IDEA PROJECT FINAL REPORT
Contract ITS-2

IDEA Program
Transportation Research Board
National Research Council

February 28, 1995

A PROTOTYPE SYSTEM FOR REAL-TIME INCIDENT LIKELIHOOD PREDICTION

Samer Madanat and Pen-Chi Liu
Purdue University

The ITS-IDEA program is jointly funded by the U.S. Department of Transportation’s Federal Highway Administration, National Highway Traffic Safety Administration, and Federal Railroad Administration. For information on the IDEA Program contact Dr. K. Thirumalai, IDEA Program Manager, Transportation Research Board, 2101 Constitution Avenue N.W., Washington, DC 20418 (phone 202-334-3568, fax 202-334-3471).

INNOVATIONS DESERVING EXPLORATORY ANALYSIS (IDEA) PROGRAMS MANAGED BY THE TRANSPORTATION RESEARCH BOARD (TRB)

This investigation was completed as part of the ITS-IDEA Program which is one of three IDEA programs managed by the Transportation Research Board (TRB) to foster innovations in surface transportation. It focuses on products and result for the development and deployment of intelligent transportation systems (ITS), in support of the U.S. Department of Transportation's national ITS program plan. The other two IDEA programs areas are Transit-IDEA, which focuses on products and results for transit practice in support of the Transit Cooperative Research Program (TCRP), and NCHRP-IDEA, which focuses on products and results for highway construction, operation, and maintenance in support of the National Cooperative Highway Research Program (NCHRP). The three IDEA program areas are integrated to achieve the development and testing of nontraditional and innovative concepts, methods and technologies, including conversion technologies from the defense, aerospace, computer, and communication sectors that are new to highway, transit, intelligent, and intermodal surface transportation systems.

A PROTOTYPE SYSTEM FOR REAL-TIME INCIDENT LIKELIHOOD PREDICTION

Executive Summary

In the first part of this research project, we focused our efforts on developing and validating the freeway incident likelihood prediction models that form the core of the prototype system. These models are based on traffic, weather, and incident data collected by the Indiana Department of Transportation (INDOT) on the Borman Expressway, a segment of Interstate 80/94 in northwest Indiana. The two major types of incidents in terms of associated traffic delays on the Borman Expressway are vehicle crashes and overheating vehicles. Therefore, we have developed likelihood prediction models for these two types of incidents. As shown in the Research Approach section of this report, these models exhibit high goodness-of-fit to data and provide accurate predictions of incident likelihoods. Moreover, they capture the effect of relevant traflic, location, weather and geometric variables on freeway incident probabilities.

In the second part of the research, an incident likelihood prediction simulator was developed. This report describes the incorporation of the incident likelihood prediction models into the siiulator. A traffic simulator (INTRAS) is used to generate the traffic characteristics (i.e., volume and speed) that are inputs to the freeway incident likelihood prediction models, whereas the environmental conditions are specified by the user. The simulator combines the sequential outputs from an existing incident detection algorithm and those of our incident likelihood prediction models through Bayesian updating. The likelihood prediction outputs are used as the initial prior probabilities. As the detection outputs are received every minute, the incident likelihoods are sequentially updated. Every fifteen minutes, new incident likelihood predictions are produced by the models and used to update the estimates of freeway incident likelihoods. This sequential fusion of incident detection and prediction probabilities produces better estimates of incident likelihoods because more accurate prior probabilities are used and because both traffic and environmental factors are taken into account. The simulator framework is shown in Figure 1.

Virtually any existing incident detection algorithm can be incorporated into our simulator. In this research, we used a Bayesian-type detection algorithm to demonstrate the significance of combining the incident likelihood predictions with the detection outputs. Based on a limited number of simulations representing a range of realistic situations, it was found that the proposed combined approach achieves early detection and high levels of confidence in incident confirmation.

The ITS-IDEA Product consists of the simulator, which includes the incident likelihood prediction models and the Bayesian updating algorithm. The simulator is designed to receive traffic outputs from the INTRAS traffic simulator as well as user-specified inputs of environmental conditions.

Final Report

1. Problem Statement

The first objective of this research project was to develop models which can be used to provide real-time predictions of freeway incident likelihoods. Such predictions will serve as the basis for a proactive corridor-wide traffic control system. In such a system, traffic stream and environmental conditions measured by surveillance sensors would be used as inputs for predicting incident likelihoods in near real-time. Traffic control strategies can thus be immediately impkmented to reduce the probability of an incident, as well as to mitigate incident-related problems if they occur.

To prove the feasibility of this concept, it was essential to demonstrate the possibility of accurate predictions of freeway incident probabilities, based on near real-time measurements of traffic and weather variables. As described in the following section of this report, we have successfully developed models for likelihood prediction of two critical types of freeway incidents: crashes and overheating vehicles. These models capture the influence of various traffic and weather factors on the probabilities of vehicle crash and overheating vehicle incidents. Furthermore, both models have high internal and external validity, as demonstrated by their fit to the data and their predictive accuracy, respectively.

The predictions given by the incident likelihood models can be used in two ways. First, they can be combined with incident detection outputs to improve the accuracy of the estimates of incident probabilities. State-of-the-art incident detection algorithms utilize only traffic information. By considering both traffic and environmental variables, it is possible to achieve a more accurate estimate of incident probability. This estimate might **be** used as an input to a sequential incident-response decision-making process. Second, they can be used as part of a proactive warning system for freeway motorists.

2. Research Approach

This section describes the approach that was used in developing the freeway incident likelihood prediction models and in combining the prediction likelihoods with an existing incident detection algorithm to obtain a more accurate incident probability estimate.

2.1. Incident Likelihood Prediction Modeling

Because the outputs of the incident prediction models are probabilities of a binary event, an appropriate methodology to **use** is binary logit. Binary logit is a powerful tool which has been widely used in transportation demand modeling studies.

Eight-and-a-half months of incident, traffic and weather data for the Borman expressway were used for model development. We sampled non-incident data from the non-incident population which comprises those time periods in which no incidents were observed. Therefore, our sample is a stratified random sample with two strata, incidents and non-incidents.

Two binary logit incident prediction models are presented in the following paragraphs. These are models for two types of incidents: (i) overheating vehicles, and (ii) crashes. In Models 1 & 2 the column entitled "Independent Variable" lists the explanatory variables used in the model. The "Estimated Coefficient" column shows the contribution of each explanatory variable to the probability of that type of incident and the "t-Statistic" column displays the statistical significance of that variable. A t-statistic larger than 1.65 in absolute value means that the variable is a significant predictor of that type of incident at the 90% confidence level. The goodness of fit of each model shown in Model 1 is represented by p^2; the larger the value of p^2, the better the fit of the model to the data. In binary logit models, a value of p^2 higher than 0.10 is considered appropriate. The statistic "percent correctly predicted" provides an estimate of the predictive accuracy of each model.

For the overheating vehicle incident likelihood model, ***the*** variables ***peak, merge, temp*** (temperature), ***rain,* and spv** (speed variance) were found significant.

Independent Variable	Estimated Coefficient	t-Statistic
constant	-2.45269	-5.25041
peak	0.40010	1.62515
merge	0.51087	2.19683
temp	0.03228	4.63282
rain	-1.06644	-2.29946
spv	-0.05907	-2.37734
number of observations		427
percent correctly predicted		73.536
	p^2=0.21 5	

Model 1: Incident Likelihood Model for Overheating Vehicles

The coefficient for the variable peak has a positive sign, which suggests that an overheating vehicle incident is more likely to occur in a peak period than a non-peak period. This is expected because traveling speeds are slower during the peak period. This variable is not significant at the 90% confidence level, as can be seen by the low value of its t-statistic (1.625), possibly because the peak period on the Borman expressway is widely spread out. The coefficient of the variable merge represents the effect of location relative to on/off ramps on the likelihood of an overheating vehicle incident. The positive sign of this coefficient indicates that an overheating vehicle incident is more likely to occur in a merge section than a mid-section. The value of the t-statistic (2.197) suggests that this effect is significant. The coefficient of the variable temp shows the effect of temperature on the likelihood of an overheating vehicle incident. The positive sign suggests that an overheating vehicle incident is more likely to occur in high temperature conditions than low temperature conditions, because high temperatures aggravate engine overheating. The high t-statistic (4.633) strongly supports this explanation. The coefficient of the variable rain has a negative sign which indicates that an overheating vehicle incident is more likely to occur in sunny (non-rainy) conditions than in rainy conditions. The t-statistic (-2.299) shows a significant effect for the variable rain. The coefficient of the variable spv represents the effect of speed variance between lanes on the likelihood of an overheating vehicle incident. The negative sign means that an overheating vehicle incident **is** more likely to occur in lower speed variance conditions than higher speed variance conditions. This is because when the speed variance is low, there are less overtaking opportunities, which increases the likelihood of an overheating vehicle incident. The t-statistic (-2.377) suggests that this result is significant. Overall, this model demonstrates good fit to the data, as can be seen from the value of p^2 (0.215) and high predictive accuracy, as measured by the high percentage of observations correctly predicted (74%).

The effects of temperature and speed variance on the likelihood of an overheating vehicle incident are shown in Fignres 2 and 3. In these figures, Case 1 represents a mid-section of the freeway during an off-peak hour on a rainy day. Case 2 represents a merging section during a peak hour on a dry day. As can be seen in these figures, an overheating vehicle is more likely to occur in peak hours, merge sections and dry conditions (case 2) than in off-peak periods, mid-sections and rainy conditions (case 1). Moreover, the figures show that the higher the speed variance, the lower the likelihood of an overheating vehicle incident and that the higher the temperature, the higher the likelihood of an overheating vehicle incident.

For the crash model, the variables *merge*, *visi* (visibility), and *rain* are found significant.

Independent Variable	Estimated Coefficient	t-Statistic
constant	-0.76655	-2.23940
merge	0.31452	1.46231
visi	-0.02779	-1.02077
rain	1.48747	3.45081
number of observations		434
percent correctly predicted		71.198
	ρ^2=0.140	

Model 2: Incident Likelihood Model for Crashes

In Model 2, the coefficient of the variable *merge* has a positive sign, which suggests that a crash is more likely to occur in a merge section than a non-merge section. Though the t-statistic (1.462) indicates that this variable is not strongly significant at the 90% confidence level, it has the correct sign, because there are more vehicle interactions and therefore a higher crash probability in the merge sections, where traffic flow is not as smooth as in the mid-sections. The coefficient of the variable *visi* has a negative sign, which indicates that a crash is more likely to occur in low visibility conditions, as expected. This variable is not strongly significant, as can be seen by its t-statistic (-1.020) possibly because, in our dataset, visibility is measured in miles, a unit which is not sufficiently precise to capture the effect of low visibility on drivers. The coefficient of the variable *rain* has a positive sign, which means that a crash is more likely to occur in rainy conditions than non-rainy conditions. This is because the presence of rain reduces visibility and lowers pavement skid resistance. The high t-statistic (3.451) supports this explanation. The fit of this model is satisfactory, as shown by its ρ^2 value (0.140), as is its predictive accuracy (71% of observations correctly predicted).

It should be noted that the estimated coefficients in these models are unbiased regardless of the use of a stratified random sampling scheme in which incidents are oversampled. The only correction that must be made is for the constant, using the method described in Ben-Akiva and Lerman (1985). The effect of this correction is to reduce the probability of an incident by a factor proportional to the log of the fraction of incident observations in the sample divided by the fraction of incident observations in the population. The predicted incident probabilities shown in figures 1 and 2 are the corrected incident probabilities.

2.2 Bayesian Combination of Incident Likelihoods

In this research, we have demonstrated the applicability of the incident likelihood prediction models by combining the predictions with a Bayesian-type incident detection algorithm (Levin and Krause, 1978). Our purpose is to demonstrate that combining the outputs from this incident detection algorithm with the incident likelihood predictions results in better incident likelihood estimates than those based on the incident detection algorithm only. The following bayesian updating formulae were used:

$$P_1 = \frac{P_0 f_1\langle Z|1\rangle}{P_0 f_1\langle Z|1\rangle + (1-P_0) f_1\langle Z|0\rangle} = \frac{1}{1+\dfrac{(1-P_0) f_1\langle Z|0\rangle}{P_0 f_1\langle Z|1\rangle}}$$

$$P_{t+1} = \frac{1}{1+\dfrac{(1-P_t) f_{t+1}\langle Z|0\rangle}{P_t f_{t+1}\langle Z|1\rangle}} \; ; t = 1, 2, \ldots$$

$$P_{t+1} = \frac{1}{1 + \dfrac{(1-P_t) f_{t+1}\langle X|0\rangle}{P_t f_{t+1}\langle X|1\rangle}} \quad ; t = 15, 30, \ldots$$

Where:

t = index for time, in minutes (we assume here that the frequency of the detectors' outputs is one measurement per minute).

P_0 = initial prior incident probability obtained by the prediction model.

P_t = incident probability at minute t.

Z = outputs from detectors (i.e., occupancy and difference in occupancy).

$f_t\langle Z|k\rangle$ = probability distribution function of detector outputs under incident (k = 1) or non-incident (k = 0) situation at minute t.

X = Measures of environmental conditions (i.e., temperature, rain, visibility, time of day, location).

$f_t\langle X|k\rangle = \dfrac{P_t\langle k|X\rangle \times P_t(X)}{P_t(k)}$ = probability distribution function of different environmental conditions under incident (k = 1) or non-incident (k = 0) situation at minute t. Where:

> $P_t\langle 1|X\rangle = \; - P_t\langle 0|X\rangle$ = incident prediction likelihood obtained by using the prediction model for period ending at t.
>
> $P_t(X)$ = historical probability of environmental conditions X.
>
> $P_t(1)$ = historical incident probability based on the sample data
>
> $P_t(0)$ = historical non-incident probability based on the sample data

The above equation can be applied sequentially to compute the incident likelihoods at any minute of the analysis period.

The above methodology can be represented by the following two flowcharts. Figure 4 is the overall framework and Figure 5 contains detailed descriptions of each component.

As shown in Figure 4, two types of inputs are specified by the user: traffic and environmental variables. The traffic data serve as inputs to INTRAS (the traffic simulator used in our study), whereas the weather data serve as inputs to the likelihood prediction models. The outputs from INTRAS, which consist of occupancies and speed variances, feed into both incident detection and likelihood prediction models. The Bayesian updating formula sequentially utilizes the outputs from both detection and prediction models and produces updated incident likelihoods. These incident likelihoods can be used as a basis for proactive warnings and as inputs for incident response decision-making algorithms.

Figure 5 provides a more detailed description of the simulator. The traffic data that feed into the simulator are volume and speed. These two factors serve as the inputs to INTRAS, which simulates time-varying traffic characteristics such as speed variances (which are inputs to the prediction model) and occupancy readings (inputs to the detection algorithm). The likelihood prediction models also utilize several environmental factors measured over fifteen minute intervals, such as rain, visibility, time of day and temperature, as well as location variables such as freeway section location. The Bayesian updating formula sequentially combines the outputs from the detection and the likelihood prediction models, and provides minute-by-minute incident likelihood estimates.

The simulator described in Figure 5 was used to evaluate the adequacy of the proposed combined algorithm for incident probability estimation. The simulation results are described in the following section.

3. Results

This section describes the results from a number of simulation runs that were performed using the simulator developed in this study. In each run, an incident is generated on the freeway. The presence of an incident is associated with: (i) a sharply increasing occupancy in the case of a single detector, or (ii) a sharply increasing discrepancy in occupancies between adjacent detectors, in the case of multiple detectors. Occupancy (%) is the indicator used to determine the presence of incident and is defined as below (Levin and Krause, 1978):

(1) the percentage occupancy at a single detector in minute t = Occ_t

(2) the discrepancy of percentage occupancies at adjacent detectors in minute t

$$= \frac{\boldsymbol{Occu_t - Occd_t}}{\boldsymbol{Occu_t}}$$

where Occ_t = minute-average occupancy measured at single detector at time t.

$\boldsymbol{Occd_t}$= minute-average occupancy measured at downstream detector at time t.

$\boldsymbol{Occu_t}$ = minute-average occupancy measured at upstream detector at time t.

Note that the location of the incident is at the downstream end of the freeway segment in the two-detector scenario. A simple conceptual illustration is shown in Figure 6. The discrepancy of percentage occupancies between two detectors results from backup traffic between the two detectors.

The probability density functions (pdfs) of Occ (in the single-detector case) and AOcc (in the double-detector case) used in the combined algorithm are portrayed in Figures 7 and 8, respectively.

We created ten scenarios based on a range of traffic and weather conditions. Since the simulation results for the overheating vehicle incident and the crash incident cases are not substantially different, we only present the results for the overheating vehicle incident scenarios. We present the comparisons between our combined algorithm, which combines incident likelihood prediction with measurements of occupancies, and the Bayesian algorithm which only includes occupancy measurements (Levin and Krause, 1978). The comparison criterion used is the time-to-detect an incident, in minutes. The time-to-detect is defined as the length of time from incident occurrence to incident detection. For both algorithms, an incident is "detected" when the computed incident likelihood crosses a certain threshold. In our combined algorithm, this threshold is computed using the Sequential Probability Ratio Test (SPRT) (Bertsekas 1987). In the Bayesian algorithm, the threshold is 0.3 for the single-detector case and 2 for the double-detector case (Levin and Krause 1978). The comparative results are listed in Tables 1 and 2. We should note that the time shown for the combined algorithm is actually the sum of detection and verification times. This implies that a certain amount of verification time (typically 3 minutes) should be added to the times shown for the Bayesian algorithm. This verification time is needed, in conventional incident detection algorithms, in order to decrease the false alarm rate. On the other hand, the use of a combined incident likelihood algorithm with the thresholds computed through the SPRT method does not require additional time for incident verification. This is because the SPRT thresholds are computed by considering the possibility of a false alarm. More detailed discussions of the SPRT method will be provided in the final report of ITS-IDEA Project ITS- 17, "A sequential-hypothesis-testing-based decision-making system for freeway incident response".

Conclusion:

Accurate incident likelihood prediction models can be used for improving the accuracy of incident probability estimates. Traditional incident detection algorithms do not account for the effects of environmental factors, and thus produce less accurate estimates. The importance of this study was to show the benefit of using the outputs of incident likelihood prediction models to enhance the accuracy of current incident detection algorithms. Specifically, it was shown that combining such predictions with conventional traffic measurements decreases incident detection and verification times. The full benefits of this combined approach will be realized by incorporating it within a dynamic incident-response system, such as the one currently being developed as part of ITS-IDEA Project #17. Therefore, one contribution of this research is to provide a critical input to incident-response systems.

References:

1. Ben-Akiva M. and ***Lerman S. Discrete Choice Analysis: Theory and Application to Travel Demand,*** the MIT Press, Cambridge, MA, 1985.

2. Wicks D.A. and Andrews B.J. "Development and Testing of INTRAS, a Microscopic Freeway Simulation Model", ***User's manual,*** vol. 2, Federal Highway Administration, 1980.

3. Levin M. and Krause G.M. "Incident Detection: A Bayesian Approach", ***Transportation Research Record No. 682,*** pp. 52-58, 1978.

4. Bertsekas D.P. ***"Dynamic Programming: Deterministic and Stochastic Models",*** Prentice Hall, 1987.

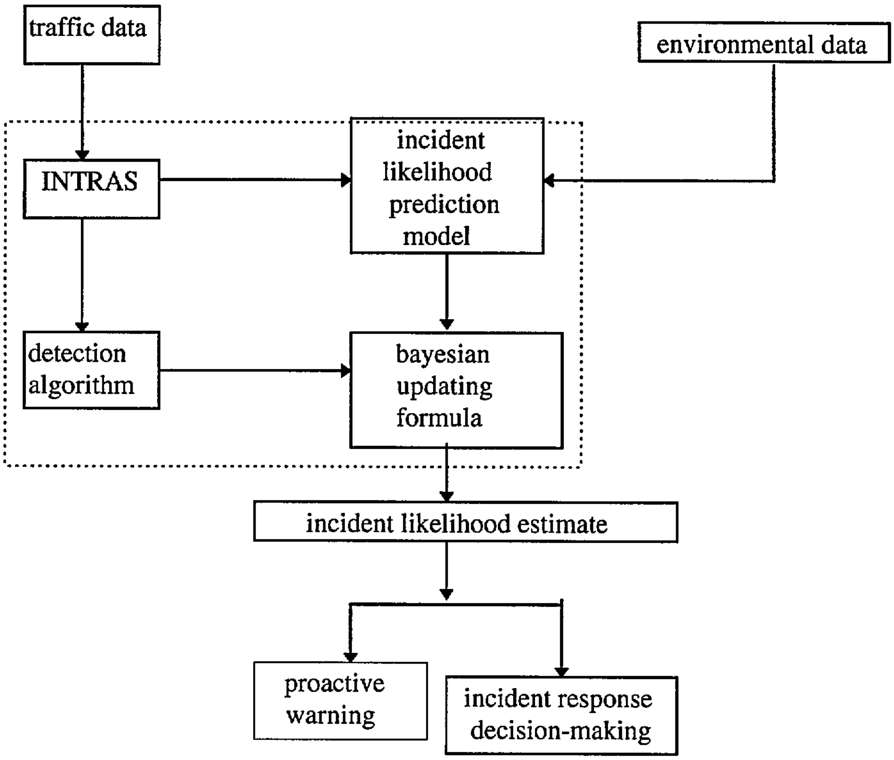

Figure 1. Framework for the incident likelihood prediction simulator

Case 1: Off-peak, midsection, rain
Case 2: Peak hour, merge section, no rain

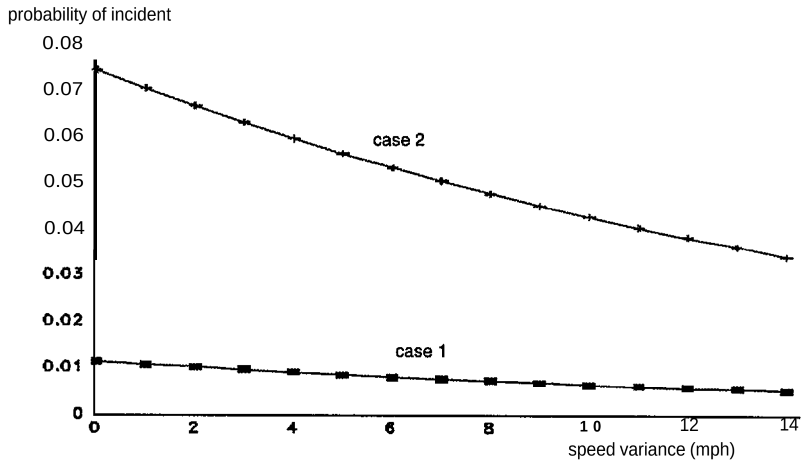

Figure 2. Effect of speed variance on the probability of an "overheating vehicle" incident

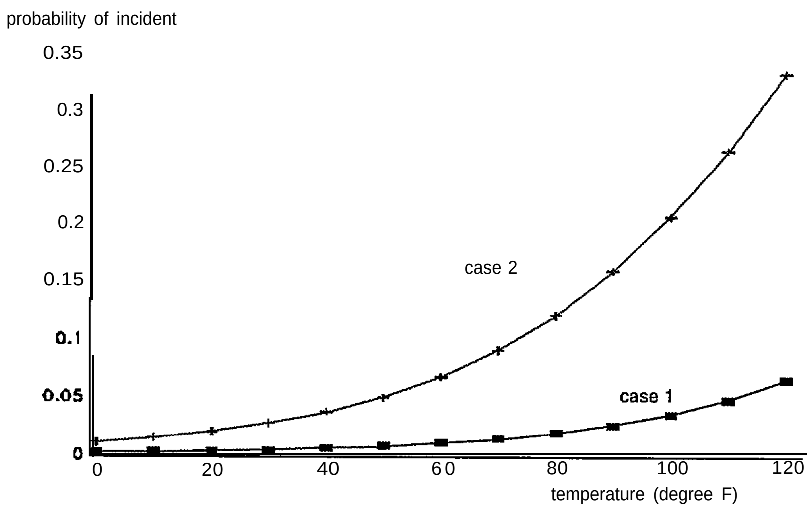

Figure 3. Effect of temperature on the probability of an "overheating vehicle" incident

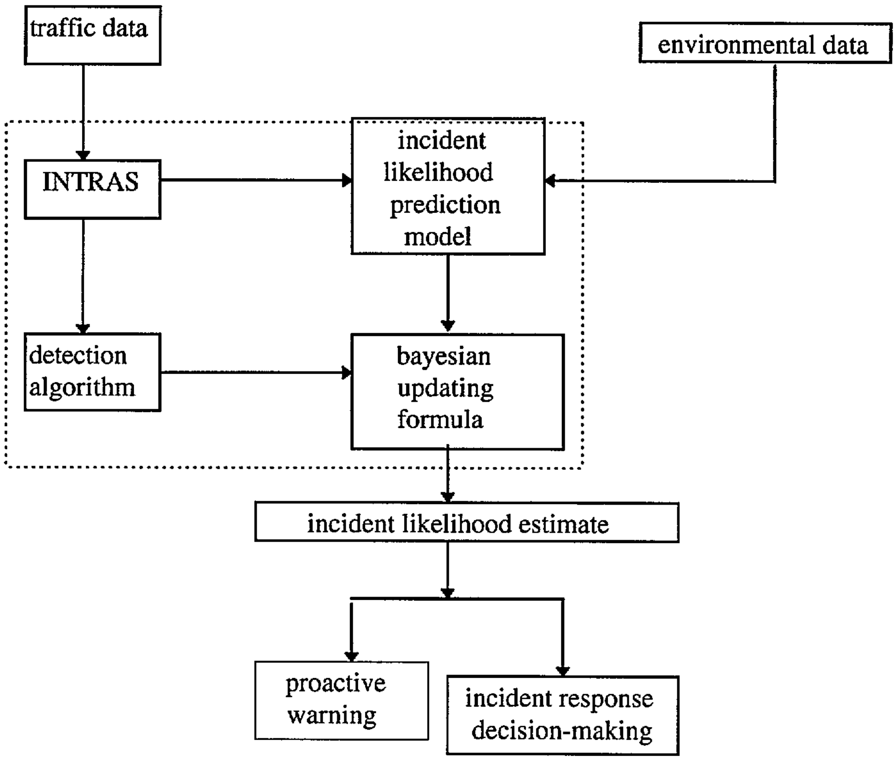

Figure 4. Framework for the incident likelihood prediction simulator

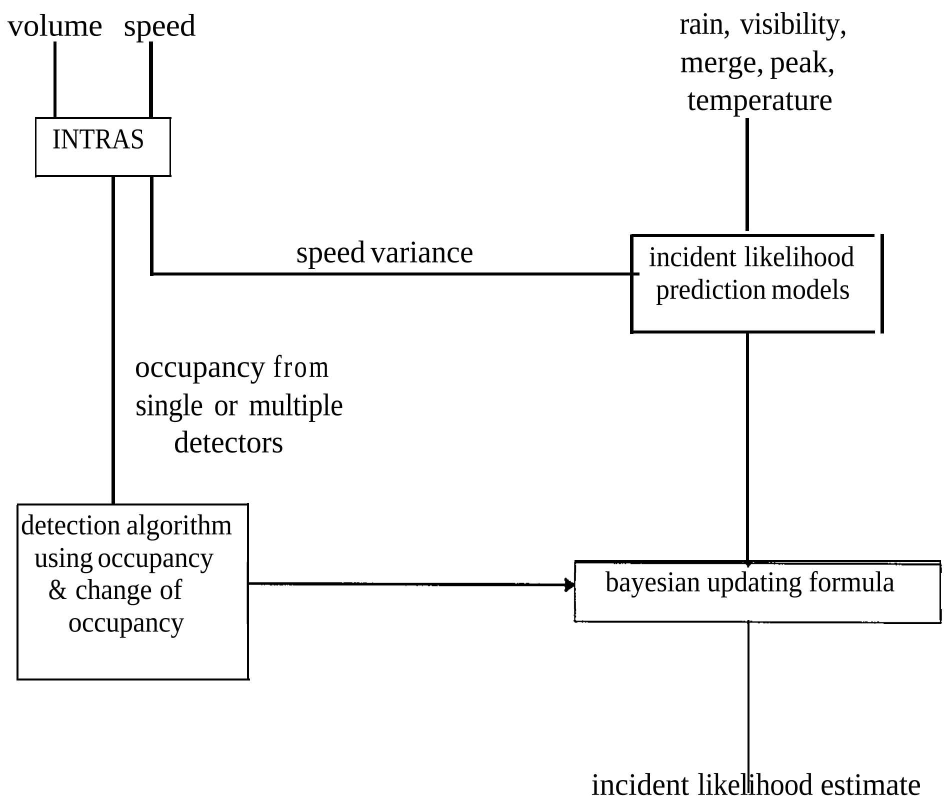

Figure 5. Detailed description of the incident likelihood prediction simulator

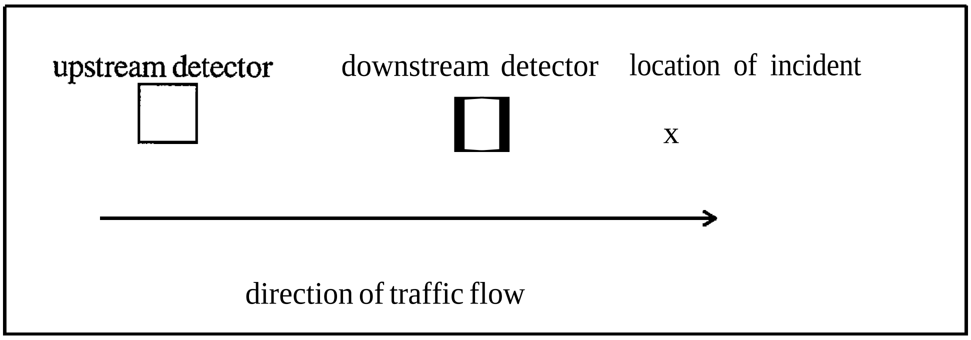

Figure 6. Location of incident in the simulation

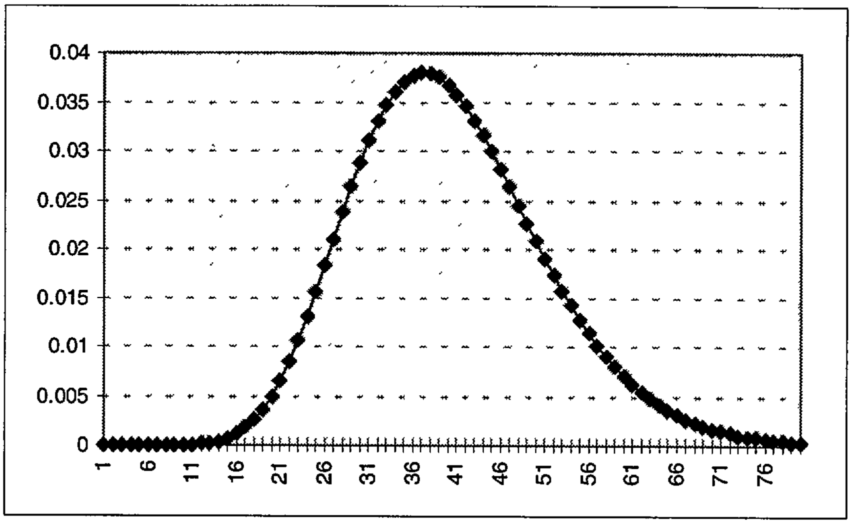

Figure 7. The pdf of occupancy (%) for the single-detector case

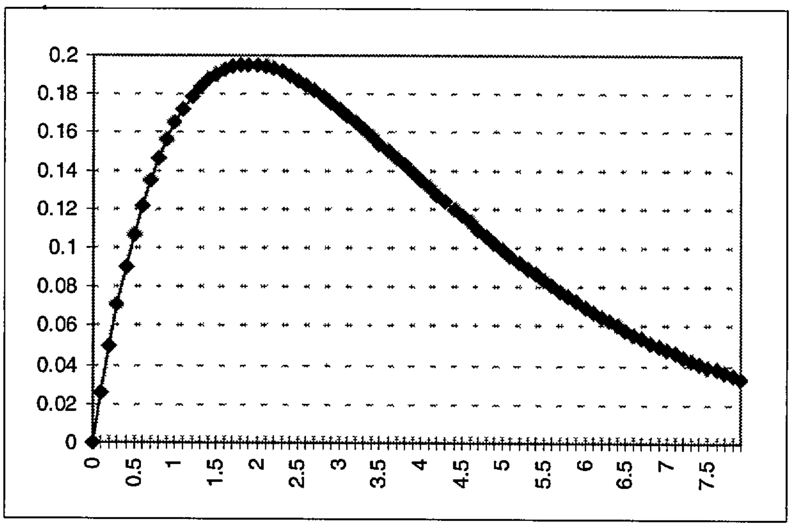

Figure 8. The pdf of change of occupancy for the double-detector case

Algorithm	scenario 1	2	3	4	5
Combined	4	4	5	3	2
Bayesian	*	4	2	2	2

Algorithm	scenario 6	7	8	9	10
Combined	2	1	2	2	6
Bayesian	2	2	1	1	6

Table 1. Time-to-detect for the single-detector case for the combined algorithm and the Bayesian algorithm

*: incident was not detected.

Algorithm	scenario 1	2	3	4	5
Combined	1	1	1	2	2
Bayesian	1	1	1	1	1

Algorithm	scenario 6	7	8	9	10
Combined	2	2	1	1	3
Bayesian	1	1	1	2	*

Table 2. Time-to-detect for the double-detector case for the combined algorithm and the Bayesian algorithm

*: incident was not detected.