Towards an AI-Based After-Collision Forensic Analysis Protocol for Autonomous Vehicles

Conference Paper · May 2020

DOI: 10.1109/SPW50608.2020.00055

3 authors, including:

Prinkle Sharma
University of Massachusetts Dartmouth
6 PUBLICATIONS **139** CITATIONS

SEE PROFILE

Towards an AI-Based After-Collision Forensic Analysis Protocol for Autonomous Vehicles

Prinkle Sharma, Umesh Siddanagaiah, Gökhan Kul
University of Massachusetts, Dartmouth
{psharma1,usiddanagaiah,gkul}@umassd.edu

***Abstract*—Safety-critical applications in the cooperative vehicular networks are built to improve safety, traffic efficiency and handle emergencies by communicating the road condition captured using data from sensors (camera, LiDAR, RADAR, etc.). These cyber-physical systems maintain records of the data received from its sensors to make decisions while driving on road. Such proliferation of data opens possibilities of scenarios where attackers can forge into the system with unrestricted access to the internal network of the vehicle and perform malicious acts. Due to the possibility of such acts, it is crucial how forensic analysis should be carried out in case of traffic accidents that include autonomous vehicles (AV).**

In this paper, we propose a forensic investigation protocol on autonomous vehicles, specifically to investigate if there was an attack that targeted the vehicle sensors. The proposed process consists of three main phases: data curation, analysis and decision making. We argue that, by using supervised deep neural network-based architecture YOLO trained in the Darknet framework and tested with SORT, an effective model to detect traffic data can be built to perform forensic investigations.

***Index Terms*—Autonomous vehicle, deep learning, digital forensics, security, sensors**

I. Introduction

Autonomous cyber-physical systems combine the physics of motion with advanced algorithms to ensure safety, privacy and improved experience without any close human supervision. Autonomous vehicles (AV), an application of cyber-physical systems, are gradually preparing the consumers for the time where they relinquish control of their vehicles. Even though driverless cars are still at the advance testing stage, partially automated technology has been around for half a decade [12]. Some manufacturers even produce cars capable of driving without any driver intervention, but prefer to brand these capabilities as *driver assist technology* due to regulations [6].

To monitor the driving environment and warn the driver of immediate dangers, AVs rely on several sensors and actuators. With the advances in sensing technologies and information fusion, the transportation industry is moving forward into the era of full autonomy (Level 5) as defined in standard SAE J3016. Unfortunately, cyber-attacks have become one of the major threats in today's IoT world, and AVs are no exception [9]. Although millions of dollars are invested by the industries for improving the robustness and accuracy of the sensors in AVs, the security threat still persists. Instances like the death of a woman bicyclist in Arizona USA, on road at night, when hit by Uber Self-Driving car [18], and successful spoofing of vehicle's LiDAR system showcased by researchers in BlackHat Europe 2015 [10] demonstrate that existing autonomous vehicle sensors cannot be completely trusted even on normal road conditions.

Sensors inside the AV help in three major tasks that include: Navigation and guidance (GPS, road maps), Driving and Safety (LiDAR, RADAR, Camera) and performance (On-Board Units, Wheel encoders, etc.) [14]. The safety of these sensors has been a focus of the prolonged debate over this technologyAttacks like sensor blinding, misidentifying the object and falsifying the sensors' wrong traffic light sign resulting in false driving actions could be life-threatening [9].

In this work, we outline an AI-based forensic analysis protocol for traffic accidents involving AVs to detect if the sensors of the vehicle is attacked. Our work utilizes the extensive research on the security concerns of sensors equipped in state-of-the-art AVs and digital forensics know-how on IoT systems. The protocol works as follows: (1) it curates visuals from storage and memory devices that may be damaged in the accident, (2) analyzes the accident through a supervised deep convolutional neural network model, and (3) identifies anomalies in curated data for law-enforcement and third party experts. Experts, then, use these anomalies as evidence of foul-play, if there is any.

The rest of this paper is organized as follows. The life cycle of sensor data inside AVs is discussed in Section II. Section III introduces the system and attacker model. The details of the protocol is given in Section IV. In Section V, we discuss the possible validation scenarios. Finally, Section VI presents the future work.

II. Sensor Data Lifecycle in Autonomous Vehicles

AVs rely on numerous sensors to ensure passenger safety. To achieve high accuracy, Original Equipment Manufacturers (OEMs) continuously develop and enhance the reliability of the perception systems. However, there have been several instances in the past where attackers were successfully able to fool the sensor systems and control the vehicle [10]. To achieve reliable automation, it is crucial to accurately create a threat model that defines the attacker profile, possible motives, and the system components that they can target. This section identifies the major assets within AVs and data fusion life-cycle.

Figure 1 presents the process of how sensor functions and data collection in AVs as represented in Petit *et al.* [10]. Sensor Data Life-cycle can be categorized into three parts: (1) Sense,

DOI 10.1109/SPW50608.2020.00055

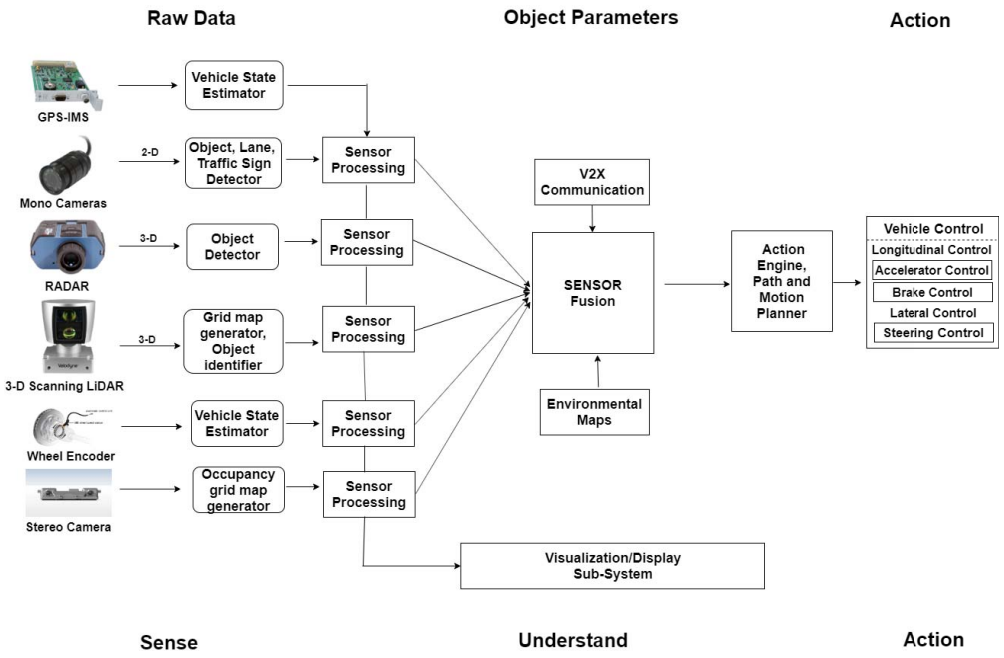

Fig. 1: Autonomous Car Sensor and Control System

(2) Understand, and (3) Action. The phases of the life-cycle are as follows.

- **Data Collection:** This is the initial state of data in AVs. Several sensors contribute to the generation of data using light intensity, radio waves, sound waves, etc. The data from several sensors is collected in raw format.
- **Data Processing:** This stage requires the manipulation of raw data collected from different sensors to machine-understandable data. The acquired data is encoded/decoded as per the system requirements.
- **Sensor Fusion:** This is one of the crucial steps for autonomous vehicles. Sensor fusion aims to precisely understand the environment and the objects as perceived by different sensors. Eg: a Camera is a great tool for detecting roads, reading signs and recognizing an object. LiDAR is better at estimating the position of the object and RADAR is good at estimating the speed of the same object. It can also be referred to as Decision Maker.
- **Vehicle Control:** This stage provides the final step after decision making. Once the AV understands its surrounding environment based on data received from several sensors, the final action (brake, speeding, turning, etc.) is made.

III. Threat Model

Cyber-physical systems (CPS) typically consist of networked embedded systems that are used to sense, actuate and control physical processes. The physical layer aspects of such systems are subject to novel attack vectors, and as well as providing opportunities for defenses that require advanced attackers' capabilities to penetrate.

As described in Figure 1, AV sensor infrastructure comprises of three phases, i.e., 'Sense', 'Understand', and 'Action'. After collecting raw sensor data from its surroundings using multiple sensors, AV generates an image of the environment by fusing the data giving directions to 'Action Engine'. In this paper, we mainly focus on the 'Sense' part. Similar to the case of the human body, inhaling polluted air leads to lungs/breathing problem, feeding/collecting bad or fake data leads to improper AV functioning.

For this work, we consider the attacker as an external entity that targets the sensor data acquisition. Therefore, in our work, we focus on **sensor jamming** that operates on the physical layer, corrupting the sensor data quality. The attacker focuses on creating an attack to provoke an accident, disrupting road traffic, risky lane changing and/or controlling the car remotely.

Criminals use physical-layer cyber-attacks to steal autonomous vehicles and their freight, cause crashes, or imperil passengers and pedestrians. In a Sensor Jamming Attack, the attacker spoofs the autonomous vehicle sensors by blocking the access to the vehicle for a short term with an unknown object. Performing such an attack blindly may not be enough to deceive machine learning models. However, there are more advanced approaches that can successfully deceive existing systems [2].

IV. Methodology

As stated before, we have divided the methodology into three phases: data curation, analysis and decision making.

A. Data Curation

Digital forensics is a branch of forensic science concerned with the use of digital information produced, stored and

transmitted by computers as a source of evidence in investigations and legal proceedings. Recently specialized analysis software systems have been made available for both the private and public sector users [8]. For example, Feng *et al.* [5] has used digital forensic methods on an AV to identify vehicle diagnostics to understand if a malfunction caused an accident. However, accessing and preserving evidence is one of the first obstacles in IoT and vehicular technology [3].

There are two types of devices that we can curate our data:

- **Memory:** In-memory data collection requires the computer system to be continuously running after the accident. However, in-memory data can reveal information on the processes running on the CPU, passenger activity, any malfunction, and many more. Furthermore, it is possible to access sensor information such as cameras and LIDAR on some operating systems (OS) such as Android [13]. Even if the onboard computer is turned off, it may also be possible to find snapshots of the memory by default on some OSes. Since the OS used on an AV is vendor specific, the forensic investigator would need to check the vendor specifications.
- **Storage Devices:** Extracting information from HDD and SSD devices is a common practice and can be achieved through a variety of software and hardware. Moreover, data curation from these devices may be performed even if the storage device gets damaged in the event of an accident depending on the extent of the damage. We also expect AVs to be equipped with black boxes.

In our case, we are using CARLA [4], an open-source simulator, to generate data and demonstrate the effectiveness of the model visually and quantitatively.

B. Data Analysis

In the second phase, after collecting data, the analysis of data is performed. There are two sub-phases: The forensic investigator should (1) build the AI model using real-world normal sensor information, and (2) identify the relevant curated data through metadata or visual analysis.

The first phase can be standardized with a sensor data corpus. With such a corpus, the investigators would not need to retrain a new model for each incident. Two aspects of any AI-model should be considered carefully: (1) designing an appropriate architecture, and (2) choosing the right learning algorithm. In this case, we have data collected from the CARLA simulator in the form of images (frames) generated at the rate of 10Hz. Since we are using (Supervised) Deep Learning Model to identify the attacker and benign vehicles, data needs to be labeled. We have used a tool named **LabelImg** [17] for data labeling. LabelImg is a graphical image annotation tool that labels multiple objects in a single image, classify and label them (in our case as benign or attacker) with their location in each frame (as a bounding box).

For our work, we have a small dataset of around 1225 labeled images with around 1000 instances containing benign vehicles (82%) and 225 images with attacker vehicles (18%). The training and testing ratio for this work is 80:20 for the complete dataset. For object detection, we have used YOLO (You Look Only Once) [11] algorithm. YOLO reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities by just looking at an image once during the whole training and testing phase. The architecture of the YOLO network has 24 convolutional layers with 2 fully connected layers to reduce the features space from preceding layers. By default, we have pre-defined the images to 416px x 416px square sizes. Further, we use Darknet [16], an open source neural network framework that has pre-trained configuration and weights, to train YOLO by providing it the architecture of the network. After training the CNN based YOLO network with 80% of the data, the Darknet framework provides the label of the object, detection confidence level and the overall time consumed for making the decision.

In the simulation, we have multiple vehicles on the road. This means that we can have both benign and attacker vehicles captured in a single frame and it is important to know if a vehicle in one frame is the same as one in a previous frame. This is also known as object tracking and requires multiple tracking to identify a specific object over time. To address this, we have used SORT [1] algorithm that uses a Kalman Filter to predict the traces of previously identified objects. Authors in [15] have demonstrated the successful use of Kalman Filtering techniques for tracking purposes. SORT focuses on frame-to-frame prediction and association for timely tracking of the objects in a frame.

C. Decision Making

The last phase utilizes the trained model to determine if the sensor data leading up to the accident includes any anomalies. The outputs of this phase are then evaluated by experts to determine if the anomaly is a product of foul-play.

In the next section, we describe 3 test scenarios that we can measure the usability of the model. We acknowledge that the machine learning models should be further evaluated with various machine learning metrics including Confusion Matrix, F-measures, Detection Accuracy and Detection Rate. We will further investigate and release our evaluation.

Many autonomous systems follow the OODA loop, which is the cycle of *observe-orient-decide-act* [7]. Another verification technique we can use is to follow the events that led to each decision that the AV made based on its OODA loop prior to collision from the logs.

V. Test Scenarios

CARLA is an open-source simulator for autonomous driving research and has been developed from the ground up to support the development, training, and validation of autonomous driving systems. We conduct three test scenarios in Carla. We also released the code on GitHub for our experiments [1]. Figure 2 includes an instance from Carla simulator demonstrating all the features equipped in the vehicles.

[1]https://github.com/PrinkleSharma/WAAS2020

Fig. 2: Carla Simulator view

We consider three scenarios:

- Autonomous normal mode driving behavior
- Accidents not involving foul-play
- Fooling sensors via the sudden appearance of another vehicles/obstacles or vanishing objects
- **Autonomous normal mode driving behavior:** The main objective here is to observe the difference between the driving behavior of autonomous and manually driven cars on road traffic. In this scenario, the input data is supplied by the various sensors of an AV.
- **Accidents not involving foul-play:** In such a scenario, the aim is to collect information about accidents where there is no sensor malfunction. This data specifically helps the model to distinguish features that leads to accidents.
- **Fooling sensors via the sudden appearance of another vehicles/obstacles or vanishing objects:** While driving, an autonomous car senses its environment and identifies objects around it. However, when sensors have tampered, they tend to miss the objects/cars around it, causing violation of traffic rules and in the worse case causing accidents. For example, an autonomous car whose sensor is tempered with sensors sees an object 100 meters away and while nearing, the object may disappear to the car sensors.

VI. Future Work

Sensor spoofing can severely impact the decision making capability of AVs, both the AV and the network of AVs. The authors are currently working on a comprehensive attack detection framework that accompanies the proposed protocol. Future work includes work focusing on identifying the specific sensor that is tampered during the attack. The results and analysis will be demonstrated in the upcoming publications.

References

[1] Alex Bewley et al. “Simple online and realtime tracking”. In: *IEEE ICIP*. 2016.

[2] Yulong Cao et al. “Adversarial Sensor Attack on LiDAR-Based Perception in Autonomous Driving”. In: *ACM CCS*. 2019.

[3] Mauro Conti et al. “Internet of Things security and forensics: Challenges and opportunities”. In: *Future Generation Computer Systems* (2018).

[4] Alexey Dosovitskiy et al. “CARLA: An open urban driving simulator”. In: *arXiv preprint arXiv:1711.03938* (2017).

[5] X. Feng, E. S. Dawam, and S. Amin. “A New Digital Forensics Model of Smart City Automated Vehicles”. In: *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. 2017.

[6] Steven Goodridge. *Autonomous Driving and Collision Avoidance Technology*. https://www.bikewalknc.org/2018/02/autonomous-driving-and-collision-avoidance-technology/. 2018.

[7] Hui-Min Huang et al. “Autonomy levels for unmanned systems (alfus) framework: An update”. In: *Unmanned Ground Vehicle Technology VII*. Vol. 5804. International Society for Optics and Photonics. 2005, pp. 439–448.

[8] Ravneet Kaur and Amandeep Kaur. “Digital Forensics”. In: *International Journal of Computer Applications* 50.5 (2012).

[9] Jonathan Petit and Steven E Shladover. “Potential cyberattacks on automated vehicles”. In: *IEEE Transactions on Intelligent Transportation Systems* 16.2 (2014), pp. 546–556.

[10] Jonathan Petit et al. “Remote attacks on automated vehicles sensors: Experiments on camera and lidar”. In: *Black Hat Europe* (2015).

[11] Joseph Redmon et al. “You only look once: Unified, real-time object detection”. In: *IEEE CVPR*. 2016.

[12] AutoPilot Review. *Cars with Autopilot in 2019*. https://www.autopilotreview.com/cars-with-autopilot-self-driving. [Online; accessed 19-September-2019]. 2019.

[13] Brendan Saltaformaggio et al. “Vcr: App-agnostic recovery of photographic evidence from android device memory images”. In: *ACM CCS*. 2015.

[14] Bill Schweber. *The Autonomous Car: A Diverse Array of Sensors Drives Navigation, Driving, and Performance*. https://www.mouser.com/applications/autonomous-car-sensors-drive-performance/. 2018.

[15] Prinkle Sharma et al. “Securing wireless communications of connected vehicles with artificial intelligence”. In: *IEEE HST*. 2017.

[16] Trinh Hoang Trieu. “Darkflow”. In: *GitHub Repository. Available online: https://github. com/thtrieu/darkflow* (2018).

[17] D Tzutalin. *Labelimg*. 2018.

[18] Daisuke Wakabayashi. *Self-Driving Uber Car Kills Pedestrian in Arizona Where Robots Room*. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html. [Online; accessed 30-September-2019].