Development and Evaluation of Lane Hazard Prediction Application for Connected and Automated Vehicles (CAVs)

Fei Ye,[1,2] *Student Member, IEEE*, Guoyuan Wu,[2] *Senior Member, IEEE,* Kanok Boriboonsomsin,[2]
Member, IEEE, Matthew J. Barth,[2] *Fellow, IEEE*, Samer Rajab,[3] and Sue Bai[3]

Abstract — **More than 37,000 fatalities occurred on U.S. roads in 2016. The number of both vehicle miles traveled and traffic accidents has increased in the past two years, and is mainly attributed to human error, mainly due to inattentiveness during hazardous driving situations. By crowdsourcing traffic data (e.g. vehicle position, speed, direction, etc.) through vehicle-to-vehicle (V2V) communication, connected vehicles (CV) can detect upcoming hazards at lane level with reasonable lead time. The work described in this paper aimed to develop and simulate an innovative V2V-based application to perform lane-level hazard prediction, and a corresponding driver response model. The concept of Lane Hazard Prediction (LHP) is to improve the mobility and safety of both individual users and the entire traffic system. LHP identifies the position of a downstream lane-level hazard (within seconds after it occurs) based on a spatial and temporal data mining and machine learning techniques. It then guides the LHP-equipped vehicles with recommended lateral maneuvers to avoid traffic jams resulting from the hazards. Simulation results demonstrate reliable hazard prediction, even when the V2V penetration rate is as low as 20%. A comprehensive evaluation of the developed LHP application from the perspectives of both user benefits and system benefits has been conducted over different CV penetration rates. The results demonstrate that the proposed LHP application can significantly improve both the safety and mobility performance of the equipped vehicles without compromising the mobility and safety performance of the overall traffic.**

I. INTRODUCTION

Traffic accidents have been increasing over the past two years and are one of the leading causes of non-natural fatalities in the United States [1]. The conventional method of detecting an accident or hazard is based on either fixed-location sensors (such as loop detectors), crowdsourced roadway data, or vehicle onboard sensors (e.g. radar, LiDAR). However, these approaches may not be effective in reducing traffic congestion and potential collision risks, due to their constrained detection range in space and time, or the fact that they only provide partial/road-level information regarding the hazard. In recent years, connected vehicle (CV) technology has been rapidly emerging worldwide as a method to enhance roadway users' safety and mobility, while reducing fuel consumption and emissions. For example, several V2V applications have already been identified by the U.S. National Highway Traffic Safety Administration (NHTSA) as possible candidates to improve roadway safety [2]. Nevertheless, most of these applications require a certain level of market penetration to be effective. A variety of V2V-based applications have been proposed and developed in various projects. Examples of such efforts in the U.S. including:

- Connected Vehicle Reference Implementation Architecture (CVRIA) [3], which summarizes a large number of CV applications developed under the U.S. Department of Transportation (USDOT);

- Safety Pilot model deployment program [4];

- Dynamic Mobility Application (DMA) program [5];

- Applications for the Environment: Real-time Information Synthesis (AERIS) program [6].

In addition to V2V activities in the U.S., the European Union, Japan, and other countries have also been actively supporting research on V2V technology [7-9]. To date, a few research efforts have been conducted using simulation to evaluate the effectiveness of V2V and V2I warnings/alerts about an event or hazard information on the highways in the assumption of the hazard message can be directly accessed through V2V and V2I communication [10-11]. Liu et al. [12] discussed the potential of using variations of instantaneous driving decisions to understand the occurrence of extreme event such as crash or hazard. Given the partial availability of vehicle trajectory information, a reliable hazard prediction model is needed to enable the safety and mobility benefits when a hazard occurs.

The research presented in this paper aimed to develop and simulate an innovative agent-based, lane-level hazard prediction application called Lane Hazard Prediction (LHP) based on partially-available vehicle trajectories data collected from the V2V environment. In addition, the corresponding driver response model was simulated to estimate the effectiveness of the LHP application under various penetration rate. The results of the research shows the potential for LHP to significantly improve the mobility and safety for both individual LHP users and the entire traffic system. The LHP application identifies the position of a downstream lane-level hazard based on a spatial and temporal data mining and machine learning technique. It then guides

[1] Department of Electrical and Computer Engineering, University of California, Riverside, CA 92507 USA.
[2] Center for Environmental Research & Technology, UC Riverside, CA 92507 USA.
[3] Honda R&D Americas, Inc. Email: fye001@ucr.edu, gywu@cert.ucr.edu, kanok@cert.ucr.edu, barth@ece.ucr.edu, SRajab@hra.com, SBai@hra.com.

the application-equipped vehicles with suggestions for proper lateral maneuvers far ahead of the hazard to avoid a traffic jam. This guidance provides the vehicle driver with a suggestion to either change lanes (out of the hazard lane) or maintain the current lane to avoid getting into the hazard lane.

The remainder of this paper is organized as follows:

- Section II introduces the framework of the proposed LHP application, followed by the details of the LHP algorithm.
- Section III presents the simulation setup and test scenarios.
- Section IV presents a comprehensive analysis and discussion of the simulation results.
- Section V concludes this paper.

II. METHODOLOGY

This section describes the overall system architecture of the LHP application, its selected features and prediction model based on binary logistic regression, the development of a driver response strategy, and an evaluation of the LHP application.

A. Framework

The framework of the proposed LHP application is shown in Figure 1. The developed LHP application contains four major modules:

- *Mobile Crowdsourced Sensing*: This module obtains the position, speed, and direction information of CVs downstream within the communication range, with respect to the host application-equipped vehicle through V2V network communication. It then partitions the spatial and temporal domain in the traffic network into lane-level longitudinal segment cells and time slices, performing integration over multiple time steps.

- *Feature Extraction*: Using the cell-based CVs' information in the spatial and temporal domain, this module identifies the key factors that are deemed to be representative and critical for detecting a potential downstream hazard or abnormality in traffic.

- *Lane Hazard Pattern Recognition*: This module runs locally for each partitioned cell and outputs a binary hazard flag (1: hazard exists, 0: no hazard) every 20 seconds. The lane hazard pattern recognition module provides prediction on a lane-level hazard position and its associated longitudinal bound with a resolution of 30 meters, which is the longitudinal segment length for each cell.

- *Lane Recommendation*: Based on the lane-level hazard prediction results, this module suggests a lane change out of the hazard lane for LHP-equipped vehicles. Alternatively, it suggests that the vehicle keep moving in the current, non-hazard lane when approaching the hazard location to avoid joining the associated queue behind it.
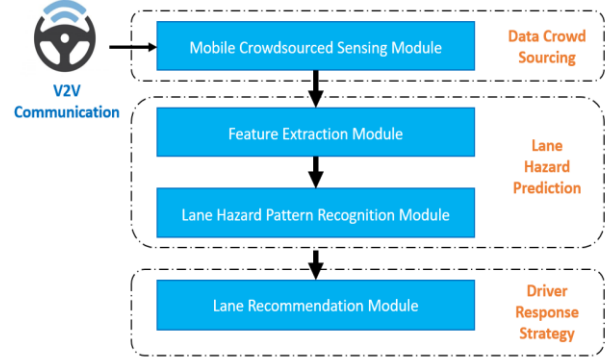

Figure 1: Framework of Lane Hazard Prediction application

B. Lane-Level Hazard Prediction Model

To perform lane-level hazard prediction, the traffic network was partitioned into spatial lane-level cells (30 meters long). The data integration and temporal resolution for the developed LHP application was 20 seconds. The information (e.g. lane-level position, speed, and direction) obtained from CVs over the V2V network were accumulated at a rate of 0.05 Hz, integrating over 20 seconds at each cell. The binary hazard predictor ran locally for each cell, which facilitated its adaptability to different geographic locations and the scalability to a larger scope. Patterns were observed that could potentially identify unusual collective behaviors for vehicles approaching the hazard location.

By combining the knowledge of traffic engineering with a data-driven approach, a total of eight features were identified as input variables for the binary logistic regression-based LHP model. For each cell (i, j) in the traffic network (where i represents the longitudinal position, and j indicates the lane number), measurements were considered from the ego-cell as well as its adjacent cells in both the upstream and downstream segments, as shown in Figure 2.
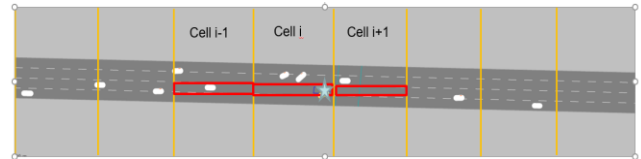

Figure 2: Illustration of scaling and partitioning on an example network

The vehicle maneuvers within a cell were categorized into five classes:

- M1—Through maneuver including both entry and exit
- M2—left lane change out of the cell
- M3—right lane change out of the cell
- M4—right lane change into the cell
- M5—left lane change into the cell

Var.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Coeff.	-2.42	-2.24	-2.21	-2.23	-2.25	-1.90	0.88	-0.03	-0.17

The input features considered for the algorithm included:

- The average speed of all vehicle travel across the cell during the 20-second time interval, calculated by the ratio of vehicle miles traveled to vehicle hours traveled
- The average speed ratio between the cell and across all the lanes at the same/upstream adjacent/downstream adjacent longitudinal segment
- The percentage of through maneuvers in the cell and lane changes into and out of the cell
- The entropy of vehicle maneuvers, which captured the diversity of all maneuvers in the cell. The entropy attained its minimum value of zero when all the vehicle maneuvers were from the same category, and its maximum value when all the vehicle maneuvers were uniformly distributed over different categories.

The logit function constrains the values of landslide susceptibility index of the model in the range [0, 1] (the index threshold was set as 0.75). The logistic regression-based LHP model is described in Equation (1):

$$logit(P_{ij}) = ln\left(\frac{P_{ij}}{1 - P_{ij}}\right)$$

$$= \beta_0 + \beta_1 \times \overline{V}_{ij} + \beta_2 \times \frac{\overline{V}_{ij}}{\overline{V}_i} + \beta_3 \times \frac{\overline{V}_{ij}}{\overline{V}_{i-1}} + \beta_4 \times \frac{\overline{V}_{ij}}{\overline{V}_{i+1}} + \beta_5 \times \frac{m_1}{m} + \beta_6 \times \frac{m_2+m_3}{m} + \beta_7 \times \frac{m_4+m_5}{m} + \beta_8 \times \sum_{i=1}^{n} \frac{m_i}{m} log\left(\frac{m_i}{m}\right)$$

(1)

Therefore, the probability of a hazard in cell (i, j) can be obtained by

$$P_{ij} = \frac{1}{1+exp\left(logit(P_{ij})\right)} \qquad (2)$$

Where,
$\overline{V}_{ij}$ is the average speed of cell (i, j)
$\overline{V}_i$ is the average speed across all the lanes of longitudinal segment i
$\overline{V}_{i-1}$ is the average speed across all the lanes of upstream adjacent longitudinal segment i-1
$\overline{V}_{i+1}$ is the average speed across all the lanes of downstream adjacent longitudinal segment i+1
m_i is the number of maneuvers at cell (i, j), belonging to predefined maneuver type i
m is the total number of maneuvers at cell (i, j)
n is the number of maneuver types
β_k is the parameters' coefficients

The parameters calibration results of the proposed Lane Hazard Prediction model based on logistic regression approach is shown in Table 1.

C. Driver Response Strategy

Using the output of the LHP algorithm, the lateral, lane selection, and driver behavior of LHP-equipped vehicles could be modified to avoid a hazard. This guidance is performed when a hazard flag is activated downstream of the ego-vehicle's current position in the same lane, within the V2V communication range (assumed in this work to be 2,000 meters). In addition, the upstream LHP-equipped vehicles in the other lanes would be guided to stay in the current non-hazard lane until they passed the hazard position.

D. Lane Hazard Prediction Evaluation

The LHP evaluation was conducted with respect to prediction accuracy, efficiency, and application effect on safety and mobility. The LHP application provided a lateral maneuver decision based on the results of a lane-level hazard prediction model. Both effectiveness and efficiency of the proposed LHP application were evaluated under different levels of penetration rate and traffic conditions.

In this study, the receiver-operating characteristic (ROC) curve and the area under the curve (AUC) were used to evaluate the performance of the developed LHP model [13]. The ROC curve illustrated the tradeoff between the true positive rate and the false positive rate of the binary logistic regression-based LHP model. The closer the ROC curve got to the top left corner of the graph, with a larger AUC number, the better the LHP model performed in terms of predicting a hazard, with limited false positives. The AUC for random guessing was 0.5 with the ROC curve following the diagonal. The reaction time of the LHP model, defined as the time in seconds from the point when the hazard occurred to the point when the first accurate prediction was triggered, was used as a measure of efficiency. In addition to evaluating the performance of the hazard prediction model, the LHP application's effectiveness in terms of safety and mobility were also assessed. The safety performance of LHP-equipped vehicles, unequipped vehicles, and the overall traffic was analyzed using a conflict frequency measure of effectiveness. The average speed measure of effectiveness was used to evaluate the mobility performance, provided in following Equation (3):

$$\bar{v} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} VMT_{i,t}}{\sum_{i=1}^{n} \sum_{t=1}^{T_i} VHT_{i,t}} \qquad (3)$$

Where,

$VMT_{i,t}$ = vehicle miles traveled for vehicle i at timestep t

$VHT_{i,t}$ = vehicle hours traveled for vehicle i at timestep t

III. SIMULATION SETUP

This section introduces the simulation tools, network model, and scenarios for a comprehensive analysis of the LHP application.

A. Simulation Tools

PTV VISSIM [14] was used in this study as the microscopic traffic simulation tool for traffic network modeling, development of the LHP application using an application programming interface (API), and evaluation of individual vehicles and the overall traffic system performance. As a state-of-the-art microscopic, timestep-oriented, and behavior-based traffic simulation tool, VISSIM is capable of simulating a large-scale road network and wireless communication network and calibrating the traffic flow and speed with real-world data. In addition, VISSIM provides an add-on programming interface, Component Object Model (COM), which was used in this study to modify the underlying simulation models, access model outputs, and override default vehicle behaviors.

For safety performance evaluation, the Surrogate Safety Assessment Model (SSAM) [15] was used as a post-processing model to perform safety evaluation by analyzing vehicle trajectory data (.trj files) generated from VISSIM. In SSAM, the safety performance was assessed through measured conflict potentials, considering both the risk of longitudinal collisions (rear-end conflicts) and lateral collisions (lane change conflicts).

B. Simulation Network Model

The real-world network used in this study was a 17-mile stretch of Interstate 270 (I-270) North, in Ohio, with seven on-ramp/off-ramp pairs (see Figure 3). The longest segment located between the second on-ramp/off-ramp pair was selected, and an accident event was created to test LHP performance. The speed limit on I-270N is 70 mph, and the traffic demands are well calibrated with real-world data from the Traffic Count Database System (TCDS) by the Ohio Department of Transportation [16].
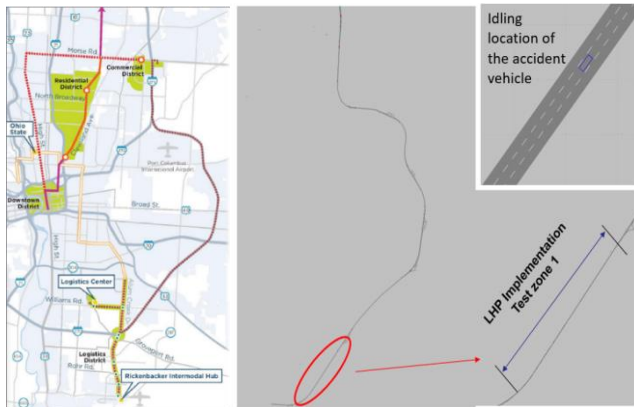

Figure 3: Road network of I-270N in real world and VISSIM

C. Simulation Scenarios

To gain in-depth insight into the developed LHP application performance, simulation tests were conducted under various V2V penetration rates. V2V market penetration rate is expected to take several years to achieve a significant level. A full penetration rate (i.e., 100%) enables the LHP application to attain the most accurate traffic measurements, leading to higher prediction accuracy and shorter reaction times. However, since 100% penetration is not currently available, the sensitivity analysis over different penetration rates is meaningful. The application provides solid performance, even at lower penetration rates, as discussed in the results section.

The simulation test was performed on a three-mile stretch of road, and the simulation duration was 1,800 seconds. The LHP-equipped vehicle percentage was set to 9% out of the total V2V-equipped vehicles. This percentage corresponds to Honda's market share. Therefore, three types of vehicles ran in the simulation network: 1) LHP- and V2V-equipped vehicles; 2) V2V-equipped vehicles; and 3) conventional vehicles. Conventional vehicles do not have V2V communications capability, and their behavior follows VISSIM default lane change and car following models. V2V-only vehicles exchange real-time information (e.g., speed, lane-level position, etc.) with other V2V vehicles without the onboard application. LHP-equipped vehicles exchange information via V2V, perform hazard predictions, and change (or maintain current) lanes to avoid hazards downstream. Seven levels of penetration rate (0%, 5%, 10%, 20%, 50%, 80%, and 100%) V2V-equipped vehicles were evaluated (in which LHP-equipped vehicles accounted for 9% of the above penetration rate levels). Both user benefits and system benefits were evaluated in terms of mobility (average speed difference) and safety (conflict frequency difference) under different penetration rate. Ten simulation runs for each simulation setting were conducted.

IV. RESULTS AND DISCUSSION

This section presents the evaluation of LHP application from three aspects:

1) Prediction analysis in terms of accuracy and efficiency of the developed LHP model
2) Safety analysis of LHP-equipped and unequipped vehicles
3) Mobility analysis of LHP-equipped and unequipped vehicles

B. Prediction Performance

As previously described, ROC curves and associated AUC values were used to assess the accuracy of the developed lane hazard predictor. In addition, the reaction time of LHP was used as the key index to evaluate the algorithm efficiency. The results of sensitivity analysis for different levels of penetration rate are shown in Figure 4 and Figure 5. With a 100% penetration rate of connected vehicles, the LHP algorithm provided the best performance with 0.98 AUC. Other results were as follows (see Figure 4):
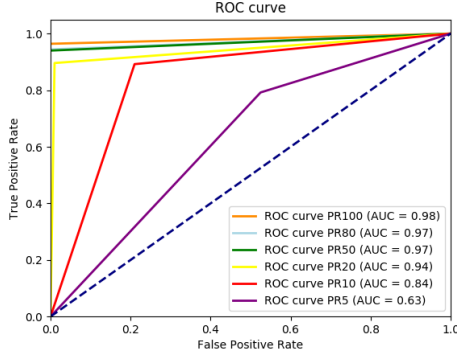
Figure 4: ROC curve of LHP under different penetration rates

As for the application efficiency, the average reaction time of LHP was less than 60 seconds across the penetration rates between 20% and 100% (see Figure 5). However, there was a rapid increase in reaction time at the 5% penetration rate, with larger variation. This result may be caused by a noisy measurement due to the low number of vehicles equipped with LHP at the 5% penetration rate. Therefore, LHP was proven to be highly efficient and reliable at a penetration rate as low as 20%. At lower penetration rates, such as 5% or 10%, the LHP application can be considered a lane advisory system.
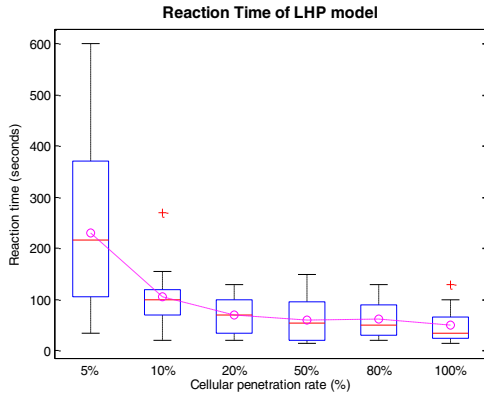

Figure 5: Reaction time of LHP under different penetration rates

C. Mobility Performance

Mobility performance was quantified by the average speed, which was calculated using Equation (3). The comparative results in terms of user benefits and system benefits were obtained using the average speed relative ratio (%), as denoted in Equations (4) and (5), respectively.

$$\frac{MOE_e - MOE_{ue}}{MOE_{ue}} * 100\% \qquad (4)$$

Where, MOE_e is the metric of equipped vehicles (i.e. average speed of equipped vehicles); MOE_{ue} is the metric of unequipped vehicles (i.e. average speed of unequipped vehicles)

$$\frac{MOE_{oa} - MOE_{bl}}{MOE_{bl}} * 100\% \qquad (5)$$

Where, MOE_{oa} is the metric of overall vehicles with LHP scenarios; MOE_{bl} is the metric of overall vehicles with 0% penetration rate as the baseline.

The boxplot and error bars of the average speed relative ratio in terms of V2V penetration rates are shown in Figure . According to Figure , average speed improvements (up to 7%) for LHP-equipped vehicles were witnessed across all penetration rates. When the penetration rate was high enough (i.e., greater than 10%), the improvement tended to be stable, with much less variation. Considering the system mobility benefits, the average speed relative ratio of overall vehicles compared to the baseline (penetration rate of 0%) varied between -0.2% and 0.7%. Overall, the applied LHP model had negligible effects on the system-level mobility, regardless of the penetration rates.
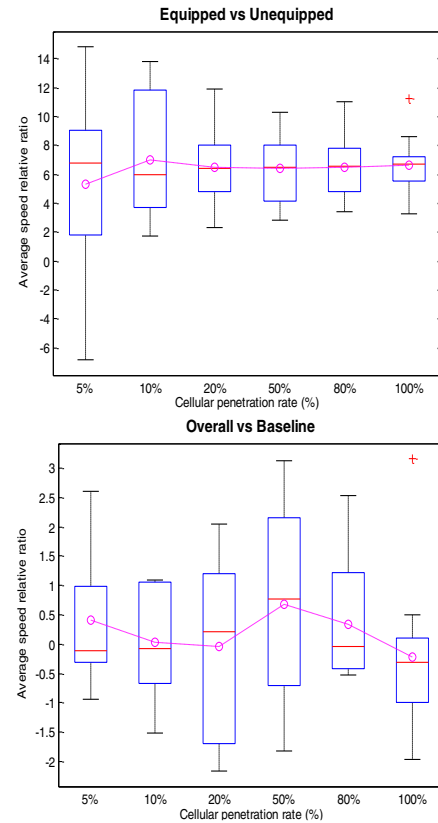

Figure 6: Sensitivity Analysis of penetration rate on mobility benefits (Top: User benefits, Bottom: System benefits)

D. Safety Performance

The rear end and lane change conflict frequency were selected to assess vehicle safety performance. The decrease in conflict frequency implied an improved safety performance. The comparison of LHP-equipped and unequipped vehicles was quantified with a rear end conflict frequency relative number, as denoted in Equations (4) and (5). As shown in Figure , significant user benefits in terms of safety could be achieved across all simulated penetration rates.

The average conflict frequency difference was decreased by a factor of -32.8% ~ -68.1%. The user benefits in terms of

safety achieved the highest point at a penetration rate of 100% and the lowest at a 5% penetration rate. The results were more stable at higher penetration rates. As indicated in Figure , positive effects on system benefits in terms of safety were also witnessed across all penetration rate levels. The average conflict frequency ratio varied between -1.4% and -3.0%. The conflict frequency was significantly reduced from both user and system perspectives under different penetration rate. The significant improvement in safety performance may have resulted from the fact that the LHP application enabled equipped vehicles to avoid sharp speed drops upstream in the hazard lane and encouraged lane change behavior earlier and in a smoother manner.
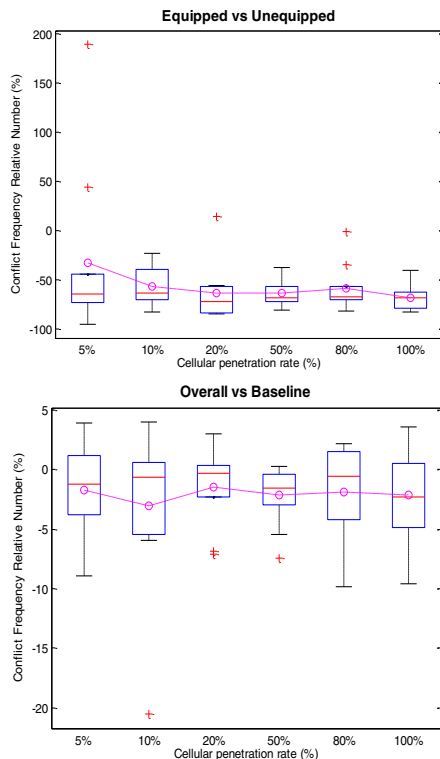

Figure 7: Sensitivity Analysis of penetration rate on safety benefits (Top:User benefits, Bottom: System benefits)

I. CONCLUSION

An innovative V2V-based lane-level hazard prediction algorithm and corresponding driver response model called Lane Hazard Prediction (LHP) was developed and evaluated. Results of a comprehensive simulation study showed that the LHP application could provide highly accurate lane-level prediction of a downstream hazard within tenths of seconds after it occurred, by crowdsourcing V2V communications information. The LHP application provided lateral maneuver guidance to LHP-equipped vehicles. A detailed simulation study was performed to assess the effectiveness of the proposed LHP application in terms of safety and mobility benefits. Results demonstrate that LHP-equipped vehicles may gain significant mobility and safety benefits without compromising the mobility and safety performance of the overall traffic. An attractive feature of the proposed LHP application is that accurate prediction within seconds and noticeable benefits in safety and mobility can be achieved, even under a relatively low V2V penetration rate.

REFERENCES

[1] The U.S. Department of Transportation's National Highway Traffic Safety Administration, https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data Oct. 2017.

[2] NHTSA, "Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application", 2014.

[3] Iteris, "Connected Vehicle Reference Implementation Architecture," http://www.iteris.com/cvria/html/applications/applications.html. 2016.

[4] US Department of Transportation (USDOT), "Connected Vehicle Safety Pilot," http://www.its.dot.gov/safety_pilot/index.htm. Oct 2, 2015.

[5] USDOT, "Dynamic Mobility Applications (DMA)," http://www.its.dot.gov/dma/index.htm. Oct 2, 2015.

[6] USDOT, "Applications for the Environment: Real-time Information Synthesis (AERIS)," http://www.its.dot.gov/aeris/index.htm. June, 2015.

[7] Krajzewicz *et al*., " COLOMBO: Investigating the Potential of V2X for Traffic Management Purposes Assuming Low Penetration Rates," 2013.

[8] Zinoviou *et al*., "Performance Evaluation of an Adaptive Route Change Application Using an Integrated Cooperative ITS Simulation Platform," Proceedings of 8th International Wireless Communications and Mobile Computing Conference (IWCMC), pp.377-381, 2012.

[9] European Commission, "PRE-DRIVE C2X: PREparation for DRIVing implementation and Evaluation of C2X communication technology," Technical Report, 2010.

[10] Nair G.S. et al. "A Microsimulation Approach to Quantify the Safety Benefits of Connected Vehicles: A Road Hazard Warnings Application", Transportation Research Board 96th Annual Meeting, 2017.

[11] Yeo, H., et al, " Microscopic Traffic Simulation of 13 Vehicle-to-Vehicle Hazard Alerts on Freeway". Transportation Research Record: Journal of 14 the Transportation Research Board, Vol. 2189, pp. 68–77, 2010

[12] Liu, J., Khattak, A., "Delivering improved alerts, warnings, and control asssitance using basic safety messages transmitted between connected vehicles", Transport.Res. Part C: Emerg. Technol. Vol.68, pp. 83-100.

[13] Hanley, J. A., and B. J. McNeil, " The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology", 1982, vol.143, pp. 29–36.

[14] PTV Vissim Group http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/ , 2018.

[15] Federal Highway Administration, "Surrogate Safety Assessment Model and Validation," Final Report, FHWA-HRT-08-051, 2008.

[16] Traffic Count Database System, http://www.ms2soft.com/products/traffic-counts