CAD3: Edge-facilitated Real-time Collaborative Abnormal Driving Distributed Detection

Ahmad Alhilal*, Tristan Braud*, Xiang Su†‡, Luay Al Asadi§, and Pan Hui*†,
*The Hong Kong University of Science and Technology, Hong Kong
†University of Helsinki, Helsinki, Finland
‡University of Oulu, Oulu, Finland
§AlBaraka Bank Syria, Damascus, Syria
aalhilal@ust.hk; braudt@ust.hk; xiang.su@helsinki.fi; alassadi.luay@gmail.com; panhui@cse.ust.hk

***Abstract*—Speeding, slowing down, and sudden acceleration are the leading causes of fatal accidents on highways. Anomalous driving behavior detection can improve road safety by informing drivers who are in the vicinity of dangerous vehicles. However, detecting abnormal driving behavior at the city-scale in a centralized fashion results in considerable network and computation load, that would significantly restrict the scalability of the system. In this paper, we propose CAD3, a distributed collaborative system for road-aware and driver-aware anomaly driving detection. CAD3 considers a decentralized deployment of edge computation nodes on the roadside and combines collaborative and context-aware computation with low-latency communication to detect and inform nearby drivers of unsafe behaviors of other vehicles in real-time. Adjacent edge nodes collaborate to improve the detection of abnormal driving behavior at the city-scale. We evaluate CAD3 with a physical testbed implementation. We emulate realistic driving scenarios from a real driving data set of 3,000 vehicles, 214,000 trips, and 18 million trajectories of private cars in Shenzhen, China. At the microscopic (road) level, CAD3 significantly improves the accuracy of detection and lowers the number of potential accidents caused by false negatives up to four times and 24 times as compared to distributed standalone and centralized models, respectively. CAD3 can scale up to 256 vehicles connected to a single node while keeping the end-to-end latency under 50 ms and a required bandwidth below 5 mbps. At the mesoscopic (driver-trip) level, CAD3 performs stable and accurate detection over time, owing to local RSU interaction. With a dense deployment of edge nodes, CAD3 can scale up to the size of Shenzhen, a megalopolis of 12 million inhabitant with over 2 million concurrent vehicles at peak hours.**

I. Introduction

Anomalous driving patterns significantly increase the risk of road accidents. For instance, a high variance in driving speed results in low predictability, more vehicle encounters, more overtaking manoeuvres, and more accidents [1]. Addressing abnormal driving is thus critical for improving road safety. However, classifying driving behaviors is a challenging task. Driving patterns vary depending on contextual information, which include the time of day, the road type, and the individual behavior of the driver. The increasing ubiquity of connected vehicles and their multitude of embedded sensors can be leveraged to measure such contextual data [2]. Connected vehicles can actively participate in anomalous driving behavior detection, leveraging computation and communication power of roadside units (RSUs). With collected vehicular data, an RSU can train models locally and classify driving behaviors using embedded computation units.

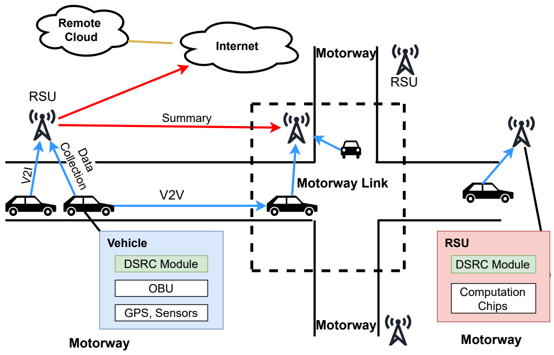


Figure 1: Spontaneous vehicular networks formed between vehicles and RSUs using DSRC. One RSU at the Motorway link and an RSU on each motorway.

Identifying abnormal driving behaviors is a challenging issue at the individual level. Scaling up to the city-scale represents a considerable effort and raises multiple questions. The amount and diversity of contextual information with the number of vehicles results in a significant network and computational load. Cloud servers provide elastic computing capabilities that can address the high computation load. However, cloud computing facilities receive updates at city or region-scale, resulting in a massive network load at a single point of the network. Comparatively, edge computing distributes the computing and network load, and provides de-facto geographic context-awareness. This comes at the cost of losing information on the vehicles's behavior outside of the coverage area of servers.

In this paper, we present CAD3, a collaborative distributed system to detect anomalous driving behavior. CAD3 leverages the distributed deployment of edge nodes co-located with RSUs to maintain contextual information about the roads, ensure low processing and network latency, and the ability to scale up to large metropolitan areas. Vehicles act as sensing nodes that deliver the collected data to the nearest RSU. CAD3 relies on standard and widely deployed technologies, such as Apache Kafka and Spark, to stream and analyze driving pattern data in real-time. CAD3 reinforces the detection process through inter-edge collaboration. At the microscopic level (road trunk), adjacent RSUs intercommunicate to share prediction messages. At the mesoscopic level (the distance/duration of a trip), upon vehicle handover, the former RSU passes a prediction summary to the next, the process which is carried on, allows the system to gain driver-awareness. We implement CAD3 on a physical testbed and apply a real driving dataset to train the models and evaluate realistic scenarios. The dataset contains 3,000 vehicles, 214,000 trips, and 18 million trajectories collected from private cars in Shenzhen, China. CAD3 improves the detection of anomalous driving behavior as compared to standalone edge or centralized cloud systems. Thanks to its collaborative capabilities combined with geographic context awareness, CAD3 significantly decreases the amount of false negatives, reducing the number of potential accidents caused by false negatives by an order of magnitude compared to centralized solutions. Finally, by relying on a decentralized deployment of edge nodes at the roadside, CAD3 can address the most demanding peak hours at the scale of a 12 million-inhabitant metropolis featuring more than 2 million concurrent vehicles, such as Shenzhen, with an end-to-end latency (the time between the transmission of a packet by a vehicle and the subsequent dissemination of a warning message) below 50 ms. Our contributions is threefold:

- We propose a *full-stack system for decentralized and context-aware data analytics*, combining a pervasive deployment of nodes at the roadside, a pipeline for real-time data streaming, an edge-assisted distributed computing platform for abnormal driving behavior detection, and a mechanism for inter-edge collaboration.
- We implement a *proof-of-concept prototype on a physical testbed*, to which we apply a real driving dataset from private cars in Shenzhen, China to train the models and evaluate the scalability of our proposed system.
- We carry out *extensive experiments and analytics*. The collaborative model improves the F1 score by 3.52% and 6.44% as compared to a standalone edge model and a centralized model, respectively. It also decreases the number of false negatives by up to **2/3**, decreasing the number of potential accidents caused by false negatives by an order of magnitude (**24** times less than the centralized model). By conserving an end-to-end latency below **50 ms** up to 256 vehicles, and a bandwidth of **20 Kb/s per vehicle**, CAD3 can handle the data load from peak hour traffic of up to 13 million vehicles spread over 55K road trunks.

The remainder of this paper is organized as follows. After summarizing the core research motivations in Section II, Section III presents the background and related works. Section IV describes the system architecture and methodology. We then describe our dataset in Section V, present the experiments and analysis in Section VI, and discuss our findings in Section VII. Finally, we conclude the paper with suggesting future research in Section VIII.

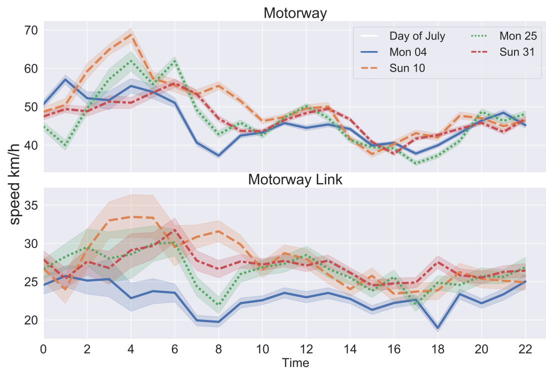


Figure 2: Speed profile of motorway vs. motorway link roads, between weekends, weekdays, and the beginning and end of July 2016. Spatio and temporal variations in speed profile which reflects the time and space changing behavior.

II. Research Motivation

Machine learning-driven anomaly detection has the potential to significantly improve road safety. However, applying such techniques at the city-scale presents numerous challenges. In this section, we summarize such challenges and justify the need for lightweight and distributed algorithms.

A. Detection Challenges

The challenges of machine learning-driven anomalous driving behavior detection are threefold:

- **Changing Patterns:** Driving behavior changes over time, owing to the day time (rush hours vs. normal hours), the day of the week (weekday or weekend), and the road type (motorway or street road), as shown in Figure 2.
- **Context-awareness:** Due to the above changing patterns, spatio-temporal context awareness is crucial for abnormal driving behavior detection. However, current solutions often fail to capture such context. On-board detection solutions lack the holistic view to assess what normal driving behavior is, based on other drivers' behavior for a specific road trunk. On the other hand, cloud solutions tend to deploy city-scale models that lack the fine-grained resolution to address road-level abnormal driving behavior detection, as well as requiring higher latency [3].
- **Continuous Big Data Stream:** Connected vehicles continuously generate massive amounts of data. Integrating data from multiple vehicles further amplifies this challenge. The latest received data is the most valuable for accurate timely decision making. It is therefore crucial to ingest the continuously incoming big data, process, consider the most recent data points, and make decisions in real time [4].

B. The Need for a Lightweight Distributed Models

The above-mentioned challenges are difficult to address with typical centralized detection models. With the continuous generation of massive amounts of data, these models burden the backhaul network with heavy loads. In our dataset, vehicles sharing only 200 Bytes of information at 10Hz result into a collective data rate of 4 GB/s. With the constant addition of new data, the size of the dataset will increase over time, proportionally increasing the training time. Finally, achieving spatio-temporal awareness at the city-scale with centralized models is non-trivial, and may result either in low accuracy or overfitting. On the other hand, lightweight, geographically-bound models allow real-time inference and require less computation and communication resources. Deploying such models at the edge relieves the already burdened backhaul network by keeping data transmission and processing within a single hop. In our dataset, each edge node receives on average 256 KB/s for 2560 events, a much more reasonable load, even on constrained access links. Moreover, by considering only a small portion of the road network, lightweight models considerably reduce the amount of data, thus the training times [5], while facilitating local and context-aware processing and minimizing latency. Lightweight distributed models at the edge of the network therefore allow low-latency data transmission and processing, while enabling de-facto fine-grained spatio-temporal context-awareness.

III. Background and Related Work

In this section, we review the related literature on computation paradigms and the proper model for anomalous driving detection.

A. Edge Architecture For Context-aware and Low-latency Data Analytics

Unsafe driving detection is a highly latency-sensitive service, where unexpected latency may cause severe accidents. Offloading the detection tasks to the cloud fails to address the rapidly changing context of driving conditions [6]. Edge computing features proximity to the end-user, dense geographical distribution, and support for mobility, context-awareness and real-time interactions. This paradigm delivers scalable, highly responsive services and masks transient cloud outages [7]. White *et al.* [5] use a stacked autoencoder with dropout on a deep edge architecture to reduce the training and request time for quality of service prediction. Edge Based Reinforcement Learning (ERL) [8] employs multi-level edge computing to optimize traffic light control in real-time and alleviate congestion in urban environments. Su *et al.* [9] presents an edge architecture for semantic reasoning and investigates how to distribute reasoning tasks on cloud and edge devices to deduce the activities of vehicles. Moreover, collaborative edge computing promises better performance. EF-dedup [10] deduplicates the data at the network edge and introduces inter-edge collaboration. It improves the throughput and lowers the cost of network storage compared to the standalone cloud-based approach. CAD3 offloads anomaly driving detection to edge nodes deployed alongside roads (RSUs), equipped with communication and computation capabilities. Each node learns the normal behavior over time and maintains contextual information of the road in its coverage. CAD3 also introduce inter-edge collaboration to improve the detection.

B. Anomalous Driving Detection

Three types of anomalies exist, including point anomalies, contextual anomalies, and collective anomalies [11]. Convolutional neural network (CNN) [12] and CapsNet [13] are potential machine learning (ML) architectures for collective anomaly driving detection. They detect the invasion of center-lines based on road image data and centralized ML models. Matousek *et al.* [14] analyze an LuST-based simulated dataset to detect driving behavior. Their approach handles the overlapping behaviors. Chen *et al.* [15] identifies specific types of abnormal driving behaviors in real-time using smartphone sensors. The above-mentioned focus on unrealistic trends and apply point anomaly detection. Trip recommendation [16] assists drivers in avoiding stressful driving. It employs a multi-task learning based neural network to predict stress level. This level and the driving behavior are used to recommend the driver to accept or reject the planned trip. However, the system is driver-centric, and thus does not ensure the holisticness nor the safety of others if he/she takes non-recommended trips. QF-COTE [17] employs mobile edge computing (MEC) to detect road anomalies in real-time (i.e., end-to-end delay>300 ms). QF-COTE enables coarse-grain geographic context-aware detection that is offloaded partially to the cloud for collective detection.

Most of these works are either centralized which lack the context-awareness and the responsiveness, or coarse-grained context-awareness. In this paper, we present CAD3, a collaborative, distributed edge computing framework for real-time abnormal driving behavior detection. This framework enables fine-grained geographical awareness and introduces inter-RSU collaboration to provide both collective and context-aware anomaly detection, and thus road and driver-aware detection. To the best of our knowledge, this is the first attempt for anomalous driving detection with leveraging collaborative and distributed edge computing. We also present a complete implementation of the proposed architecture, complemented with a comprehensive evaluation. We use a real driving dataset *(i.e., trips and trajectories data collected from private cars in Shenzhen, China)* to train and evaluate the models, and investigate the system's performance in terms of scalability and accuracy in light of providing collaboration between RSUs from the ML perspective.

IV. System Architecture and Methodology

We propose to distribute computation to enable context-awareness, run anomaly detection on the distributed computing units on RSUs, and enable collaboration between them. In this section, we first discuss the system infrastructure, followed by the distributed computing architecture. Finally, we focus on enabling technologies for road-and-driver aware and collaborative anomalous driving detection.

A. System Infrastructure

We consider the scenario described in Figure 1. To offload the driving activity analysis tasks, we adopt the Edge Computing paradigm. By installing the RSUs in proper locations, they can communicate directly with vehicles, i.e., vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (I2V) within a localized area. We compare alternative communication technologies and consider Dedicated short-range communication **(DSRC)** outperforms long term evolution **(LTE)** in latency-sensitive applications [18], and fulfills the active safety requirements without involving a centralized network infrastructure [2]. Therefore, we propose to equip the RSUs and vehicles with DSRC communication modules to intercommunicate. RSUs also feature either a wired connection (either coaxial or optical Ethernet) for fast and reliable intercommunications, or cellular communication (5G or LTE) as the latency requirements and data volumes are lower. The onboard vehicle Inertial Measurement Unit (IMU) measures the speed, acceleration, deceleration, angular rate, and the orientation of the body, using a combination of accelerometers, gyroscopes, and magnetometers. Then, the onboard DSRC module transmits this information several times per second over a range of a few hundred meters. Each RSU collects this information from the vehicles in its range. The RSUs contain a computationally powerful chip to process the collected data, and they disseminate any notifications using V2I communications afterwards. Owing to the dedicated nature of the spectrum band and the low latency feature of DSRC, unsafe driving activity alerts can be delivered in near-real-time to the drivers.

B. Distributed Computing Architecture

Streaming and analyzing high-velocity continuous data flows in real-time is a challenging problem, especially in decentralized systems. The MapReduce streaming architecture, which extends the consistent hashing function, is a solution to support run-time elasticity and fault-tolerance [19]. Apache Spark[1] achieves high performance for both batch and streaming data, up to 100 times faster than Hadoop MapReduce. Moreover, Spark offers over 80 high-level operators that make it easy to build parallel apps. To provide real-time data streaming and analytics, we integrate the most promising technologies, i.e., Apache Kafka[2] and Spark. Apache Kafka is used for distributed event streaming and Spark for batch-based data streaming and data analytics. Each RSU trains a model that learns the normal driving behavior locally, and maintains road contextual information. The detection algorithm imports the model, recognizes the driving behaviors and detects any anomalies. We create three data topics, i.e., "IN-DATA" for ingesting the incoming vehicular data, "OUT-DATA" for writing detected anomalies, and "CO-DATA" for writing detection summaries.

We propose a two-stage framework, the offline pre-processing and training stage and the online streaming, data analytics, and dissemination stage.

In the **offline stage (i.e., outliers detection, labelling, pre-processing, and training)**, we identify abnormal data points in the dataset and label them as abnormal or otherwise normal.

[1]https://spark.apache.org
[2]https://kafka.apache.org/

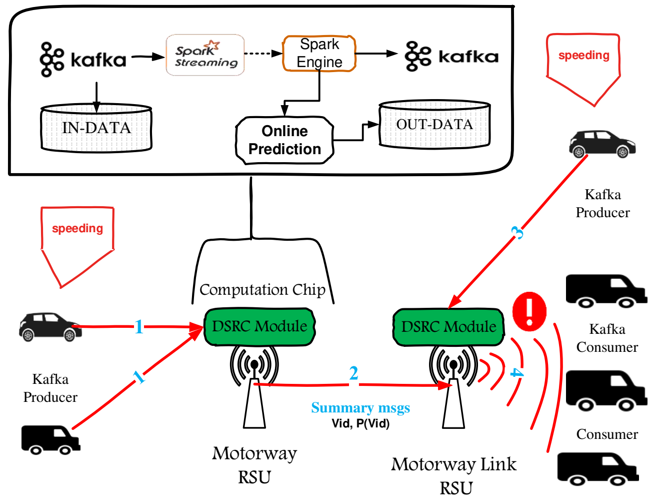


Figure 3: Microscopic level. Data collection, analytic, and inter-RSU collaboration. 1) Vehicles send data to Motorway RSU. 2) Motorway RSU detects the anomolies and shares summary with Motorway Link RSU. 3) Motorway Link RSU detects anomalies based on collected vehicular data and summaries. 4) The latter RSU disseminates the warning.

From our dataset II, we consider four features, including the instantaneous vehicle speed and acceleration (InstSpeed, accel), the hour of the day (Hour), and the road type (RoadType). The speed data of each road type is Gaussian-like; therefore, we use the standard deviation as a cut-off for identifying outliers. We label a data point as normal (class=1), if it exhibits a speed and acceleration in the range $[\mu - 1 * \sigma, \mu + 1 * \sigma]$. Otherwise, we label it as abnormal (class=0). In **Model Training**, we divide our dataset into sub-datasets based on the road type, i.e., motorway, motorway link. We fit each subset with a Naïve Bayes model, resulting in a model per road type. Each RSU uses the model that complies with the road type it is responsible for.

In the **online streaming and detection stage**, shown in Figure 3, we implement a pipeline of the following:

Data Streaming and Analytics. Vehicles use DSRC to push their data via Kafka Brokers to "IN-DATA". Spark streaming then consumes the continuous stream, divides it into micro-batches of 50 ms, and creates a resilient distributed dataset (RDD). The RDD is then passed to the Spark engine for processing to detect any deviation from normal driving, and write the anomalies into "OUT-DATA".

Data Sharing. The RSUs account for vehicular mobility by transferring summaries of the model to adjacent RSUs. Motorway RSU transmits the summary to motorway link RSU (writes to its "CO-DATA" topic).

Data Dissemination. The vehicle consumes any warnings by pulling the messages of "OUT-DATA". These warnings can then be conveyed audibly, visually, or haptically using an infotainment system to raise the driver's awareness and allow them to react in-time.

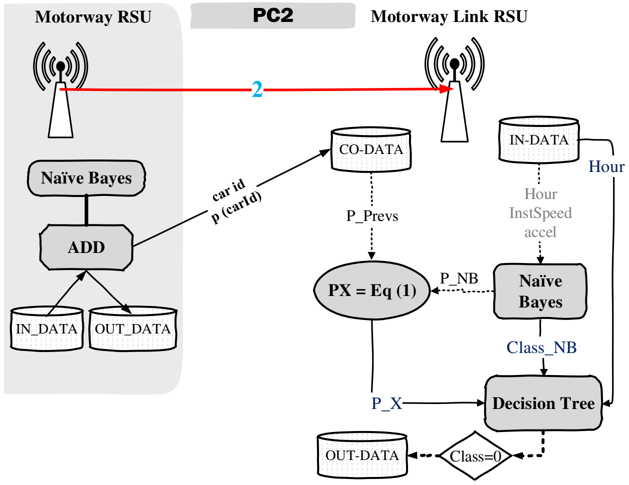


Figure 4: Collaborative Anomaly Detection (CAD3). Motorway RSU detects anomalous driving activity using a Naïve Bayes model and transmits a summary to Motorway link RSU. The latter combines this summary with its prediction using Decision Tree.

C. Context-aware Anomaly Driving Detection (AD3)

Each RSU receives vehicle status information (e.g., location, time, speed, and acceleration), utilizes contextual information (i.e., road type, hour of the day, and speed profile), runs driving activity analysis on them, and notifies the drivers in range of potential danger. The activity is abnormal, if it deviates from the normal distribution [20]. For instance, a driver speeding at 90 km/h on an motorway link where most vehicles drive between 0 km/h and 35 km/h (see Figure 2) will be classified as abnormal. Context-awareness allows for fine-grained distinction between normal and abnormal driving behaviors to avoid any overlap or confusion. For this purpose, Naïve Bayes classifier is trained locally to learn the normal profile, provide road-awareness, and classify accordingly.

D. Collaborative and Context-aware Detection (CAD3)

Neighboring RSUs collaborate to share a summary of driving activity analysis and carry on this process to provide collective driver-aware detection. As a vehicle moves from a motorway to a motorway link, its previous predictions (P_{prevs}) are forwarded to the motorway link RSU, written to its "CO-DATA" topic, (see Figure 4). As part of the collaborative model, the vehicle data is passed to the Naïve Bayes model to predict, resulting in a prediction probability of P_{NB}. A probability P_X is then computed through the following equation:

$$P_X = 0.5 * \overline{P}_{prevs} + 0.5 * P_{NB} \tag{1}$$

P_{NB} is the prediction probability computed by Naïve Bayes, and $\overline{P}_{prevs}$ is the average of prediction probabilities along the motorway. Afterwards, the Decision Tree classifier uses the feature vector [Hour, P_X, $Class_{NB}$] to predict the driving activity, whether normal (class=1) or not (class=0). Once the motorway link RSU detects anomalous driving (speeding, sudden acceleration), it produces a message into the "OUT-DATA" topic. The vehicles in the RSU range consume the message and interpret it as an aggressive driving warning (warning dissemination), as shown in Figure 3.

E. Potential Accidents

For each data point labeled as abnormal, we compute a binary value, fn_i, indicating whether the model recognizes the point as abnormal (1) or normal (0). We also compute δ, the proximity of vehicle's speed to the normal speed. According to the Nislon formula [21], the number of injury-causing accidents A_2 after changing the road speed v_r from v_{1r} to a safer one, v_{2r}, is equal to the number of injury-causing accidents before the change A_1 multiplied by the square of the speed change ratio (see Equation 2).

$$A_2 = A_1 \left(\frac{v_{2r}}{v_{1r}}\right)^2 \tag{2}$$

This formula can be applied directly to an instantaneous speed $v_r(i)$ of a vehicle i on road r, given that v_r is the target speed that is the road's normal speed. $v_r(i)$ deviates from the normal speed v_r, causing abnormal speeding or slowing.

$$A_2 = \begin{cases} A_1 \left(\frac{v_r}{v_r(i)}\right)^2 & \text{if speeding,} \\ A_1 \left(\frac{v_r}{v_r+(v_r-v_r(i))}\right)^2 & \text{if slowing.} \end{cases}$$

A smaller difference between v_r and $v_r(i)$ reflects a driver's commitment to safe speeds, whereas a higher difference may lead to an increase in potential accidents. We denote the proximity of this difference to 1 by δ as follows:

$$\delta = \begin{cases} 1 - \left(\frac{v_r}{v_r(i)}\right)^2 & \text{if speeding,} \\ 1 - \left(\frac{v_r}{v_r+(v_r-v_r(i))}\right)^2 & \text{if slowing.} \end{cases}$$

Ultimately, we calculate the vector of the difference between driving speed and road speed $\vec{v}_\delta$, and a binary false negative vector $\vec{v}_{FN}$. When δ tends to 1 ($\delta \to 1$) and the model does not detect ($fn_i = 0$), an accident has a higher probability to occur. As such, the number of potential accidents is estimated based on the falsely unreported FN speed violation as follows:

$$E(\Lambda) = \sum v_{\vec{FN}}.\vec{v_\delta} \tag{3}$$

V. Dataset Description

In order to evaluate our system, we use real-world driving data. We use a dataset collected from 68,000 privately owned vehicles for one month (1-31 July 2016) in China [22]. The dataset contains trips and trajectories of the cars (see Table I). Using Shenzhen's boundaries, we extract the trips and trajectories within the city and map them onto its road network using a map-matching algorithm [23].

We compute the instantaneous vehicle speed and average road speed v_r as follows:

$$v_r(i) = \frac{Dist(l_{p_i}, l_{p_{i+1}})}{t_{p_{i+1}} - t_{p_i}}, \qquad \overline{v_r} = \sum_{i=1}^{m} v_r(i) \tag{4}$$

Table I: Attributes of a trip and its trajectories.

ObjectID	StartTime	StopTime	StartLon
556605	01/07/2016 09:27:33	01/07/2016 11:10:00	109.82224

StartLat	StopLon	StopLat	Mileage	Fuel	Period
40.64159	110.522649	40.597578	85,611	6,979	6,147

ObjectID	Lon	Lat	GPSTime	AcMileage
556605	109.8223	40.6416	01/07/2016 09:27:33	6,383

Table II: Preprocessed dataset used in the analysis.

CarID	RdID	accel	Speed	Hour	Day	RdType	$\overline{v_r}$
1235	5636	30	36	18	0	motorway	43

Table III: Dataset statistics after filtering the erroneous values.

Region	#Cars	#Trips	$\overline{Speed}$	#Trajectories
Shenzhen	3,306	214,718	23.7	17,926,810
Motorway	2,986	69,182	160	1,338,552
Motorway Link	2,824	48,030	115	445,482

where m is the number of measurements for a car on a road segment r of normal road speed $\overline{v_r}$. Each car moves over a distance $Dist(l_{p_i}, l_{p_{i+1}})$ from point p_i to point p_{i+1} at speed $v_r(i)$. $Dist$ is a function calculating the great-circle distance between two geographical points, t_{p_i} is the time at GPS location l_{p_i} (i.e., longitude and latitude). Vehicle acceleration is thus the difference in speed over time between the specified points. By applying a map-matching algorithm, we obtain the road ID, road type, and tag ID that define the context of each vehicle displacement on the roads. The instantaneous and average speeds determine the speed profile of the vehicle. These derived attributes (see Table II) are used as the features for the anomalous driving detection algorithm. After we filter out erroneous measurements, we get our second input dataset II to be used as vehicle status information. Table III illustrates some statistics about this dataset after filtration, including some context-dependent anomalous records. It shows the number of private cars, trips, trajectories, and average speed in Shenzhen and some types of roads there.

VI. System Evaluation

To validate our system, we implement a prototype system consisting of two RSUs on two adjacent roads. We evaluate our system performance in terms of network and bandwidth scalability, computation scalability, and model accuracy.

A. Use Case

In this section, we consider a microscopic scale use case, vehicular movement from a motorway to a motorway link, and its corresponding workflow, as presented in Figure 3. During movement, the motorway RSU collects status information from vehicles in its range, and runs anomaly detection on this data in real-time. Upon handover, the data collection and anomaly detection results are offloaded to the latter RSU.

B. Experiment Setup and Implementation

We install and configure Apache Spark and Kafka on PC2 (Ubuntu 18.04 LTS, Intel i7-5820K CPU @ 3.30GHz). We configure Apache Spark to run a cluster of 6 worker nodes,

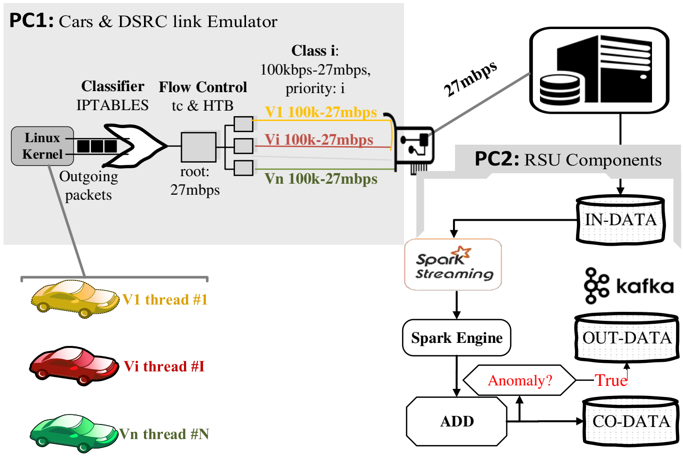


Figure 5: Experimental Testbed. PC1 emulates the vehicle usage of the shared bandwidth of DSRC (27Mbps), and PC2 emulates the RSU components (Kafka, Spark) and the workflow. In PC1, vehicles are emulated using *tc* tool and Hierarchy Token Bucket (HTB).

and Apache Kafka to run 2 servers (Brokers) to act as motorway and motorway link RSUs. **In Kafka**, we create three topics, "IN-DATA" to ingest the data from Kafka Producers, "OUT-DATA" to save the warning messages generated by the anomaly detection algorithm, and "CO-DATA" to ingest the summary messages handed over from Motorway RSU to Motorway Link RSU. We assign three partitions for each topic to speed up reading and writing. We develop the anomaly detection algorithm using **Spark MLLib2**, serializer and deserializer to send and read the vehicular data. To keep the processing latency minimized, we create micro-batches of 50 ms (RDDs) to read data from the topic "IN-DATA", on which we apply the algorithm.

On PC1 (Ubuntu 18.04 LTS, Intel(R) Core(TM) i5-4590 CPU @ 3.30GHz), we emulate the contributing vehicles by running several Kafka producers that read from the dataset II and publish to "IN-DATA", and Kafka consumers to subscribe to "OUT-DATA" through Kafka Brokers. We configure the producers and consumers to push/pull messages instantaneously. To the best of our knowledge, there is currently no readily available DSRC chips for PCs in the consumer market. PC1 and PC2 are connected through a 1Gb/s Ethernet link. On PC1, we emulate the DSRC bandwidth using the hierarchical token bucket (htb) feature of `netem`. We rely on `iptables` to mark the packets from different producers. We then use `netem` to construct a hierarchy that sets the bandwidth of each producer to a minimum of 100 Kbps, up to a maximum shared bandwidth of 27 Mb/s (bandwidth of DSRC). Such a setting allows us to emulate vehicles occupying different shares of the bandwidth of DSRC based on their distance to the RSU and the amount of data to send.

C. Experiment Scenario

In our experiment, we attempt to mimic the scenario illustrated in Figure 3, with the architecture in Figure 4. We extracted two real roads using OpenStreetMap, a motorway road, and a motorway link road, for which we extracted the vehicular data from our dataset. PC1 emulates vehicles by running Kafka Producers and Consumers, while PC2 emulates two adjacent RSUs: one RSU located on a motorway, and another located on a motorway link. The motorway RSU collects and processes the data sent by Kafka Producers, read from the motorway subdataset (see Table II). Similarly, the motorway link RSU collects and processes data from Kafka Producers, read from the motorway link subdataset. The anomalous driving detection algorithm on the motorway RSU forwards summary messages to the motorway link RSU by writing to "CO-DATA" topic of the Motorway Link. We emulate the handover by migrating a portion of Kafka Producers from the motorway RSU to the motorway link RSU. This portion of producers start reading from the motorway link subdataset. Thus, we emulate the mobility of the vehicles in practice and provide a continuous process of driving behavior analysis. As a result, we ensure collective anomalous driving behavior detection.

D. Results

Latency, scalability, and accuracy are the primary concerns for driving behaviour detection systems. The latency between the detection of an aggressive driving event and the delivery of a warning must be small enough to support an in-time reaction. The system should be scalable with an ever increasing number of concurrent vehicles, and address peak-hour traffic. The bandwidth usage must also scale and never exceed the transmission medium's capability (27 Mb/s for DSRC). Finally, the detection model should be accurate enough to minimize the false negatives.

*1) **Scalability with the Number of Vehicles**:* The total number of vehicles that are simultaneously offloading driving activity detection to an RSU, varies from road to road. The safety system must deliver warnings in-time once unsafe driving is detected. Moreover, it must scale up with an increasing number of vehicles, and consider events, such as rush hours. We assess the scalability of the system with an increasing number of vehicles, from 8 to 256. Each vehicle transmits records of the dataset at a frequency of 10 Hz. Given the density of 30 vehicles/km/lane corresponds to fluent traffic with a high density of vehicles [24], and assuming the road comprises of eight lanes, we regard 256 as the maximum number of vehicles within an RSU range. As Figure 6a shows, the system maintains an average total latency below 50 ms, which is small enough for the receiving drivers to react and avoid accidents. Offloading the computation to the RSUs requires total latency of 39.7 ms for 8 vehicles and up to 48.1 ms for 256 vehicles. The average processing time ranges between 7.3 ms, and 11.7 ms, respectively. The remaining latency encompasses the queuing and dissemination that are collectively 31.8 ms up to 36.4 ms, for 8 to 256 vehicles, respectively. The lower error of total latency error bars shows that dissemination of warnings is almost instantaneous after processing the event, whereas the upper error reaches 40 ms at maximum. The testbed emulates multiple concurrent access to a shared access link at the physical and MAC layers. DSRC physical standard (IEEE 820.11p) controls the access to the shared medium using CSMA/CA. Assuming 256 vehicles on a motorway link, sharing the status information at load update rate of 10 Hz (every 100 ms), and each packet is approximately 200 bytes. Each vehicle listens to the medium for DIFS time, and sends a packet if the medium is free, otherwise waits for backoff time, then sends the packet. The time required to access and get all vehicles' packets through the medium, is given by:

$$t_v = t_{Backoff} + num_v * (DIFS + t_{pkt}), \quad (5)$$

$$t_{bakoff} = p_c * cw_{max} * t_{slot}, DISF = SIFS + 2 * t_{slot}. \quad (6)$$

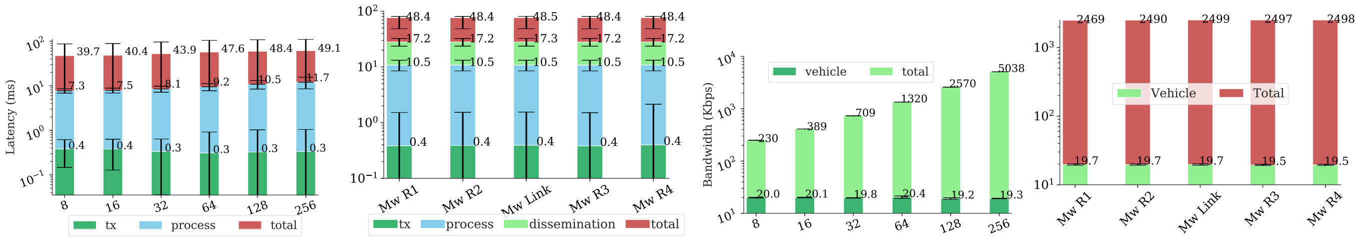
Given the IEEE 802.11p modulation and coding schemes (MCSs) [25], the time is 92.62 ms, using MCS 3 and 54.28 ms using MCS 8, where $t_{slot} = 9\mu$s, SIFS=16 μs [26], [27], $cw_{max} = 255$, and $p_c \leq 0.03$ [24], the collision probability that is proportional to vehicle density and the distance to RSU. Each vehicle waits backoff time at maximum. It is thus possible for 256 vehicles to send at 10 Hz while preventing the buildup of packets at the sender side, as all packets are sent before the next packets are generated, 100 ms later.

The system scales particularly well with road traffic conditions, from almost empty roads (8 vehicles) to peak hour (256 vehicles) with an increase of only 10 ms of the end-to-end latency. *With an average end-to-end latency below 50 ms for all traffic conditions, the system can process and disseminate potential accident warnings in real-time even during the busiest peak hours.*

*2) **Bandwidth Usage with the Number of Vehicles**:* Figure 6c illustrates the average bandwidth usage per vehicle and the total bandwidth with an increasing number of vehicles from 8 up to 256 vehicles. The concurrent number of vehicles at the city-scale (i.e., Shenzhen) can reach over 2 million on the road in the morning rush[3]. According to Figure 6c, each vehicle uses 20 Kbps on average. Two million vechicles would thus require over 40 Gbps. In contrast, the decentralized deployment of RSUs allows CAD3 to scale with increasing traffic with a load of 5 mbps per RSU for 256 vehicles, much lower than the DSRC bandwidth of 27 mbps. Our dataset contains 51,129 individual road trunks for the city of Shenzhen. With a single RSU per road trunk, CAD3 can support a total of 13 million concurrent road users supported by the system, while exploiting only 1/5 of the DSRC bandwidth. *By leveraging a decentralized deployment of edge computing nodes, CAD3 can therefore scale far beyond the current records of peak hour traffic in one of the largest megalopolis in the world.*

*3) **Performance of multiple RSUs**:* We emulate the scenario illustrated in Figure 1. We set up 5 Kafka Brokers as 5 RSUs on PC2. Among them, a Motorway Link RSU connects to 4 Motorway RSUs. On PC1, we run 5 sets of 128 Kafka Producers, each set transmits to an RSU. We assess

[3]https://www.eyeshenzhen.com/content/2020-05/15/content_23155107.htm



(a) Latency vs vehicle increase (b) Latency components vs RSUs (c) Bandwidth vs vehicle increase (d) Bandwidth usage vs RSUs

Figure 6: Given vehicle's data sharing at 10 Hz, (a) The transmission (Tx), processing, and total latency increases slightly with an increase in vehicles, less than 10 ms, with standard error 40 ms at maximum. The total latency remains below 50 ms. The black bars represent their standard error. (b) Dissemination latency vs RSU type, Motorway (MW) or MW Link RSU. (c) Average vehicle bandwidth and total received bandwidth. The total bandwidth $<$ DSRC bandwidth, 27 mbps. (d) The bandwidth received by each RSU, each with total bandwidth $<$ DSRC bandwidth, 27 mbps. The bandwidth received by Motorway Link RSU, Mw Link, is slightly higher than the bandwidth of other Motorway RSUs, Mw R#.

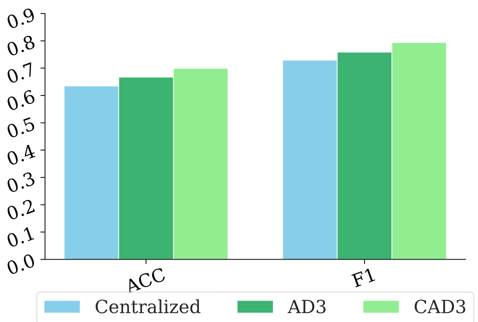


Figure 7: Centralized vs distributed standalone (AD3) vs collaborative (CAD3) model. CAD3 increases the F1 and accuracy by 0.0352 and 0.0322, compared to AD3, and increases them by 0.0644 and 0.0644, compared to the centralized.

Figure 8: Mesoscopic (driver-trip) level illustration. Detection of car abnormally slowing using CAD3 (green), AD3 (white), and centralized (orange). Given the car is moving southwest to northeast, the red car icon corresponds to driving behavior being detected as abnormal, and a blue icon as normal. Only CAD3 detects abnormal driving accurately and stably, while AD3 still fluctuates, and centralized remains unpredictable.

the dissemination latency for each involved RSU. This latency measures the delay between the detection and the delivery of the warning. Figure 6b illustrates all the components of end-to-end latency. The average dissemination time is in the range $[17.2, 17.3]$ ms with standard error ≈ 4.4 ms. Each Kafka consumer pulls every $10\,$ms to avoid consuming the bandwidth. Thus the lower bound of dissemination delay can be composed as $10 + 7.2 \pm 4.4$ ms. Figure 6d illustrates the bandwidth usage for each RSU. Owing to its collaboration with 4 adjacent RSUs, the motorway link RSU, named Mw Link, consumes slightly higher bandwidth as compared to Motorway RSUs, albeit much lower than the DSRC bandwidth.

The system efficiently decentralizes the workload at large scales through local interactions adjacent RSUs, affordable load and real-time latency.

*4) **Collaborative (CAD3) Model Performance**:* We apply a hybrid model using Naïve Bayes and Decision Tree classifiers to improve the detection. Both models are binary classifiers, implemented under the Spark Mllib. Besides being reasonably lightweight classifiers, these models are explainable. As we are dealing with road safety, where human lives are at stake, explaining the algorithms' decisions is critical in case of potential lawsuits. We believe that the improvement in performance brought by black-box models such as neural networks is not worth the loss in explainability and accountability.

Table IV: True positive (TP), false negative (FN) rate for 89K records and estimated accidents $E(\Lambda)$, Equation. 3, for 500K

Model #89K records	**TP Rate**	**FN Rate**	$E(\Lambda)$ #500K records
Centralized	49.2%	19.9%	9,004
AD3	52.3%	11.8%	1,475
CAD3	57.9%	6.2%	371

As baselines, we implement a **distributed standalone (AD3) model**, using only Naïve Bayes, and **centralized model** that assumes training all road vehicular data at once. We split the Motorway and Motorway Link data (see Table III) into 80% training data and 20% online testing data. We use Naïve Bayes to train the motorway model using the motorway sub-dataset and train motorway link model using the motorway link sub-dataset. In the online testing, we use PC1 to send the vehicular data. Kafka producers send vehicle data from the motorway test data to Motorway Kafka Brokers, and from the test data of motorway link to Motorway Kafka Brokers on PC2. The **collaborative model (described in Subsection IV-D)** uses the previous prediction probabilities passed from Motorway RSU, Naïve Bayes prediction probability, class, and hour of the day, to predict the driving activity using Decision Tree, as seen in Figure 4. Figure 7 illustrates that CAD3 outperforms both AD3 and the centralized model in the motorway link RSU. The collaborative model improves the F1 score and accuracy by 3.52% and 3.22%, respectively, compared to the standalone model, and both of them by 6.44%, compared to the centralized model. Figure 8 illustrates what happens at a mesoscopic level (driver-trip) for a car moving southwest to northwest. We compare the detection capabilities of CAD3, AD3 and Centralized solutions. CAD3 detects abnormal points correctly, with a high stability over time, owing to the transfer of reports between edge servers. On the other hand, although AD3 detects most of the abnormal driving points, the detection accuracy still fluctuates. Finally, the centralized model results in unpredictable results as it lacks context-awareness.

To evaluate the three models (centralized, standalone AD3, and collaborative CAD3) in terms of number of avoided potential accidents (see Equation 3), we train and test each model over a dataset of 500K data samples extracted from Table II. 35% of the samples exhibit abnormality. For each data point detected, we compute a logical value (FN), whether the model recognizes the point as abnormal or not. Our model clearly reduces the expected number of potential accidents caused by false negatives by leveraging road- and driver-awareness to improve the prediction accuracy. CAD3 results in only 371 expected potential accidents, 4 times less than its edge counterpart, and 24 times less than the centralized model.

Improving the correctly detected anomalous behaviors ensures less disturbance to other drivers with false warnings. CAD3 increases TP rate by 5.6% and 8.7%, compared to AD3 and centralized, respectively, as seen in Table IV. Additionally, the lower the fake negatives (FN), the less potential accidents if the drivers act in-time in response to the warning. CAD3 reduces FN rate by 5.6% and 13.7%, compared to AD3 and centralized, respectively. Thanks to its low FN rate, CAD3 was more reliable than both the edge-assisted standalone model and the centralized model. *The inter-RSU collaboration allowed the system to address both road and driver awareness, consolidating the prediction accuracy and significantly lowering the number of potential accidents.*

VII. Discussion

This paper presents CAD3, a full-stack framework for collaborative anomaly detection at the edge of the network. In this section, we analyze how CAD3 compares to alternative systems in literature. We then discuss our data analytics model, limitations, and suggest future works.

A. Comparison with state-of-the-art

The existing literature presents few systems comparable to CAD3. QF-COTE [17] is an MEC system that detects road anomalies in over 300 ms, using the cloud for inter-node collaboration. In comparison, by distributing the collaboration directly at the edge, we can achieve a latency as low as 50 ms, with a comparable accuracy and F1 score. Other works focus on the detection and do not consider how to collect and disseminate data in real-time.

B. Scalability

CAD3 achieves high scalability through inexpensive local interactions. Adjacent RSUs collaborate to improve the detection performance at a large scale. In this paper, we demonstrated that CAD3 can scale up to 256 vehicles connected to a single node, with end-to-end latency below 50 ms. In this section, we consider the practical implications of city-scale scalability under two conditions, i.e. scalability of the access link for dense deployment of edge computation nodes and practical deployment of our system in urban areas.

Access-link scalability: Owing to a high vehicle density, the system may require deployments of many RSUs in a small area, which causes signal overlapping between adjacent RSUs. Therefore, proper installation, setting, and techniques are required to avoid interference.

Thoughtfully positioning the RSUs: By considering the geographical characteristics of the area, it is possible to install RSUs with such a way that their ranges do not interfere while covering the vehicles on the road without interruption [28]. Heavily congested roads can deploy more RSUs located at smaller distances from each other and set the modulation coding rate to a mode with higher performance, i.e., higher data rate and smaller range. For example, two RSUs at a distance 125 m, using 64-QAM 3/4 modulation (MCS 8) [24], and updating at 10 Hz, can serve up to 400 vehicles under 85 ms.

High-level management scheme: A manager can change the operating service channel and use a different SCH when the interference level increases [29]. Active channel management would allow more vehicles to be served with lower interference.

C. Data Analytic Performance

Thanks to the usage of well-known data analytics platforms, including Apache Kafka and Spark, our framework allows reusing a multitude of existing data analytics algorithms and considerably simplifies the development of new algorithms. In this study, we demonstrate the potential of CAD3 by using a combination of Naïve Bayes (for anomaly detection), and Decision Tree (for collaboration between RSUs). Even with such simple algorithms, CAD3 displays a significant improvement over the edge standalone model and the centralized model with an increase in F1 score of 3.52% and 6.44%, respectively. CAD3 also significantly decreases the false negative rates, dividing the number of potential accidents by 4 compared to the edge standalone model and by 24 compared to the centralized model. Our performance evaluation shows that CAD3 has a good potential deploy sophisticated algorithms while keeping a low end-to-end latency. With such algorithms, we expect CAD3 to bridge the gap between standalone distributed models and centralized models in terms of latency, scalability, and accuracy.

D. Real-world Feasibility

Proposing a massively-distributed edge computing solution requires consideration of the physical feasibility of deployment. Smart cities have already deployed roadside infrastructure along roads, co-located for instance with lamp poles and traffic lights. We use the Overpass API[4] and OSMPythonTools wrapper[5] to extract their locations from OpenStreetMap[6] on Shenzhen's road network. Figure 9 illustrates the spatial placement of traffic signs (blue colored dots), lamp posts (cyan colored), and presents the vehicular density as a red heatmap overlay. Given the vehicular density and the existing infrastructure, Table II assesses the number of RSUs required to deploy. For cost efficiency, the deployment considers frequently used roads, having vehicles traversing on them, and takes into account both DSRC range and average road length. Moreover, Table VI presents statistics about the relative distances between the already deployed roadside infrastructure (traffic signs and lampposts). Except the regions marked by gray colored circles in Figure 9, the existing roadside infrastructure almost covers the entire city with respect to the maximum distance and DSRC range. Colocating edge computing facilities with such infrastructure simplifies the deployment and minimizes the cost. They can embed units to allow communication to vehicles within each localized area[7] COTS ships can also be installed to perform the required computation, such as Raspberry Pi 4[8]. However, the challenge is to implement inter-RSU collaboration where RSUs are not connected (due to long distance). LTE and 5G are potential technologies to support distant collaboration where needed. Interestingly, the roadside infrastructure are envisioned to function as small cells in soon-to-com 5G era. Raspberry Pi 4 or NeoGLS RSU[7] can plug LTE/5G modem to reach RSUs beyond DSRC range (where needed). 5G addresses ultra-reliable low latency use cases, making it an efficient candidate to replace DSRC on the forward link, vehicle-to-RSU or RSU-to-RSU link, and accommodate considerably the traffic volume in any region.

E. Limitations and Future Works

Our study presents several limitations. First, CAD3 is evaluated through a testbed emulating a DSRC network. Although we provide theoretical performance calculations to account for the physical and MAC layer impact on performance, these results should be further verified through a real-life implementation involving DSRC-equipped vehicles moving at typical speeds in an urban environment. Moreover, other communication technologies, such as 5G, should be investigated to further enhance the scalability of the system. Additionally, as this paper primarily focuses on providing a scalable architecture and collaborative detection at the network edge, we employ simple detection algorithms to highlight the improvement in detection performance over centralized and non-collaborative edge models. In future works, we will implement complex anomaly detection algorithms to operate within CAD3 and validate the system performance through experiments on real-life vehicles.

[4]https://overpass-turbo.eu/

[5]https://github.com/mocnik-science/osm-python-tools

[6]www.openstreetmap.org

[7]They can embed RSU-301U rsu communication unit, and vehicles embed OBU-301U DSRC onboard units unex.en.taiwantrade.com/product-catalog.

[8]www.raspberrypi.org/products/raspberry-pi-4-model-b/

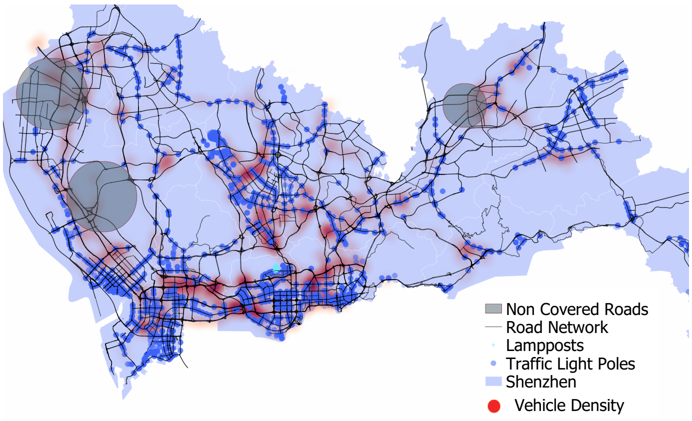


Figure 9: Macroscopic level illustrating the feasibility using traffic signals to embed roadside units. Spots marked with gray circles require RSU installation for city-scale coverage. Red-colored heatmap illustrates the density of vehicles on the corresponding roads on Monday 8:00am.

Table V: RSUs required as function of vehicle density and average road length, and classified based on road type.

Road	**Density**	**# road**	**Mean** (m)	**STD** (m)	**RSUs**
motorway	7.7%	435	3357	7652	1460
motorway link	2.8%	159	596	1626	94
trunk	11.6%	656	1622	5520	1064
trunk link	4.4%	247	339	1931	83
primary	25.2%	1431	668	2939	956
primary link	3.4%	191	211	169	40
secondary	20.1%	1140	561	2337	639
tertiary	18.8%	1064	522	2592	555
residential	5.3%	303	334	1470	101
secondary link	0.3%	36	186	156	6

Table VI: Relative spatial distance of traffic signs and lamps.

RSU	**count**	**AVG** (m)	**STD** (m)	**75%** (m)	**MAX** (m)
Traffic light	3,278	244.57	299.7	444.2	999.5
Lamp poles	116	71.9	100	82.8	520

VIII. Conclusion

Slowing, speeding, and sudden acceleration are the leading causes of severe accidents on highways. However, detecting such patterns at the city-scale is extremely challenging. We proposed CAD3, a collaborative edge-assisted framework to detect anomalous driving in real-time. CAD3 leverages the decentralized deployment of computing nodes at the roadside (RSUs) to distribute the detection. RSUs collaborate by sharing aggregated data to account for vehicle mobility. We implemented CAD3 and evaluated the system with a real driving dataset of private cars collected in Shenzhen, China. In addition to the microscopic level (an intersection connecting four motorways), we illustrated the system performance at the mesoscopic level (over a vehicle trip on multiple roads), and investigated the system's feasibility at macroscopic scale. CAD3 presents significant scalability capabilities, allowing up to 256 vehicles on a single RSU to be served with an average end-to-end latency below 50 ms and a total bandwidth around 5Mb/s. It outperformed other MEC-based standalone systems such as QF-COTE (300 ms latency). The collaborative mechanisms at the core of CAD3 improves the percentage of correctly detected unsafe behaviors, and significantly reduces the false negatives, leading to a number of potential accidents 24 times lower than a centralized model. CAD3 is also cost-effective, lightweight, and deployable on large scale with affordable infrastructure, making it accessible to municipalities.

Acknowledgment

This research has been supported in part by project 16214817 from the Research Grants Council of Hong Kong, and the 5GEAR and FIT projects from Academy of Finland.

References

[1] Z. Tian, "Speed-accident relationship at urban signalized intersections," *Procedia-social and behavioral sciences*, vol. 96, pp. 1383–1388, 2013.

[2] 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.

[3] M. Satyanarayanan, G. Lewis, E. Morris, S. Simanta, J. Boleng, and K. Ha, "The role of cloudlets in hostile environments," *IEEE Pervasive Computing*, vol. 12, no. 4, pp. 40–49, 2013.

[4] E. Mehmood and T. Anees, "Challenges and solutions for processing real-time big data stream: A systematic literature review," *IEEE Access*, vol. 8, pp. 119 123–119 143, 2020.

[5] G. White, A. Palade, C. Cabrera, and S. Clarke, "Autoencoders for qos prediction at the edge," in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2019, pp. 1–9.

[6] G. Orsini, D. Bade, and W. Lamersdorf, "Context-aware computation offloading for mobile cloud computing: Requirements analysis, survey and design guideline," *Procedia Computer Science*, vol. 56, pp. 10–17, 2015.

[7] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[8] P. Zhou, T. Braud, A. Alhilal, P. Hui, and J. Kangasharju, "Erl: Edge based reinforcement learning for optimized urban traffic light control," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2019.

[9] X. Su, P. Li, J. Riekki, X. Liu, J. Kiljander, J. Soininen, C. Prehofer, H. Flores, and Y. Li, "Distribution of semantic reasoning on the edge of internet of things," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2018, pp. 1–9.

[10] S. Li, T. Lan, B. Balasubramanian, M.-R. Ra, H. W. Lee, and R. Panta, "Ef-dedup: Enabling collaborative data deduplication at the network edge," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 986–996.

[11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[12] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322–3337, 2017.

[13] M. Kim and S. Chi, "Detection of centerline crossing in abnormal driving using capsnet," *The Journal of Supercomputing*, vol. 75, no. 1, pp. 189–196, 2019.

[14] 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.

[15] Z. Chen, J. Yu, Y. Zhu, Y. Chen, and M. Li, "D 3: Abnormal driving behaviors detection and identification using smartphone sensors," in *2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2015, pp. 524–532.

[16] R. Verma, B. Mitra, and S. Chakraborty, "Avoiding stress driving: Online trip recommendation from driving behavior prediction," in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2019, pp. 1–10.

[17] Z. Zheng, M. Zhou, Y. Chen, M. Huo, and D. Chen, "Enabling real-time road anomaly detection via mobile edge computing," *International Journal of Distributed Sensor Networks*, vol. 15, no. 11, p. 1550147719891319, 2019.

[18] Z. Xu, X. Li, X. Zhao, M. H. Zhang, and Z. Wang, "Dsrc versus 4g-lte for connected vehicle applications: A study on field experiments of vehicular communication performance," *Journal of Advanced Transportation*, vol. 2017, 2017.

[19] A. Kumbhare, M. Frincu, Y. Simmhan, and V. K. Prasanna, "Fault-tolerant and elastic streaming mapreduce with decentralized coordination," in *2015 IEEE 35th International Conference on Distributed Computing Systems (ICDCS)*, 2015, pp. 328–338.

[20] 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.

[21] G. Nilsson *et al.*, "Traffic safety dimensions and the power model to describe the effect of speed on safety," Ph.D. dissertation, Univ., 2004.

[22] T. Li, A. Alhilal, A. Zhang, M. A. Hoque, D. Chatzopoulos, Z. Xiao, Y. Li, and P. Hui, "Driving big data: A first look at driving behavior via a large-scale private car dataset," *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, pp. 61–68, 2019.

[23] P. Newson and J. Krumm, "Hidden Markov Map Matching Through Noise and Sparseness," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '09. New York, NY, USA: ACM, 2009, pp. 336–343. [Online]. Available: http://doi.acm.org/10.1145/1653771.1653818

[24] A. Bazzi, B. M. Masini, A. Zanella, and I. Thibault, "On the performance of ieee 802.11 p and lte-v2v for the cooperative awareness of connected vehicles," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 419–10 432, 2017.

[25] A. Bazzi, B. M. Masini, and A. Zanella, "Cooperative awareness in the internet of vehicles for safety enhancement," *EAI Endorsed Transactions on Internet of Things*, vol. 3, no. 9, 2017.

[26] K. Bilstrup, E. Uhlemann, E. Ström, and U. Bilstrup, "On the ability of the 802.11 p mac method and stdma to support real-time vehicle-to-vehicle communication," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, p. 902414, 2009.

[27] B. Bilgin and V. Gungor, "Performance comparison of ieee 802.11 p and ieee 802.11 b for v-to-v communications in highway, rural, and urban areas," *International Journal of Vehicular Technology*, 2013.

[28] M. Klapez, C. A. Grazia, and M. Casoni, "Minimization of ieee 802.11 p packet collision interference through transmission time shifting," *Journal of Sensor and Actuator Networks*, vol. 9, no. 2, p. 17, 2020.

[29] J. Almeida, M. Alam, J. Ferreira, and A. S. Oliveira, "Mitigating adjacent channel interference in vehicular communication systems," *Digital Communications and Networks*, vol. 2, no. 2, pp. 57–64, 2016.