Autonomous Vehicles: Sensor Vulnerabilities, Attack Vectors, and Solutions

Submitted by: Ryan C. Barron
Masters Student, Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, United States of America
ryanb4@umbc.edu

Submitted to: Gerald S. Tompkins
Professor, Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, United States of America
gerald@umbc.edu

Abstract—**The demands of transportation by way of wheeled vehicles on roadways is increasing as society becomes increasingly digital. More people and products become connected and obligated, then subsequently must be moved to meet those obligations. As demand to move those people and goods outpaces reasonable prices, automation becomes the apparent solution. First and foremost is the concern of safety for the people in and around the automated vehicles. The first ring in safe operation of autonomous vehicles is decision making from the presented data, usually represented by a machine learning model. Even if the decision-making process of the car is impossibly perfect given input data, the data itself comes from the environment, which means it is an external attack vector. Such attacks are made on input data to the autonomous vehicle's sensors and must be thoroughly examined and cleaned of malicious manipulation in order to ensure the safe operation of the autonomous vehicle in public spaces. This can be done through a panel of machine learning models to label data on specific attributes, flagging malicious data. A second solution is adding multiple sensors from different angles and depths to verify the veracity of incoming signals and observations. A third solution is to build infrastructure into the roadways to act as supplemental sensors to make observations on behalf of the car. A fourth solution is to allow inter-car communication as a form of environmental input verification.**

I. Introduction

As the capabilities of computing increase with time, a greater ability to perform tasks in the real world also increases. One such task is the automation of transportation, specifically the passenger vehicle. Throughout the papers, this is referred to as Advanced Driving Assistance Systems (ADAS). To make automated driving possible, a car must be mounted with sensors or have built-in embedded sensors. These sensors take in the readings of the surrounding world, then synchronize internally to create a unified world view. This world view of the immediate environment is then processed through machine learning models such as object detection to find useful information with which to make decisions. With the environment details, the car can process the processed information through a decision-making program to decide things such as speed, braking, navigation paths, etc. While hacking into the physical systems of the car is difficult, it is more simple to influence the sensor data read by the car prior to entering the protected systems of the car– that is, sensors read from the environment, so manipulating the environment can be the avenue for an attack on an autonomous vehicle.

II. MOTIVATION

As the population of society increases, more and more people will require transpiration to get to their work or have things shipped to them such as food, and other quality of life items. Transportation to professions causes traffic jams, which incentives public transportation. People, however, in more remote locations of the world may not have ready access to public transportation and still need to get to work. Similarly, the population sends items such as food through shipping trucks. While the truck drivers are valuable members of society, it is more beneficial for society to provide the transportation of goods through automation, and train the would-be truck drivers of the next generation in more fulfilling Jobs, for themselves and for society, As more people use automated cars, the greater the incentive there is for attackers to find weaknesses in the system, particularly when high-profile people such as celebrities and political figures begin using them. Having secure systems will ensure all of the above-mentioned people and goods will be safe.

III. CHALLENGES

A. Cost

With many state-of-the-art systems, implementing new technology will have higher costs than older implementations, if any exist. With automated driving, the costs of making them more able to detect issues are related to adding either software in the form of models, hardware in the form of sensors, or systems in the form of Similarly, the ADAS are usually not open source, so researchers have to take data samples on the input-output to make predictions about the true operation of the internals of the systems. This means the systems must be purchased and observed to discover some of the underlying attributes rather than referring to a repository as is the case for other major societal undertakings.

B. Unified Regulation

Given that automated driving systems are on the cusp of modern innovation, they will not be regulated like many of the long-known functions of society, such as the regulation of wall street, or the regulation of conduct for law enforcement, or even for the regulation of human drivers. As such, the safety requirements for automated driving are only just beginning to emerge. Given that drivers have the ability to freely drive across state lines, placing the regulation of the vehicles under interstate travel, and under the purview of congress, the states usually defer to guiding regulation from congress, however, congress attempted to pass regulation in 2017, called the H.R.3388 - SELF DRIVE Act, where it passed the House and then was given to the senate's Committee on Commerce, Science, and Transportation [7]. The act then failed due to the following members: Tammy Baldwin (Wis), Richard Blumenthal (Conn), Roy Blunt (Mo), Maria Cantwell(Wash), Shelley Moore Capito (WVa), Ted Cruz (Texas), Tammy Duckworth (Ill), Deb Fischer (Neb), Cory Gardner (Colo), Maggie Hassan (NH), Dean Heller (Nev), James M Inhofe (Okla), Ron Johnson (Wis), Amy Klobuchar (Minn), Mike Lee (Utah), Ed Markey (Mass), Catherine Cortez Masto (Nev), Jerry Moran (Kan), Bill Nelson (Fla), Gary Peters (Mich), Brian Schatz (Hawaii), Dan Sullivan (Alaska), Jon Tester (Mont), John Thune (SD), Tom Udall (NM), Roger Wicker (Miss), Todd Young (Ind) [6]. The delay is increasingly pushing the responsibility of regulation to the states, exacerbating the fractured expectations on developers and car producers in the United States [7].

IV. Vulnerabilities

An ADAS is a system that is mounted to a car to perceive the world. The perceptions of the world come through sensor data, such as LiDar, monocular cameras, sonar, Inertial Measurement Units, among others. The car then must go through a process of sensor fusion to synchronize all of the data. Some of the data overlap in perception, such as when the monocular cameras read in front of the car, and then the LiDar scanner also reads the same location. this allows there to be a form of verification of the sensor input data over the region. However, not all of the car's information will be verifiable with all types of sensors, and the data may even conflict in the cases of a malicious attack, as the range sensor may not be able to determine the existence of a phantom, but cameras can. the environment through systems such as Synchronous Localization and Mapping Systems (SLAM) in robotics.

Nearly all of the attacks in the following sections are prepared and constructed using a type of neural network machine learning model called convolutional neural networks. A convolutional neural network will start with its training image and a kernel. The kernel will be a subarea of the image size and have a stride to move across the image, which forming the convolution– that is, the image is convolved over with the predefined kernel and aggregates the results in pooling layers. The process allows information to be extracted, relating to the definition of the values into the kernel. At the end, the convolutions are flattened and then passed to linear layers where determinations about image contents can be decided, such as classification [13].

A. Phantoms

The first type of attack is a phantom attack or an image that does not actually exist in the physical reality the car is in. This type of attack creates a ghost that monocular cameras will perceive and then process. This attack can be produced by using a projector, either placed on the road near existing signs, or the projector may be mounted on a drone and flown remotely to the attack location [1][3]. The latter version, mounting to a drone, allows the attacker to conceal their identity more effectively.

1) Attack: The mounted projectors display an image of various objects to elicit specific behaviors. Some of the objects demonstrated have been road signs, such as stop signs and speed limit postings, a person, or lane markings. These phantom objects fool the cars into stopping on highways, swerving into oncoming traffic, or changing to dangerous speeds[1][3].

2) Suggested solution: Since humans are able to make distinctions with the naked eye to determine the difference between an actual physical sign and a phantom sign, so too would a machine learning model, or a series or them in tandem, be able to as well. A combiner model of convolutional neural networks can e paired to verify specific aspects of the information received. In the paper, "Phantom of the ADAS: Phantom Attacks on Driver-Assistance Systems" by Nassi et al., the combiner model was selected such that one model examined the context of the image, another model examined the surface of the sign, and a third examined the lighting. The models scored and then rendered a result of real or false, and decisions were made based on the findings. [3] Additional solutions suggested to phantom signs is the use of QR codes on each sign as something the car could scan and verify as it is driving to verify the location and veracity of the message on the sign, 3-dimensional signs, which could not be projected, or navigational applications with navigation built into them, such as Waze. [1].

B. Perturbations

A printed perturbation is an image of a traffic sign that has been run through a machine learning model in order to produce changes such that the autonomous vehicle system would recognize the sign as different from that of the original sign, and people see the changes as sun spots or wear [2].

1) Attack: The perturbations can be printed out and placed over regular signs, or displayed from a TV. The car will mistake the perturbations for a higher speed in slow sections, or slow speeds in high-speed sections [2]. The attack of printing the signs out and displaying them onto a track in "Fooling a Real Car with Adversarial Traffic Signs" by Morgulis et al. found that 40% of the signs they trained were successful in attacking the systems, with one of the most extreme being the change from a sign a human would see as 30 km/hr to 80 km/hr perceived by the car.

C. Injected Signals

Radio-frequency identification Sensors have actuators that reach out to probe the energy in the environment, which then come back to the sensor to be read as a signal. This projected signal can be listened to by attackers. The signals are electromagnetic sine waves that can be counteracted by other signals, as is the case in a malicious attack[4].

1) Attack: There are three attacks: listening, blind injection, and canceling the actual signal, and then uses a calculated signal to make the car believe a different reality is true. The first attack is achieved through a sensor to listen to the car's signals, the second through an actuator to send a signal, and the third is used through a sensor to detect the car's signal and an actuator to send the new signal back. [4]

2) Suggested solution: In "PyCRA: Physical Challenge-Response Authentication For Active Sensors Under Spoofing Attacks" by Shoukry et al., the Physical Challenge-Response Authentication is a method to issue a challenging phase of the signal, then a confusion phase, and finally end with a silent phase. The attacker will not be able to distinguish the phases fast enough to be able to hide their attack, making their signal apparent to the autonomous car's sensor. The first step, the challenge signal is the standard actuator signal. The confusion phase exists because the malicious sensor will mistake the signal as noise since all sensors have to account for noise. The attacker will then keep their signal going when the car sensor enters the silent phase due to the confusion, enabling the car to detect the attack.[4]

D. Camouflage Sign Stickers

Small stickers can be attached to surfaces, particularly road signs in a way that obstructs part of the signs actual message.

1) Attack: The camouflage stickers work as perturbations, and are calculated through a convolutional neural network machine learning model to find obstructive patterns in order to achieve a mask such that human will not perceive the disruptive symbols maliciously, but the autonomous cars will interpret the stickers in conjunction with the existing sign as an alternate meaning. The stickers are actually placed onto the sign, and so will have realistic lighting as well as context [5].

2) Suggested solution: Since the sensors are reading the environment, and sensors account for noise, the stickers on signs will produce an amount of irregularity that will be less than the tolerable noise. As a result, detecting the noise of the physical objects as protection alone in sensors is not enough to prevent attacks on autonomous cars , and should be avoided as a preventative measure [5].

V. THREAT MITIGATION

Security of the autonomous cars in the transportation systems is not a small undertaking, but nonetheless, solutions exist. Several of the following measures may be taken to mitigate the risk of succumbing to a malicious attack.

A. Immediate solutions

For immediate solutions, there are several options that developers of the vehicle system can take. First, increase the number of sensors so that more angles, distances, and time differences in the signal can be used to cross references incoming signals. Having more sensors will make some of the equipment redundant, but it will provide increased requirements to attack. Given that each additional sensor will have a definite cost to the car, the solution may be costly when many sensor systems, such as LiDar, cameras, sonar, and other systems are installed on the vehicle. Multiple sensors are a solution to the manipulation of a sensor because while one sensor may pick up the fraudulent signal and read it as valid, the second or third sensor built-in for redundancy will be able to read the signal from a different angle and verify the veracity the integrity of the signal if true. The second direct solution is to implement the combiner machine learning model on the sensor data to verify multiple aspects of all of the incoming data in order to make the best judgments about real and fake objects and threats in the real world. This in itself is an additional data verification, and it inspects the quality of the object data, whereas more sensors increase the quantity of in-taken data. In order to make a well-rounded system of safety on the roadways, there should be a federal Autonomous Driver Assistance System Inspection for all interstate cars as part of the U.S. Department of Transportation, and similar state inspection taking guidance from the U.S. Department of Transportation for intrastate cars. The inspection should have a battery of tests such as the perturbations, phantoms, and signal injection, such that only cars implementing a combination of the above solutions will be able to pass and safely drive on the roadways.

B. Long-Term

After the internal systems have been optimized for the safety of the travelers, an additional layer of safety can be built into the infrastructure itself through the Internet of Things and blockchain. For instance, the Algorand block-chain has a project called planet-watch that uses air sensors to report air quality to the blockchain, which is incentivized by planet tokens, which can be bought and sold [9]. The tokens encourage the community to participate in monitoring the air quality, and up-keep of the sensors. Similarly, the readings the helium block-chain incentives localized internet hot-spots by providing helium tokens for operating internet nodes [14]. In a sense, the operation of the internet node, or air sensor work in the same way as mining hashes in proof of work blockchains, such as Bitcoin. In a similar way to the air sensors and Internet hot-spots, road sensors could be placed along the roadways to publicly owned and verified nodes using the Algorand block-chain, which would combine the helium concept of of direct funding for public and local ownership, and the Planet sensor project's method of low-cost decentralized reporting. The sensor readings are publicly available because the Algorand blockchain is a public ledger blockchain [8]. When sensors are placed along the roadway, automated readings would be placed on the block-chain and could be accessed through API calls to the blockchain through AlgoExplorer [11], or vector Software Development Kit, for connecting Internet of Thing devices devices to blockchains [12] in the car in real-time, and verify the signals through authenticated public sensors, in conjunction with the internal sensor attack remediation steps described in the preceding subsection.

Acknowledgment

Instruction on autonomous vehicles and sensor manipulation were explained and expounded on by Wei Wang in conjunction with a project for CMSC 611 - Computer Architecture with Maksim Eren and Cahrels Varga to build an autonomous vehicle sensor fusion system of autonomous cars, called pyCarDisplay [10] in the Python Package Index.

References

[1] Dudi Nassi, Raz Ben-Netanel, Yuval Elovici, Ben Nassi, “MobilBye: Attacking ADAS with Camera Spoofing”, 2019, https://arxiv.org/pdf/1906.09765.pdf, [Online; accessed 23-April-2021]

[2] Nir Morgulis, Alexander Kreines, Shachar Mendelowitz, Yuval Weisglass, “Fooling a Real Car with Adversarial Traffic Signs”, 2019, [Online; accessed 23-April-2021]

[3] Ben Nassi, Dudi Nassi, Raz Ben-Netanel, Yisroel Mirsky, Oleg Drokin, Yuval Elovici, “Phantom of the ADAS:Phantom Attacks on Driver-Assistance Systems”, 2020, https://eprint.iacr.org/2020/085.pdf, [Online; accessed 23-April-2021]

[4] Yasser Shoukry, Paul Martin, Yair Yona, Suhas Diggavi, and Mani Srivastava, “PyCRA: Physical Challenge-Response Authentication For Active Sensors Under Spoofing Attacks”, 2015, http://www.cs.wm.edu/ ksun/csci680-f15/papers/PyCRA%20CCS2015.pdf, [Online; accessed 23-April-2021]

[5] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song, “Robust Physical-World Attacks on Deep Learning Visual Classification“, 2018, https://arxiv.org/pdf/1707.08945.pdf, [Online; accessed 23-April-2021]

[6] Center for Responsive Politics, ”Senate Commerce, Science and Transportation Committee”, 2018, https://www.opensecrets.org/cong-cmtes/profiles?cmte=SCOM&cmtename=Commerce%2C+Science+and+Transportation&cong=115&cycle=2018, [Online; accessed 21-April-2021]

[7] National Conference on State Legislatures, ”Autonomous Vehicles — Self-Driving Vehicles Enacted Legislation”, 2020, https://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx, [Online; accessed 21-April-2021]

[8] Jing Chen, Silvio Micali, ”ALGORAND”, 2017, https://algorandcom.cdn.prismic.io/algorandcom%2Fece77f38-75b3-44de-bc7f-805f0e53a8d9_theoretical.pdf, [Online; accessed 21-April-2021]

[9] PlanetWatch, ”White Paper”, 2021, https://planetwatch.io/white-paper/index-h5.html?page=1#page=10, [Online; accessed 21-April-2021]

[10] Ryan Barron, Maksim E. Eren, Charles Varga, and Wei Wang, ”pyCarDisplay”, 2021, https://pypi.org/project/pyCarDisplay/ [Online; accessed 21-April-2021]

[11] Rand Labs, ”Indexer API v2 enhanced by AlgoExplorer API”, 2021, https://algoexplorer.io/api-dev/indexer-v2, [Online; accessed 21-April-2021]

[12] Cyril Fougeray, Ted Nivan, ”Vertices SDK”, 2021 https://github.com/vertices-network/c-vertices-sdk, [Online; accessed 21-April-2021]

[13] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ”ImageNet Classification with Deep ConvolutionalNeural Networks”, NIPS’12: Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, 0December 2012 pp. 1097–1105

[14] Andrew Allen, Andrew Thompson, Marc Nijdam, Rahul Garg, ”Helium A Decentralized Wireless Network”, 2018, http://whitepaper.helium.com/, [Online; accessed 21-April-2021]