This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

A Machine Learning Based Defensive Alerting System Against Reckless Driving in Vehicular Networks

Lan Zhang, Li Yan, Yuguang Fang, *IEEE Fellow*, Xuming Fang, and Xiaoxia Huang

***Abstract*—Reckless driving severely threatens the safety of innocent people, which accounts for around 33% of all fatalities in major vehicle accidents. However, most existing efforts focus on the detection and adjustment of a vehicle's own driving behavior, whose effectiveness is very limited. In this paper, we develop a defensive alerting system to detect and notify the threats posed by reckless vehicles. By integrating the computation capability of a cloud server with that of vehicles nowadays, we propose a cooperative driving performance rating (CDPR) mechanism to automatically and intelligently detect reckless vehicles based on machine learning algorithms. To support such a defensive alerting system, a three-tier system architecture is developed from existing vehicular networks, which consists of vehicles, road-side units (RSU) and a cloud server. Moreover, given the fact that most vehicles can be trusted to drive safely, to further reduce the transmission load of the CDPR mechanism, we design our scheme in such a way that every vehicle only uploads the data of driving maneuvers with reckless potential to RSUs. Based on the aggregated data, the cloud server globally rates every vehicle's driving performance by using support vector machine (SVM) and decision-tree machine learning models. We finally implement the proposed CDPR mechanism into a popular traffic simulator, Simulation of Urban MObility (SUMO), for evaluation. Simulation results illustrate that our defensive alerting system can accurately detect reckless vehicles and effectively provide timely alerts.**

***Index Terms*—Cooperative driving behavior rating, Defensive reckless driving alert, SUMO simulation, Machine learning.**

I. Introduction

Driving safety is no doubt the most critical concern behind the wheel. According to the statistics from the National Safety Council (NSC) of United States (U.S.), more than 40,000 roadway fatalities happened in 2017 [1]. A major reason leads to these road tragedies is the human factor, where reckless driving is the most considerable one [2], [3]. According to the Department of Motor Vehicles (DMV) in U.S., any driving behavior disregarding the safety of other individuals can result in a charge of reckless driving. Even with corresponding law enforcements, reckless driving still accounts for around 33% of all fatalities in major car accidents [4]. Therefore, in addition to the law enforcement, how to proactively detect and avoid the safety threats from reckless driving becomes a vital concern.

One effective mechanism is to monitor and regulate every vehicle's own driving performance, which has been intensively studied. According to several survey papers [5]–[8], a vehicle can collect the driving information of itself through its own sensing devices, such as the electroencephalograph (ECG) sensor embedded in the steering wheel, the dome camera, the inertial measurement unit, and even the smartphone, etc. By analyzing the collected sensing data, either a vehicle's driving performance or a driver's state of mind that may reflect the driving performance is evaluated. For example, the biological measure such as the ECG signal can be used to evaluate the attentiveness of a driver [6]; the driving state such as the velocity and acceleration can be used to determine the driving performance [5]; and the external contexts such as the traffic and weather conditions can be combined with the internal contexts, such as the velocity of the vehicle and the level of alcohol in the driver's blood, to detect the abnormal driving behavior [9]. Based on the detection results, the system may provide either passively or actively corrective feedback to the driver, where the audio or video advisories are regarded as passive feedbacks and the direct interventions such as modifying the pedal force are regarded as active feedbacks [5]–[7]. However, even with accurate detection, the feedback suitability as well as the effectiveness are still inadequate. Besides, a driver, no matter whether he/she is a reckless driver or not, will not voluntarily share his/her behavior information with others. Thus, a reckless driving behavior detection should use other drivers' observation to draw a conclusion. Moreover, we unfortunately cannot change the driving behavior of others. Given the fact that the human reaction time in vehicular accidents is around 1.5 seconds [10], it is difficult to avoid a sudden crash caused by a reckless driving vehicle within such a short time. Therefore, a reckless driving defensive alerting system is imperative to help a driver to proactively avoid the safety threats from the approaching reckless vehicle.

Apparently, the aforementioned self-rated driving performance [5]–[9] is not reliable nor trusted for others. In order to provide accurate alerts of reckless vehicles, every vehicle's current driving performance needs to be objectively rated.

This work was partially supported by US National Science Foundation under grants CNS-1717736 and IIS-1722791. The work of X. Huang was supported by NSFC-Guangdong Joint Program under Grant U1501255 and Guangdong Special Support for Young Innovation Talents under grant 2016TQ03X279.

L. Zhang and Y. Fang are with Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA (email: lanzhang@ufl.edu, fang@ece.ufl.edu.

L. Yan and X. Fang are with Key Lab of Information Coding & Transmission, Southwest Jiaotong University, Chengdu 610031, China (email:liyan12047001@my.swjtu.edu.cn, xmfang@swjtu.edu.cn).

X. Huang is with School of Electronics and Communication Engineering, Sun Yat-Sen University, Guangzhou, China (email: xiaoxiah@gmail.com).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

Therefore, to develop a fair rating mechanism, a trustworthy third party needs to be involved to monitor every vehicle's driving performance and provide the rating results. An intuitive monitoring mechanism is to utilize the surveillance cameras deployed along the road [11]–[13]. However, on the one hand, there are not enough deployed camera devices to seamlessly cover every corner, let alone the cost of the massive installation and maintenance. On the other hand, the technologies of vision-based vehicle monitoring have not been well-developed yet [11], [12], [14]. Even the vehicle tracking, the most fundamental task of vision-based driving performance monitoring, is still challenging due to the abrupt object motion, appearance pattern change, non-rigid object structures, etc., especially in the large-volume traffic and highly-dynamic driving environments [12], [14]. Based on these observations, in the paper, we propose a cost-effective monitoring mechanism, where a vehicle's driving performance is monitored by its neighbor vehicles. Similar idea on utilizing the cooperation among vehicles has also been applied to design a stable routing protocol in vehicular networks [15].

With the rapid development of automobile manufacturing [5]–[8], vehicles nowadays exhibit the capability to monitor their surrounding driving environments. For example, the ultrasonic sensors are used to detect the surrounding obstacles for parking assistance; the radar sensors are used to sense the road ahead for partially automated driving through adaptive cruise control (ACC); the inexpensive low-range and low-resolution versions of lidar sensors are used for forward collision prevention [8], [16]. Based on these equipped sensor devices, we can reasonably assume that a vehicle can locally monitor the driving performance of its neighbor vehicles. However, due to the highly-dynamic driving environments as well as the complicated external factors, such as the influence from other reckless vehicles, a vehicle cannot acquire adequate sensing data of a neighbor vehicle and thus cannot provide an accurate rating result. Therefore, we propose to acquire the driving performance data of a vehicle by aggregating the monitored data from its previous and current neighbor vehicles. Recently, machine learning algorithms have been widely utilized in networking areas, such as network traffic control, due to its efficiency and effectiveness to cope with the dynamic, large-volume and complicated data in a more intelligent and autonomous fashion [16]–[21]. In this paper, based on the aggregated multi-modal monitored data, we utilize machine learning algorithms to more accurately and efficiently rate the driving performance of a vehicle and detect the reckless driving behavior. To support such sensing, aggregating and rating functions, we design a three-tier system architecture based on existing vehicular networks. To the best of our knowledge, none of existing works establishes a reckless driving defensive alerting system by integrating the computation capabilities of neighbor vehicles with that of the cloud server.

Our main contributions are summarized as follows:

- To better understand the consequences of reckless driving, we first theoretically evaluate the crash probability of a typical reckless driving maneuver, i.e., reckless lane changing. Our analytical result indicates a long enough distance is required to against reckless driving, which motivates us to devise a defensive alerting system to proactively enlarge the driving distance gap towards reckless vehicles.
- We propose a reckless driving defensive alerting system to proactively detect and notify the approaching of reckless vehicles. By integrating the computation capabilities of the cloud server and vehicles nowadays, we propose a cooperative driving performance rating (CDPR) mechanism to automatically and intelligently detect reckless vehicles. To facilitate such a defensive alerting system, a three-tier system architecture is developed from existing vehicular networks, where neighbor vehicles of any vehicle monitor its driving performance and then upload their monitored data to a local road-side unit (RSU) to be forwarded to the cloud server if needed. Given the fact that most vehicles can be trusted to drive safely, we further reduce the transmission load by only uploading the driving performance data with reckless potential.
- We devise the machine learning based driving performance rating algorithms in the cloud server to achieve accurate detection based on more comprehensive aggregated driving data. To evaluate the performance of the proposed alerting system, we implement our design into a popular trafc simulator, Simulation of Urban MObility (SUMO). Experimental results demonstrate the effectiveness of our defensive alerting system, where the accuracy of reckless vehicle detection and the timeliness of reckless vehicle alerting are comprehensively evaluated.

The rest of this paper is organized as follows. In Section II, we describe reckless driving behavior and theoretically analyze the consequence of a typical reckless driving maneuver, i.e., reckless lane changing, which further motivates our analysis. In Section III, we design the reckless driving defensive alerting system, where the CDPR mechanism, system architecture and signaling process are presented. The machine learning based driving performance rating models are analyzed in Section IV. In Section V, we evaluate the performance of our defensive alerting system through simulations. Section VI finally concludes this paper.

II. Reckless Driving

To better understand the consequences of reckless driving, we first describe the characteristics of reckless driving, and then theoretically analyze its consequences through a case study.

A. *Reckless Driving Performance*

Although more advanced safety mechanisms are utilized in modern automobiles nowadays than ever, the number of driving fatalities is still alarming high. In fact, any driving behavior disregarding the safety of other individuals can be regarded as reckless driving. Although the charge of reckless driving may vary in different states in USA based on their corresponding law enforcement [4], there is still general consensus of reckless driving, where several driving maneuvers can be regarded as

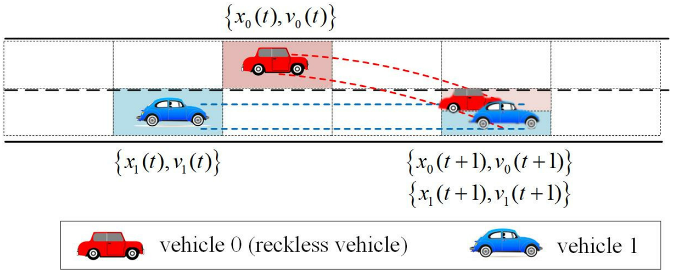


Fig. 1: A typical scenario of reckless lane changing maneuver and its corresponding cellular automata model.

"reckless". In the following, we list some of the most common reckless driving maneuvers reported annually:

- Tailgating the leader vehicle, i.e., following too close;
- Intentionally failing to yield the right-of-way, such as reckless lane changing;
- Excessive speeding at a considered dangerous velocity;
- Placing others in danger by running red lights/stop signs;
- Driving under the influence (DUI) or driving while intoxicated (DWI), such as lane drifting under drunk driving;
- Distracted driving, such as texting while driving.

These driving maneuvers, each dangerous in its own right, are responsible for the majority of reckless driving caused accidents, which lead to serious injuries and medical costs for innocent people. In this paper, we focus on how to proactively detect and avoid the safety threats from the vehicles that constantly exhibit reckless driving behavior and put themselves and others in danger. To better understand the driving characteristics of reckless vehicles, we utilize a theoretical model to analyze its consequences. Due to the complexity of modeling human driving behavior, we focus on a typical reckless driving maneuver, i.e., reckless lane changing.

B. Case Study: Reckless Lane Changing

According to the beginner driver's guide, a proper lane changing should follow these procedures: turn on the blinker, then check the rear view, side mirror and the blind spot, and finally maintain the speed to smoothly and quickly move to the desired lane. However, a reckless vehicle may aggressively move to the other lane regardless of these proper procedures, which may result in serious accidents [22]. As illustrated in Fig.1, a crash happens when vehicle 0 (assumed as a reckless vehicle) changes to the lane of the vehicle 1 without checking the blind spot. Besides this scenario, the crash between vehicle 0 and 1 may also happen in the following scenarios: without using the right blinker, vehicle 1 may accelerate and crash into the suddenly drifting vehicle 0; vehicle 0 may mis-estimate the distance gap or relative velocity towards vehicle 1 and fail to yield to vehicle 1; vehicle 0 may decelerate during its lane changing or slowly drift to the target lane; etc.

To theoretically analyze the aforementioned scenarios, we utilize a symmetric two-lane cellular automata (CA) model for highway scenario, where each lane is divided into multiple cells [23]. As illustrated in Fig. 1, each cell is either empty or occupied by a vehicle. The status of a cell is updated at each discrete time step $t \in \{1, 2, 3, ...\}$. During the time $t \to t+1$, a vehicle located at $x(t)$ with velocity v^t will move to

$$x(t+1) \to x(t) + v^t, \tag{1}$$

where v^t is the value of horizontal velocity along the road at time step t.

Without loss of generality, in our lane changing scenario, we assume the reckless vehicle 0 located at $x_0(t)$ with velocity v_0^t starts changing to the lane of vehicle 1 at time step t, where vehicle 1 is the following vehicle located at $x_1(t)$ with velocity v_1^t, i.e., $x_0(t) > x_1(t)$. Note that the scenario that vehicle 1 is the ahead vehicle of the target lane can be similarly analyzed. Since the lane changing process is finished within 1 time slot [23], a crash happens when vehicle 0 is not ahead of vehicle 1 at time step $t+1$, i.e., $x_0(t+1) \leq x_1(t+1)$. Therefore, according to (1), a crash happens when the distance gap $d = x_0(t) - x_1(t)$ satisfies the following condition:

$$d \leq v_1^t - v_0^t. \tag{2}$$

Under the symmetric lane changing setting [23], we do not distinguish fast and slow lane, and thus v_0^t and v_1^t are assumed to be independent random variables. Denote $f_{v^t}(v)$ as the probability density function (PDF) of v^t. At time step t, given the value of v^{t-1}, we assume the mean value of the random variable v^t is the given value of v^{t-1}, which matches the fact that a vehicle's current velocity is closely related to its velocity at previous time step. The probability that a crash happens between vehicle 0 and vehicle 1 can be given by

$$\begin{aligned} P_C &= \Pr\left\{d \leq v_1^t - v_0^t\right\} \\ &= \int_d^{+\infty}\int_{-\infty}^{+\infty} f_{v_1^t}(v_1 + v_0)\, f_{v_0^t}(v_0)\, \mathrm{d}v_1 \mathrm{d}v_0 \\ &= \int_d^{+\infty} f_{v_0^t}(v) * f_{-v_1^t}(v)\mathrm{d}v. \end{aligned} \tag{3}$$

Given the PDF of v_0^t and v_1^t, we have the PDF of $v_0^t - v_1^t$, i.e., $f_{v_0^t - v_1^t}(v) = f_{v_0^t}(v) * f_{-v_1^t}(v)$. Thus, according to (3), the crash probability P_C decreases with the increasing of distance gap d, which reveals that vehicle 1 should keep distance to vehicle 0 to reduce their crash probability. Thus, the utility function of safety can be given by $U_{\text{safe}}(d) = \alpha(1 - P_C(d))$, where α is the weighting factor. However, on the other hand, the increased distance gap d reduces the transportation efficiency, which cannot be ignored especially in crowded scenarios such as in rush hours. Define the utility function of road efficiency by $U_{\text{eff}}(d) = \beta/d$, where β is the weighting factor. Thus, the total utility of a lane changing maneuver by a reckless driver is composed of $U_{\text{safety}}(d)$ and $U_{\text{eff}}(d)$.

On the other hand, for the normal driving vehicles based on the aforementioned proper lane changing procedures, we assume the crash probability is small enough to be ignored. Therefore, the utility function of a lane changing maneuver can be given by

$$U_i(d) = \begin{cases} \alpha(1 - P_C(d)) + \beta/d, & i = \text{reckless driving} \\ \beta/d, & i = \text{normal driving}. \end{cases} \tag{4}$$

According to the utility function in (4), it is noticeable that the optimal distance gap d^* is different for reckless and normal lane changing maneuvers. Specifically, the d^* in reckless driving is larger than that in normal driving due to the safety threat by reckless vehicles, i.e., the safety utility U_{safe} increases with d. When a reckless vehicle is correctly identified, a larger distance gap is absolutely preferred, which can be realized by a proactive reckless vehicle alert. However, in contrast, when a normal vehicle is mis-classified as reckless, a larger distance gap will lead to a smaller utility as well as less transportation efficiency. Therefore, the defensive alert of reckless driving is imperatively necessary, and meanwhile an accurate alarm is expected to improve the overall system performance, which motivates our further analysis.

III. Reckless Driving Defensive Alarming System

Given the severe consequences of reckless driving, we propose a machine learning based defensive alarming system to proactively detect and alert reckless vehicles. Specifically, we design a cooperative driving performance rating (CDPR) mechanism to automatically detect reckless vehicles, whose detection results are sent to the drivers nearby. These functions are supported by a three-tier system architecture. In the following, we first introduce the CDPR mechanism, and then present our system architecture followed by the signaling process.

A. The CDPR Mechanism

As a complementary safety monitoring mechanism besides the traffic polices and video surveillance, the CDPR mechanism utilizes the monitoring capability of vehicles. On the one hand, more and more sensor devices are installed on vehicles, such as inertial sensors, ultra-sonic, radar, lidar, infrared sensors, vision sensors, etc., which can be utilized to acquire the information of driving performance of neighbor vehicles and provide a way for neighbor vehicles to monitor themselves, such as the distance gap to neighbor vehicles [8]. On the other hand, supported by vehicular networks, vehicles can communicate with each other as well as with the infrastructures, i.e., road-side units (RSU). Thus, in this paper, it is reasonable to assume that a vehicle is capable of independently sensing and reporting the driving performance of its neighbor vehicles. In addition, considering the limited local sensing data as well as the complicated external factors, such as the influence by nearby reckless vehicles, a vehicle can hardly provide accurate rating results. Therefore, we propose the CDPR mechanism to aggregate a vehicle's monitoring data from its current and previous neighbor vehicles. In addition, given the fact that most vehicles can be trusted to drive safely, to reduce the data volume of uploading, only the monitored information of vehicles with reckless potentials, i.e., reckless maneuvers, are uploaded. Since a reckless vehicle with constant reckless driving behavior can be continuously reported by multiple neighbor vehicles, our scheme can intuitively realize a timely detection.

Since the CDPR mechanism is based on the cooperation among neighbor vehicles, every vehicle is encouraged to participate into our defensive alerting system. Within this system, every participant is responsible for reporting the driving behavior of neighbor vehicles. As a return, these participants can acquire the valuable alerts of reckless vehicles. Besides, other incentives might be considered to attract a large-scale participation for a safer driving environment in future smart cities [24], [25]. For example, some quota of vehicular communication services can be granted to a vehicle based on its participation. Considering the privacy of participants, instead of the personal information of a driver, only the vehicle description is involved in an alert, such as the position, heading direction, current velocity, color, body style and makes, etc. Note that a vehicles dynamic description, such as the position and heading direction, can be derived by its periodic message exchanges in the vehicular communication system (detailed in Section III.C). A vehicles static description, such as the color and body style, is recorded by the system during its registration, which can be easily extracted based on that vehicles communication ID[1]. To avoid the interference of unrelated alerts, a vehicle can only receive the alert of approaching neighbor reckless vehicles. By combining the received alert with its own location, a vehicle can clearly recognize the approaching reckless vehicle, which is informed to the driver through audio or video advisory, similarly as the audio alarm of a fire. Given the effectiveness of proactive reckless driving alerts as well as the proper incentives, it is reasonable to believe that each vehicle would like to take part in the CDPR mechanism, where the more participants, the more accurate rating results are. To clearly present our design, we assume that every vehicle is enrolled in the following analysis. Besides sending the alert to neighbor vehicles, the transportation department may be informed simultaneously to timely check the detected reckless vehicles. Based on the CDPR mechanism, a harmonic driving environment can be finally realized.

B. System Architecture

To support such a defensive alerting system, we develop a three-tier system architecture from existing vehicular networks. As illustrated in Fig.2, our system architecture consists of vehicle monitoring tier, local aggregation tier and global management tier. In the following, we describe the functions of these three tiers, receptively.

- *Vehicle monitoring tier* is composed of vehicles moving on the road. The functions of this tier include the driving environmental sensing, reckless driving maneuver detection and report uploading, and the reckless driving alert reception. Specifically, every moving vehicle uses its equipped sensing devices to acquire the information of surrounding driving environment. These multi-modal sensing information including the distance gap towards a neighbor vehicle, the relative velocity and acceleration, etc., is used to reveal the driving behavior of neighbor vehicles. By using a light-weight algorithm, a vehicle can quickly filter the normal driving maneuvers and

[1] A vehicle's communication ID is uniquely assigned by the cloud server, like the subscriber identification module (SIM) ID used by smartphones in cellular system, to identify the vehicle entity in vehicular communications.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

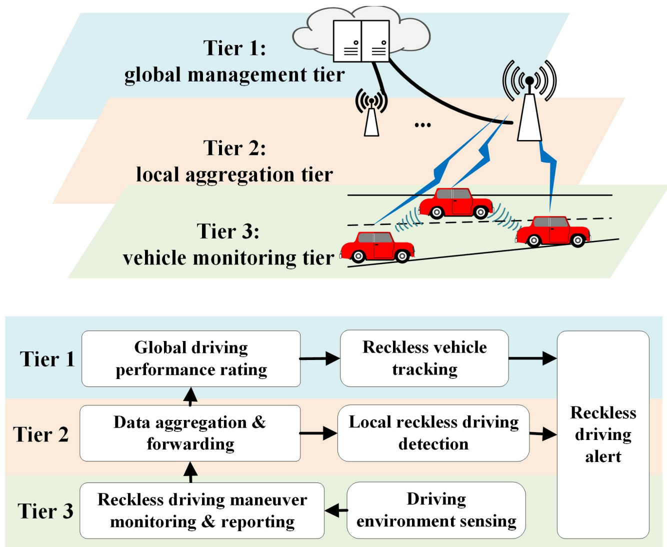


Fig. 2: The three-tier system architecture and the corresponding functions of each tier.

coarsely detect the reckless ones. To derive a more accurate rating result in the upper tier, the data of driving performance during the duration of a reckless maneuver will be collected and uploaded to the second tier, i.e., local aggregation tier. The communication between a vehicle and the second tier is based on dedicated short range communication (DSRC) technology [26], [27]. To guarantee the timeliness of monitoring, the driving performance data with reckless potentials will be uploaded within time T_{th} since its generation. In addition, each vehicle will listen to the broadcasted reckless driving alerts, and provide audio or video advisory to inform the connected drivers. The detail operation procedures of each participating vehicle are presented in Algorithm 1.

- *Local aggregation tier* is composed of road side units (RSU)[2], which are the communication infrastructures in vehicular networks. This tier connects the first and the third tier, respectively, through wireless and wired links. The functions of this tier are to aggregate and forward the uploaded driving behavioral data, to locally detect reckless vehicles, and to broadcast reckless driving alert. For the local reckless vehicle detection, we should provide timely alert in some emergency scenarios, such as the identification of a hogged drunk driving. Nevertheless, in reality, a reckless driver may perform normally in most time, such as when waiting at a traffic light or driving along with very few neighbor vehicles. Thus, the number of detected reckless driving maneuvers in a single road segment will be limited, and instead the driving performance can be accurately rated in the third tier by aggregating more comprehensive data. However, the global alert needs longer detection time, which is not suitable for some emergency scenarios, and thus we introduce the local detection function in the second tier. Only when the percentage of a vehicle's reckless maneuvers exceeds an empirical threshold, the local alert is broadcasted. To achieve an accurate detection, this alert is meanwhile reported to the third tier, and the updated detection result will be sent back to support or correct the current alert in the second tier.

Algorithm 1 Procedures at participating vehicles

```
when a vehicle is started;
build up connection with system through a nearby RSU;
repeat
    report its current location to its service RSU;
    monitor the surrounding driving environment;
    if a reckless driving maneuver is detected then
        report the corresponding data to its service RSU;
    end if
    listen the alerting channel;
    if an alert is received then
        alert the driver through audio advisory;
    end if
until its engine stop;
```

- *Global management tier* is the cloud center with strong computation and storage capabilities. The functions of this tier include globally rating a vehicle's driving performance and sending the detected alert to the corresponding RSUs. Specifically, the cloud center aggregates data of reckless driving maneuvers from all RSUs through the wired connections, and analyzes these data by using machine learning algorithms. The machine learning based driving performance rating models are trained by historical data with reckless or normal labels, which can be periodically updated by the newly arrived data. The label of a reckless vehicle can be derived by the record from the transportation department, where reckless vehicles are the ones intercepted and charged by traffic polices. Once a reckless vehicle is detected, it will be tracked based on the previous connections with the system. Note that a vehicle will periodically communicate with the system, which will be explained in details in the following signaling process, and thus the system can acquire its position and heading direction. The global alert is finally sent to the RSUs based on the tracking results. Meanwhile, the data center may contact the transportation department to further check or intercept the detected reckless vehicles. The detail operation procedures are illustrated in Algorithm 2.

C. Signaling Processes

To realize the communications among the aforementioned three tiers, we describe their signaling processes. In our defensive alerting system, each tier only connects with its neighbor tiers. Specifically, two upper tiers are wire connected, and two lower tiers are wireless connected based on vehicular networks, i.e., DSRC technology [26]. According to DSRC standard, the on-board unit (OBU) equipped in a vehicle can

[2]The local aggregation tier can be composed by any local entities, which exhibits both communication and computation abilities. We utilize RSUs as an example.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

Algorithm 2 Procedures at the cloud server

Initialization: register participating vehicles with their corresponding vehicle and driver information;
repeat
  periodically update the vehicle/driver information;
  for each vehicle currently driving on the road:
    record its driving location;
    record/accumulate its reckless driving maneuvers;
    if meet the detection requirement **then**
      execute the global driving rating algorithm;
      if it is a reckless vehicle **then**
        track its current location;
        announce alert to the corresponding RSU;
        announce alert to the traffic office;
      end if
    end if
until the system is shut down;

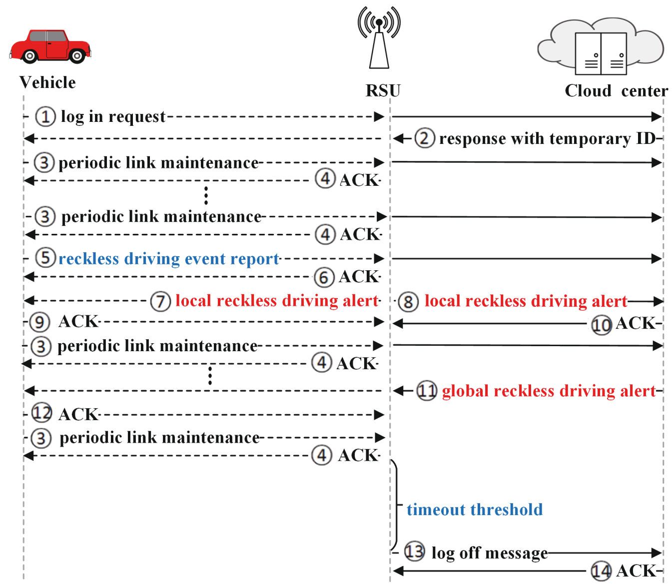


Fig. 3: The signaling process of CDPR service for a vehicle.

communicate with an RSU through 5.9 GHz frequency bands. As illustrated in Fig.3, the signaling processes are described as follows.

When a vehicle starts engine, it automatically sends a request to log in the defensive alerting system in signaling ①. After the authentication due to privacy and security protection, data center sends back a response with a temporary identity in signaling ②. Meanwhile, the data center records the participation of this vehicle to accumulate its incentive, such as service quota. Notice that the temporary identity corresponds to a vehicle's real identity, including the plate number and the title information, which is recorded in the cloud center. A vehicle periodically communicates with its serving RSU through signaling ③ to maintain the connection, and an acknowledgement (ACK) will be sent back in signaling ④ for each signaling ③. The signaling ③ carries the temporary identity of a vehicle, its current position, velocity and heading direction, which can also be utilized in other vehicular services, such as the driver assistant services like cruise control and navigation [8]. Meanwhile, signaling ③ is broadcasted to its neighbor vehicles to distinguish the identity of different neighbors and thus mark their corresponding driving performance. Once a reckless driving maneuver is detected, the corresponding information is immediately reported in signaling ⑤, including the identity of the monitored vehicle, the maneuver duration, and during this duration the monitored distance gaps and relative velocities, etc. These data is recorded in both the RSU and data center, and an ACK is sent back in signaling ⑥. Once a reckless vehicle is locally detected by an RSU, the local alert is broadcasted in signaling ⑦ and meanwhile reported to data center in signaling ⑧. The ACKs are sent back in signaling ⑨ and ⑩, respectively. Once a reckless vehicle is detected in the cloud center, a global alert is sent to its serving RSU in signaling ⑪ to alert the vehicles in that road segment. An ACK will be sent back from the listened vehicles in signaling ⑫. Both the local and global alert do not include personal information but the temporary identity and the description of the reckless driving vehicle. When the time without the updated signaling ③ exceeds the timeout threshold, a logoff message will be sent to cloud center in signaling ⑬, and an ACK is sent back in ⑭. When none of the RSUs updates signaling ③ of a vehicle, the vehicle is considered to be logged off the system.

IV. Machine Learning based Driving Performance Rating Model

In order to provide accurate reckless driving alert, this section focuses on analyzing the driving behavior rating model based on machine learning algorithms. In the following, we first overview the driving performance dataset, and then present the driving performance rating model.

A. Driving Performance Dataset

Due to the difficulty of collecting adequate dataset of driving performance, we implement a popular traffic simulation toolkit, Simulation of Urban MObility (SUMO) for data generation [28]. The simulator based dataset can be our first step to test the availability of our defensive alerting system, which can be extended to real-world data in our future analysis. As an open-source microscopic traffic simulator, SUMO enables repeatable controlled experiments at a per-vehicle level based on real-world traffic topologies. To reproduce mobility patterns close to reality, we utilize Luxembourg SUMO Traffic (LuST) scenario based on the City of Luxembourg, whose traffic patterns are verified to be consistent with that of Google Map over 24 hours of simulation [29]. In LuST scenario, we consider different types of vehicles, which are distinguished by different parameter settings, including color, size, length, shape, maximum speed, acceleration, deceleration, minimum distance gap, etc. In addition to the 6 normal vehicle types provided in LuST scenario, we insert reckless vehicle types in our simulation. Based on the characteristics of reckless driving mentioned in section II that are continuously aggressively

TABLE I: Parameter setting for different vehicle types.

vType	accel	deccel	minGap	maxSpeed	shape
Nv1	2.6	4.5	1.5	70	sedan
Nv2	3.0	4.5	1.5	50	hatchback
Nv3	2.8	4.5	1.0	50	hatchback
Nv4	2.7	4.5	1.5	70	wagon
Nv5	2.4	4.5	1.5	30	van
Nv6	2.3	4.5	2.0	30	delivery
Rv1	8.0	8.0	0.5	140	sedan
Rv2	9.0	9.0	0.5	110	hatchback
Rv3	9.0	9.0	0.3	100	hatchback
Rv4	8.0	8.0	0.2	105	wagon
Rv5	9.5	9.5	0.7	130	van
Rv6	10.0	10.0	0.8	120	delivery

disregarding safety of other individuals, we modify the vehicle parameter settings to generate 6 reckless vehicle types as illustrated in Table I, where 'Nv' and 'Rv' represent normal and reckless vehicle type, respectively. Similar parameter setting has been used in [30] to define aggressive driving.

According to the CDPR mechanism, the data of driving performance is collected by aggregating that of reckless driving maneuvers, which are detected by the detection function of vehicles mentioned in Section II.B. To realize this function, we utilize the surrogate safety measure (SSM) devices provided in SUMO. An SSM device exhibits the ability of sensing and computing, which can be equipped on a vehicle. The function of an SSM device is to detect conflict encounters, i.e., reckless driving maneuvers, and meanwhile log their corresponding SSM values. Several SSMs are considered including time to collision (TTC), deceleration rate to avoid a crash (DRAC), post encroachment time (PET), brake rate (BR), etc., whose exact definition can be found in [31]. Once the value of an SSM exceeds the corresponding threshold, the current encounter is regarded as a conflict encounter, where different conflict encounters may last for different time duration. Conflict encounters include crossing, merging, tailgating situations, etc., where different calculation procedures of SSMs may be implemented. Note that different conflict encounters correspond to different reckless driving events.

In the output file of an SSM device, the driving data of each conflict encounter is recorded, including its duration, encounter situation, velocities of the ego and the conflict neighbor during this duration, positions of the ego and the conflict neighbor during this duration, minimum TTC, maximum DARC and PET. Due to the complexity of analyzing the impact of different conflict encounters, we exclude the features affected by different conflict encounter types and derive the following features to characterize the driving performance:

- the time duration of a conflict encounter,
- the distance gap over this duration based on the position of the ego and conflict neighbor vehicle,
- the relative velocity over this duration based on the velocity of the ego and conflict neighbor vehicle.

In addition, to utilize these features in our rating model, we preprocess the data with normalization, which is necessary for distance-metric based machine learning algorithms, such as support vector machine (SVM).

B. Training the Driving Performance Rating Models

Based on the aforementioned dataset, we train the driving performance rating model by using machine learning algorithms [32]–[35]. We utilize two machine learning algorithms [32]–[34], SVM and decision-tree, which are typically used in classification problems that make them appropriate for classifying normal vehicles versus reckless vehicles. For SVM model, among several useful theoretical and practical characteristics, the following two reasons motivate our selection: 1) Since the training of SVM involves a convex optimization problem, the optimal solution is a global optimum; 2) The upper bound of the generalization error is independent of problem dimensionality. For decision-tree model [32], [33], [35], the nonparametric nature and easy interpretation make it popular in a variety of application fields. Its advantage over many other models is the effectiveness to construct classifications through segmenting a data set into smaller and more homogeneous groups.

To rate a vehicle's driving performance, the input data of SVM and decision-tree models is in the form of $\langle$21-dimensional features, label$\rangle$. Specifically, the 21-dimensional features are obtained from the normalized driving behavior features, including the time duration (1 dimension), the percentage of the distance gap (dividing the overall range into 10 dimensions) and the percentage of the relative velocity (dividing the overall range into 10 dimensions). In the following, the SVM and decision-tree models are presented, respectively.

1) SVM model [32]–[34]: In general, the SVM model maps the points of two categories in the $n-$dimensional space into a target $m-$dimensional space, where a hyperplane in the new space can better separate two categories of points. Specifically, define our training set by $\{\mathbf{x}_i, y_i\}$ with N training samples $i = 1, ..., N$, where $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^{21})$ is the aforementioned 21-dimensional features, and $y_i \in \{-1, +1\}$ is the label of normal and reckless vehicle. In SVM models, the mapping is based on a kernel function, which has multiple types as the inner product between the mapped pairs of points in the feature space. Since kernel selection is essential to obtain satisfactory classification results, we consider the following well-known kernel functions in our design:

- Linear kernel:
$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j, \tag{5}$$
where $\mathbf{x}_i^T$ is the transposition of matrix $\mathbf{x}_i$.
- Polynomial kernel:
$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + 1)^d, \gamma > 0, \tag{6}$$
where γ is an empirical parameter, and d is the degree of the polynomial.
- Gaussian radial basis function (RBF) kernel:
$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right), \tag{7}$$
where γ is an empirical parameter, and $\|\mathbf{x}_i - \mathbf{x}_j\|$ is the norm of the point differences, representing the distance between vector $\mathbf{x}_i$ and $\mathbf{x}_j$ in the corresponding Hilbert space.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

In the target space mapped by a kernel function, SVM classifies data by finding the best hyperplane that separates all data points of the two classes. The best hyperplane is the one with the largest margin between the two classes, which is trained with solving the following optimization problem:

$$\begin{aligned} &\max_{\boldsymbol{\theta}} \ C(\boldsymbol{\theta}) = \sum_{i=1}^{N} \theta_i - \frac{1}{2} \sum_{i=1,j=1}^{N,N} \theta_i \theta_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &\text{s.t.} \ \sum_{i=1}^{N} \theta_i y_i = 0, \quad \theta_i \geq 0. \end{aligned} \tag{8}$$

By utilizing the optimal $\boldsymbol{\theta}$, when a test sample with feature vector $\mathbf{z}$ arrives for rating, this sample is rated by the following decision function [33]:

$$D(\mathbf{z}) = \operatorname{sgn}\left[\sum_{i=1}^{N} \theta_i y_i K(\mathbf{z}, \mathbf{x}_i) + B\right], \tag{9}$$

where B is the bias term.

2) Decision-tree model [32], [33], [35]: A decision tree is a flow-chart-like tree structure, which derives a classification decision through a sequence of tests along a path of nodes. The root node contains the entire dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_,.., \mathbf{x}_N\}$. The tree grows through splitting data at the nodes, where the outgoing branches of a node correspond to the partitioned data subsets. When the data of a node cannot be splitted any further, it becomes a terminal node indicating a classification decision, i.e., the normal or reckless rating decision of a vehicle. In our decision-tree model based classifier, we consider three popular selection algorithms to find the best split predictor: standard classification and regression tree (CART), curvature test, and interaction test [35]–[37]. The standard CART selects the split predictor that maximizes the split-criterion gain over all possible splits of all predictors; the curvature test selects the split predictor that minimizes the p-value of Chi-square tests of independence between predictors and the response; and similarly interaction test based on the calculation criterion of curvature test considers the interaction between each pair of predictor and response. We evaluate the performance of these three algorithms in our evaluation part. To present a concrete introduction, we utilize the standard CART algorithm as the selection algorithm to describe the decision-tree model training in the subsequent development.

The principle of *tree growing* step is to recursively partition target variables so that the data in the descendant nodes are always purer than that in the parent node. When a training data enters the root node, a test is performed to search for all possible splits based on a splitting criterion that measures the splitting quality. In CART, the Gini-index splitting criterion is applied. According to the definition of Gini-index, the impurity of node n is given by

$$I(n) = 1 - \sum_{j} p_j^2(n), \tag{10}$$

where $p_j(n)$ is the probability that a sample in subset $\mathbf{X}_n$ belongs to class j. $p_j(n)$ can be easily estimated by N_n^j/N_n, where N_n^j is the number of vectors in subset $\mathbf{X}_n$ belongs to class j and N_n is the total number of vectors in $\mathbf{X}_n$. The decrease of node impurity $\Delta I(n)$ is given by

$$\Delta I(n) = I(n) - \sum_{j=1}^{2} p_j(n_j) I(n_j), \tag{11}$$

where n_j is the jth descendant node of node n. By exhaustively searching for all possible splits, the one with the maximum impurity decrease is selected. When the probability $p_j(l)$ is more than the threshold P_0, node l is regarded as a terminal node with the classification decision $j^* = \arg\max_{j\in\{1,2\}} p_j(l)$.

In the *tree pruning* step, the established decision tree is pruned to avoid overfitting the training data due to its large size. We apply the most common pruning rule, minimal cost-complexity pruning, due to its computation efficiency. The cost-complexity of decision tree T measures its misclassification rate and complexity (the number of terminal nodes), which is given by

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|, \tag{12}$$

where $R(T)$ is the corresponding misclassification rate, which is defined as the percentage of misclassified vehicles by this decision tree; α is the complexity factor; $\tilde{T}$ is the set of all terminal nodes in T and $|\tilde{T}|$ is the size of $\tilde{T}$. The pruning starts from the terminal nodes to examine each node and subtree. When the replacement of a subtree by a terminal node can give a lower cost-complexity, the subtree is pruned.

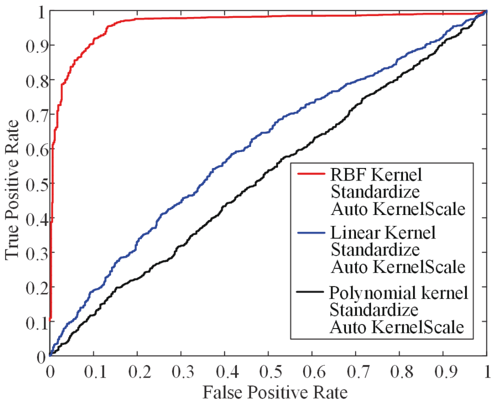
V. Performance Evaluations

In this section, we evaluate the effectiveness of our defensive alerting system, where the dataset preparation is firstly described, followed by the performance evaluation involving two critical aspects. Specifically, we first evaluate the accuracy of our driving performance rating, as well as the performance comparison between our machine learning based method and the conventional statistical based method. Then, we comprehensively investigate the timeliness of our reckless vehicle alerting, involving the delay of the whole alerting process.

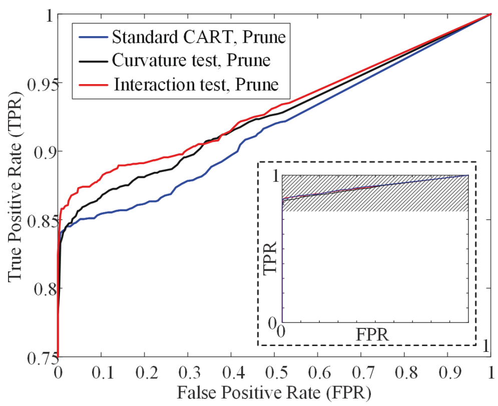
A. Dataset Preparation

Based on our defensive alerting system design in Section III and IV, we implement the CDPR mechanism in SUMO platform to collect the dataset of driving performance. In LuST scenario, we consider the 12 vehicle types as illustrated in Table I, and the percentage of reckless vehicles is set to be 10%. Our simulation involves 15,000 vehicles equipped with SSM devices, whose total arrival time is 50 minutes between 8:00 AM and 8:50 AM. The number of our samples, i.e., the number of vehicles with reported conflict encounters, is 6977.

Based on the output file of SSM devices, we collect the 21-dimensional driving features mentioned in Section IV.B, including the time duration of conflict encounters (1 dimension), the percentage of the distance gap (dividing the overall range into 10 dimensions) and the percentage of the relative velocity (dividing the overall range into 10 dimensions). To avoid the effect due to irrelevant features, we further test



(a) SVM models under different kernel functions.



(b) Decision-tree models under different predictor models.

Fig. 4: The receiver operating characteristic (ROC) curves of different models for driving behavior rating.

feature combinations and finally delete some features that are constant values, i.e., close to 0 for most training samples, such as the percentage of the last segment in relative velocity range (the relative velocity segment $[28.98, 32.60]$ to the overall relative velocity range $[0, 32.60]$). Therefore, the size of our final dataset is 6977×10. In our driving behavior rating model, we utilize 70% samples as the training data and the remaining 30% samples as the testing data.

B. Overall Performance

Based on the prepared dataset, we first evaluate the performance of the machine learning based driving performance rating models. According to Section IV.B, we explore the performance of the SVM models under different kernel functions, i.e., RBF, linear and polynomial kernel, as well as that of the decision-tree models under different predictor selection algorithms, i.e., standard CART, curvature test and interaction test. As illustrated in Fig. 4, we illustrate the receiver operating characteristic (ROC) curves of the above models, respectively. A ROC curve presents the true positive rate (TPR) against the false positive rate (FPR) under various threshold settings, whose analysis is a direct and natural way to select possibly optimal models and to discard suboptimal ones. A good prediction model yields the points in the upper left corner of the ROC space, i.e., coordinate (0,1), which represents no false negatives (FN) and no false positives (FP). In contrast, the points of a random guess are along the diagonal line from the left bottom to the top right corners. Thus, according to our experiment results illustrated in Fig. 4, the ROC curve of SVM model with RBF kernel apparently outperforms that of the other two, while the ROC curves of the three decision-tree models have very close performance. To clarify the difference of the ROC curves in decision-tree models, the ROC curves with y-axis ranging from 0.75 to 1 is presented in the main figure, where the whole curves with y-axis ranging from 0 to 1 is illustrated on the bottom right. Based on our simulation results, we apply the SVM model with RBF kernel and the decision-tree model with standard CART algorithm as our driving performance rating models in the following.

TABLE II: Accuracy Evaluation.

rating model	accuracy(%)	precision(%)	recall(%)	F1(%)
SVM	92.73	89.53	97.05	93.14
decision-tree	93.29	89.61	99.98	93.61
statistical method	47.95	49.32	1.5	16.53

Before we further analyze the accuracy and the impact of parameters for the two selected machine learning based rating models, we take the conventional statistical rating method into account. In the following, we evaluate the accuracy of the three rating methods, i.e., SVM, decision-tree and statistical driving performance rating methods. In the CDPR mechanism, we utilize multiple sensing devices to fully characterize different reckless driving maneuvers, where the machine learning based driving performance rating models are introduced due to the difficulty of analyzing the multi-modal sensing data. Since the consequence of a crash comes from the geographically proximity between vehicles, a reckless vehicle with high crash probability intuitively drives closer to neighbor vehicles. Therefore, instead of the multi-modal sensing features, we should evaluate the availability of a single feature, i.e., the distance gap, to rate a vehicle's "reckless" level. By using the single distance gap feature, the conventional statistical method can measure a vehicle's driving performance through comparing the vehicle's statistical distance gap, i.e., mean distance gap, with the empirical threshold derived from previously detected reckless vehicles with low computation complexity. Based on the definition of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), we examine the following metrics in the comparison:

- Accuracy: the probability that the classification of a vehicle's driving performance is the same as the ground truth.
- Precision: the probability that the classification for reckless driving is exactly a reckless vehicle in ground truth.
- Recall: the probability that all reckless vehicles in ground truth are classified correctly as reckless driving.
- F1 score: a measure of rating accuracy that is the harmonic average of the precision and recall, i.e., $F_1 = 2/(\text{recall}^{-1} + \text{precision}^{-1})$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

To conduct a fair comparison, the dataset of the statistical method is the same as our machine learning based methods, and the testing results are illustrated in TABLE II. We can observe that the machine learning based methods apparently outperforms the statistical one, which verifies that a vehicle's driving behavior cannot be simply measured by only using the distance gap feature. Thus, due to the highly dynamic nature of driving environments, the machine learning based driving performance rating models by using multi-modal sensing data are imperatively necessary. By comparing the four metrics, we can observe that the precision and recall metrics are with the lowest and the highest values, respectively. According to their corresponding definitions, their evaluation results verify that although a few false alarms exist in the system, nearly all reckless vehicles can be identified, which can be accepted for the safety consideration in our defensive alerting system. In addition, to fully utilize our dataset for a more accurate comparison, we conduct a 10-fold cross-validation for our two machine learning based rating models, where the loss of the SVM and decision-tree models are 0.0708 and 0.0606, respectively. Therefore, these two machine learning based rating models can provide accurate rating to detect reckless vehicles, and the decision-tree model performs better in most conditions.

C. Timeliness of Reckless Vehicle Alert

Since the delay requirement of defensive alerting system is stringent, our design aims at providing accurate detection and reducing alerting delay at the same time. However, as an alerting system, there are inevitable delay for alerting procedures, such as the hazard warning and driver's reaction. Thus, our design makes efforts on reducing the delay from the vehicle sensing, data transmission, and data processing in the cloud. Specifically, each vehicle on road keeps monitoring the driving performance of neighbor vehicles, and immediately reports the detected reckless driving maneuvers to reduce the sensing delay. The sensing delay can be further reduced with more advanced sensing technologies embedded on vehicles, which is out of the scope of our analysis. For the data transmission delay, instead of reporting all sensing data, vehicles in our design only report the data with reckless potential to fasten data transmissions, and meanwhile to avoid possible transmission congestion. Besides, we utilize the RSUs to forward the sensing data, which provides fast uploading by shortening the transmission distances between vehicles and RSUs. The collected sensing data will be forwarded from RSUs to the cloud server through fast and reliable wired connections. For the data processing delay in the cloud, our design only collects the sensing data with reckless potentials to reduce the processing data amount. Meanwhile, instead of massive image or video processing, the cloud server uses simple machine learning algorithms, such as SVM and decision tree, to quickly process the low-dimensional sensing data. By conducting an Intel i7-2600 CPU with 16GB memory, our rating models, i.e., using SVM and decision-tree algorithms, with the same setting of TABLE II can rate the driving performance of more than 900 vehicles per second. With much more powerful computation capability, the data processing delay in the could server will be further reduced.

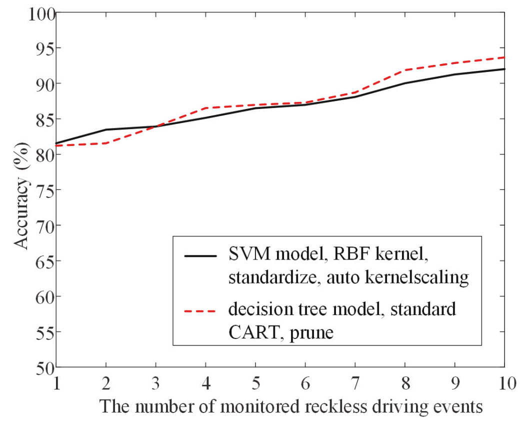


Fig. 5: The detection accuracy as a function of the number of monitored reckless driving events.

In fact, the most significant delay in our design comes from the sensing data accumulation, i.e., accumulating the reckless driving maneuvers. Since the driving environment is highly dynamic and complicated, a vehicles driving performance can be hardly judged by a short time observation. For example, a vehicle with a single or a few reckless driving maneuvers may be a normal vehicle, whose abnormal reckless maneuvers are affected by other environmental factors, such as its reckless neighbor vehicles. Therefore, a certain observation time is required to accumulate enough sensing data, and there is a trade-off between the accuracy and the timeliness. As illustrated in Fig.5, the accuracy is formulated as a function of the number of monitored reckless driving maneuvers respectively with the SVM and decision-tree rating models. Since the dataset of driving performance comes from the accumulation of reckless driving maneuvers reports, the more accumulated maneuvers means the longer monitoring time. Thus, we utilize the number of reported reckless driving maneuvers to measure the length of the monitoring time. We can observe that both the two machine learning based rating models perform better under a larger number of maneuvers, which means the longer monitoring time the more accurate rating results are. In addition, within a short monitoring time, i.e., only a few reported reckless driving maneuvers, the accuracy of our rating is around 85%, which will be improved to around 90% when the number of the monitored reckless driving events is more than 5. Thus, a quick rating will sacrifice some accuracy, but still with a high accuracy level. For some areas with historically high crash probability, such as the bar area, a reckless driving alert may be sent in advance within a few reported data to improve the local driving safety. In addition, due to the increase of accuracy with more reported reckless driving maneuvers, the rating results should be timely updated to terminate the false alarm and identify the real reckless vehicles.

VI. Conclusion

Observing the severity of reckless driving on the road nowadays, we proposed a defensive alerting system to proactively detect and notify the threats from approaching reckless vehicles. To better understand the consequence of reckless driving, we first theoretically evaluated the crash probability of a typical reckless driving maneuver, i.e., reckless lane changing. Based on the verified the need for defensive reckless driving alerts, we further developed a machine learning based cooperative driving performance rating (CDPR) mechanism by integrating the computation capabilities of neighbor vehicles and a cloud server. By utilizing the monitored information from neighbor vehicles, the CDPR mechanism automatically and intelligently detects reckless vehicles. To support such a defensive alerting system, a three-tier system architecture was developed from existing vehicular networks. Moreover, we devised a transmission load reduction scheme by only uploading the data of driving maneuvers with reckless potential. Based on the aggregated monitoring data, the cloud server globally rates every vehicle's driving performance by using SVM and decision-tree machine learning algorithms. We finally implemented the proposed CDPR mechanism into SUMO simulator. Extensive simulation results illustrated that our defensive alerting system can accurately detect reckless vehicles and provide timely alerts.

References

[1] "Driving just got a lot deadlier," https://www.cbsnews.com/news/traffic-fatalities-jump-by-5-6-percent-in-2016/.

[2] C. Saiprasert and W. Pattara-Atikom, "Smartphone enabled dangerous driving report system," in *2013 46th Hawaii International Conference on System Sciences*. IEEE, 2013, pp. 1231–1237.

[3] J. Yu, Z. Chen, Y. Zhu, Y. J. Chen, L. Kong, and M. Li, "Fine-grained abnormal driving behaviors detection and identification with smartphones," *IEEE transactions on mobile computing*, vol. 16, no. 8, pp. 2198–2212, 2016.

[4] U. S. Department of Motor Vehicles (DMV), "Reckless driving 101," https://www.dmv.org/safety-laws/reckless-driving.php.

[5] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 666–676, 2017.

[6] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE transactions on intelligent transportation systems*, vol. 12, no. 2, pp. 596–614, 2010.

[7] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[8] K. Bengler, K. Dietmayer, B. Färber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems."

[9] S. Al-Sultan, A. H. Al-Bayatti, and H. Zedan, "Context-aware driver behavior detection system in intelligent transportation systems," *IEEE transactions on vehicular technology*, vol. 62, no. 9, pp. 4264–4275, 2013.

[10] M. Green, "Roadway human factors: from science to application," 2017.

[11] D. C. Luvizon, B. T. Nassu, and R. Minetto, "A video-based system for vehicle speed measurement in urban roadways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1393–1404, 2016.

[12] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8971–8980.

[13] V. A. Memos, K. E. Psannis, Y. Ishibashi, B.-G. Kim, and B. B. Gupta, "An efficient algorithm for media-based surveillance system (eamsus) in iot smart city framework," *Future Generation Computer Systems*, vol. 83, pp. 619–628, 2018.

[14] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking: A review," *Neurocomputing*, vol. 74, no. 18, pp. 3823–3831, 2011.

[15] T. Taleb, E. Sakhaee, A. Jamalipour, K. Hashimoto, N. Kato, and Y. Nemoto, "A stable routing protocol to support its services in vanet networks," *IEEE Transactions on Vehicular technology*, vol. 56, no. 6, pp. 3337–3347, 2007.

[16] Q. Luo, Y. Cao, J. Liu, and A. Bensilmane, "Localization and navigation in autonomous driving: Threats and countermeasures," *IEEE Communications Magazine*, 2019.

[17] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrows intelligent network traffic control systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432–2455, 2017.

[18] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE wireless communications*, vol. 24, no. 3, pp. 146–153, 2016.

[19] J. Wang, L. Zhao, J. Liu, and N. Kato, "Smart resource allocation for mobile edge computing: A deep reinforcement learning approach," *IEEE Transactions on Emerging Topics in Computing*, 2019.

[20] J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1243–1274, 2018.

[21] L. Zhao, J. Wang, J. Liu, and N. Kato, "Routing for crowd management in smart cities: A deep reinforcement learning perspective," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 88–93, 2019.

[22] B. Sen, J. D. Smith, W. G. Najm *et al.*, "Analysis of lane change crashes," United States. National Highway Traffic Safety Administration, Tech. Rep., 2003.

[23] X.-G. Li, B. Jia, Z.-Y. Gao, and R. Jiang, "A realistic two-lane cellular automata traffic model considering aggressive lane-changing behavior of fast vehicle," *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 479–486, 2006.

[24] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, "Incentives for mobile crowd sensing: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 54–67, 2015.

[25] J. Gao, L. Zhao, and X. Shen, "Network utility maximization based on an incentive mechanism for truthful reporting of local information," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7523–7537, 2018.

[26] J. B. Kenney, "Dedicated short-range communications (dsrc) standards in the united states," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, 2011.

[27] K. A. Hafeez, L. Zhao, B. Ma, and J. W. Mark, "Performance analysis and enhancement of the dsrc for vanet's safety applications," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3069–3083, 2013.

[28] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of sumo-simulation of urban mobility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, 2012.

[29] L. Codeca, R. Frank, and T. Engel, "Luxembourg sumo traffic (lust) scenario: 24 hours of mobility for vehicular networking research," in *2015 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2015, pp. 1–8.

[30] M. Matousek, M. Yassin, R. van der Heijden, F. Kargl *et al.*, "Robust detection of anomalous driving behavior," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–5.

[31] S. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS research*, vol. 41, no. 4, pp. 153–163, 2017.

[32] P. Harrington, *Machine learning in action*. Manning Publications Co., 2012.

[33] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[34] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[35] N. Elmitiny, X. Yan, E. Radwan, C. Russo, and D. Nashar, "Classification analysis of driver's stop/go decision and red-light running violation," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 101–111, 2010.

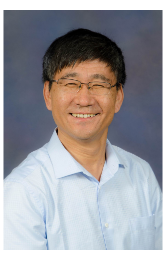
[36] W.-Y. Loh, "Regression tress with unbiased variable selection and interaction detection," *Statistica Sinica*, pp. 361–386, 2002.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2945398, IEEE Transactions on Vehicular Technology

[37] W.-Y. Loh and Y.-S. Shih, "Split selection methods for classification trees," *Statistica sinica*, pp. 815–840, 1997.

Lan Zhang received the B.S. and M.S. degrees from University of Electronic Science and Technology of China, Chengdu, China, in 2013 and 2016, respectively. She is currently pursuing the PhD degree in the Department of Electrical and Computer Engineering at University of Florida, Gainesville, FL, USA. Her research interests include wireless networking, vehicular communication, and security and privacy for various cyber-physical systems.

Li Yan received the B.E. degree in communication engineering and the Ph.D. degree in communication and information systems from Southwest Jiaotong University, China, in 2012 and 2018, respectively. She was a Visiting Student at the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, from 2017 to 2018. She is currently a Lecturer with Southwest Jiaotong University. Her research interests include 5G communications, mobility managements, network architecture, and HSR wireless communications.



Yuguang "Michael" Fang (F'08) received an MS degree from Qufu Normal University, Shandong, China in 1987, a PhD degree from Case Western Reserve University in 1994, and a PhD degree from Boston University in 1997. He joined the Department of Electrical and Computer Engineering at University of Florida as an assistant professor in 2000, was promoted to an associate professor in 2003 and a full professor in 2005, and has been a distinguished professor, since 2019. He holds a University of Florida Research Foundation (UFRF) Professorship (2017-2020, 2006-2009), University of Florida Term Professorship (2017-2019), a Changjiang Scholar Chair Professorship (Xidian University, Xian, China, 2008-2011; Dalian Maritime University, Dalian, China, 2015-2018), Overseas Adviser, School of Information Science and Technology, Southwest Jiao Tong University, Chengdu, China (2014-present), and Overseas Academic Master (Dalian University of Technology, Dalian, China, 2016-2018).

Dr. Fang received the US National Science Foundation Career Award in 2001, the Office of Naval Research Young Investigator Award in 2002, the 2015 IEEE Communications Society CISTC Technical Recognition Award, the 2014 IEEE Communications Society WTC Recognition Award, and the Best Paper Award from IEEE ICNP (2006). He has also received a 2010-2011 UF Doctoral Dissertation Advisor/Mentoring Award, a 2011 Florida Blue Key/UF Homecoming Distinguished Faculty Award, and the 2009 UF College of Engineering Faculty Mentoring Award. He was the Editor-in-Chief of IEEE Transactions on Vehicular Technology (2013-2017), the Editor-in-Chief of IEEE Wireless Communications (2009-2012), and serves/served on several editorial boards of journals including Proceedings of the IEEE (2018-present), ACM Computing Surveys (2017-present), IEEE Transactions on Mobile Computing (2003-2008, 2011-2016), IEEE Transactions on Communications (2000-2011), and IEEE Transactions on Wireless Communications (2002-2009). He has been actively participating in conference organizations such as serving as the Technical Program Co-Chair for IEEE INFOCOM2014 and the Technical Program Vice-Chair for IEEE INFOCOM'2005. He is a fellow of the IEEE and a fellow of the American Association for the Advancement of Science (AAAS).

Xuming Fang (SM'16) received the B.E. degree in electrical engineering in 1984, the M.E. degree in computer engineering in 1989, and the Ph.D. degree in communication engineering in 1999 from Southwest Jiaotong University, Chengdu, China. He was a Faculty Member with the Department of Electrical Engineering, Tongji University, Shanghai, China, in September 1984. He then joined the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, where he has been a Professor since 2001, and the Chair of the Department of Communication Engineering since 2006. He held visiting positions with the Institute of Railway Technology, Technical University at Berlin, Berlin, Germany, in 1998 and 1999, and with the Center for Advanced Telecommunication Systems and Services, University of Texas at Dallas, Richardson, in 2000 and 2001. He has, to his credit, around 200 high-quality research papers in journals and conference publications. He has authored or coauthored five books or textbooks. His research interests include wireless broadband access control, radio resource management, multi-hop relay networks, and broadband wireless access for high speed railway. Dr. Fang is the Chair of IEEE Vehicular Technology Society of Chengdu Chapter.

Xiaoxia Huang (M'07) received the B.E. and M.E degrees in electrical engineering both from Huazhong University of Science and Technology, Wuhan, China, in 2000 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2007. She is currently a Professor with School of Electronics and Communication Engineering at Sun Yat-Sen University. Her research interests include intelligent wireless networks, ambient backscatter networks, and energy harvesting.