Part 2. On-Line Evaluation

Moshe Levin, Gerianne M. Krause, and James A. Budrick, Bureau of Materials and Physical Research, Illinois Department of Transportation, Oak Park

Five algorithms were evaluated on-line by using the facilities of the Traffic Systems Center of the Illinois Department of Transportation. Three of the algorithms developed by Technology Services Corporation (TSC), were of a pattern-recognition nature. The other two—a pattern-recognition and a probabilistic or Bayesian algorithm—were developed locally. Thresholds for the features used in each of the pattern-recognition algorithms were developed by TSC. The thresholds for the probabilistic algorithm were developed by using accident data on the Eisenhower Expressway. The measures of effectiveness in the evaluation were detection rate, false-alarm rate, and mean-time-to-detect. The three TSC algorithms were evaluated twice on the Eisenhower Expressway at the 80 and 90 percent levels of detection thresholds, and then problem areas showing high false-alarm rates were represented by the 50 percent level. The three TSC algorithms were then evaluated on a section of the Dan Ryan Expressway that was free of geometric problems, for comparison purposes. Statistical analysis showed no difference in detection rate, false-alarm rate, and mean-time-to-detect among the three TSC algorithms at any of the evaluated detection levels. Introduction of the 50 percent level improved certain measures of effectiveness. Algorithm 7, the best of the TSC algorithms, showed overall superiority to the two local algorithms. The false-alarm rate was shown to be related to geometric and other features of the problem areas and yielded algorithm 8, which uses a shockwave-suppressor mechanism and requires the least effort in developing appropriate thresholds.

This paper discusses the on-line evaluation of five incident-detection algorithms that were all evaluated off-line in the preceding paper to obtain the optimal threshold sets used in the on-line evaluation.

The specific goals of this research were

1. To determine the on-line efficiency of algorithms proved effective in the off-line evaluation,
2. To correlate algorithm efficiency parameters derived from the on-line evaluation with those derived from the off-line evaluation, and
3. To evaluate combinations of thresholds with respect to geometric conditions on the freeway.

ALGORITHM DESCRIPTION

Consider an n-lane freeway section of length L between two fully detectorized stations. At each station a set of flow characteristics for occupancy, volume, and speed is measured at specific time intervals. Suppose that at time \( t_1 \), an incident occurs at a certain point on one of the lanes in section L. A shock wave will develop and travel upstream of the incident with an intensity that is dictated by the severity and lateral location of the incident and by environmental and geometric conditions. At time \( t_0 + \delta t \) an incident-detection algorithm, by continuously measuring and comparing the flow characteristics upstream and downstream of the incident with predetermined thresholds, will detect the incident.

This section describes the structure of the incident-detection algorithms evaluated in this research. Of the five algorithms evaluated, three of the pattern-recognition type were developed by TSC (2) and the other two, one pattern-recognition and one probabilistic (7), were developed locally in the course of this research.

The research effort of TSC included the development of 10 incident-detection algorithms that could be grouped into three categories. The first, comprising algorithms 1-7, is composed of variations on the classic California algorithm (2). The second consists of algorithms 8 and 9, which are characterized by suppression of incident detection after a compression wave is detected. Finally, algorithm 10 represents an attempt to detect those incidents that occur in light-to-moderate traffic but do not lower capacity below the volume of oncoming traffic.

Of these 10 algorithms 7 were selected for evaluation, 1 from each category. The algorithms selected (7, 8, and 10) were chosen for a number of reasons. Preliminary investigation by TSC had indicated algorithm 7 to be a superior form of the California algorithm. Algorithm 8 is identical to algorithm 9 except for an added persistence check. According to TSC's preliminary investigation, algorithm 8 has a slightly lower FAR but a longer MTTD than algorithm 9. Although algorithm 10 did not perform especially well in TSC's view, it was included in the on-line evaluation because it represents a first attempt to solve the problem of detecting incidents...
that do not produce marked traffic-flow discontinuities. The TSC algorithms are in binary decision-tree form; at each node of the decision tree a feature value is compared with a user-specified threshold value to determine whether an incident is to be signaled. Clearly, the effectiveness of the algorithm depends on the thresholds chosen. The program TSC developed for optimizing threshold selection has been described in the first part of this paper. It uses a random-number generator that produces increments to be added to the current optimal threshold vector to produce a new threshold vector for evaluation. After a predetermined number of iterations, and given a certain level of detection, the threshold vector that has the lowest FAR is termed the optimal threshold vector at that level.

The thresholds obtained by using the CALB program for the evaluated detection levels were used in this analysis. Also evaluated, in addition to the above three TSC algorithms, were 16-14, the pattern-recognition algorithm, and the Bayesian or probabilistic algorithm, both developed locally. Threshold selection for algorithm 16-14 was accomplished by using the CALB program; calibration of the Bayesian algorithm used accident data collected on the Eisenhower Expressway.

The meaning of the features involved in each algorithm and the tree structures of these algorithms are given in Part 1 of this paper.

ON-LINE INCIDENT-DETECTION SYSTEM
TSC controls 360 directional km (224 miles) of expressways through its Freeway Traffic Management System (Figure 6). The backbone of the system is the detector subsystem that uses full detector stations (5 km (3 miles)) and single-detector stations (800 m (0.5 mile)).

The major function of the on-line incident-detection system is to detect a capacity-reducing incident through its incident-detection logic, which uses three algorithms simultaneously and then delivers a message to the monitor. Another function is to provide continuous evaluation of algorithm performance, refinement of thresholds, and evaluation of response to incidents. Figure 7 presents the basic on-line incident-detection system.

The basic programs for both functions of the on-line system are the incident-determination logic program (ST) and the incident message program (S). The former uses appropriate thresholds obtained from previous analyses to determine the incident status of each of the main-line detectors. A status matrix is used for recording the status and is updated every minute. At the end of the update, S scans the matrix for detected incidents and generates an appropriate message. The generated messages include information on detector subsection, upstream occupancy, downstream occupancy, time of incident, day, and date and are maintained in a disk-based file.

Once the incident message is produced it becomes possible to monitor the incident file through the display as part of the incident message management phase, which the display program (E) directs. Appropriate parameters are passed into programs Q and O for operation. Program Q controls the queuing and the displaying of the incident messages. Queue manipulation enables the operator to inspect the incident file and delete old messages, because new messages are ignored when the queue is full. These messages consist of six elements. Three describe location: expressway name, direction (inbound (IB) or outbound (OB)), and detector station; the others are vector number, incident file number, and earliest detection time.

Program O can handle various options initiated by the operator. In the future, these options could include communications between the Traffic Systems Center in Oak Park and the IDOT Communication Center in Schaumburg.

Other related programs in the on-line software are
program H, which produces a hard copy of the incident file; program D, which records the time it takes the operator to respond to the message and displays the clock time on the screen; and program E, which is an existing program extended to include the input required by the display program (E).

DEVELOPMENT OF DATA BASE

The major site chosen for the study was the Eisenhower Expressway (I-90) between I-94 and Wolf Road (Figure 8). This expressway contains various characteristics along its 8-km (13-mile) length. Geometrically, the expressway is four lanes wide between I-94 and Austin Boulevard and then drops to three lanes from Austin to Wolf Road. This lane drop is the major bottleneck area for westbound traffic. For eastbound traffic, First Avenue is the major problem area. Here the degree of curvature, change in grade, and volumes of traffic are the main causes of congestion. Both sections are quite a challenge for the on-line incident-detection algorithms, especially during peak hours. For comparison purposes, another expressway (the Dan Ryan between 65th and 95th Streets) was chosen for study. This section of expressway is a straight section, four lanes wide, with no major bottlenecks between its terminal points.

The time period picked for the survey was 3:00-5:00 p.m. Monday-Friday. During this period, four capacity-reducing incidents are expected on the Eisenhower Expressway.

A helicopter aerial survey of the study section was made to collect the incident data. The information obtained for each stopped vehicle included time of spotting, longitudinal location (I8 or OB), lateral location (a cross street), lane, vehicle description, reason for stopping (if ascertainable), type of aid present (if any), and comments to describe or explain traffic operations.

The helicopter was able to maintain an average speed of 180 km/h (110 mph), which allowed one trip along the entire length of expressway, i.e., terminal points of the study, to be made in about 7.5 min. In reality, however, each point was viewed nearly every 5 min because of the visibility from the helicopter flying at about 200-250 m (700-800 ft) above the expressway.

At the completion of each day of data collection, the aerial survey data were correlated with the incident information produced by the on-line operating algorithms. This recorded information included longitudinal location, lateral location, lane, detection time of each individual algorithm being tested, termination time, computer and actual duration times, type of incident or congestion-causing situation, comments, and actual time of occurrence, detection, and termination.

After completing this correlation of computer-recorded incident messages and actual recorded incidents, various statistics were determined. These were DR, FAR, missed incidents, and so forth, calculated for each day for each individual algorithm.

A total of 29 days of data on the Eisenhower Expressway and 4 days on the Dan Ryan Expressway were collected.

ALGORITHM EVALUATION

Based on the off-line evaluation of the algorithms it was decided to conduct the on-line evaluation by using optimal
thresholds developed for the 80 and 90 percent DRs as obtained in that evaluation.

In the first phase, algorithms 7, 8, and 10 were evaluated on the Eisenhower Expressway during the afternoon rush. Preliminary analysis of the data suggested that problem areas (bottlenecks and curves) were producing a considerable number of FARs, and it was decided to run an evaluation after having introduced less-sensitive thresholds—the off-line 50 percent DR—into the problem areas. Then an evaluation of the algorithms on the Dan Ryan study section was conducted with thresholds representing the off-line 90 percent DR.

In the second phase, the apparent best algorithm among the three above was selected to operate simultaneously with algorithm 16-14 and the Bayesian algorithm on the Eisenhower Expressway. Each of the study cases referred to in Table 6 was analyzed for differences in DR, FAR, and MTTD among the algorithms. Algorithm efficiency at the 80 percent detection level was compared with that at the 90 percent level, and the efficiency at that level was compared with algorithm efficiency at the 90-50 percent detection level, which was represented by a set of thresholds derived for the 90 percent and 50 percent detection levels at nonproblem and problem sections, respectively.

The most promising algorithms at the detection levels of 90 percent and 90-50 percent were selected for further analysis. In this analysis the cumulative distributions of the message duration of false alarms and real incidents were compared in order to give an indication as to the change with time of the probability that an incident message is true. Also, the distribution of false alarms with respect to time during the rush period was investigated to yield an indication of the need for threshold refinement.

To clarify the relationship between numbers of false alarms and geometric features of the problem section, an analysis was conducted at the 90 and 90-50 percent levels of detection. In this analysis the number of false alarms for each problem section for one detection level was compared with that for the other detection level. This was done for algorithms 7, 8, and 10.

Tables 7 and 8 present the types of problems on the various sections of inbound and outbound Eisenhower. These problems had a tendency to produce a high number of false alarms. The sections that were operating with thresholds related to the 50 percent detection level during the 90-50 percent detection level evaluation period are also indicated. No attempt was made to find the relation between DR and the geometric features of each section because of the relatively low number of incidents (16) during the 90-50 percent detection-level evaluation period.

Comparative Analysis of Algorithm Efficiency

Tables 9, 10, and 11 present DR, FAR, and MTTD for algorithms 7, 8, and 10 for the off-line detection levels of 80, 90, and 90-50 percent, respectively.

As can be seen from these tables, the on-line DRs are lower than the off-line rates. However, the positive correlation between DR and FAR, which was found in the off-line analysis, seems to exist in the on-line analysis, as shown for the off-line 80 and 90 percent detection levels in Tables 9 and 10, respectively.

The statistical t-tests conducted for each off-line detection level for differences in the measures of effectiveness among the algorithms did not indicate any signifi-
significant (0.05 level) differences for any of the measures of effectiveness for any of the detection levels. Differences in MTTD values between the off-line and on-line evaluations were also noted. The on-line evaluation yielded MTTD values ranging between 2.7 and 8.9 min for thresholds representing the 90-50 percent detection level. The off-line evaluation yielded MTTD values ranging from 2 to 4 min. The large MTTD values obtained in the on-line evaluation could be attributed to some inherent inaccuracies in determining the exact time of occurrence of an incident because of the obvious limitations of the aerial survey. Taking this into consideration, as far as the MTTD was concerned, both the

### Table 7. Relation between FAR and geometric features on I-88 Eisenhower Expressway.

<table>
<thead>
<tr>
<th>I-88 Eisenhower Section</th>
<th>Problem Description</th>
<th>90 Percent Threshold</th>
<th>90-50 Percent Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm No.</td>
<td>Algorithm No.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>0.28</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>FAR</td>
<td>0.87</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>8.8</td>
<td>9.3</td>
<td>9.0</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.

### Table 8. Relation between FAR and geometric features on OB Eisenhower Expressway.

<table>
<thead>
<tr>
<th>OB Eisenhower Section</th>
<th>Problem Description</th>
<th>90 Percent Threshold</th>
<th>90-50 Percent Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm No.</td>
<td>Algorithm No.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>0.37</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>FAR</td>
<td>0.86</td>
<td>0.73</td>
<td>0.86</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>8.9</td>
<td>6.8</td>
<td>5.3</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.

### Table 9. On-line algorithm efficiency for off-line 80 percent detection level for RD conditions on Eisenhower Expressway.

<table>
<thead>
<tr>
<th>Measure of Effectiveness</th>
<th>Algorithm No.</th>
<th>Apparent Best Algorithm</th>
<th>Statistically* Best Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>0.28</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>FAR</td>
<td>0.87</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>8.8</td>
<td>9.3</td>
<td>9.0</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.

### Table 10. On-line algorithm efficiency for off-line 90 percent detection level for RD conditions on Dan Ryan Expressway.

<table>
<thead>
<tr>
<th>Measure of Effectiveness</th>
<th>Algorithm No.</th>
<th>Apparent Best Algorithm</th>
<th>Statistically* Best Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>0.37</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>FAR</td>
<td>0.86</td>
<td>0.73</td>
<td>0.86</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>8.9</td>
<td>6.8</td>
<td>5.3</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.

### Table 11. On-line algorithm efficiency for off-line 90-50 percent detection level for RD conditions on Eisenhower Expressway.

<table>
<thead>
<tr>
<th>Measure of Effectiveness</th>
<th>Algorithm No.</th>
<th>Apparent Best Algorithm</th>
<th>Statistically* Best Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>0.56</td>
<td>0.41</td>
<td>0.56</td>
</tr>
<tr>
<td>FAR</td>
<td>0.63</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>7.5</td>
<td>5.3</td>
<td>5.2</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.

### Table 12. On-line algorithm efficiency for off-line 90 percent detection level for RD conditions on Dan Ryan Expressway.

<table>
<thead>
<tr>
<th>Measure of Effectiveness</th>
<th>Algorithm No.</th>
<th>Apparent Best Algorithm</th>
<th>Statistically* Best Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>FAR</td>
<td>0.58</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>MTTD, min</td>
<td>10.0</td>
<td>11.0</td>
<td>13.5</td>
</tr>
</tbody>
</table>

*At the 0.05 level of significance.
were no significant differences in DR and MTTD but there was one in FAR.

The results of the limited algorithm evaluation on the Dan Ryan Expressway at the 90 percent detection level are presented in Table 12. Statistical analysis at the 0.05 level of significance for differences among algorithms 7, 8, and 10 indicated no significant differences for any of the measures of effectiveness.

During the second phase of the study algorithm 7, which was found to be the apparent best for the 90-50 percent detection level, was compared with algorithm 16-14 and the Bayesian algorithm. Table 13 presents the results of this evaluation. Statistical analysis at the 0.05 level of significance indicated that, as far as the detection rate was concerned, no best algorithm could be found. Algorithm 7 and the Bayesian algorithm were superior to algorithm 16-14 with respect to the FAR, while algorithms 7 and 16-14 were superior with respect to the MTTD.

**Duration of Incident Messages**

To increase decision credibility regarding an incident message, one could require the message to have a certain duration, the assumption being that a false message will terminate after a short while. Thus, if the distributions of durations of true and false messages are determined, it should be feasible to relate message duration to the probability of a message's being true.

Cumulative distributions of duration of false alarms and incident messages for algorithm 7 are shown in Figures 9 and 10 at the 90 and 90-50 percent detection levels. From these figures it can be seen that the distribution of duration of false-alarm messages is such that for both levels of detection, nearly 50 percent of the messages endure 30 min or more. This, of course, indicates a weakness in the algorithm that experienced between 0.60 and 0.70 FAR.

The distribution of false alarms with time (by 30-min intervals) during the daily study period (3:00-5:00 p.m.) was found to be uniform. This suggests that no change in thresholds with time was necessary for any particular location.

**Relationship Between FAR and Geometric Features**

The introduction of problem-section-related thresholds representing the 50 percent detection level led to some improvement in the efficiency of the algorithms. The relationship between the number of false alarms and geometric features that resulted from the operation of algorithms 7, 8, and 10 is presented in Tables 8 and 9 for the 90 and 90-50 percent detection levels for both directions of the Eisenhower Expressway.

Algorithm 7 showed the most improvement in terms of reduction of false alarms because individualized thresholds were incorporated. The other algorithms did not show consistent improvement. For example, the introduction of thresholds representing the 50 percent DR at the lane drop at Austin (Figure 8) did not change the FAR of algorithm 8 but rather increased it (not...
necessarily significantly) for algorithms 7 and 10. This lane drop causes the most severe shock waves on the facility for most of the afternoon rush period.

The long duration of false alarms in this section is a major cause of the high percentage of messages of long duration in the cumulative distribution of incident-message duration (Figures 9 and 10). When shockwaves are less severe, as in the case of the sun effect on traffic on the outbound freeway near Des Plaines Avenue, the individualized thresholds (related to the 50 percent detection level) seemed to improve the false-alarm situation considerably for all algorithms. Another problem section inducing false alarms and rendering the individualized set of thresholds there ineffective was the bridge near Addison Creek between 25th Avenue and Mannheim Road, where only algorithm 8 showed improved operation. The effect of other problem sections inducing nonincident shock waves resulting in false alarms can be determined from the above figure.

FINDINGS, OBSERVATIONS, AND RECOMMENDATIONS

Based on the analyses conducted in the course of this research the following are the major findings and observations.

1. No statistically significant differences at the 5 percent level of significance in DR, FAR, and MTTD were found among algorithms 7, 8, and 10 for the 80, 90, and 90-50 percent detection levels, when they were operated on the Eisenhower Expressway.

2. The introduction of individualized thresholds at problem sections did not affect algorithm 8 but improved DR and FAR of algorithm 7 and improved DR and MTTD for algorithm 10.

3. As far as the MTTD was concerned, no apparent differences between the on-line and off-line evaluations were observed.

4. The efficiency of algorithms 7 and 8 remained statistically the same for the 90 and 90-50 percent detection levels.

5. When compared with the locally developed algorithms (16-14 and Bayesian) at the 90-50 percent detection level, algorithm 7 showed overall superiority.

6. Nearly half of all incident and false-alarm messages lasted longer than 30 min.

7. The introduction of individualized thresholds at problem sections could reduce the number of false alarms generated in these sections.

8. DR obtained by algorithms in the off-line evaluation are considerably higher than those obtained in the on-line evaluation.

9. The shockwave-suppressor mechanism of algorithm 8 seemed to be quite effective; required less effort to prepare thresholds for this than for any other algorithm.

10. FARs are quite high, and reducing them poses the biggest challenge in refining present algorithms or developing new ones.

11. The distribution of false alarms over time seemed to be uniform for the 90 and 90-50 percent detection levels, which indicates that no changes in thresholds at any particular section with time during rush hour were necessary.

12. Algorithms 7 and 8 seem to operate quite similarly, but algorithm 7 was apparently better.

The recommendations for further action are

1. To investigate the behavior of traffic features at bottlenecks during incidents in order to be able to distinguish between incident- and non-incident-related shockwaves,

2. To develop an effective and inexpensive supportive incident-verification system to minimize FAR, and

3. To develop an improved nonincident shockwave-suppressor mechanism and to incorporate it into the efficient pattern-recognition algorithms.

ACKNOWLEDGMENT

The research reported in this paper, like that in the preceding paper, was conducted within the framework of IDOT and with the cooperation of FHWA. We wish to acknowledge the support of the staffs of the Bureau of Air Operations and the Traffic Systems Center in the data-collection and programming phases of this study. The contents of this report reflect our views, and we alone are responsible for the facts and the accuracy of the data presented here. The contents do not necessarily reflect the official views or policies of IDOT or FHWA. This report does not constitute a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Freeway Operations.

Development of a Transport System Management Planning Process in the Delaware Valley Region

Rasin K. Mufti and James J. Schwarzwalder, Delaware Valley Regional Planning Commission, Philadelphia

The joint Federal Highway Administration and Urban Mass Transportation Administration (FHWA-UMTA) guidelines require cities to develop a transportation system management (TSM) element, a short-range element of the transportation plan. The metropolitan planning organizations (MPOs) initially responded to these requirements by preparing a plan report that includes a composite list of projects from the highway and transit capital programs (reverse process). Then, the MPOs began to improve on their initial submissions and to create a process for developing the TSM elements of the plans. This paper presents the Delaware Valley's experience, the outcome of the first