to reduce the clarity of the uncongested flow-occupancy function. In these tests, only median-lane data have been used, although work is also underway to investigate the use of a second left-hand lane.

In order to fully appreciate the conceptual basis for the logic, it is instructive to focus on the pattern upstream of a specific incident. Figure 3 isolates the time period during which an incident approximately midway between NB–7 and NB–8 (see Figure 1) affected operations at NB–7. In this figure, the boundary line is taken from Figure 2 while the speed corresponding to each flow-occupancy point is printed beside the point, and the time sequence of the data is indicated by the arrows. The letter *A* denotes the data observed at NB–7 when the incident was reported. It became obvious from Figure 3 and corresponding figures for other incidents that, for several time intervals at the start of an incident, operations lie in Area 2 but at occupancies lower than what is normally considered to be a threshold occupancy for uncongested operation. For the data shown in Figures 2 and 3, this critical occupancy is about 26 percent. Hence, to assume that all low-occupancy data represent uncongested operation, or (phrased another way) that it is only after a critical occupancy has been reached that congestion is present, is not valid.

Locating the Congestion

One of the distinctive features of the logic is that it also detects congestion that has happened upstream of a station. In fact, if the station is just downstream of the incident, operations at that station might be affected first, before the queue reaches the station upstream of the incident, as was the case for several of the incidents examined. It is also conceivable (and appears to happen, albeit infrequently) that queue buildup is so slow the queue never reaches the upstream station. Under such a circumstance, neither a comparative logic nor a logic based simply on high occupancies will find the incident, although it can still be identified with this new logic at the downstream station.

On reflection, and after noting observations by Dudek and Messer (*12*), the detection of congestion downstream of an incident should not be surprising. Even if vehicles start accelerating on passing an incident located between two stations, they might still be traveling slowly when they pass the downstream station. That this condition is in fact so is evident from Figure 4, which corresponds to the incident on which Figure 3 is based, but pertains to operation at Station NB–8, which is the first station downstream of the incident. The downstream speeds have clearly decreased. The data corresponding to the reported start of the incident is again marked by the letter *A*. For this station, the calibrated boundary between Areas 1 and 2 is slightly different from that for Station NB–7 (see Figures 2 and 3). The movement to Area 2 in the downstream data (Figure 4) might have occurred as much as 3 min earlier than for the upstream data (Figure 3). The downstream pattern naturally depends on the distance of the incident to the detector station, but the common element in an examination of the downstream pattern for several incidents is that, as shown in Figure 4, operations tend to have a speed drop and to move to Area 2 and lower volumes at

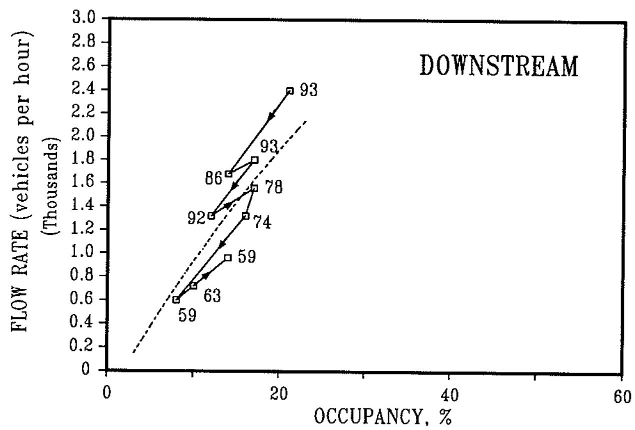

FIGURE 4 Median-lane 30-sec flow-occupancy-speed sequence at Station NB–8 for an upstream incident.

the start of an incident. The closer the incident is to the downstream station, the larger is the speed drop and the more to the right of Area 2 is the flow-occupancy value.

Other Features of the Logic

In order to translate the conceptual framework into an operational detection logic, it was necessary to incorporate two other features, namely, the treatment of incident data at adjacent stations after an incident has been declared, and the identification of the end of an incident.

For all of the incidents examined in initial off-line tests, an incident affected operations at more than one station. In some cases, operations at two or more upstream stations were affected as the queue traveled upstream. In other cases, downstream operation was affected as well. In order to prevent an incident alarm being given falsely while incident management resources are focused on an incident at an adjacent station, the current implementation of the algorithm assumes that as long as an incident exists at one station, an alarm at an adjacent station is related to the first incident; under this assumption, only a secondary alarm is given for the purposes of incident management. Because of the distinctive patterns upstream and downstream of an incident, it will be possible to replace this assumption with additional tests later.

The logic to detect the end of an incident at a station is the reverse of the incident start detection logic. In all cases, it is necessary for speed to return to a higher level and for operations to return to Area 1, and for these levels to be maintained for a number of consecutive periods (Q) before the incident state is declared over.

CALIBRATION

The main tasks in calibration for this new algorithm are distinguishing between congested and uncongested flow-occupancy regions (i.e., defining the boundary between Areas 1 and 2 in Figure 2); identifying a speed threshold to distinguish congested from uncongested speeds; and establishing the duration of the persistence checks (P and Q). The trade-offs in calibration are simple. A clockwise rotation about the origin (0,0) of the boundary between Areas 1 and 2 will result in fewer false alarms, but this result is achieved at an increase in detection time and the possibility of missing incidents with minor impact on traffic. Lowering the speed threshold or raising the value of P produces similar outcomes.

Calibration of the first two items has gone through several stages as testing has proceeded. Part of the task in refining the calibration procedure is to develop automated methods for doing the calibration, so that it can be accomplished easily whenever new stations are brought on line. The discussion that follows treats early methods only briefly, and focuses on the final method for identifying the speed threshold and the flow-occupancy boundary. The persistence check has been left at $P = 2$ (i.e., two 30-sec intervals) for the beginning of an incident and at $Q = 3$ for the end of an incident during this testing, although the results suggest slightly higher values will be more effective in practice. The other topic discussed in this section is that of adapting the boundary between the flow-occupancy regions for varying operating conditions, such as might be caused by the onset of a heavy rainstorm.

Defining the Boundary Between Uncongested and Congested Data

Given that identification of the flow-occupancy boundary would require mathematical estimation, the speed boundary was used as the first-cut identifier of congested versus uncongested data points. (This procedure will not be available for single-loop systems, although there will be other ways to identify uncongested data.) For the off-line tests, the speed boundaries were set arbitrarily after inspection of large quantities of data. Uncongested data were deemed to be those occurring at speeds above 70 km/hr, and data below 60 km/hr were defined to be congested. Operations at speeds between 60 and 70 km/hr were left undefined.

During the on-line tests, the lower limit of uncongested speed varied by location along the roadway, and an automated procedure was developed to identify this speed, as follows. At locations where congestion occurs regularly, frequency diagrams for speeds have a near-normal distribution at the higher speeds, plus a variety of much lower frequencies of midrange speeds. Hence, the mean and standard deviation of speeds were calculated for the station, and an initial threshold set as the mean minus three standard deviations. Observations below this limit were eliminated from calculation, and the mean and standard deviation were recomputed. This procedure was repeated until two consecutive threshold speeds differed by less than 0.5 km/hr. The resulting values are presented in Table 1. For all stations, the value calculated agrees visually with the point on the frequency graph where the high-speed normal distribution would most likely meet the horizontal axis.

The approach used to identify the boundary between Areas 1 and 2 (Figure 2) was to fit a function to the full set of uncongested data and then to determine the lower bound on the set from that function. The boundary lines in Figures 2–4 were based on earlier work (*13,14*) in which it was suggested that uncongested flow-occupancy operation at a station can be described for specified operating conditions by a function

TABLE 1 CALIBRATION RESULTS FOR 90-DAY ON-LINE TEST

Station	lower limit of uncongested speed (km/h)	equation for 30-second volume as a function of occupancy (x)	constant difference
SB1	91	$0.4 + 1.96x - .031x^2$	3.4
SB2	93	$0.6 + 1.67x - .016x^2$	3.5
SB3	92	$0.6 + 1.69x - .016x^2$	3.5
SB4	84	$0.7 + 1.16x - .006x^2$	3.5
SB5	96	$0.7 + 1.11x - .001x^2$	2.5
SB6	85	$0.5 + 1.31x - .011x^2$	2.7
SB7	92	$0.7 + 1.56x - .013x^2$	2.5
NB7	91	$0.7 + 1.29x - .007x^2$	2.9
NB8	78	$0.7 + 1.06x - .002x^2$	2.5
NB9	80	$0.6 + 2.18x - .063x^2$	4.7
NB10	61	$0.4 + 1.60x - .024x^2$	4.3
NB11	81	$0.6 + 1.58x - .020x^2$	4.0
NB12	84	$0.8 + 1.54x - .030x^2$	5.4
NB13	77	$1.0 + 1.27x - .024x^2$	5.4

Note: the constant difference is subtracted from the volume calculated by the function, to identify the lower bound on uncongested flows for a given occupancy.

of the form

$$\text{Flow} = a * (\text{occupancy})^b$$

The off-line test and initial on-line tests used this function, and set the lower bound as a specified percentage of the value of the function. Values between 70 and 80 percent were used, with 75 percent applied in the on-line tests.

Close inspection of the on-line test results indicated that this function often overestimated at higher flows. Originally, the function had been selected to force the relationship through the origin—zero occupancy implies zero flow. However, this relationship turns out to be empirically incorrect. Zero occupancy is associated with flows of 0, 1, or 2 vehicles per 30-sec interval frequently, and sometimes even with flows of 3 vehicles per 30-sec interval. Figure 5 shows this with the frequency of occurrence of flow-occupancy data points for uncongested data for 1 week for Station NB–7. Given truncation rather than rounding of the initial raw measurements, nonzero flows for zero occupancy turn out to be quite plausible. To force the empirical function through the origin is therefore not appropriate, although it would seem to be theoretically correct. Hence, other functional forms were considered.

At all stations, much of the data occurs at occupancies of 0 or 1 vehicle (e.g., see the 3,353 observations in the week of data in Figure 5). Unless some corrective measures are taken in the estimation procedures, this nonuniform distribution of occupancy values will not provide the best estimate for an equation. Observations with an occupancy of 0 were omitted from the estimation of the flow-occupancy function. Any value of flow associated with these occupancies would be acceptable. Indeed, for most stations the same holds true for occupancies of 1 or 2 percent as well. Hence, the lower bound of uncongested flow, as a function of occupancy, will only become positive at an occupancy of 3 percent or greater. The function to be estimated, then, is not applicable below 3 percent occupancy. There is no need to include data for lower occupancies in estimation of the function. On this basis, a quadratic function was found to fit the data best, of the form

$$\text{Flow} = c + d_1(\text{occupancy}) + d_2(\text{occupancy})^2$$

Values of the coefficients for the several stations are presented in Table 1.

Inspection of the scatter of flow values about the mean for each value of occupancy, such as is shown in Figure 5, also called into question the use of a percentage of the function as the boundary. These diagrams suggested that the range of flow values is roughly constant with increasing occupancy, rather than increasing as the percentage would imply. Hence, the final test version of the algorithm subtracted a constant difference from the quadratic function to obtain the boundary between Areas 1 and 2, rather than taking a percentage of it. This difference (Table 1) was obtained by taking half of the average range of flow values, over those occupancy values having sufficient observations.

Hence, the final form of the algorithm used for the vast majority of the testing was calibrated as follows. The minimum uncongested speed was estimated separately for each station. This value was used to isolate the uncongested flow-occupancy data. From that data, a quadratic equation was estimated to describe flow as a function of occupancy at each station, and a constant flow was identified to subtract from that function to create the lower bound for the uncongested flow-occupancy data. The values used for the bulk of the on-line testing are presented in Table 1.

Accommodation of Varying Operating Conditions

The calibration of the speed thresholds and flow-occupancy boundary seems to assume implicitly that the same values can

Volume \ Occupancy	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
3480																				
3360																				
3240																				
3120																				
3000																				
2880																				
2760																				
2640																				
2520																	5	1		
2400															1	4	2			
2280															10	9	1	1		
2160													3	3	18	14	1			
2040													9	36	38	4	1			
1920											1	11	61	45	9	1				
1800											8	96	107	26	2					
1680										2	68	155	40	4						
1560										74	237	90	8							
1440								2	43	251	144	12								
1320								53	319	228	24	1								
1200							22	348	300	36	3									
1080						14	308	433	50	5										
960					3	192	639	88	7											
840					114	720	143	10												
720				54	699	270	9	2												
600			19	577	449	18														
480		2	352	714	32	2														
360		134	978	43	1	1														
240	27	1032	49	4																
120	1129	58	2																	
0	3353																			

FIGURE 5 Frequency of volume versus occupancy at Station NB-7, May 5, 1989.

be used under all weather and roadway conditions. However, earlier work has shown that this is probably not correct. For example, the onset of rain causes a clockwise rotation of the regression function about the origin (*9,14*), so will undoubtedly affect the correct location of the boundary line as well. Unless the function is modified, legitimately uncongested data might fall in Area 2 and falsely indicate an incident. To require the operator to recognize that conditions have changed and then to choose an appropriate flow-occupancy function from a library of such functions is not practical. A more sensible approach is to continuously update on line the way the uncongested flow-occupancy function is applied in the algorithm.

The method used for this on-line updating is to multiply the predicted flow at each time interval by an updating factor based on a smoothed average of recent ratios of observed (uncongested) flows to predicted flows. Hence, if observed flows fall below predicted values for an extended period, the resulting predicted values used by the algorithm will also be reduced, and so will the location of the boundary line.

TESTING

The Skyway FTMS is staffed 24 hours a day, 7 days a week. The system operators log each event, including stalled vehicles on the shoulder as well as accidents or other incidents that affect the traffic flow. For testing of the algorithm, the operators' log has been used as the description of reality, including in particular the time of occurrence of each incident. The operator usually learns of an incident in one of three ways: from the Ontario Provincial Police, who have been notified through CB radio; by observing unusual traffic conditions on the CCTV monitors; or through the existing algorithm. Most incidents are found through CB radio and the police. The existing algorithm identifies few. Regardless of the way the incident was discovered, it may in fact have begun some time before the operator became aware of it, but that information is unobtainable. Hence, the time identified in the operators' log is taken to be the start time of the incident, against which algorithm efficiency is evaluated.

Off-Line Tests

The off-line tests were run on stations NB-7, NB-8, NB-9, SB-5, SB-6, and SB-7, which are on or near the bridge over the Burlington Canal. Even with only two lanes in each direction, the traffic flow and geometrics of the Skyway were such that there were no natural bottlenecks (other than the grade itself) even though flows were occasionally close to 2,000 veh/hr per lane. On average, at least one incident per day resulted in congestion on some section of the bridge. Hence, it was possible to extract several segments of data before, during, and after incidents, for the initial tests of the algorithm. These initial tests helped to identify appropriate sets of parameters and thresholds for later on-line testing, and were encouraging in several respects, despite the fact that they used a 70-km/hr cutoff for all stations, the power function for relating flow to occupancy, and a percentage of that function as the lower bound of uncongested data.

The algorithm detected all 12 of the incidents included in the data sets, with an average detection time of 0 min measured from the time the incidents were reported in the FTMS operators' log.

Times to detection were compared with the optimal times found in an earlier evaluation of a comparative algorithm (*15*), which was based on later versions of the TSC algorithms (*1*). In the data set used for that evaluation, there were 10 incidents. The comparative algorithm found 9 of them, at an average of 1.6 min after the operators' log recorded them. The McMaster algorithm detected all 10 incidents, with an average time to detection of 0.2 min before the times recorded in the operators' log.

In order to examine the performance of the algorithm in updating the flow-occupancy function, a set of data was assembled by alternating 1-hr periods of clear and rainy weather data at station NB–7. In the 10 hr covered by this data set, one incident occurred, and this was during a rainy period. The result of this test was that the incident was detected and that, depending on the threshold values, the number of false alarms during rainy weather was reduced with the on-line updating facility.

First On-Line Tests

On-line tests commenced on the Burlington Skyway in the spring of 1989. By that time, as mentioned earlier, the Skyway over the bridge consisted of four lanes in each direction. This change considerably reduced traffic in each lane. There were also fewer incidents and fewer occurrences of nonincident congestion after the opening of the new bridge. In some respects, then, the Skyway FTMS might seem to constitute an easy test for an incident detection algorithm. In practice, however, it seems not to be. The comparative algorithms currently used there all give many false alarms, apparently because of the considerable changes in grade occurring through the system and the presence of an exit or entrance ramp between some stations. Hence, false-alarm performance is one of the key criteria for the on-line tests.

The results of 29 days of on-line testing were positive, but showed considerable room for improvement. There were two accidents logged during this time. (There were also numerous instances recorded of vehicles requiring assistance, but on the shoulder of the road. Because no incident detection algorithm would have detected these, as others (*1*,*3*) have also noted, they were ignored in evaluating the algorithm.) The first accident was detected by the algorithm at the same time as by the operator. The second accident occurred downstream of the section where detectors were currently in operation and backed up traffic through the operating detectors, so comparison of start times is not possible. However, this incident demonstrated that the aspect of the algorithm that tracks the queue upstream works quite well, as four working detectors were affected.

The room for improvement arose from the performance of false alarms. There were 127 false alarms during the 29 days of the test. (There were an additional 21 alarms at two stations during roadwork, but these were to be expected. A final algorithm would have a facility for the operator to disable a section during such roadwork.) Although 127 alarms over 14 stations and 29 days, from 30-sec data, can be reported as a false-alarm rate of 0.01086 percent, in terms of the load on the operators it is still too high, at roughly 4.4 alarms per day. [Payne and Tignor (*1*,p.36) in evaluating several versions of comparative algorithms reported false alarm rates of 0.01 percent for detection rates of 35 to 40 percent.]

Inspection of the nature of the alarms shows that 61 of the total would be eliminated by increasing the persistence check from 2 to 3 intervals. This fact makes the trade-off between detection time and false alarm rates clear: a 50 percent increase in detection time (from 1 to 1.5 min) will produce roughly a 50 percent reduction in false alarms (from 127 to 61).

Of the remaining 66 false alarms, 43 occurred during the nighttime, when traffic was sparsest. These turn out to have been triggered by low speeds, but the low speeds occurred when the reported flow and occupancy values were both 0. These data are therefore spurious, and the alarms can be avoided by including a screen for such data in the algorithm. Thus, there were only 23 unexplained false alarms, which over 29 days was approaching an acceptable number.

Two major changes were made to the algorithm for the second round of on-line evaluation: changes to the estimation of the boundary between uncongested and congested operation, as discussed under calibration; and the inclusion of a number of screening tests for bad data. One has already been mentioned, namely a test for observations with zero flow and occupancy but positive speeds. Other tests checked for consistency between the upstream and downstream loops of a speed trap. If they disagreed by more than a stated amount, then neither detector was to be used. The persistence check P was kept at two intervals, to be able to test whether the screening and recalibration reduced the number of false alarms.

Second On-Line Test

The second on-line test contained two parts. For the 1st week, the previous version of the algorithm was run in parallel with the version containing the revisions just described. Following that, the revised portion was run alone. Some unexpected problems were encountered during this testing, but continued operation has provided nearly 90 days of testing. The problems arose from brief power failures to the Ministry Control Centre that occurred intermittently for several days.

The comparative test of the initial on-line algorithm and the revisions ran from the afternoon of June 21 until the morning of June 28. False alarm performance was definitely improved by the revisions, but at the expense of missing one accident. The original version triggered 26 alarms in 6.5 days. Of these, 1 was an accident (identified 30 sec earlier by the algorithm than by the operator), 7 occurred during road work so were not false, and 18 were false. The revised version triggered 6 alarms: 3 during the roadwork, and 3 false. Because of the power failures, the data for the day with the one accident was lost, so it has been impossible to investigate why one version caught the accident and the other did not. However, work with other data suggests that the additional screening tests included in the revised version were too stringent, leading to the exclusion as bad data of some of the observations that identified the incident for the initial version.

The second part of the second on-line test has now run from June 28 through November 23. Because of system shutdowns for various reasons, this period provides a total of just under 90 days of information. The results of this testing can be summarized under three categories: incidents properly identified, incidents missed by the algorithm, and false alarms.

Thirteen incidents were properly detected. For 8 of these, the algorithm also identified congestion effects at a second station, indicating that it can be effective for tracking queue development, and need not be disabled at stations adjacent to an incident. The time of detection by the algorithm varied from 33 min before the operators' log time to 7 min later. Three incidents were recorded earlier by the operator; 9 were recorded earlier by the algorithm. The midpoint of the distribution of the time differences was 1 min earlier for the algorithm. (The mean would be misleading because of the 33-min value.)

Seven incidents were recorded by the operators that the algorithm did not identify. However, three of these were identified in the operators' log as having no significant impact on traffic, and a manual inspection of the data for the appropriate times confirmed the operators' comments—there was no suggestion in the data that there was a traffic problem. An additional two incidents occurred upstream of the first detector station currently working, and might therefore not be reflected in the downstream data. Of the remaining two, one occurred at 4:00 a.m., during light traffic, and the other was identified as being on the shoulder. Hence, there is a possibility that no algorithm would have been able to identify any of these incidents.

A number of alarms occurred during roadwork, or were identified by the operator as congestion coming into the system from further downstream. These are not included in the discussion of false alarms. During the 90 days, there were 231 alarms that were labeled false. However, careful scrutiny of the alarms raises doubts that all of them were in fact false. Certainly, the alarms that lasted only 1.5 min were false: there were 105 of these. The revisions to the algorithm did not overcome this problem, but raising the persistence check to three periods will clearly do so. Of the 126 remaining false alarms, only 6 occur between 10:00 p.m. and 6:00 a.m., demonstrating that the improved screening techniques worked well.

The 126 longer-duration false alarms seemed to arise for several reasons. One may simply be a loop-tuning problem. For example, in the period before August 15, Station NB–8 produced nearly one-third of the false alarms. After August 15, Station NB–8 produced only one-tenth of the false alarms. A more important possibility is that a number of these seemed to be instances of congestion that went unrecorded on the operators' log. This assessment seems plausible because several of the alarms occurred at sequential stations, within 5 or 10 min of each other, which was entirely consistent with the way queues formed during the recorded incidents. At worst, then, there were 126 false alarms during 90 days of test (for a false alarm rate of 0.00012 percent), and some of these may in fact have been real events that the operator should know about.

Summary of Test Results

The off-line tests showed that the new algorithm is a feasible approach, and suggested that its results are better than one implementation of the comparative approach to incident detection. The on-line test, with results from a total of about 126 days, demonstrates excellent overall performance. Some fine tuning, particularly of the data screening procedures, needs to be done, but the algorithm seems ready for implementation. Implementation will provide the final test, by providing for operator identification of all of the alarms, rather than results' being based on after-the-fact matching of operator records with algorithm alarms.

CONCLUSIONS

Each incident detection approach in use today has its advantages. The balance between strengths and weaknesses makes one approach preferred over another. In that context, it is appropriate to summarize four strengths of the proposed logic.

First, because it uses speed as well as flow and occupancy data, for systems that provide reliable speed data the chances of detecting incidents are increased. The two criteria are rarely first met at the same exact interval. Also, if one variable is missing, the proposed logic still has the ability to detect incidents that are based on the other test. (This feature means the logic can also be used on systems without speed traps, although there has not yet been any test of how well it will work under those conditions.)

Second, as is done with the comparative algorithms based on occupancy only, to examine occupancy values at adjacent stations is not necessary. As is the case for all single-station algorithms, this is a great advantage when conditions vary between successive detector stations. (The comparative occupancy-based logic is known to have difficulties when there are natural changes in occupancy because of grade or geometry.) Also, if there is a temporary malfunction of detectors at a station, the California-type algorithm has difficulty, depending on the arrrangement of redundant detectors, in flagging incidents between stations upstream and downstream of the malfunctioning one—a problem that is considerably mitigated by single-station logic.

Third, because the new algorithm can raise incident flags at occupancies less than critical occupancy, incidents can be detected with a larger probability, and earlier than with those approaches that wait until occupancies are larger than the critical value before flagging an incident. An examination of the data for several incidents indicated that this distinction appears to matter, particularly if congestion is first indicated downstream of the incident.

Finally, with the proposed single-station logic, detection at a station close to an incident is not suppressed. Instead, when congestion reaches that station, a secondary alarm is given. This procedure is advantageous in estimating queue length and in cases when an incident affects only one or two stations, but for a long time. By contrast, in most applications of the

California-type algorithms, incident detection is automatically suppressed at several stations surrounding a declared incident.

DIRECTIONS

The algorithm is basically congestion detection logic and can be used where congestion is mainly incident related, as is the case for the Burlington Skyway FTMS, and where it is not important to identify automatically the cause of the congestion. For other situations, it is necessary to incorporate the separately developed logic (*8*) for distinguishing between incident congestion and the recurrent congestion caused by demand's exceeding capacity. In addition, three other features need to be added to the algorithm.

First is the issue of how to deal with alarms caused by compression waves that appeared to cause problems in the development of earlier algorithms. These waves are slowdowns of traffic that, because of the close spacing of vehicles, are propagated upstream and often result in a complete stoppage momentarily. These were not apparent in the tests conducted so far. There are three possible explanations for their absence. It may be that no compression waves occurred on the system, or that the algorithm is effective in avoiding an alarm caused by them. Or, it may be that some of the brief false alarms were in fact caused by compression waves. This aspect is one that requires the algorithm to be reporting to the operator for resolution. However, several system managers have indicated in discussions that they would prefer compression waves to generate an alarm that could be investigated, so that the operator could make the decision on what is happening, rather than the computer. This preference is reasonable because compression waves often result in accidents, and in this sense are worthy of the operator's attention.

Second, screening of data is an integral part of algorithm effectiveness, as suggested by the one accident picked up by the original on-line version and missed by the version that included stringent screening tests. In these evaluations of the algorithm, somewhat more sophisticated tests were used than just the simple single-variable range tests, but more work needs to be done to establish the most effective set of screening tests.

Third, the output of the algorithm needs to be enhanced to identify whether the incident can be expected to be upstream or downstream of the station at which the alarm has been triggered. As indicated earlier, the pattern in the data looks different in the two cases. Finally, and related to this third feature, is the detection of incidents during already-congested operation. This possibility was identified by Gall and Hall (*8*), and is a possibility within this single-station approach that is not available within all approaches.

ACKNOWLEDGMENTS

Much of the work for this paper was performed under a contract with the Ontario Ministry of Transportation. The authors gratefully acknowledge this financial support as well as the assistance of Colin Rayman, Mike Young, Phil Masters, Hoi Wong, Sui Wong, Emmanuel Morala, Brian Delsey, John Proctor, and others in the Freeway Traffic Management Section in making the data available to them. Brian Delsey deserves special thanks for his work in programming the on-line versions of the algorithm. The authors also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada for the fundamental research that seeded the development of the McMaster algorithm, and for supporting Lisa Hall through an undergraduate student research award.

REFERENCES

1. 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. 30–36.
2. K. G. Courage and M. Levin. *A Freeway Corridor Surveillance, Information, and Control System*. Research Report 488–8, Texas Transportation Institute, Texas A&M University, College Station, 1968.
3. 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.
4. 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.
5. M. Levin and G. M. Krause. Incident Detection Algorithms, Part 1: Off-Line Evaluation. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 49–58.
6. M. Levin and G. M. Krause. Incident Detection Algorithms, Part 2: On-Line Evaluation. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 58–64.
7. Peat, Marwick, Mitchell & Co. *Evaluation of the Howard Frankland Bridge Surveillance and Control System in Florida*. Report FHWA-RD-85-061. FHWA, U.S. Department of Transportation, 1985.
8. A. 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., 1989, pp. 1–8.
9. P. Athol. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. In *Highway Research Record 72*, HRB, National Research Council, Washington, D.C., 1965, pp. 58–88.
10. B. N. Persaud and F. L. Hall. Catastrophe Theory and Patterns in 30-Second Traffic Data—Implications for Incident Detection. *Transportation Research A*, Vol. 23A, No. 2, 1989, pp. 103–113.
11. F. L. Hall and B. N. Persaud. Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. In *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 1989, pp. 9–16.
12. C. L. Dudek and C. J. Messer. Detecting Stoppage Waves for Freeway Control. In *Highway Research Record 469*, HRB, National Research Council, Washington, D.C., 1973, pp. 1–15.
13. B. N. Persaud and F. L. Hall. Effect of Grade on the Relationship Between Flow and Occupancy on Freeways. In *Transportation Research Record 1173*, TRB, National Research Council, Washington, D.C., 1988, pp. 33–38.
14. F. L. Hall and D. Barrow. Effect of Weather on the Relationship Between Flow and Occupancy on Freeways. In *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1988, pp. 55–63.
15. DelCan. Evaluation of Incident Detection Algorithms for Highway 401 FTMS. Ontario Ministry of Transportation, Canada, March 1987.

Publication of this paper sponsored by Committee on Freeway Operations.