**Real Time Traffic Incident Detection Fusing Multiple Data Streams on Freeways System Network in Iowa**

**Pranamesh Chakraborty**

PhD Student, Department of Civil, Construction and Environmental Engineering Iowa State University, Iowa, USA

Phone: (515)-357-9076

[pranames@iastate.edu](mailto:pranames@iastate.edu)

**Bijan Vafaei**

MS Student, Department of Civil, Construction and Environmental Engineering Iowa State University, Iowa, USA

Phone: (515)-708-2332

[bvafaei@iastate.edu](mailto:bvafaei@iastate.edu)

**Shefang Wang**

PhD Student, Department of Civil, Construction and Environmental Engineering Iowa State University, Iowa, USA

Phone: (515)-708-2332

[bvafaei@iastate.edu](mailto:bvafaei@iastate.edu)

**Anuj Sharma (Corresponding Author)**

Associate Professor, Department of Civil, Construction and Environmental Engineering Iowa State University, Iowa, USA

[anujs@iastate.edu](mailto:anujs@iastate.edu)

Phone: (765)-430-0023

Fax:

Word Count: XXX words of text + XX figures/tables×250 words (each) = XXX words Submission Date: March 28, 2017

**ABSTRACT**

Near real time traffic detection of incidents is one of the key step to reduce incident-related congestion. With the increasing usage of GPS based navigation, promising data-scalable crowdsourced probe data is now available which can provide near-real time traffic speed information.Various methods to flag accidents and congestions on freeways have been introduced over the course of past ten years; however, false alarms rate on those methods can be potentially reduced by using multiple data sources and fusing them together. Early traffic incident detection with higher reliability resulted from using multiple data streams which reduces the number of false alarm aims to help police department, emergency services, and road agencies to response to the incident in a more effecitive time frame. This study utilizes such extensive historical datasets (approximately 500 GB) to gain useful insights on the normal traffic pattern of each segment. The insights come in the form of speed threshold for different time of the day and days of week for each segment. Thereafter, the anomalous traffic behaviour are classified as incidents. The dynamic thresholds developed for each segment simplifies the calibration steps that is often required when applying a model to a different dataset. For the purpose of this project, probe data, radar based data, in combination of user based incidents data collected through the navigation application WAZE were selected to use. This study aims to analyze large traffic data (approximately 100 GB of INRIX data and 10 GB of Wavetronix data), and also introduce two alternatives of the traditional Standard Normal Deviate (SND) based incident detection algorithm. The proposed algorithms can handle the masking effect of SND method where the outliers inflate the mean and standard deviation values and result in lower threshold values and in turn, lower detection rate. The high detection rate (94-97%) obtained by these algorithms compared to the SND method (83%) shows the efficacy of the models. Although higher false alarm rate (FAR) are observed for these models, but their values (4 false alarms/day) are quite lower than the acceptable FAR (10 false alarms/day) reported in previous literature (*5*). Big data analytics such as MapReduce, Hadoop, and Python programming will be used as the purpose of this study.

Keywords: Traffic incident management, outlier mining, big data, fusing multiple data sources, standaed normal deviate, false alarm rate, map reduce, hadoop

**INTRODUCTION**

Real time traffic incident detection, is a key component to reduce incident-related congestion, alleviate the waste of vehicles’ fuel and passengers’ time as well as to provide appropriate information in an earliest time frame for the field operation troops including highway maintenance, police department, and emergency group, in addition to the infrastructures and vehicles that would be connected and telecommunicated in near future. Traffic congestion has been defined by US Department of Transportation (USDOT) as "one of the single largest threats" to the economic prosperity of the nation (*1*). The cost of congestion in the year 2014 was calculated to be $160 billion for the top 471 urban areas in the United States. This included 6.9 billion hours of wasted time and 3.1 billion gallons of wasted fuel (*2*). A major contributor to this congestion are traffic incidents. Schrank and Lomax (*3*) showed that imple- mentation of improved incident management procedures in 272 out of 439 urban areas resulted in reduction of 143.3 million hours of incident-related congestion and $3.06 million.

Early detection of incident is one of key step for improved incident management. Hence, significant efforts have been devoted in the past for development of accurate and fast automatic incident detection (AID) algorithms. Researchers have used pattern recognition algorithms, outlier mining methods, artificial neural networks, fuzzy set theory, genetic algorithms, wavelet transfor- mation and other machine learning methods for traffic incident detection (*4*). However, a nation- wide survey on deployment of AID algorithms in Traffic Management Centers (TMC) showed that 90% of survey respondents feel that the current AID algorithms are inappropriate for use either in present (70%) or in future (20%) (*5*). The two major reasons behind disabling of AID algorithms in TMCs are difficulty in algorithm calibrations and unacceptable false alarm rates when deployed in large scale. The complicated and time consuming calibration of AID algorithms make it difficult to use them by local TMC personnel. Thus, there is a significant need to revisit the AID algorithms and develop an algorithm which can address these major issues.

Automation of calibration process of AID algorithms can resolve one of the major hin- drances of deployment of AID algorithms in TMCs. However, as pointed out by Castro-Neto et al. (*6*), development of an incident dataset with accurate start and end time of incidents is time- consuming and often requires manual investigation. This makes the calibration of AID algorithms even more difficult for TMC personnels. In this paper, the main goal is to develop an AID algorithm that can extract maximum information from the traffic data to generate the normal travel pattern of each segment. Thereafter, the anomalous behaviour can be classified as incidents and hence sidestep the need for algorithm training with incident dataset. In the era of big data, traffic parameters (e.g. speed, volume, etc.) are stored for each and every segment across 24 *×* 7 hours and 365 days. For example, in Iowa State, probe vehicle data of 23,000 segments spread across the entire state are archived every day in one minute interval. This results in generation of approximately five gigabytes of daily traffic data, which in turn produce around two terabytes of traffic data in an annual basis. And, for traffic incident detection, traffic data needs to be collected and processed continuously for each segment. With the cheap data storage technologies now available, it makes more sense to store the entire dataset and use it to gain useful insights on the performance of the road network. These insights can help in developing more efficient AID algorithms. Thus, incident detection turns out to be an important field in the area of transportation which can get direct benefits from the big data analytics. This paper proposes detecting incidents considering them as outliers or anomalies in the continuous traffic data stream. The next section gives an overview of the past research done on AID algorithms and performance measures used to evaluate the algorithm. The third section provides description of the data used in this paper. Section 4 gives the details of the research methodology followed by the detailed results in Section 5. The final section provides a summary of the paper and outlines the future work.

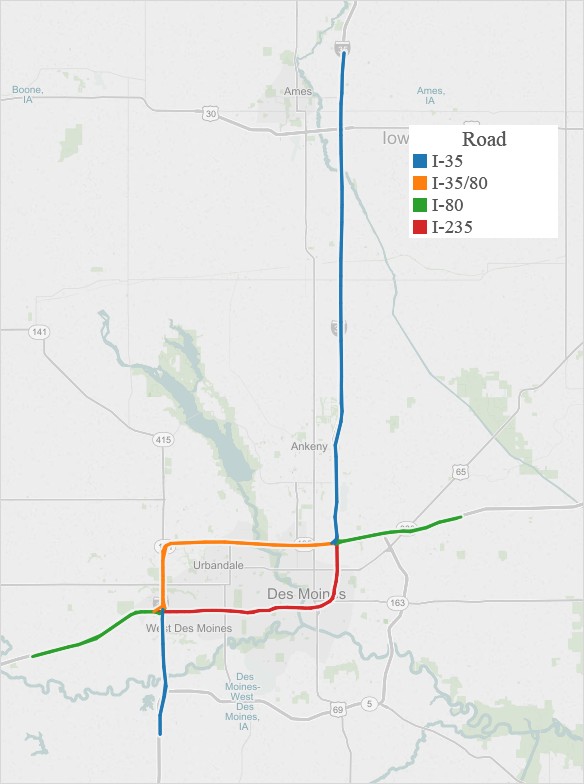
**Literate Review**

Significant research efforts have been devoted to develop efficient automatic incident decetion (AID) algorithms since the last five decades. AID algorithms can be divided into three general categories based on the type of traffic data collection: roadway-based, probe-based and driver-based (4). Roadway-based algorithms use fixed detector data installed at specific points in the road segments whereas probe-based algorithms use probe vehicle data for detecting incidents. AID algorithms can be further classified into two broad categories based on the methodology used to detect incidents (a) algorithms that compare the present traffic parameter values with the historical values observed under similar conditions (e.g., time of day, day of week) and (b) present traffic parameters are compared with the immediate previous intervals to trigger an incident alarm. In either of these cases, the feature vectors are compared with a predetermined threshold for incident detection. Also, a persistence test is usually performed to confirm the preliminary detected incidents before triggering incident alarm (7). This is done to eliminate the false alarms caused due to sudden spurious traffic fluctuations. Probe-based AID algorithms that use historical traffic parameter values for incident detection are presented next followed by a discussion of the algorithms that utilizes sudden change of immediate traffic values during an incident to trigger an alarm. Arterial traffic incident detection algorithms developed in the ADVANCE operational test by Sethi et al. (8) and Sermons and Koppelman (9) used discriminant analysis techniques for incident detection. Linear relationship of predictor variables were developed to distinguish incident conditions from incident-free ones. These algorithms use travel time and speed of a particular link and its immediate upstream link to trigger incident alarms. Balke et al. (10) considered traffic incidents as outliers in data stream and used the principle of standard normal deviates (SND) to indicate the confidence intervals for incident-free travel time conditions. Historical average travel time were computed for each link by time of day (in 15-min intervals) and day of week. Algorithms were also developed to detect traffic incidents comparing the present conditions with the immdeiate past. For example, Parkany and Bernstein (11) algorithms were based on the principle that temporal and spatial discrepancies of travel time and headways and frequent lane switch maneuvers can be observed when traffic switches from incident-free to incident conditions. Waterloo algorithm proposed by Hellinga and Knapp (12) were based on the assumption that the travel time are log-normally distributed, rather than normally distributed as assumed by Balke et al. (10). And, the confidence limit in Waterloo algorithm is based on the travel time observed in previous N intervals instead of using the historical average travel time for the required interval of the day. The bivariate analysis model (BEAM) developed by Li and McDonald (13) use average travel time of probe vehicles and differences in travel time between adjacent time intervals to distinguish an incident condition from incident-free one. Zhu et al. (14) and Zheng et al. (15) used speed differences between adjacent sections and adjacent time intervals as feature vector for mining incidents as outliers from non-incident conditions. In 2013, Li et al. designed an extention of SND algorithm which improved mathematical model for calculationg time intervals’ mean and standard deviation (4)., Li et al. extended the SND algorithm by introducing two modifications: (a) weighted average and standard deviates of the traffic parameter values are used based on the traffic flow, and (b) in order to eliminate the false alarms caused by acute fluctuations of SND values, if the coefficient of variation of the traffic parameter is below a predetermined threshold, the the SND value of the previous time interval is used to replace the SND of the current time interval. In this paper, traffic incidents are considered as anomalies/outliers in continuous traffic data stream and are detected by comparing them with the historical averages. The basic reason behind adopting this technique is that it will allow to utilize the massive historical dataset to gain useful insights of the traffic pattern of each link thereby helping in detecting incidents. With the increasing usage of navigation applications installed in mobile phones, promising data-scalable crowdsourced probe data is now available which provide near real-time traffic speed information. Li et al. (16) used such crowdsourced probe data provided by INRIX to identify shockwave boundaries while Park and Haghani (17) developed models for detecting secondary incidents utilizing same data source. So, it makes sense to also develop AID algorithms utilizing such extensive data source. In traditional AID algorithms, sample data are used for developing the models hoping that the model could be generalised and applied to every other segment. However, this makes the calibration and fine tuning of the model parameters even more difficult. Utilising data of each and every segment will help making the parameters dynamic and can be continuously trained from new incoming data. Also, in this study, alternatives of the traditional SND algorithm are applied to detect out- liers. A basic disadvantage of SND algorithm is that it is impacted heavily by the presence of outliers. So, in this study, two other outlier detection methods are applied and compared with the traditional SND algorithm to find out the efficacy of the proposed methods.

Previous research studies shown Dempster–Shafer (D-S) evidence theory which is an assimilation classification algorithm, is cable of considering the divergence of information since it provides a uniform structure to aggregate multiple data streams. D-S is utilized when various data sources cannot be trusted with a 100 percent likelihood of accuracy. (18) Zheng et al. used Dempster–Shafer evidence theory to mix multiple multi-class probability support vector machines (MPSVM) and showed MPSVM results in more reliable adaptive classifier for traffic incident detection. (18)

**DESCRIPTION OF DATA**

This project fuses traffic speed data obtained from radar based sensor data (Wavetronix) and probe vehicle based data (INRIX). The study region comprises of the Interstates 35, 80 and 235 in Des Moines region of Iowa and shown in Figure 1. The Des Moines region is the busiest region on Iowa roadways experiencing significant amount of road congestion and traffic incidents throughout the year. Besides this, video cameras are also installed in this region which helps in verification of incident data.

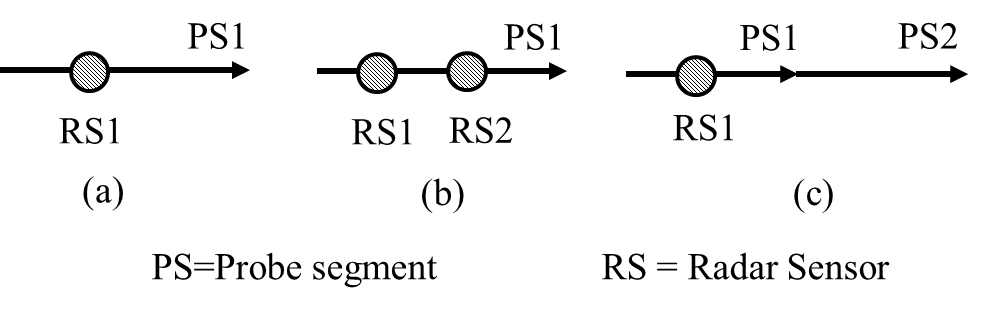


**FIGURE 1 Location of the segments used**

The one-hundred and sixty-four miles long study region is divided into two hundred and fify-four segments for collection of probe based speed data. The length of the segments vary from 0.2 miles to 1.5 miles. Average speed of each of this segment are reported in one-minute interval. A total of one hundred and seventy-five radar-based sensors are also placed in the region which sends average speed, volume and occupancy data in every 20-seconds interval. The duration of the study is from April, 2016 to October, 2016.

**Matching sensor data and segment data**

Figure 2 shows three cases that can occur while matching radar sensor (RS) data with probe segment (PS) data. In Case a, one RS is present in one PS. Hence, there is a one-to-one correspondence between RS and PS. In Case b, two RS (can be more than two also) are present in a single PS. Thus, it is a many-to-one correspondence between RS and PS. In this case, average speed data obtained from all RS in the given PS is taken as the corresponding RS data. In Case c, no RS is present in PS2 i.e., one-to-many corresepondence case. In this case, the nearest upstream RS (withing one mile distance) is assigned to the corresponding PS. If the distance is more than one-mile between end of a PS segment and RS, then the PS shouldn’t be assigned with any RS. However, such case is not present in our study. The reason for choosing upstream RS is because traffic effects (in case of incidents) spreads upstream and hence no effect can be seen in downstream sensors.



**FIGURE 2 Matching radar sensor (RS) data with probe segment (PS) data**

The probe-based speed data used in this study is provided by INRIX (16) with a reporting frequency of one-minute. Details of this cloud-based speed data can be found in Li et al. (15) study. Reliability of the speed data is dependent on the number of probe vehicles available, which in turn depends on the flow volume. Confidence score and C-value are two parameters provided by INRIX to indicate the data quality of the reported average speed of a particular segment. Confidence score of 30 indicates that the data is generated exclusively from real-time data sources while a score of 10 indicates that historical data is used to report the speed. When a mix of the two sources are used, a score of 20 is provided. The C-value is used to provide an additional degree of confidence to the real-time data. The C-value is reported only when the confidence score is equal to 30. In this paper, the reported speed data is considered to be reliable real-time speed data and used for further analysis only when the confidence score is equal to 30 and C-value is also greater than 30 (as suggested by Haghani et al. (17)).

**METHODOLOGY**

Traffic incidents have been often considered as outliers/anomalies in the continuous data stream. The common strategy applied to detect the anomalous traffic behaviour is using the SND algorithm. However, the SND algorithm is impacted heavily by the presence of outliers or incidents. This issue can be resolved by removing all incident-related data points before calculating the average and the standard deviation values. However, this will lead to application of semi-supervised learning instead of unsupervised learning which requires information of all incidents occurring in the study region over the entire study period. This is difficult because development of an accurate incident dataset is very time-consuming and cumbersome manual investigation is required in most cases**.** Particularly, information of the accurate start time and end time of incidents are often hard to get which makes the calibration process very difficult. However, alternate outlier analyses methods exist which can cater the affect of outliers for calculating the threshold. Detailed description of such a outlier method and it’s modifications to make it work as an AID algorithm is discussed next.

**Univariate Outlier Analysis: Inter-Quartile Distance Method**

Univariate outlier analysis is the simplest method of detecting outliers where the output depends only on a single variable. Fundamentally, univariate outlier detection procedures involve selecting a reference value *x0* and a measure of variation ζ from the data sequence {*xk*} (19). Then, data point *xk* is said to be an outlier if it satisfies,

 (1)

Different univariate outlier detection procedures exist depending on the choice of *x0* and ζ. In this study, *x0* is chosen to be sample median () and ζ is chosen as Inter-quartile distance (IQD), the difference of the 75th and 25th percentile. The main advantage of IQD method is that it is robust to outliers compared to the traditional SND algorithm which takes *x0* and ζ as sample mean and standard deviation. However, the IQD method suffer from a different phenomenon, namely swamping. In swamping, the ζ value becomes zero if more than 50% of the data values *xk* have same value. This will lead to declaring any value different from the median as an outlier, irrespective of its distance from the median. For example, if the median speed value is 60 mph, the current speed value of 59 mph will also be declared as outlier since the ζ is zero in case of swamping. This will result in a very high value of *FAR* in the case of traffic incident detection. However, for AID algorithms, we can take advantage of the fact that an alarm should be triggered only in cases when congestion has occurred. As per FHWA guidelines, congested conditions is said to occur in freeways when the speed is less than 45 mph (20). So, typically alarm should not be triggered when speed is higher than 45 mph. Thus, it eliminates the false alarms which can trigger in swamping cases, where the ζ parameter is zero and the median speed value is quite high (greater than 45 mph). Previous study have shown that IQD method have higher detection rate (DR) compared to SND method and also a reasonable false alarm rate (FAR) (21).

Normal traffic condition for each segment varies depending on the time of day, day of week, weather conditions, etc. For univariate outlier analysis, the *x0* and ζ values are computed from historical speed data of each segment for each 15-min period for each day of the week (similar to Balke et al. (10) study). These are denoted by  and where, *s* denotes the segment, *d* denotes day of the week (e.g. Monday, Tuesday, etc.) and *p* denotes time period of the day divided in 15 minutes interval (e.g. 12:00 PM to 12:15 PM, 12:15 PM to 12:30 PM, etc.). Thus, for the IQD rule, the *x0* and ζ are denoted as  and  respectively. This method assumes that for each *s*, *d* and *p*, the speed values are normally distributed and their paramters are assumed to its corresponding median,  and inter-quartile distance, instead of the mean and standard deviation, assumed in Balke et al. (10)**.** Therefore, the degree of outlierness is given by:

 (2)

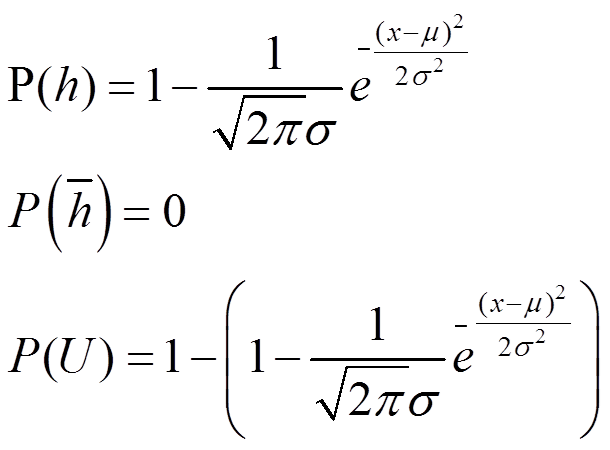
In this paper, Apache Pig Latin is used for computation of these parameters for each segment from the respective historical data of last two months. This required processing of approximately 500 GB of data which is not possible to process via traditional single CPU machines. For this reason, Pig Latin is used. It is a high level Map-Reduce (MR) language to run MR jobs on Hadoop cluster.

**Data Fusion:Dempster-Shafer Theory (DST)**

Dempster-Shafer theory combines evidences obtained from multiple sources (in this case, traffic speed data from radar sensors and probe data). DST is usually applied in cases when there is an uncertainity associated with an event. To incorporate this uncertainity, it introduces a third alternative: “unknown”, along with the measure of confidence for each alternative. In DST, there are three states of knowledge assigned for each state: *incident*, *not incident* and *suspicious* (unknown). However, DST doesn’t perform well in cases when there is a high degree of conflict among evidences. An extension of DST has been suggested recently to overcome this problem.

DST is a mathematical evidence theory based on belief and plausibility. Belief is the lower bound of uncertainity interval and plausibility is its upper bound. It assumes a set of mutually exclusive and exhaustive possibilities, called as the Universe of Discourse or Frame of Discernment. For traffic incdeint detection problem, the Frame of Discernment U consist of two possible values for every minute of incoming traffic data, given as *U*={*incident,¬incident*}. For this U, the power set has three possible elements: hypothesis h = {*incident*} implying that the given interval of traffic data for the particular segment belongs to incident, hypothesis ={*¬incident* } implying that it is not, and universe hypothesis U that is suspicious. The Basic Probability Assignment (BPAs) for each data stream can be given as follows:

*BPA for Incident Detection*: For a time interval detected as outlier by IQD method, the following basic probability assignments are made using the degree of outlierness given in Equation 3.

 (3)

Here, the zero probability of odesn’t indicate impossibility of incident occurrence. It implies that the given data stream doesn’t give any support that the time interval belongs to any incident.

Given the basic probability assignments obtained from the two data streams, radar sensor (RS) and probe segment (PS) data, *PRS(h)* and *PPS(h)*, they are combined into a third basic probability assignment *PPS,RS(h)*, given by

 (4)

Simplying Equation 4 gives the following form:

 (5)

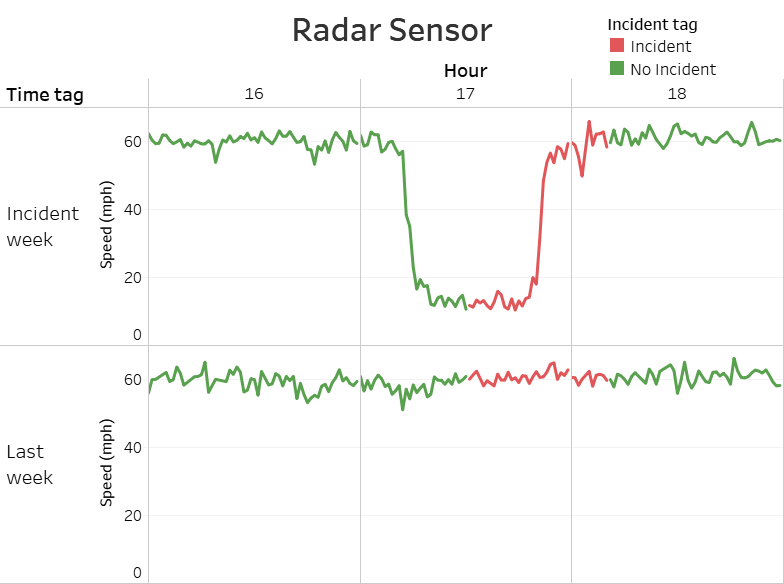
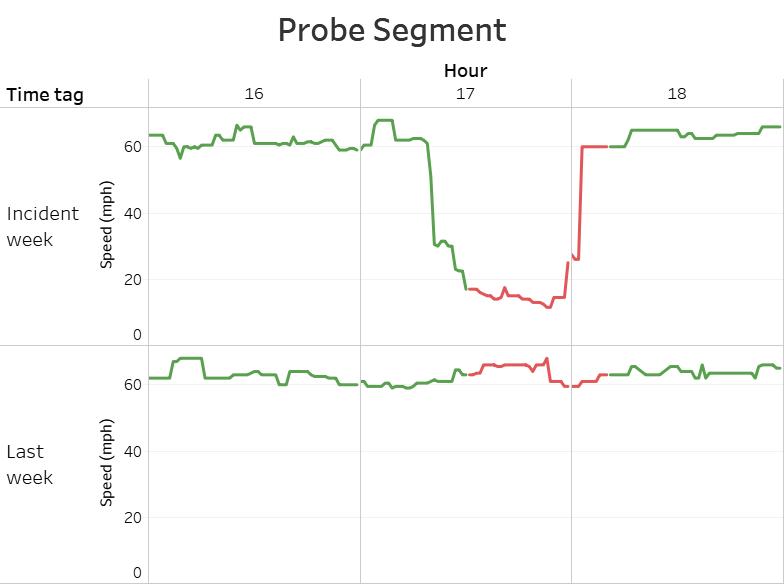
where *pnorm* can be obtained fromEquation 2.

When *PPS,RS(h)* goes above a threshold (to be determined from cross-validation set), then incident alarm can be trigerred.

**RESULTS**

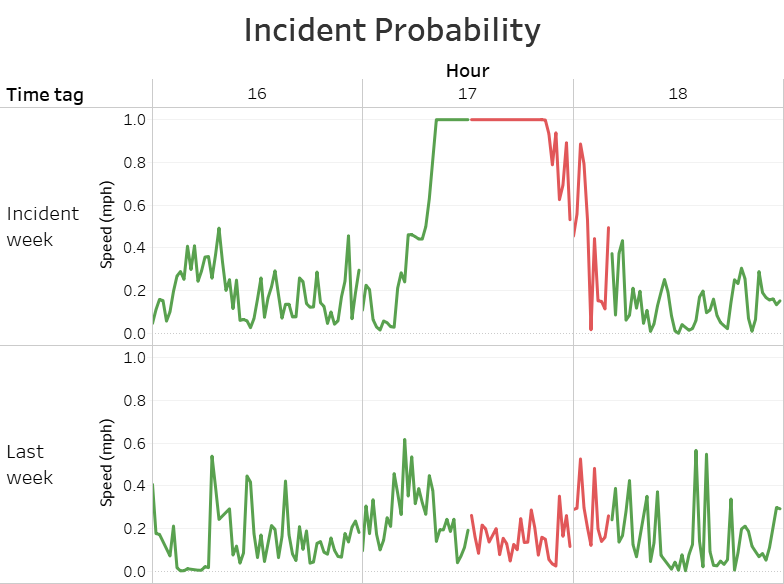
The speed data obtained from RS and PS are used to obtain the corresponding median,  and inter-quartile distance, . Thus, these values are computed every week with the lastest two months data These are then used to find out the *pnormPS* and *pnormRS* and finally the *PPS,RS(h)*. Using Apache Pig Latin, the the combined probability values are obtained for every segment for each time interval from July, 2016 to October, 2016.

Figure 3-5 shows a sample of incident detection by each of the data streams, RS, PS and also fusing the data streams. Figure 3a shows the traffic speed of the RS near the incident during incident time and also in similar time last week. As can be seen, due to the impact of the incident, the traffic speed came down. The red portion indicates the incident time reported by local Traffic Management Center (TMC). It clearly shows that the incident occurred much earlier compared to what was reported. Similarly, Figure 3b shows the speed plot obtained from the PS data. Figure 4 shows the incident probability obtained fusing the two data streams. Very high incident probability (near 1.0) clearly shows the efficacy of the method in detecting incidents. Figure 5 shows the timeline of the incident. The incident started approximately at 5:10 PM while the incident detected by RS, PS and the fusing theory are 5:13 PM, 5:21 PM and 5:14 PM. However, the incident reported by local TMC was at 5:31 PM. It shows that using automatic incident detection algorithm could have resulted in significantly quicker incident detection.

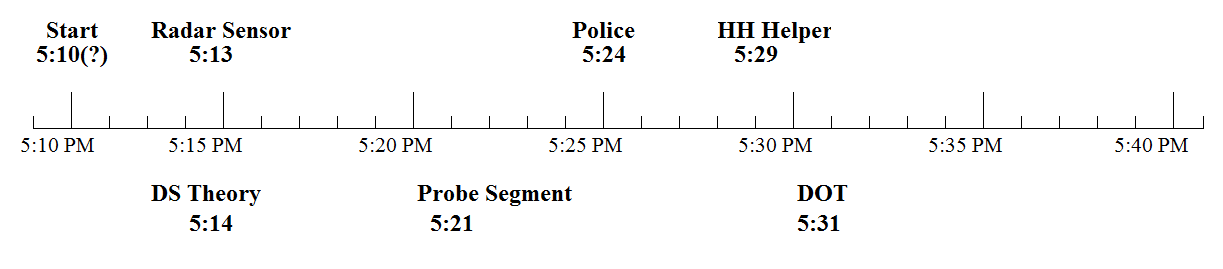


(a) (b)

**FIGURE 3 Traffic speed observed during incident and non-incident day by (a) Radar Sensor and (b) Probe Segment**



**FIGURE 4 Incident probability obtained by Demspter-Shafer Theory**



**FIGURE 5 Timeline of a sample incident**

**CONCLUSIONS**

This paper attempts to fuse traffic data obtained from different streams data for incident detection. Each data stream have its own limitations. Fusing data streams helps in detecting incidents with more confidence. Since large scale data volume needs to be handled for the given purpose, big data analytics is used in this paper. Apache Pig Latin, a high level map-reduce language is used to dereive the parameters and therby detect incidents. Presently, radar sensor data and probe segment data are fused for incident detection. Case study of a given incident detection using the proposed algorithms shows the efficacy of the method used. Future study include determing the threshold of proabability value for trigeering incident alarm. This will be done using validation data of traffic indeints obtained from local traffic management centers and also verifying from camera images. Also, incident reports obtained from waze data will be included in the fusion platform. This will help in detecting incidents which had no effect on traffic stream and hence not possible to detect using only traffic data from sensors or probe segments.

**REFERENCES**

[1] Owens, N., A. Armstrong, P. Sullivan, C. Mitchell, D. Newton, R. Brewster, and T. Trego, Traffic incident management handbook, 2010.

[2] Schrank, D., B. Eisele, T. Lomax, and J. Bak, 2015 urban mobility scorecard, 2015.

[3] Schrank, D. L. and T. J. Lomax, The 2007 urban mobility report, 2007.

[4] Parkany, E. and C. Xie, A complete review of incident detection algorithms & their deployment: what works and what doesn’t, 2005.

[5] Williams, B. M. and A. Guin, Traffic management center use of incident detection algorithms: Findings of a nationwide survey. Intelligent Transportation Systems, IEEE Transactions on, Vol. 8, No. 2, 2007, pp. 351–358.

[6] Castro-Neto, M., L. Han, Y.-S. Jeong, and M. Jeong, Toward Training-Free Automatic Detection of Freeway Incidents: Simple Algorithm with One Parameter. Transportation Research Record: Journal of the Transportation Research Board, No. 2278, 2012, pp. 42–49.

[7] Li, X., W. H. Lam, and M. L. Tam, New automatic incident detection algorithm based on traffic data collected for journey time estimation. Journal of Transportation Engineering, Vol. 139, No. 8, 2013, pp. 840–847.

[8] Sethi, V., N. Bhandari, F. S. Koppelman, and J. L. Schofer, Arterial incident detection using fixed detector and probe vehicle data. Transportation Research Part C: Emerging Technologies, Vol. 3, No. 2, 1995, pp. 99–112.

[9] Sermons, M. W. and F. S. Koppelman, Use of vehicle positioning data for arterial incident detection. Transportation Research Part C: Emerging Technologies, Vol. 4, No. 2, 1996, pp. 87–96.

[10] Balke, K., C. Dudek, and C. Mountain, Using probe-measured travel times to detect major freeway incidents in Houston, Texas. Transportation Research Record: Journal of the Transportation Research Board, No. 1554, 1996, pp. 213–220.

[11] Parkany, E. and D. Bernstein, Design of incident detection algorithms using vehicle-to roadside communication sensors. Transportation Research Record, Vol. 1494, 1995, p. 67.

[12] Hellinga, B., Knapp, G., Automatic Freeway Incident Detection using Travel Time Data from AVI Equipped Vehicles, Paper No. 1059, Department of Civil Engineering, University of Waterloo

[13] Li, Y., McDonald, M., Motorway incident detection using probe vehicles, Proceedings of the Institution of Civil Engineers – Transport, ISSN 0965-092X | E-ISSN 1751-7710, Volume 158 Issue 1, February 2005, pp. 11-15

[14] Zeng, Y., Lan, J., Ran, B., Jiang, Y., A Novel Multisensor Traffic State Assessment System Based on Incomplete Data, ScientificWorld Journal. 2014; 2014: 532602.

[15] Zhu, T., Wang, J., Lv, W., Outlier mining based automatic incident detection on urban arterial road, Proceeding Mobility of the 6th International Conference on Mobile Technology, Application & Systems, Article No. 2, September 2009.

[16] Li, H., Remias, S. M., Day, C. M., Mekker, M.M., Sturdevant, J. R., Bullock, D. M., Shock Wave Boundary Identification Using Cloud-Based Probe Data, Transportation Research Record: Journal of the Transportation Research Board, Volume 2526, DOI: 10.3141/2526-06

[17] Park, H., Haghani, A., Optimal number and location of Bluetooth sensors considering stochastic travel time prediction, Transportation Research Part C: Emerging Technologies, Volume 55, June 2015, Pages 203–216.

[18] Zeng, D., Xu, J., Data fusion for traffic incident detection using DS evidence theory with probabilistic SVMs, Journal of Computers, Vol. 3, No. 10, October 2008.

[19] Pearson, R. K.. Mining imperfect data: Dealing with contamination and incomplete records. SIAM, 2005.

[20] Systematics, C.. Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation. Vol. 6. Federal Highway Administration, 2005.

[21] Chakraborty, P., Hess, R., Sharma, A., Knickerbocker, S. Outlier mining based traffic incident detection using big-data analytics. In Compendium of Papers *Transportation Research Board 96th Annual Meeting*, , DVD-ROM. Washington, D.C., January 8-12, 2017.