Internet of Vehicles: Enabling Safe, Secure, and Private Vehicular Crowdsourcing

Joshua Joy
UCLA
Email: jjoy@cs.ucla.edu

Vince Rabsatt
UCLA
Email: rrabsatt@cs.ucla.edu

Mario Gerla
UCLA
Email: gerla@cs.ucla.edu

***Abstract*—Safely avoiding accident prone routes is vital to maintaining safe and intelligent transportation. Real-time crowd sourcing of accident causing factors (e.g., icy roads, rain slicked road patches) combined with historical accident information can be utilized by intelligent navigation systems to avoid accident prone routes. A vehicle cloud can compute such safe routes and react faster than a centralized service given crowdsourced data such as images, sensor readings, etc. However, in addition to the intelligent routing, the security and privacy of each data owner must be provided as well. Additionally, crowdsourced information must be validated in the vehicular cloud. In this paper, we examine approaches to ensure safe, secure, and private vehicular clouds.**

I. Introduction

We are witnessing a revolution in the way that mobile communication and computation are occurring. Traditional vehicles are being transformed by an array of sensors that collect a broad range of information and turn vehicles into data sources. This information provides situational awareness (e.g. road conditions, road hazards) computed by a vehicular cloud. Vehicles communicate and coordinate via vehicle to vehicle (V2V) communication and collaboratively compute the shared sensor data. Computed results are shared within vehicle cloud clusters in addition to being propagated towards edge clouds (as opposed to straight to the Internet cloud) as shown in Figure **??**. The edge cloud is similar to fog computing, where cloud functions are performed by servers at the edge (e.g., access point or cellular base station). This ability to gather large amounts of data will lead the way to autonomous driving vehicles.

The first prominent and probably the most important application for autonomous vehicles, is prompt delivery of passengers with maximum safety and comfort, while minimizing the impact on the environment. We are witnessing today in the vehicle fleet the same evolution that occurred a decade ago in the sensor domain from Sensor Web [**?**](i.e., sensors are accessible from the Internet to get their data) to Internet of Things (IoT), where computers with embedded sensors are networked with each other to make intelligent use of their sensors.

The IoT in the intelligent home, formed by the myriad of sensors and actuators that cover the house internally and externally, can manage all the utilities in the most economical way, with maximum comfort to residents and virtually no human intervention. Similarly, in the modern energy grid, the IoT consisting of all components large and small can manage power loads in a safe and efficient manner, with the operators now playing the role of observers.

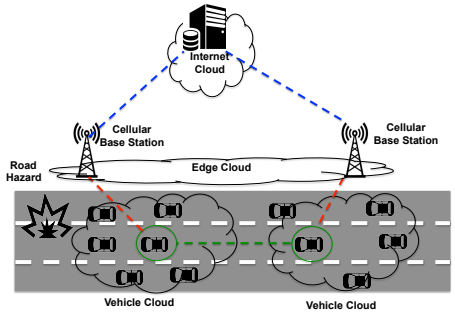


Figure 1. The vehicle cloud computes and communicates safe routes. Security and privacy protection is required as vehicle to vehicle communication occurs without reliance on the Internet cloud.

In the vehicular grid, the Internet of Vehicles (IoV) is more complex than the smart home and smart energy grid IoTs. In fact there are many different "Things" in the IoV. Namely:

- External sensors (GPS, cameras, lidars, etc.)
- Internal automotive sensors and actuators (brakes, steering wheel, accelerator, etc.)
- Internal cockpit sensors (driver's state of health, alertness, tone of voice, health sensors like the Ford heart monitor seat, etc.)
- The driver's messages (e.g., tweets, Facebook) are also measurable sensor outputs that characterize the state of the system and of the driver.
- Vehicle's beacons, alarm reports on the vehicle state; say, position, key internal parameters, possible dangers, etc.

The rest of this paper is organized as follows. In Section II, we identify characteristics that distinguish IoV from IoT. We then present the network technology that enables vehicle clouds. Section IV describes potential applications for IoV. We address possible security and privacy concerns in Section V. Finally, the paper concludes in Section VI.

II. Internet of Vehicles Distinguishing Characteristics

This complex picture (of sensors and stakeholders) tells us that IoVs are different from other IoTs. What sets them apart

are the following properties and characteristics:

1) **High Mobility**: IoVs must manage the high mobility of vehicles and its impact on wireless communication
2) **Safety critical Applications**: this implies low latency and high reliability requirements
3) **Vehicle-to-Vehicle Communication**: short-range communication and limitations in wireless environments pose many challenges
4) **Security**: false data propagation (from hackers and from malicious agents) is a threat to vehicle clouds
5) **Privacy**: driver behavior and vehicular sensor data must be privately crowdsourced

In the Internet of Vehicles, like with all the other IoTs, when the human control is removed, the autonomous vehicles must automatically, transparently, and efficiently cooperate to maintain smooth traffic flow on roads and highways. Visionaries predict that self-driving vehicles will perform much better than human drivers, handling more traffic with lower delays, less pollution and improved driver and passenger safety. However, the complexity of the distributed control of hundreds of thousands of cars cannot be taken lightly. If a natural catastrophe suddenly happens, say an earthquake, the vehicles must be able to coordinate the evacuation of critical areas in a rapid and orderly manner. This requires the ability to efficiently communicate with each other and also to discover where the needed resources are (e.g., ambulances, police vehicles, information about escape routes, images about damage that must be avoided, etc.). Moreover, the communications must be secure, to prevent malicious attacks that in the case of autonomous vehicles can be deadly since there is no standby control and split second chance of intervention by the driver, who may be surfing the web.

All of these functions, from efficient communications to distributed processing over various entities, will be provided by an emerging compute, communications and storage platform specifically designed for vehicles—the *Vehicular Cloud*. The Vehicular Cloud [**?**] is justified by several observed trends:

1) Vehicles are becoming powerful sensor platforms (e.g., GPS, video cameras, pollution, radars)
2) Spectrum is becoming scarce $=>$ Internet upload of all the sensor outputs is expensive and infeasible
3) Cooperative data processing by vehicles rather than uploading to the Internet (e.g., pedestrians crossing, shock wave mitigation, platoon coordination)

To support the above functions, the mobile Vehicle Cloud provides several basic services from routing to content search, through standard and open interfaces that are shared by all auto manufacturers.

III. Connecting Vehicles

Wireless communication technologies play a central role in enabling vehicle to vehicle (V2V) or vehicle to infrastructure (V2I) communication. In regards to V2V Dedicated Short-Range Communication (DSRC), which has been defined by IEEE 802.11p in the U.S. and ETSI ITS-G5 in Europe, provide a low-cost, low-latency, means to connect vehicles in a distributed manner. However, due to the short-range and high mobility of vehicles the ability to communicate within a large coverage area is necessary. This need can be addressed by a cellular technology such as 4G/LTE or soon to come 5G.

These heterogeneous communication technologies need to be utilized in a manner that is most efficient for the overlaying applications and the underlying wireless mediums. Spectrum for both cellular and vehicle communication is limited. However, DSRC is unlicensed and cellular is licensed, which imply that it is controlled by an operator and can be costly to use. Therefore, use of cellular communication should be restricted to as needed.

Efficient use of heterogeneous technologies can be achieved by grouping vehicles into clusters. Vehicles clusters can be formed with respect to geographical locations [**?**] or observed channel quality at vehicles [**?**]. Vehicle clusters enable a single vehicle within the group, Cluster Head (CH), to collect data from Cluster Members (CM) and aggregate this information, which will lead to less demand for cellular resources. These clusters of vehicles form the Vehicle Cloud.

This architecture raises security concerns especially with DSRC since the channel is open to all vehicles within the network, and is not coordinated by a central operator as in cellular networks. A few key concerns are 1) data integrity: how can a vehicle ensure that the received data has not been modified? 2) Driver authentication: how can users ensure that the vehicles they are communicating with are who they say the are? 3) Privacy: how to ensure messages are not intercepted by others?

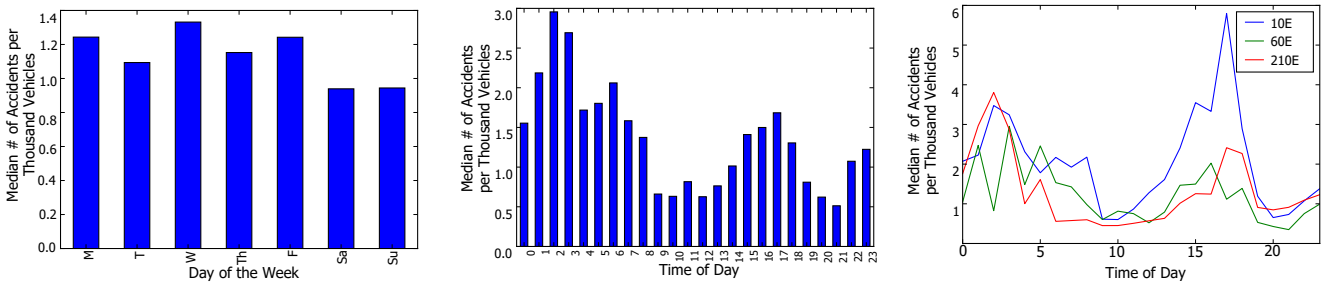
IV. Safe and Efficient Navigation

There is a wealth of information produced by vehicles daily. We are now able to collect, process, and share this information in ways that have the ability to increases safety as well overall travel time. Advanced Driver Assistance Systems (ADAS) are a key enabler to overcoming traffic issues, and improving drivers experience.

Sensors in roadways, as well as in vehicles, have the ability to detect traffic state, such as vehicle flow, speed, density, or traffic accidents. This information can be shared between vehicles, or with infrastructure to be routed to backend severs to analyze traffic conditions. This granularity of data facilitates applications that can taken into account the overall safety of various routes, and provide drivers with the ability to include route safety in their navigation decisions.

In [**?**], the authors found that the large amounts of data collected from highways can show trends in traffic patterns and accidents, which is information that can be used to enhance navigation applications. Figure **??** shows how the number of accidents per thousand vehicles vary over different times for a subset of highways in Los Angeles, CA. The accident dataset was collected from the California Performance Measurement System (PeMS) over the year of 2015.

Vehicular shock waves are a traffic phenomenon that results from unexpected, seemingly unprovoked, patches of slow



(a) Day of the week (b) Time of day (c) Freeway

Figure 2. Number of accidents per thousand vehicles at various times for a subset of roads in the Los Angeles freeway system.

moving vehicles that lead to nonuniform distribution of traffic along highways. V2V can be leveraged in this instance to inform drivers of traffic irregularities and provide recommendations such as reduced speed adjustments or lane change maneuvers to improve the overall traffic state [**?**]. This application relies on the ability to communicate between vehicles, however sufficient traffic improvements can be achieved with low penetration rates of V2V technology. Traffic information can be processed locally within vehicle clouds to provide vehicles in a geographic location with traffic awareness beyond line of sight, which will offer the vehicles the ability to adapt to traffic demands in advance.

V. Security and Privacy

A. Authenticated Crowdsourced Data

A key concern when relying on distributed crowdsourced sensor data is the vulnerability to falsified data. Both the Internet cloud [**?**] and vehicular cloud [**?**] are vulnerable to spoofed data attacks. Malicious actors are able to generate phantom vehicles, giving the appearance of increased congestion and forcing targeted vehicles along specific routes. The Internet cloud must perform anomaly detection while the vehicular cloud must drop spoofed messages, ultimately collaborating together.

The question naturally arises, how to detect spoofed data in the vehicular cloud? More importantly, is it possible to ensure the authenticity and validity of critical safety messages when critical safety decisions must be made in the matter of seconds?

One such approach is to collaboratively validate and achieve local consensus within the cluster itself. For example, securely logging and auditing all safety broadcasts limits a nefarious vehicle that broadcasts a sudden obstacle and instructs vehicles to suddenly brake causing a rear end accident. All nearby vehicles observe the nefarious braking broadcast while simultaneously observing there was in fact no such obstacle. The malicious vehicle will be quickly identified and corresponding action can be taken, such as investigation into faulty sensors or revocation of certain vehicular privileges by the authorities.

The general flow proceeds as follows for each vehicle [**?**].

1) Collect and process sensor data (e.g., video streams, lidar, radar, wireless signals)
2) Generate keys (rotation for privacy)
3) Sign computed sensor data (to share with neighbors)
4) Upload signed data to edge cloud (for persistent storage)
5) Broadcast signatures to surrounding vehicles (pointers to data)
6) Verify incoming signatures (each vehicle validates neighbors do in fact exist)
7) Access edge cloud (download signed data)
8) Validate neighbor vehicles data against local records
9) Broadcast to neighbor vehicles agree or disagree message

Naturally, the vehicular cloud is a perfect fit and is fully capable of satisfying these resource needs. Joy et. al. has shown the the above steps can be completed using vehicular computing on the order of hundreds of milliseconds [**?**].

Computation traditionally advances at a faster rate than the network improves. Thus, local vehicle clusters avoid the network bottleneck that arises when transporting and verifying large amounts of data in the Internet cloud. Rather, the vehicular cloud verifies the data and propagates the signed data towards the edge cloud in a delay-tolerant fashion. Eventually, large scale learning models are updated with the verified data.

B. Scalable Privacy

Driver and passenger privacy will become increasingly important with the deployment of the Internet of Vehicles. Drivers will only crowdsource and share their personal data upon the acceptance and guarantee of privacy.

Differential privacy has emerged as the gold standard [**?**] of privacy. Roughly speaking, differential privacy says that the ability of an adversary to inflict harm should be independent of whether any individual participates or not in the dataset. Thus, differentially private mechanisms protect a data owner allowing each individual to control insight into their personal information.

A multitude of differentially private mechanisms have been proposed to balance the trade-off of privacy and accuracy. However, these mechanisms increase privacy at the cost of reducing accuracy (adding more noise making the results unusable) or they increase accuracy yet decrease the privacy guarantee (removing noise thus reducing privacy).

Rather than sacrifice both accuracy and privacy, we would like a notion of privacy that strengthens as the queried pop-

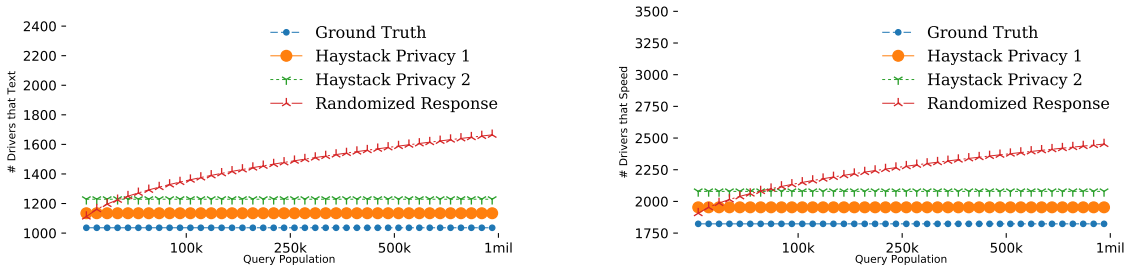


Figure 3. Absolute error measured by the difference of the ground truth and proposed strategy estimate, expressed in the number of drivers that admitted to (a) texting while driving (b) excessive speeding as surveyed in the American Automobile Association Foundation's annual traffic safety culture index. The query population is increased to strengthen privacy by increasing the adversarial uncertainty regarding the specific drivers that admitted to the violation. The randomized response mechanism quickly grows in absolute error from the ground truth while Haystack Privacy maintains constant error from the ground truth.

ulation increases yet in the worst case the accuracy remains constant, i.e., *scalable privacy* [**?**]. The larger query population provides adversarial uncertainty as the underlying distribution becomes distorted making it difficult to determine a particular data owner's truthful response. Thus, each data owner is effectively "hiding in a haystack".

Figure **??** compares the absolute error (95% confidence interval) as measured by the difference between ground truth and the proposed strategy estimate, expressed in the number of drivers excessively speeding and texting while driving [**?**]. The adversarial uncertainty regarding a particular data owner's truthful response increases with the query population, though affects the accuracy of the randomized response mechanism. Haystack Privacy is able to maintain constant error as the multi-round protocol removes the majority data owner noise, yet preserves privacy by sampling a subset of the data owners.

VI. Conclusion

In this paper, we present an architecture of heterogeneous communication technologies that enables vehicle-to-vehicle and vehicle-to-infrastructure communication to support the Internet of Vehicles. Vehicles capture and generate a large range of data. The vehicular cloud data is shared and processed within the vehicle cloud, or propagated to the Internet cloud to provide a range of services such as classifying routes based on risk of accidents. However, when sharing vehicular data it is essential that security concerns such as preventing falsified data and protecting privacy are addressed. Overall, we present various approaches of achieving secure and private vehicular clouds that will enable safer and more comfortable vehicular transportation.