Analysis of Sensor Attacks against Autonomous Vehicles

Jakobsen, Søren Bønning; Knudsen, Kenneth Sylvest; Andersen, Birger

Published in:
Proceedings of the 8th International Conference on Internet of Things, Big Data and Security - IoTBDS

Link to article, DOI:
10.5220/0011841800003482

Publication date:
2023

Document Version
Peer reviewed version

Analysis of Sensor Attacks against Autonomous Vehicles

Søren Bønning Jakobsen
DTU Engineering Technology
Technical University of Denmark
Ballerup, Denmark
s195161@student.dtu.dk

Kenneth Sylvest Knudsen
DTU Engineering Technology
Technical University of Denmark
Ballerup, Denmark
s190971@student.dtu.dk

Birger Andersen
DTU Engineering Technology
Technical University of Denmark
Ballerup, Denmark
birad@dtu.dk

***Abstract*—Fully Autonomous vehicles (AV) are estimated to reach consumers widely in the near future. The manufacturers will have to be completely sure that AVs can outperform human drivers, which first of all requires a solid model of the world surrounding the car. Emerging trends for perception models in the automobile industry seems to be towards combining the data from LiDAR and camera in Multi-Sensor Fusion (MSF). Making the perception model reliable in the event of unforeseen real world circumstances is tricky enough, but the real challenge comes from the security issue that arises when ill-intentioned people try to attack sensors. In this article, we take a deep dive into the possible attacks and countermeasures for LiDAR and camera. We discuss it in the context of MSF, and provide a simple framework for further analysis, which we conclude will be needed in order to conceptualize a truly safe AV.**

***Index Terms*—Autonomous vehicles, LiDAR, camera, sensors, attacks, countermeasures, security, Multi-Sensor Fusion.**

I. Introduction

Due to advances in technology in recent years, autonomous vehicles are becoming more and more realistic on public roads. Car manufacturers have already dealt with extremely complex challenges such as self-navigation and collision prevention. Some places in the US are even beginning to allow self-driving cars under specific circumstances [14]. There are different levels describing how autonomous a car is [17], and the aforementioned cars are examples of category 3 autonomous cars. Today, the race between manufacturers to become the first company to release a fully autonomous category 5 car continues. Considering that the end goal means giving the AV complete control with no oversight it is important that no stone is left unturned during testing, to make sure that the consumer is completely safe. This means that the AVs will have to withstand not only rigorous testing, but also the challenges that arise in the real world, where people might intentionally try to cause crashes.

Researchers, as well as white-hat hacker groups, have been conducting tests on these cars and their flaws. For example Keen Security Labs in China demonstrated flaws in security in a Tesla Model S, that allowed them to remotely hack into the car, and make it change to the reverse lane [18]. Another example is found in [10], where researchers managed to confuse the perception algorithm to have a stop sign classified as a speed limit sign using stickers.

In this article, we will take a deep dive into attacks, specifically attacks aimed at the AV's ability to perceive the world around it through sensors. Considering all the different technologies involved in self-driving cars, the complexity of attacks on AV's sensors vary wildly. We will focus on remote attacks on the physical sensors, but also the underlying algorithms working the sensor data. We assume that attacks have no access to the vehicle, yet aim to attack the perception model of the AV, since any attacks here will cascade into all other decision-making that an AV does.

II. How an AV models its' surroundings

There are many variables and obstacles on public roads like pedestrians, other cars, turns, traffic lights, bicycles and much more. Before an AV can safely drive on these public roads, the car first and foremost needs to see better than humans in order to drive better. This has been a major hurdle for development. By combining different sensing technologies developers have created detection systems which can "see" better than human eyes [15]. The common technologies in AVs used for mapping the environment are LiDAR (Light Detection and Ranging), radar (Radio Detection And Ranging) and cameras. Radar and LiDAR are somewhat interchangeable, as they offer a lot of the same information with different pros and cons. This has led to debates in the industry about which is best. Currently, Google, Uber and Toyota all rely heavily on LiDAR, while Tesla are the only real advocates for Radar [19]. Because of this, our focus will be on camera and LiDAR.

A. Camera

AV perception is achieved by many sensors and sensor systems. One of the first sensors used in driverless cars was the camera. For AV the camera is used to visualize its surroundings. The camera is used for lane detection, horizon/vanishing point detection, object detection and tracking of vehicles and pedestrians, traffic sign recognition and headlight detection, as demonstrated by the colored boxes in Fig 1. Cars can have multiple cameras covering a 360 degrees view of their environment. Cameras are very efficient at detecting texture of objects. For implementation, cameras are more affordable than lidar and radar sensors. The high pixel quality obtained

by the camera comes with a price of computational power. Today cameras can take pictures with millions of pixels in each frame and about 30-60 frames each second. Each of these frames needs to be processed in real time in order for the car to make real time decisions, this requires a lot of computational power [5]. Image quality is important in order for the system to classify the objects. The quality can be affected by lenses, windshield, vibration environmental conditions like snow, rain, fog and light. All these image disruptions can result in unnoticed objects and increasing image correction processing time. Some cars use a multi camera setup where some cameras overlap each other [1]. The camera creates a good representation of the environment, however the depth perception is not nearly as good as that of other sensors, which is why LiDAR technology is used.

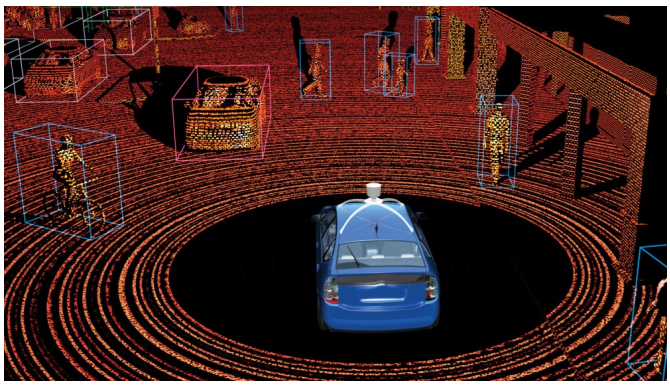

Fig. 1: An image of an autonomous car using LiDAR as distance measure and camera as object detection [22]

B. LiDAR

The LiDAR sensor system fills the existing gap between radar and camera sensors [5]. LiDAR works by emitting pulses of infrared light and measuring the time taken to reflect on distant surfaces. These reflections return a point cloud that represents objects from the environment. Most common LiDAR lasers use light in the 900 nm wavelength, longer wavelengths will perform better in poor conditions, such as fog and rain. Because the LiDAR sensor has a more focused laser beam it can create a more dense point cloud, resulting in a high resolution map of the environment [13]. The resolution obtained by LiDAR is much higher than in radar because of this more focused laser beam. Precision is important in LiDAR systems, as lower precision LiDAR sensor originate noisy point clouds. How precise a LiDAR is says how close the estimated point is compared to a point in the real world. There are different architectural techniques for creating the surroundings of the car. They can be categorized in different groups of spinning and non spinning (solid state). To get a horizontal 360 view LiDAR sensors can be combined with a mechanical part to spin around while measuring the distance of the surrounding objects, as shown in Fig. 1. This is the most common LiDAR application currently. If LiDAR technology is used but there are no moving parts this is called solid state sensors. With solid state sensors we get a more narrow angle but usually they are cheaper.

C. Multi-sensor fusion (MSF)

Multi-Sensor fusion, or sometimes just called Sensor Fusion, is a technique where the input from multiple sensors are combined, in order to leverage the best of both inputs. While it is a possible to combine several sensor types, current trends in the automobile industry trends have gone towards combining camera and LiDAR [5]. This minimises hardware complexity, as only two sensortypes are involved, and the information from these complements each other nicely. The obvious advantages here is the detailed vision of the camera, allowing for object classification, and LiDAR for accurate object detection and detailed range measurements.

MSF have been a major factor in helping researchers make reliable models of an AVs surroundings. Researchers in [5] cites application of MSF in the detection of objects, grid occupancy mapping for placing these objects in a model around the vehicle and lastly for tracking the objects movements within the model. Their example of an MSF algorithm is the PointFusion network, used for 3D object detection. This algorithm achieves sensor fusion by processing each sensors data with a different neural network (NN), and then feeding the representations into a new neural network, achieving high-level fusion as shown in Fig. 2. As seen in the figure, the separate NNs are PointNet and ResNet handling the pointcloud and RGB (Red-green-blue) image respectively. Their results are fed to dense fusion algorithm that for each input point predicts the spatial offset of the corners relative to the input [21]. The PointNet and ResNet information is also fed to a baseline model that directly regresses the box corner locations. Together, the dense fusion predictions and baseline model results in the predicted 3D boxes, that an AV would have to navigate it's way around.

While this is just one way of doing this, there are many, many ways to go about it. It really comes down to balancing the complexity of the mathematics, the computational power in the vehicle and other factors to design the 'perfect' sensor fusion algorithm.

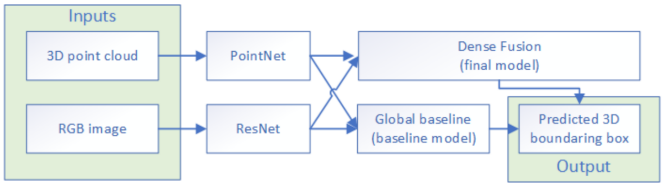

Fig. 2: Overview of the neural networks involved in the PointFusion MSF algorithm [5].

III. Attacks and countermeasures

In this section we will explain different attack types, documenting examples of successful attacks and suggesting possible countermeasures. Since our focus is remote attacks,

our assumption of possible setups to perform these attack will be categorized into one of two types, described by Petit et. al. in [1]:

Front/rear/side attacks which involve an attacker who installs hardware in his own vehicle to perform an attack. This allows the attacker to keep the hardware within distances of a target vehicle for longer time.

Roadside attacks which involve a mounted stationary setup, that allows for greater precision. This type of attack is not limited to one installation point, but can have several if need be.

A. Attacks on cameras

1) Blinding Attack: Cameras have built in functions to decide how much light is let into its' shutter, in order to take better photos in all light conditions. This attack type abuses this function, by shining a strong light source into the camera, in order to either completely blind the camera, or partially blind it, causing it to miss objects. In experiments performed in [6], researchers managed to completely blind a camera for up to 3 seconds by pointing a laser light directly at the camera, and caused irreversible damage by doing so for several seconds from less than 0.5 meters away. They also managed to cause partial blindness by directing an LED (light emitting diode) light matrix at the camera, inhibiting the object recognition and proving that there's multiple tools to perform this attack. The researchers performed the testing in stationary setups, but they also simulated a front/rear/side attack by wobbling the laser at the camera, which was still successful, though the blinding was less effective.

2) Adversarial Example (AE): While not physically attacking the camera, this attack type attacks they way the information from the camera is processed. The perception models based on machine learning (ML) and deep learning (DL) have proven vulnerable to carefully crafted adversarial perturbations. Generally, AEs are classified as 'Appearing Attacks' and 'Hiding Attacks'. There have been published several attacks of this type, where most are evaluated on stop signs as they are a critical part of decision making in driving. Researchers in [7] executed such a hiding attack, where they designed specific black and white stickers that caused misclassifications of stop signs. In order to fool any human onlookers, they designed the stickers to look like graffiti and still managed to make the sign be classified as a speed limit sign in 87,5% of the tests. In the opposing category, [10] managed to make innocent looking stickers be classified as stop signs. There are other examples of real world applications, such as [4], where researchers 3D printed a traffic cone that was ignored by cameras, and [12] who designed a billboard that causes malfunction in the steering angle of AVs. This attack type appears mostly as a roadside attack.

B. Attacks on LiDAR

1) Spoofing by relaying attack: LiDAR sensors are what is called active sensors [11]. This means that the LiDAR sensor in order to detect an object emits light intentionally from its own sensor and then listen to the echo. Because the speed of light is constant the LiDAR sensor can calculate the distance by measuring the ping time of the signal. In spoofing attacks the attacker uses this signal created by the victim but relays it back from a different position. The goal of spoofing is to deceive the victims LiDAR sensor and to create fake point clouds. The creation off fake points could potentially cause the AV to make sudden erroneous decisions. These active LiDAR sensors use particular waveform to differentiate echoes from the other inbound signals. This means before the attacker can perform the attack, he needs to obtain the ping waveform. When the waveform of the victims LiDAR sensor is obtained, the attacker can now perform the attack by relaying the signal back to the LiDAR sensor. This attack is effective because the victim car has a hard time to distinguish what signals are real or fake and is unaware of the attack, potentially providing seemingly legitimate but actually erroneous data. There are two ways to perform this attack. One way is to place the attacking device on the roadside and then aiming towards the victim lane [8], however it could also be performed as a front/rear/side attack, using computer vision to keep track of the vehicle and to aim precisely. This would however be significantly more difficult to execute. First some of LiDAR sensors used in cars are spinning LiDARs. This means that the victims LiDAR sensor has to be facing the attacking direction. In order for the attack to work the echos has to hit victims sensor at the right receiving angle. The second problem is that LiDAR will only accept echos within a certain delay time. Because of the delay threshold the distance between attacker and victim has a big influence on the attack window. Article [1] performs a spoofing attack, and relates timing to the success of spoofing.

2) Saturation attack: Typically sensors has a lower and upper bound for input signals [11]. If signals arrive at the sensor with a low signal power, the sensor will ignore the signal, this is also called "limit of detection". When the signal increases it eventually exceed the upper threshold, at this point the sensor cannot reflect the input changes well. Therefore the principal of saturation is to expose the target sensor to a signal of high, making the sensor unable to work properly. Because the sensor is unable to receive any new signals while under a saturation attack, this attack is also called DoS (Denial of Service) or a blinding attack, since it uses light as a medium. Saturation attacks are powerful because they are unavoidable, and though it is easy to detect it cannot be prevented. Human drivers and pedestrians will be unaware of the attack because LiDAR operate in eye safety wavelength namely infrared light. In [11] they performed this saturation attack with both weak and strong light sources. The experiments made was successful and had different outcome for the weak and strong light. When the attack was performed with weak light they observed randomly located fake dots, while with the strong light source, they observed that the sensor became completely blind in a sector of FoV (Field of View).

Attack type	Target sensor	Method	Impact	Feasibility
Blinding attack	Camera	Blinding the camera with some kind of light source, making the camera unable to guide the Vehicle	Low	Easy
Adversarial Examples	Camera	Introduce objects with adversarial perturbations, to confuse perception model	High	Hard
Spoofing attack	LiDAR	Relaying light pulses in a different position, creating fake obstacles.	High	Medium
Saturation attack	LiDAR	Jamming or blinding LiDAR sensors by emitting strong light in same wavelength as the LiDAR sensor.	Low	Easy

TABLE I: Overview of attacks

Countermeasure	Sensor	Method	Preventing attack	Feasibility
Redundancy	Camera	Adding more camera with a significant overlap in view	Blinding attack, assists MSF attack detection	High
Optics and materials	Camera	Adding smart materials, that can filter out harmful light	Blinding attack	High
Making AE robust perception models	Camera	Using advanced techniques to make AE attacks more difficult	Advarsarial attack	Low
Saturation detection	LiDAR	Builtin Fail-safe mode. Under attack slows down and pull to the side.	Saturation	High
Redundancy and Fusion	LiDAR	Multiple LiDAR setup with overlapping FoV. By comparing input from multiple overlapping sensors it is possible to detect and prevent some attacks	Saturation and spoofing	High
Random probing	LiDAR	By randomizing LiDAR pulse interval makes spoofing very difficult to perform	Spoofing	Medium
Side-channel authentication	LiDAR	Using side-channel information as authentication. Authentication makes it very difficult to spoof a LiDAR, not knowing the secret key	Spoofing	High
Multi-Sensor Fusion	LiDAR and Camera	Use smart MSF to model both LiDAR and camera data, checking for inconsistencies	Camera blinding, LiDAR Spoofing, saturation, relay and rotation	Medium

TABLE II: Overview of countermeasures

C. Countermeasures on cameras

1) Redundancy: Petit et. al. [1] argues the simple benefit that more cameras with overlapping view will at the least make a blinding attack harder to execute. In [6] researcher emulated a handheld blinding attack by laser, which would be very hard if several cameras were present. The argument for more cameras could also be supported by the findings in [2], which suggests an algorithm for recognizing perception error attacks in MSF, which requires stereo cameras. This will be further elaborated on in section III-E. Though this solutions seems simple, it is important to remember the extra space and cost associated with it, which are very essential factors in the highly competitive automotive industry. There is also an argument that integrating more cameras introduces more complexity in synchronizing the capturing of frames and maintaining the same exposure [1], though this should easily be overcome with todays' technology.

2) Optics and materials: Petit et. al. [1] argues that removable on-demand near-infrared-cut filter, a feature commonly found in security cameras, could serve as a defence against blinding attacks. They argue that such a defence would only be usable during daytime, as the filter would have to be removed during nighttime, in order to make use of infrared light for night vision. Intelligent applications of this countermeasure should be considered, perhaps by implementing software with thresholds for when it considers the camera to be under attack.

Another defensive material is photochromic lenses, which is a type of lenses that change color to filter out specific types of light. Several types of lenses or coating of lenses could be considered, but as an example vanadium-doped zinc telluride will turn more opaque when hit by high-intensity beams, automatically filtering these without effecting image quality in low light conditions [1]. Once again the hardware and development costs should be considered.

3) Making AE robust perception models: Researchers in [9] provide a very thorough examination of how to make adversarially robust ML/DL solutions and their efficiency. This is a very wide and highly technical topic, which we consider outside the scope of this article. We will therefore refer the interested reader to the article, and quickly mention solutions in simple concepts, as well as the conclusions.

They suggest the possibility of re-training a given classifier on images including adversarial attacks, thought this approach is easily criticized for being reactive and vulnerable to attacks that simply generate new attack types. They also suggest training an auxiliary model, whose sole purpose is to detect features commonly found in pictures with AEs and classify a frame as an outlier if it contains it. They conclude that there are multiple directions that solutions can go, and that they often can be combined. More research is needed in order to facilitate a solid defense, but it is feasible, and a valuable addition to making safer AVs.

D. Countermeasures on LiDAR

1) Saturation detection: As described in section III-B2 saturation can easily be detected by the sensor system. A victim vehicle could have an inbuilt fail-safe mode, so when the car detects saturation, it slows down the car and pull to the side [11]. This countermeasure could on a crowded road lead to a dangerous situation. If the car has to pull to the side while having a jammed LiDAR sensor, its like a person driving with closed eyes.

2) Redundancy and Fusion: One countermeasure could be by having multiple LiDARs overlapping some FoV angles [11]. With redundant LiDAR sensors, the victim car could under saturation attacks abandon input from the attacked sensor until the attack is done. Though the car knows when it is exposed to saturation, it is significantly harder to detect spoofing. The redundant setup will still work better against spoofing by cross validating the malicious points. If the attacker creates fake points in the non-overlapping zone, redundancy will have no effect. LiDAR sensors are expensive so using multiple sensors will increase the overall cost a lot. This solution is not bullet proof, because the attacker is still able to attack multiple sensors at the same time. Another option proposed in [1] is to take advantage of data intercepted by neighboring AVs. Victim vehicle could cross-validate its data with neighboring data to observe inconsistencies. This method only works if there are other vehicle on the road. V2V (vehicle to vehicle) solution opens up for more hacking opportunities because one neighboring vehicle could share incorrect data.

3) Random probing: When making a spoofing attack on LiDAR sensors, a hacker will be interested in the pulse interval. This interval is the timing for when the attacker needs to fire back attacking pulses [16]. By randomizing the interval it makes it hard for the attacker to synchronize the attack. This method is problematic for spinning LiDAR systems, as they require a constant rotation speed and angle of transmission needs to be known [1]. Another option here is to skip some pulses, as this will only require some software modification. When the sensor skip a pulse it is still able to listen to incoming pulses, making it possible to detect possible spoofing attacks. If the sensor is skipping some of the pulses, it has to run with a higher rotation speed, to keep the same resolution [11]. It is important that the skipped pulses are chosen in a pseudo-random fashion, where the attacker cannot predict the skipped pulse.

4) Side-channel authentication: To understand this countermeasure it is essential to know what is meant by side-channel information. Side-channel information is physically leaked information, which could for example be power consumption or electromagnetic radiation. The suggested side-channel information in the article [3] comes from a cryptographic device in the car. The device is making heavy calculations using AES (Advanced Encryption Standard) on a cryptographic key, and the electromagnetic radiation during these calculations are read. This information is then used to modulate and demodulate the amplitude of the laser. It will then only accept returning echoes with exactly this modulation. Though feasible, it becomes very difficult for the attacker to send fake echoes with the correct modulation, and the car can simply change the cryptographic key once in a while to have varying side channel information.

E. Countermeasures via MSF

Evaluating countermeasures via MSF can be difficult, as it is an emerging research topic with many approaches mathematically. As such, we will base this section on [2], a newly released paper with meta reflections and criticisms of the current state of MSF algorithms, as well as a suggested new approach. An important point they raise, is that most of the aforementioned attacks are evaluated on a single sensor type. They argue that today's AVs does not build their model of their surroundings based on a single sensor type, but rather through the combination of data through the use of MSF. This immediately raises the abstraction level of the discussion, as physical attacks on sensors needs to be evaluated based on the whole perception system. As an example, they criticise the design of some MSF algorithms, as their design seems too focused on working in non-adversarial settings. The algorithm F-PointNet uses a cascade approach to fuse LiDAR and camera data, by generating 2D proposals on the image data, and then projecting these onto 3D space, refined by the LiDAR data. This makes it especially vulnerable to camera attacks, as the detection failures accumulate through to the LiDAR steps of the fusion algorithm. The researchers conclude that any MSF built around the idea of projecting either LiDAR or camera data onto the other, will be significantly more vulnerable to attacks at the sensor considered to be the 'primary'. To combat these issues, the researchers designs their own sensor fusion algorithm, which uses CV and ML to map features on both the camera and LiDAR data, and analyzing any features that cannot be mapped to both. The results of their design for camera attacks yields a 100% detection rate, and for LiDAR attacks their detection rate for spoofing and saturation are

97% and 96% respectively. Thus proving that MSF can be leveraged for a significant defence against attacks on the sensors providing the data. They do however make a point, that AEs attacking sensory data processing algorithms of camera or LiDAR data would not be detected through their algorithm, as they would not appear as sensor malfunctions.

IV. Discussions

There is not doubt that AVs will arrive in the near future, but it is clear that one of the major hurdles is not just beating the difficult tasks of modeling the surroundings and navigating them, but also making the perceived information resistant to attacks. As noted in [14] a recent study showed that only 14% of drivers trust an AV to do all the driving, while 54% are too afraid to try it and 32% are unsure. Convincing people will require delivering a product that completely delivers on all safety measures, before public opinion deems AVs to be too dangerous. In table I and table II we have summed up our attacks and countermeasures, but it is still difficult to concretely say that a consumer is sufficiently safe. How do you come up with guarantees in a field that moves so fast in so many directions?

In [9] they discuss the possibility of threat modeling, as a fundamental approach to safety analysis. Here they mention several important aspects including, but not limited to, *Adversarial Knowledge*, *Adversarial Capabilities* and *Adversarial Specificity*.

Adversarial Knowledge refers to the required knowledge for executing an attack. Typically, adversarial attacks are referred to as either white-box, gray-box, or black-box. White-box assumes that the attacker has full knowledge of the underlying systems, be it hardware or software. Gray-box assumes partial knowledge of the underlying mechanisms and black-box assumes no knowledge, to the point where they might not even know which ML algorithm the perception model they are trying to attack is built on. This is arguably the most important dimension to consider, especially if you factor in who's the perceived target. With the increased tendency towards cyber-warfare that we are seeing internationally, high-priority targets like heads-of-state can expect to be targets of incredibly sophisticated attacks that regular people would have no need to fear.

Adversarial Capabilities defines the assumed capabilities of the attacker, which is important scope to consider. This leans itself towards knowledge as well, but also economic as in [1], where they specifically focus on attacks requiring only commodity hardware. They do this under the assumption, that the attacks requiring the least economic and technical knowledge will be most common, an argument that can definitely be extended to our definition as well.

Adversarial Specificity means how specifically targeted an attack is. This could be a consideration of whether or not the laser damage to the cameras tested in [1] would translate to the damage on other cameras, or as mention in [9], that black-box AEs for one ML/DL model are assumed to affect other models trained on datasets with a similar distribution as the original one. It is in the interest of car manufacturers, that attacks do not generalise well across hardware and software.

With this information in mind, one can start reflecting on the future of safety precautions, though nothing is set in stone. We can expect that simple hardware attacks will happen, be it blinding of cameras and attempts at tricking the LiDAR, considering just how low knowledge, capability and specificity requirements are especially for the camera attacks. Whether it is smartest to adopt some of the novel approaches suggested in segment III-C and III-D, or trust in the higher level protection of an MSF that checks for physical attacks is hard to tell. As with many IT-security questions, the answer probably lies somewhere in middle by combining both.

State-of-the-art research on MSF have proven that it can be a valuable tool to detect attacks on sensors [2], but one should not see this as a excuse to leave them wide open. They note that attacks are still feasible, however, they would need to attack both camera and LiDAR simultaneously and gradually, as sudden shifts would be detected. This raises the knowledge and capability requirements enormously, to the point where one could argue that any attacker with such capabilities should just ram another car into it's victim to get the same results with way less effort.

IT-security often becomes an arms race, as researchers have already tested attacks through adversarial examples, that trick both camera and LiDAR [4]. This should beat any MSF that cross-validates, as none of it's sensors will detect the object. Researchers in [2] cites exactly adversarial examples as a weakness, and though this attack appears to have had huge knowledge and capability requirements to create, the finished 3D model could theoretically be sold on a black market for very little. This would quickly lower the availability of the attack to anyone with a 3D printer.

It is clear that an attack like above certainly complicates things, and it won't be the last to do so. Though the complexity of the topic might seem disheartening, we have no doubts that further research will help increasing safety, to a point where it's safe enough. Together, researchers will have to come up with a definition of what 'safe enough' even means, and only the future can tell whether or not it will ever be completely safe.

V. future work

Considering the already mentioned importance of safety and security, it will be necessary for the automobile industry to adopt some sort of safety standards for the AVs perception models, just as you would consider standards for seat belts. As the topic we are trying to concretize are quite more complex than seatbelts, there is a dire need for a solid test environment. E.g. it needs to be measurable for how many frames an AE actually tricks a camera, and whether or not angle has any influence.

In the future, we would like to test a framework which could help define these standards. The idea would be to not require

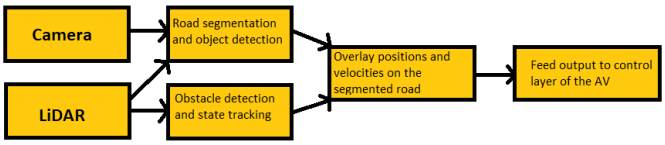

Fig. 3: Overview of the suggested pipeline for the data [20].

a hundred thousand dollar AV just for testing. Instead we will build a model of an AV. One such test-setup could potentially be done with the following components:

- Leo Rover Developer Kit 2, to have something driveable to mount the sensors on
- A sufficiently high resolution camera
- A LiDAR sensor
- Signal processing and computational hardware

Our suggested pipeline of the data follows that of [20] and considers the processing tasks as follows: First camera video frames and the depth channel information from the LiDAR are sent to a Deep Neural Network (DNN) to do the object detection and road segmentation, since DNNs such as a Fully Convolutional Network (FCN) have shown better accuracy for computer vision tasks compared to ML [20].

The second step is to overlaying the point cloud from the LiDAR with the fused data, before feeding the output to the control layer of the AV. The entire pipeline of the data can be seen in Fig. 3. This would set a foundation for testing possible attacks to see if the AV drives as expected or deviates.

VI. Conclusion

We are moving towards a future, where we will soon see driverless cars available to the common consumer. It is necessary to secure the safety and well-being of the consumers, as well as earn their trust. To do this, the AVs need to outperform human driving, which first and foremost require that the car can build a reliable model of it's surrounding, before the higher level algorithms can navigate them. This requires sensor data that are valid and sensors resistant to attacks, so adversaries cannot manipulate them to cause accidents. We have given our resume of the literature concerning attacks on cameras and LiDAR, two major factors in the current perception models, and also the two sensors most often combined together in multi sensor fusion. We have discussed these and their countermeasures in depth, before reviewing them in the context of multi-sensor fusion.

We have opened a discussion into attack complexity and suggested a framework to review them in, in order to better grasp the challenging factors of this issue. It is clear that more countermeasures are needed, however advancements in MSF are looking very promising, and some of the state-of-the-art solutions will be a huge leap in making attacks too complex to be feasible. It is still important to keep in mind that since everyone, including heads-of-state, will be driving around in these cars, complexity alone cannot be seen as a sufficient. Solid standards need to be adopted.

References

[1] J. Petit, B. Stottelaar, M. Feiri, F. Kargl, "Remote Attacks on Automated Vehicles Sensors: Experiments on Camera and LiDAR".

[2] J. Liu, J. Park., "Seeing is Not Always Believing": Detecting Perception Error Attacks Against Autonomous Vehicles", IEEE Transactions on Dependable and Secure Computing, IEEE, Vol 18, pp. 2209-2223, May 2021.

[3] R. Matsumura, T. Sugawara, K. Sakiyama, "A Secure LiDAR with AES-Based Side-Channel Fingerprinting", Sixth International Symposium on Computing and Networking Workshops, IEEE, Nov 2018.

[4] Y. Cao et al., "Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attacks", IEEE Symposium on Security and Privacy, IEEE, August 2021.

[5] Jelena Kocić, Nenad Jovičić, and Vujo Drndarević., "Sensors and Sensor Fusion in Autonomous Vehicles", Telecommunications Forum, November 2018.

[6] C. Yan, W. Xu, J. Liu, "Can You Trust Autonomous Vehicles: Contactless Attacks against Sensors of Self-driving Vehicle", Def con 24, Def con, 2016.

[7] K. Eykholt et. al., "Robust Physical-World Attacks on Deep Learning Visual Classification", IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018.

[8] Yulong Cao et. al., "Adversarial Sensor Attack on LiDAR-based Perception in Autonomous Driving", ACM SIGSAC Conference, ACM, pp. 2267-2281, November 2019

[9] A. Qayyum, M. Usama, J. Qadir, A. Al-Fuqaha, "Securing Connected and Autonomous Vehicles: Challenges Posed by Adversarial Machine Learning and the Way Forward", IEEE Communications Surveys Tutorials, Vol 22, pp. 998-1026, February 2020.

[10] K. Eykholt et. al., "Physical Adversarial Examples for Object Detectors", published online, accessed January 10 2022. https://arxiv.org/abs/1807.07769

[11] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae, "Illusion and Dazzle: Adversarial Optical Channel Exploits Against Lidars for Automotive Applications",International Conference on Cryptographic Hardware and Embedded Systems, CHES, August 2017

[12] H. Zhou et. al., "DeepBillboard: Systematic Physical-World Testing of Autonomous Driving Systems", published online, accessed January 13 2022. https://arxiv.org/abs/1812.10812

[13] R. Roriz, J. Cabral, and T. Gomes., "Automotive LiDAR Technology: A Survey", IEEE Transactions on Intelligent Transportation Systems, IEEE, Early Access, pp. 1 - 16, 15 June 2021.

[14] D. Jones, "California Approves A Pilot Program For Driverless Rides", accessed 10 January 2022. https://www.npr.org/2021/06/05/1003623528/california-approves-pilot-program-for-driverless-rides?t=1641824140634

[15] K. Burke, 'How Does a Self-Driving Car See?', accessed 13 January 2022. https://blogs.nvidia.com/blog/2019/04/15/how-does-a-self-driving-car-see/

[16] Y. Deng, "Deep Learning-Based Autonomous Driving Systems: A Survey of Attacks and Defenses", IEEE Transactions on Industrial Informatics, IEEE, Vol 17, pp. 7897 - 7912, April 2021.

[17] F. Petit, 'What are the 5 levels of autonomous driving?', accessed 14 January 2022. https://www.blickfeld.com/blog/levels-of-autonomous-driving/

[18] Tencent Keen Security Labs, 'Experimental Security Research of Tesla Autopilot', Published online, Tencent Keen Security Lab, March 2019.

[19] A. Neal, "LiDAR vs. RADAR", accessed 19 January 2022. https://www.fierceelectronics.com/components/lidar-vs-radar

[20] B. S. Jahromi, T. Tulabandhula, S. Cetin, "Real-Time Hybrid Multi-Sensor Fusion Framework for Perception in Autonomous Vehicles", Published online, MDPI, October 2019.

[21] D. Xu, D. Anguelov, and A. Jain, "PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation", IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, December 2018.

[22] O. Cameron, "An Introduction to LIDAR: The Key Self-Driving Car Sensor", accessed 5 Januar 2022. https://news.voyage.auto/an-introduction-to-lidar-the-key-self-driving-car-sensor-a7e405590cff