Australasian Transport Research Forum 2015 Proceedings
30 September - 2 October 2015, Sydney, Australia
Publication website: http://www.atrf.info/papers/index.aspx

Privacy in the Era of Mobile Sensors: A Case Study on the Data Collection in ADAS and Autonomous Vehicles

Zhitao Xiong, Vinayak V. Dixit, S. Travis Waller

Research Centre for Integrated Transport Innovation (rCITI)

UNSW Australia

z.xiong@unsw.edu.au

Abstract

Ever since the application of probe cars and instrumented vehicles for transport planning or driving behavioural studies, data collection in public road networks have become common. With the availability of Advanced Driver Assistance Systems (ADAS) and the advancing of Autonomous Vehicle (AV) deployment, more and more sensor-equipped vehicles will be available shortly and due to their dependency on accuracy data source, there can be a serious privacy concern on what data will be gathered and how privacy cannot be violated. An investigation is firstly needed to identify how those data can be used, such as surveillance, misbehaviour reporting and unauthorised large-scale data fusion. In this paper, based on the data collected by an instrumented vehicle, we examine the content of the data that can be collected by ADAS and AVs and compare them to other data collection techniques in public, including CCTV at intersections. This paper will finally propose measures to minimise the risk of privacy violation raised by ADAS and AVs.

Keyword: privacy; data exhaust; mobile sensors; ADAS; autonomous vehicles

1. Introduction

An autonomous vehicle requires no or partial human inputs to operate and is believed to benefit not only individuals but also the transport system as a whole. With the continuous achievement in technology, Autonomous Vehicles (AVs) are to appear in our road networks in about 5 to 20 years (Somers & Weeratunga, 2015). This revolution has the potential to influence many aspects of our lives, e.g., how we interact with cars, how our transport system will be managed and how our road networks will be designed.

As some initial attempts, road users have begun to experience some Advanced Driver Assistance Systems (ADAS) that can partially achieve the functionality of an AV. For instance, car manufacturers Subaru and Volvo have EyeSight and CWAB-PD (Collision Warning with Full Auto Brake and Pedestrian Detection) respectively for the market to detect lane and objects in front and perform emergency stop if needed (Hamdane et al., 2015).

As a result, road users can be optimistic about the advantages of those new systems, which can, e.g., enhance road safety. However, they always neglect one fact about those systems, including AVs and ADAS: they are equipped with varieties of sensors, and thus they can collect personal data, which "*shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*" (Directive, E. U.,1995).

In general, ever since the application of probe cars and instrumented vehicles, which are equipped with varieties of on-board sensors for transport planning or driving behavioural studies, (personal) data collection in public road networks have become common. The data collected can help people to measure traffic characteristics such as travel times and thus help us plan our transport systems. However, they can raise some privacy concerns as varieties of information can be recorded, e.g., surrounding vehicles' license plates and pedestrians.

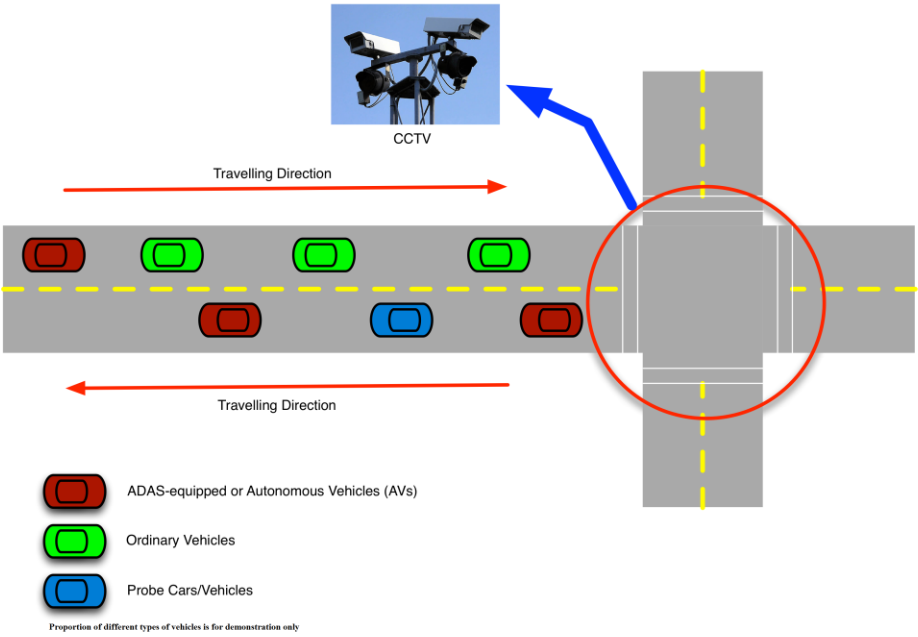
As a result, with the availability of Advanced Driver Assistance Systems (ADAS) and Autonomous Vehicles (AV), more and more sensor-equipped vehicles will be available shortly and due to their dependency on accurate data source, there can be a serious privacy concern on what data will be gathered and how the data will be used. Privacy violation will thus occur when those data are used for purposes beyond their original design.

In this paper, based on the data collected by an instrumented vehicle, we examine what we know about sensors in our travel activities in Section 2, and then in Section 3, we will identify the data that can be specially collected by autonomous vehicles. Based on some discussions in Section 4, this paper will finally propose measures from a technical perspective in Section 5 to deal with privacy violation raised by ADAS and AVs.

2. Background

In present transport networks, we may not be exposed to personal data leak as the collected data can be anonymous and aggregated, such as vehicle counts that show the traffic flow at a specified road/lane (Leduc, 2008). However, with the application of vehicles with on-board sensors and relevant data fusion techniques, personal data can be easily collected and analysed. As illustrated in Figure 1, there are mainly three data collection methods at the moment: fixed traffic monitor, mobile traffic monitor and ADAS & AVs.

Figure 1: Main Data Collection Methods in Travel Activities



Proportion of different types of vehicles is for demonstration only

With the advancing of image processing techniques, CCTV (Closed-circuit television) has been widely used to carry out tasks such as vehicle tracking & guidance (Michalopoulos et al., 1990) and incident detection (Blissett et al., 1993). By utilising the images captured through CCTV, the following information can be gathered: 1) identification of vehicles, such as their license plates; 2) movements of vehicles, such as their lane-changing manoeuvres and 3) traffic flow characteristics such as incidents or congestion.

Moreover, toll tag can be used for data collection purposes rather than just toll management. When a car passes a toll gate, the toll tag will be detected and recorded for tolling purpose with relevant billing information. As a result, the driver of the vehicle that comes with a specific toll tag can be identified. As shown in Hill (2013), such information can be easily used for other purposes: in New York, toll tags can reveal the driver’s location based on receptors installed throughout the city, through which, the relevant agency can better understand traffic flow in real time. Therefore, such fixed sensor technology allows automobile tracking rather than toll management.

Apart from those fixed sensors, mobile sensors are also being used in road networks. Probe cars utilize instrumented vehicles in the traffic to collect travel times. The probe cars can be private or commercial vehicles and report travel times to a transportation management centre in real-time. Probe cars may be equipped with several different types of sensors, which may include the monitor of cellular telephone activity and GPS receivers in order to measure how long it takes for a specific probe car to traverse a certain distance or road.

Another type of mobile sensors come from ADAS and AVs, which collect data for human driver assistance or direct driving decision makings, so they require detailed and highly accurate data such as lane markings, surrounding objects, traffic signs, etc. The task of those devices are to mimic the perception capability of a human driver, so they have to collect data that can be used to reconstruct the surrounding environment. These sensors typically acquire road users' states (such as speed and position), road characteristics (such as speed limit and lane marking) and ego status (such as speed and pedal usage).

With autonomous vehicles present in an urban environment, the DARPA (Defense Advanced Research Projects Agency) urban challenge in 2007 required teams to build autonomous vehicles capable of driving in daily traffic. If we take the Cornell Skynet team as an example, we can see the following information is needed (Miller et al., 2009):

- Map Information for route planning: this was provided by the DARPA Challenge committee based on the ENDF and MDF formats. The former indicates Route Network Definition File, showing all legal lanes that an autonomous vehicle can traverse and the latter indicates Mission Data File, showing all checkpoints that should be achieved in the competition. They were both defined with GPS coordinates;
- Scene representation for driving decision making: this is to reconstruct local map with surrounding objects based on relevant on-board sensors: LIDAR (Light Detection and Ranging), radar and vision;
- Ego status for reference: this is to estimate ego pose (such as position and yaw rate) based on GPS, IMU (Inertial Measurement Unit) and odometer.

Moreover, apart from the aforementioned data regarding driving behaviours, the ones regarding individual features can be also needed, e.g., the drivers may request services ranging from voice activated restaurant recommendations nearby (Strom & Brodsky, 2014) to automated searches for parking spots (Kopecký & Domingue, 2012). As a result, those rich data are of interest not only to a single ADAS device or an AV, but also to some organisations, such as insurance companies, which can use such information to tailor insurance options for individuals (O'Neill, 2015).

It is therefore necessary to examine carefully how those data can be used and what can be done to protect privacy. In the following section, we will examine what data can be collected. Moreover, we assume that the data transmission is secure; in which case, we don't consider any cyber security challenges in this paper raised by the application of any networks, e.g., V2V (Vehicle to Vehicle) communication.

3. Data Exhaust from an Example Vehicular Platform

Mobile sensors in a car, which can be an ADAS-equipped or an Autonomous Vehicle (AV), are used for perception. After the data has been used for any decisions such as a lane-changing carried out by an AV, it can be collected or erased. If the data is collected in the car or any other places, it is called data exhaust, which can be used for other purposes rather than on-board decision making. Privacy violation is therefore based on the existence of data exhaust.

This vehicle of interest, which is a 2012 Toyota Yaris, is being developed for a project within rCITI (Research Centre for Integrated Transport Innovation) at UNSW Australia together with Goget CarShare in Sydney. The aim of this project is to use technology to influence human behaviour and promote safer and more fuel efficient behaviour through incentive mechanisms, and is expected to influence transportation and insurance policies. Most importantly, this project will contribute towards a broader framework for Intelligent Transport Systems that involve autonomous vehicles. As a result, this vehicle is being developed towards full autonomy. Therefore, the first step for developing this vehicle of interest is, without doubt, to make it “perceive” the outside world.

The outside world is represented as driving contexts, which are mainly about road conditions including lane tracks and other road users around when it comes to structured urban environment. Three steps have been taken to build a new autonomous vehicular system: initial sensor deployment for perception, actuation module installation for driving interface manipulation and algorithm development for driving decision making. In this section, we will concentrate on the first part to examine what data are needed for such vehicular system.

Table 1: Driving-related Tasks based on Decision Levels

Decision Level	Activity	Example
Strategic Level	Set goals for driving such as the destination	Plan a route from Sydney to Melbourne.
Tactical Level	to fulfil the goal from Strategic level by driving safely and focusing on the relative positions between a vehicle and other objects in the traffic	Carry out overtaking, lane-changing or lane keeping manoeuvres.
Operational Level	To execute the decisions from Tactical level and attempt to maintain a speed or a safe distance to other road users	Avoid obstacles.

As illustrated in Table 1, data perceived by a car can be used to carry out several main tasks, including strategic, tactical and operational, in order to reach a destination safely with humans in the loop. Based on some easily accessible off-the-shelf sensors, the vehicle of interest has the following sensors to collect data as summarised in Table 2:

Table 2: Data Needed for Driving Tasks

Decision Level	Activity	Generic Sensor Technique	Sensors that have been used
Strategic Level	Set goals for driving such as the destination	GPS and IMU	Nax 5.1
Tactical Level	to fulfil the goal from Strategic level by driving safely and focusing on the relative positions between a vehicle and other objects in the traffic	Camera and radar for lane tracking and gap detection; Fuel sensor for fuel level examination.	IFV, LLS-ARF
Operational Level	To execute the decisions from Tactical level and attempt to maintain a speed or a safe distance to other road users	Radar for obstacle detection.	ESR and RSDS

As illustrated in Figure 2, an *Intelligent Forward View (IFV)* camera from Delphi, is used to track lanes and recognise speed signs. This has been implemented with the in-house development of an offline visualization module to conduct ex-poste evaluations. An *Electronically Scanning Radar (ESR)* sensor from Delphi is used to provide detailed velocity and position information of road users in front, while *the Rear and Side Detection System (RSDS)* from Delphi, is used to detect vehicles in adjacent lanes and in the rear of the vehicle, especially in blind spot regions. *Nax 5.1*, an IMU /GPS unit from Navexlorer4 is a MEMS (Microelectromechanical systems) unit that collects data regarding the dynamics of the vehicle of interest. In addition, there is also a fuel sensor (*LLS-ARF* from Omnicomm) that is being used to record fuel consumption data. The integration of the sensors is shown in

Figure 3. A mobile computer is also needed to coordinate all sensor systems and record all sensor data as well as initial data filtering and processing.

Figure 2: Hardware Setup of the Instrumented Vehicle of Interest

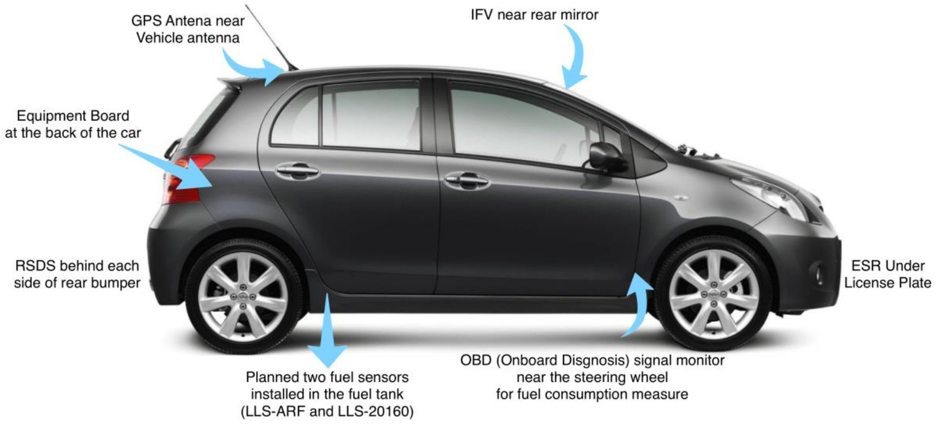
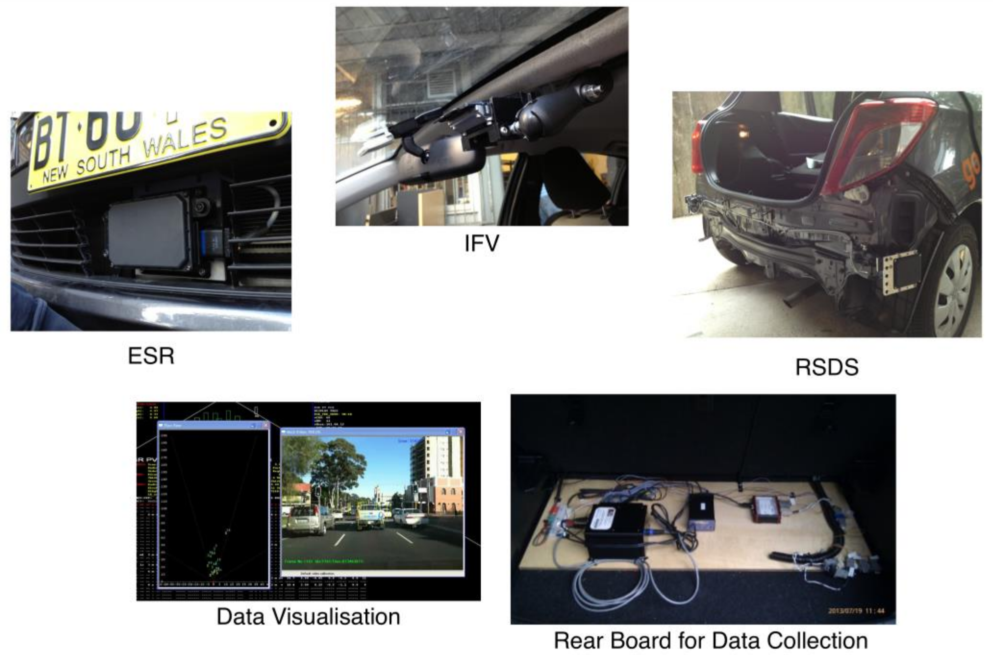


Figure 3: Sensor Setup and On-board Mobile Computer



All sensors are integrated in the mobile computer using *CAN* bus (*ESR, IFV* and *RSDS*) or serial port (*Nax 5.1, LLS-ARF*). A program, written in C++, has been hosted in the mobile computer to control all sensors and perform data collection and analysis. As a result, the data exhaust that can be collected in this system are shown in Table 3.

Table 3: Data Exhaust Generated by the Vehicle of Interest

Sensor	Range	Data Available	Visual Image
GPS and IMU unit	NA	Positions and postures of the vehicle of interest in real-time	No
IFV	45-degree horizontal field of view 29-degree vertical field of view	Lane marking, speed limit, and status of road users (type and the position in an image)	Yes
ESR	120 m with 64 tracking targets maximum	Information regarding objects in front (speed and position). The data can be fused with the one from IFV, such as associating the type of road user with its speed, lane occupancy and relative position.	No
RSDS	30 m and 80-degree range	Obstacle information in blind spots and behind the vehicles of interest.	No
LLS-ARF	NA	Real-time fuel consumption reading	No

4. Discussion

As show in Table 4, it is clear that with more and more mobile sensors in the road networks, accurate and real-time data can cover bigger areas with richer information compared to fixed sensors. Data mining for surveillance, misbehaviour reporting and large-scale data fusion, can be then carried out. There are thus concerns regarding unauthorised usage of the aforementioned services, such as unauthorised large-scale data fusion.

Table 4: Comparison between Different Data Collection Methods

Data Source	Example	Real-time	Level of Detail	Real-time Data Fusion	Person Identifiable?	Areas covered
Fixed Sensors	Intersection CCTV	Yes	Low	No, but possible	Yes, but with low resolution images	Small, local coverage
Mobile Sensors	Probe Cars	Yes	Medium	No, but possible	No, if it is designed for travel time collection	Medium, local coverage
ADAS and AVs	ACC and Google Cars	Yes	High	Yes	Yes, with high resolution images	Large, unrestricted coverage

First of all, large scale data fusion is very common as research topics, because those data can give insights into people's behaviours, such as short-term trip destinations, long-term driving behaviours and individual features (e.g., personal background, personality and risk attitudes). As a result, the following two scenarios may happen:

- A driver just received a specific advertisement regarding a car insurance offer, which was pushed to her phone based on the following information: 1) her short-term destination (a NRMA branch); 2) her personal background, e.g., she had three demerits and 3) her driving skill, e.g., her driving history shows that she had 10 recorded speeding offences and an unstable lane keeping profile;
- A driver was contacted by a transport agency, because he/she had one speeding offence in the last month according to the data recorded by the black box in his/her car. That black box was originally from the agency and records videos, speed and GPS coordinates while driving.

Basically, as shown in the two scenarios, data fusion have advantages in, e.g., improving road safety, however, due to the presence of personal data, they can also lead to serious privacy violation with respect to unauthorised tracking and surveillance.

What makes things worse is the availability of image-based sensor data with geo-tags from GPS. It is therefore possible to obtain multi-facet information from those mobile and fixed sensors, including 3D models of surroundings and personnel exposed in each image. Even without time stamps and geo-tags, it is still possible to obtain such information based on image processing techniques.

As a result, with more and more data from varieties of sources, especially those from on-board sensors, privacy violation is very hard to avoid.

5. Conclusion

To sum up, with the advancing of relevant techniques (image processing and data mining), privacy violation is very hard to avoid with more and more data available from different sources. Their applications have great potential to influence not only low-level individual activity but also high-level strategic planning. As a result, when it comes to the solution for such potential privacy violation, the first step we can try is to minimise the risk of such violation, as we never came across such issue before and it may take time for the government, industries and individuals to understand this era of mobile sensors. As a start, the following measures can be taken to help people understand and thus minimise the possibility of privacy violation by making relevant information or data accessible, anonymous, public and law-protected:

- Standardise and open the format for data storage and transmission in order to identify what have been used in real-time and what have been collected as data exhaust;
- Erase any times stamps, geo-tags, IDs in the data stored, which tries to make people unidentifiable;
- Provide relevant people details of those sensors and data recorded, by carrying out mass advertisement, report, training and education;
- Identify law and security liabilities;

As a result, future work will concentrate on examining the aforementioned measures by carrying out relevant surveys (for item 3), case studies (for item 4) and algorithm development (for item 1 and 2). Moreover, progress is also being made to discuss the measures based on the guides proposed by the Office of the Australian Information Commissioner (OAIC).

References

Somers, A., & Weeratunga, K. (2015). Automated vehicles: are we ready? Internal report on potential implications for Main Roads WA.

Hamdane, H., Serre, T., Masson, C., & Anderson, R. (2015). Issues and challenges for pedestrian active safety systems based on real world accidents.*Accident Analysis & Prevention*, *82*, 53-60.

Directive, E. U. (1995). 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the EC, 23(6).

Leduc, G. (2008). Road traffic data: Collection methods and applications. Working Papers on Energy, Transport and Climate Change, 1, 55.

Michalopoulos, P. G., & Wolf, B. (1990). Machine-vision system for multispot vehicle detection. Journal of transportation engineering, 116(3), 299-309.

Blissett, R. J., Stennett, C., & Day, R. M. (1993). New techniques for digital CCTV processing in automatic traffic monitoring. In *Vehicle Navigation and Information Systems Conference, 1993., Proceedings of the IEEE-IEE* (pp. 137-140). IEEE.

Hill, K. (2013). E-ZPasses Get Read All Over New York (Not Just at Toll Booths). *Forbes. Forbes*, *9*.

Miller, I., Campbell, M., Huttenlocher, D., Nathan, A., Kline, F. R., Moran, P. & Fujishima, H. (2009). Team cornell's skynet: Robust perception and planning in an urban environment. In *The DARPA Urban Challenge* (pp. 257-304). Springer Berlin Heidelberg.

Strom, D., & Brodsky, I. (2014). Good Food and Drink and Connected Technology.

Kopecký, J., & Domingue, J. (2012). ParkJam: crowdsourcing parking availability information with linked data.

O'Neill, M. (2015, March 10). Data retention: AAMI Safe Driver app could see information handed to police, premiums go up. Retrieved June 1, 2015, from http://www.abc.net.au/news/2015-03-10/aami-safe-driver-app-data-retention/6292198