Combined Safety and Cybersecurity Testing Methodology for Autonomous Driving Algorithms

Mohsen Malayjerdi
Department of Mechanical and Industrial Engineering
Tallinn University of Technology
Tallinn, Estonia
mohsen.malayjerdi@taltech.ee

Andrew Roberts
FinEst Centre for Smart Cities
Tallinn University of Technology
Tallinn, Estonia
andrew.roberts@taltech.ee

Olaf Maennel
Centre for Digital Forensics and Cybersecurity
Tallinn University of Technology
Tallinn, Estonia
olaf.maennel@taltech.ee

Ehsan Malayjerdi
Department of Mechanical and Industrial Engineering,
Tallinn University of Technology
Tallinn, Estonia
ehsan.malayjerdi@taltech.ee

ABSTRACT

Combined safety and cybersecurity testing are critical for assessing the reliability and optimisation of autonomous driving (AD) algorithms. However, safety and cybersecurity testing is often conducted in isolation, leading to a lack of evaluation of the complex system-of-system interactions which impact the reliability and optimisation of the AD algorithm. Concurrently, practical limitations of testing include resource usage and time. This paper proposes a methodology for combined safety and cybersecurity testing and applies it to a real-world AV shuttle using digital twin, software-in-the-loop (SiL) simulation and a real-world Autonomous Vehicle (AV) test environment. The results of the safety and cybersecurity tests and feedback from the AD algorithm designers demonstrate that the methodology developed is useful for assessing the reliability and optimisation of an AD algorithm in the development phase. Furthermore, from the observed system-of-system interactions, key relationships such as speed and attack parameters can be used to optimise testing.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

automotive cybersecurity, safety testing, autonomous driving

ACM Reference Format:
Mohsen Malayjerdi, Andrew Roberts, Olaf Maennel, and Ehsan Malayjerdi. 2022. Combined Safety and Cybersecurity Testing Methodology for Autonomous Driving Algorithms. In *Computer Science in Cars Symposium (CSCS '22), December 8, 2022, Ingolstadt, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3568160.3570235

CSCS '22, December 8, 2022, Ingolstadt, Germany

ACM ISBN 978-1-4503-9786-5/22/12…$15.00
https://doi.org/10.1145/3568160.3570235

1 INTRODUCTION

Testing autonomous driving (AD) algorithms for performance under safety test cases is a predominant focus for developers to assess the reliability of the algorithm and for optimisation. AD algorithms are also susceptible to manipulation from cyber threats which target the advanced hardware technologies sensor telemetry which serves as an essential input for perception, detection, and control decisions [2, 12, 20]. Existing methods [3, 8] for testing are challenged by the complexity of evaluating system-of-system interactions to identify key relationships and parameters, and limitations of testing inherent to real-world AV programs, resource usage and time. The main idea of this paper is to establish a method for combined safety and cybersecurity testing of developmental AD algorithms to evaluate system-of-system interactions to identify and investigate parameters that impact safety and the effect of cyber attacks, and to develop future ideas for optimisation of testing. To this end, the paper focuses on three research questions aligned with the challenges of combined safety and cybersecurity for AD algorithms.

- RQ1 How can AD algorithm designers evaluate the reliability and optimisation of the AD algorithm to both safety and cybersecurity test cases?
- RQ2 Cybersecurity testing is predominantly conducted on well-established AD algorithms. How can combined safety and cybersecurity testing be conducted on a developing AD algorithm?
- RQ3 What key relations and parameters can we identify that can optimise safety and cybersecurity testing?

To evaluate these research questions, we apply our methodology to a developing AD algorithm in a digital twin, software-in-the-loop (SiL) simulator and real-world AV testing environment. Cybersecurity testing and safety testing are often conducted separately, reducing our understanding of the relationship between failures of the algorithm caused under normal safety scenarios and failures caused by the impact of cyber attacks. For AD algorithms in the development stage, where the reliability and optimisation of the AD algorithm to safety scenarios have not been established, this exploration of the relationship between safety and cybersecurity can offer novel insights to improve the awareness of the AD algorithm designer to shortcomings in the algorithm.

The major contributions of this paper are the following:

- Methodology for combined safety and cybersecurity testing
- Safety and cybersecurity test cases conducted on an AD algorithm under development, and with feedback from the AD algorithm designer
- An analysis of the combined safety and cybersecurity test cases that identifies key relations and the sensitivity of parameters.
- All the code, our AV simulation configurations and research data used in the combined safety and security testing will be available for the research community on GitHub.

2 TARGET SYSTEM

2.1 Low-Speed AV Shuttle for Public Transportation

The target AV for this study, iseAuto (see Fig. 1), is a real-world AV shuttle for public transportation, operating in numerous EU countries. The shuttle was developed as part of a project at Tallinn

Figure 1: iseAuto autonomous shuttle

University of Technology's AV research group. The objective of this project is to build an open-source AV shuttle that provides a smart city test bed within the university campus, enabling different types of urban mobility research. Currently, this SAE level 4 and 5 shuttle is operating on the campus for experimental and study purposes. iseAuto uses a multi-LiDAR sensor system for perception and localisation. Two Velodyne LiDARs are mounted at the top front (VLP-32) and the back (VLP-16) of the vehicle, in addition to two Robosense RS-Bpearl at both sides (left and right), to decrease the sensor blind zone around the car.

2.2 Autonomous Driving Algorithm

The AV uses Autoware.ai [11] autonomous software stack which is an open-source AD software. This software enables us to employ different algorithms for each main part of the autonomous system including localization, sensing, detection, and navigation. OpenPlanner navigation planning algorithm.

In this study, we focused on OpenPlanner as one of the most widely used path-planner modules in the AD software. In the latest version of this algorithm, which is currently 2.5, the module has become noticeably more advanced in terms of supporting various high-definition map formats, predicting the trajectories of other actors, and using a kinematics-based trajectory generator [5]. This version is compatible with Autoware.ai 1.15. Open-planner combines global and local planners that jointly utilize the road network map to generate local waypoints based on a global route and manage discrete behaviours such as avoiding dynamic obstacles and following traffic lights.

The local planner module generates tracks parallel to the main path defined by the global planner. These tracks are named rollouts (see Fig. 2). The trajectory evaluator assesses all possible rollouts in case an obstacle blocks the path. Then, the behaviour selector will lead the AV to the new safe rollout. Figure 2 shows how open-planner selected rollout number 6 in order to pass the non-player character (NPC). It also detects the curb lines and avoids those rollouts which intersect the curbs.

The algorithm uses the output of the kf_contour_track algorithms to consider all the perceived objects based on the LiDARs point cloud in its local path planning. Earlier, the euclidean clustering algorithm received the filtered point cloud data and prepared point clusters, which is the input of the kf_contour_track. This combination of cluster and contour tracking is done in each sequence for the open-planner to evaluate possible trajectories and create the behaviour based on that. Figure 3 shows the diagram of how the open-planner module works under the AD software package.

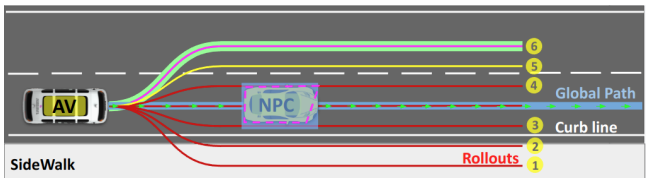

Figure 2: How open-planner generates different trajectory to pass an object

3 COMBINED SAFETY AND CYBERSECURITY TESTING METHODOLOGY FOR AD ALGORITHMS

The architecture of the proposed combined testing methodology is presented in Figure 3. This method takes advantage of a high-fidelity software in the loop (SiL) simulation [16] approach to validate and verify the performance of a AD software under critical cyber security conditions. This method consists of three main following elements:

- Attack script: which simulates a critical security condition.
- High-fidelity simulator: It is a game engine environment that provides the physics for modeling sensors and motion.
- AD software: It is the autonomous driving software that controls the AV.

The combined safety and cybersecurity methodology consisted of the following iterative steps:

- **Scenario Selection**
- **Analysis of the scenario to extrapolate the safety evaluation criterion applicable**
- **Safety Test Case Setup**
  - Initialisation of the SiL high-fidelity simulator and configuration to the real-world AV

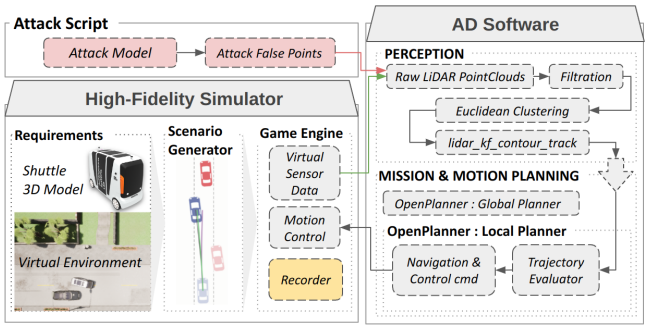

Figure 3: Architecture of the testing platform

  - Initial scenario testing using the safety test cases to assess the reliability of the algorithm and the quality of the test data
  - Optimisation of the safety test cases to select a subset of the scenario tests to assess the reliability of the algorithm
  - Run of the safety test case scenarios
  - Selection of distinct safety test case scenarios which provide most stable results in terms of success of mission and safety violation
- **Cybersecurity Test Case Setup**
  - Analysis of the scenario to determine cyber attack strategy for test cases
  - Development of the code for adversary generation in the SITL high-fidelity simulator
  - Selection of attack parameters
  - Optimised the cybersecurity test cases
  - Evaluate cybersecurity test cases in SiL high-fidelity simulator
  - Real-World AV Testing for safety and cybersecurity
- **Results Analysis**
  - Analysis of the performance of AD algorithm to safety criteria
  - Analysis of sensitivity of attack parameters and driving parameters

3.1 Testing Environment

All tests are conducted in a virtual environment powered by the "Unreal game engine" (Unreal) [4]. Carla simulator [6] is one of the open-source high-fidelity vehicle simulators capable of connecting to different AD software and scenario generator applications. In this study, we use Carla 0.9.13 as the high-fidelity simulator. Figure 3 illustrates the requirements for the high-fidelity simulator to conduct simulation testing which are two components, the digital twin of our AV and the virtual replication of our target environment. These replicated components help us to gain more accurate results of the proposed platform [14]. The AV digital twin is a 3D model of our real-world world AV shuttle, designed in Blender, a graphical 3d modelling software, and imported and built in Unreal for deployment in Carla. This model uses the same dimension and sensor configuration (model, position, and orientation) from the real AV shuttle. The environment digital twin, in our case, is identical to the location where we are testing and operating our shuttle, this includes the urban details and vegetation. The next module in the simulator is a scenario generator that produces the desired scenario based on the user input specification. Finally, the simulator engine generates sensor data from sensors, including LiDARs, cameras and others and publishes it for other blocks (see Fig. 3 the simulator block). Then, the AD software receives this data as raw LiDAR point-cloud information and processes the data as mentioned in the diagram (Figure 3).

This simulation setup was implemented on a desktop computer with the following configuration:

- Intel® Core™ i7-11700K @ 3.60GHz × 16 cores
- NVIDIA GeForce RTX 3080 10 GB
- RAM: 128 GB

3.2 Scenario Selection

To evaluate the combined safety and cybersecurity testing, we chose a simple overtaking maneuver, which is one of the most safety challenging operations [13]. Figure 4 shows the functional level of the planned scenario. To generate a variety of distinct scenarios, we opt for the initial relative distance to the NPC D_x and the NPC constant speed S_{NPC} as the distinct scenario parameters.

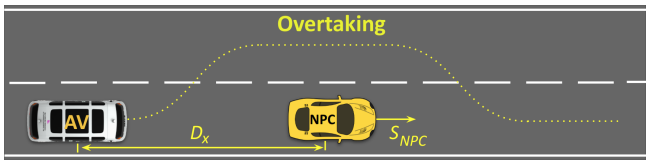

Figure 4: Dx and S_{NPC}, define the initial relative distance to the NPC and the constant NPC speed in each scenario

Table 1: Target scenarios definition

Actor	Speed	D_x	Goal
AV	$[0{:}6]m/s$	0 (m)	overtake the NPC safely
NPC	[1 1.4 1.8 2.1 2.5]	[15 20 25](m)	keep moving

3.3 Safety Evaluation Criteria

In determining the evaluation criteria for AV safety we considered two conditions, 1) mission success and 2) safety violations. A safety violation consists of a collision and dangerous driving behaviour. In determining which criteria to apply, we considered the EuroNCAP [1] and ISO26262 [10] standards as well those used in composite studies [3, 7, 8]. We derived that the safety goal of the AD algorithm is to execute the overtaking mission without colliding or interfering with other ego vehicles or objects and without exhibiting driving behaviour which is dangerous to the AV passengers. Table 2 details the safety criteria applied in our experiments.

3.4 Safety Test Case Setup

To evaluate the reliability and optimisation of the AD algorithm for the overtaking manoeuvre, we, firstly, initiated a run of 50 distinct scenarios in the high-fidelity simulator, repeating 6 times. Each scenario was repeated 6 times to ensure the reproducibility

Table 2: Safety Evaluation Criteria

Safety Condition	Data Label	Description	Metric
Succeed	Suce	AV Successful complete the mission	Pass/Fail
Not Finished	NotF	Failure to finish the mission	Pass/Fail
Distance-to -Collision	DTC	Violation of the safe distance between AV and NPC	AV within 0.5m of other vehicle
Break on Driving Lane	BrD	AV initiates emergency break on driving lane	Pass/Fail
Break on Passing Lane	BrP	AV initiates emergency break on passing lane	Pass/Fail
Collision	Col	AV collides with NPC	Pass/Fail
Violation	V	Safety Violation	

of the outcome. With the mentioned desktop configuration, it took approximately 100 *sec* for each scenario and, in total, 8.3 hours for 300 runs. The purpose of the first scenario run was to provide a general overview of the performance of the algorithm. We targeted a range of 1 to 3 m/s for the NPC speed and 15 to 30 m for the initial relative distance to the NPC for selecting the 50 distinct scenario parameters. The results showed that the AD algorithm could not safely overtake the NPC at an NPC speed higher than 2.5 m/s and a distance (D_x) of more than 25 m.

Although a high number of scenario variations shows better coverage in the scenario space to find corner cases, it will lead to an increase in the time duration of the runs. Furthermore, the number of each scenario repetitions was not sufficient to statistically explain the occurrence of each safety violation. Finally, it is worth mentioning that, as our primary study focus is not just the validation of the AV performance, we need to use an optimum number of trials for both safety and cyber test cases. Due to this, we limited the scenario parameters space to the intervals listed in Table 1 that regressed the test set to 15 distinct cases in a full factorial setup. This enabled us to repeat the simulation of these test cases 50 times and apply the full set of safety criteria: collision, DTC, break in passing lane, break in driving lane, failure to finish, and mission success.

Each scenario is generated by the Carla scenario runner utilizing the Python behaviour trees to handle series and parallel events in the scenario. Figure 5 depicts the scenario scheme starting with the main sequence behaviour. This series begins with transforming the actors into the environment and finishes by destroying the actor block. A parallel behaviour (Driving Toward Intersection) is defined to run the attack and the scenario stop block while the NPC follows the defined waypoint. For safety test case scenarios, the attack block is skipped, and the scenario waits till the stop criteria are satisfied.

3.5 Cyber Test Case Setup

To determine the cyber attack strategy for implementation in this test scenario, we analysed the overtaking scenario and its applicability to state-of-the-art attacks on AD algorithms. We selected

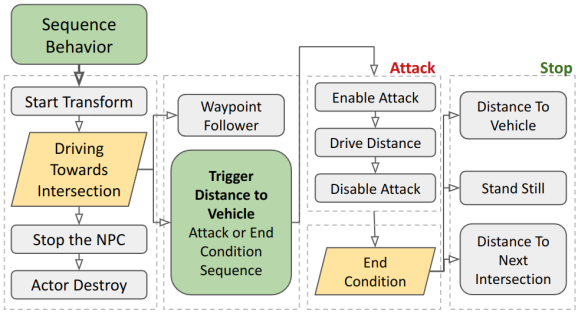

Figure 5: Flow-graph of how each scenario is processed in the simulation platform

LiDAR spoofing as it is a realistic attack in the driving environment of our real-world AV shuttle [3] and its impact is relevant to safety outcomes due to the likelihood that the manipulated driving behaviour will result in collisions, emergency breaking, and lane violations [20]. Attacks on LiDAR perception predominantly focus on spoofing LiDAR 3D point-clouds through the following means: 1) injection of adversarial LiDAR 3D point cloud data to add adversarial objects to the driving environment inducing a *false positive result* of the AD perception [3, 17] 2) removal of LiDAR 3D point cloud data to perturb the ability of the perception algorithm to detect objects in the driving environment, also known as a *false negative result* [8, 9] 3) manipulating LiDAR 3D point cloud data to obfuscate the true distance of environmental objects (Other road vehicles, pedestrians, other road objects) from the AV, causing the perception to *fail translation* 4) implementation of adversarial mesh in the driving environment to introduce manipulated points into the LiDAR 3D point cloud and create unpredictable perception events [19]. The aim of the attacker, in adversarial LiDAR threat models, is to induce the victim AV to perform dangerous driving maneuvers, which include; emergency breaking, collisions, and exceeding the limits of the driving lanes. Variables that have been shown to influence attack success include; angle of attack of the adversarial point cloud vector, density of the spoofed points, duration of the broadcast of spoofed points, distance of the point cloud to the target [3, 8, 17, 20]. We implemented a variation of the attack suggested by Yang et al. [20], where the adversary creates an adversarial roadside object to inject spoofed, malicious LiDAR point clouds into the target AV LiDAR. In our attack, an adversary has configured a LiDAR on the roadside to inject malicious point cloud data into the AV as it is conducting the overtaking manoeuvre. Figure 6 demonstrates the implementation of our attack.

Using the knowledge gained from literature [8, 17, 20], the parameters we chose to generate our attack are: density of the LiDAR point clouds, frequency (the publishing rate of the fake points), duration of the adversarial point cloud broadcast, and location, which is the relative location between the target vehicle and NPC. As an infinite number in the range of each of the parameters can be chosen, we decided to limit our testing to parameter values that had demonstrated utility to investigate the impact of cyberattacks on AD algorithms. For example, Hallyburton et al. [8] found that the success of cyber attacks increased when spoofed point density were over 80. Therefore we chose a range for spoof point density from 50 to 300.

3.5.1 Taguchi Analysis. In this study, we use the Taguchi method for statistical evaluation [18] of the attack parameters effect on each safety criterion. The number of tests with four parameters and 3 levels for each in full factorial mode would become unrealistic to perform, noting that each experiment should repeat 50 times (81x50 = 4050 distinct scenarios). A design of the experiment is recommended in order to avoid full factorial tests and reduce the number of tests without compromising accuracy [18].

A Taguchi design of experiment (DOE) technique [18] was applied to quantify the influence of four proposed attack parameters; the false points (FP) density, the FP frequency, the attack duration, and the attack location. In total, 9 experiments were designed with 3 different values for the four parameters. The analyses hence possess four factors and three levels for the Taguchi L9 matrix. Table 3 lists the configuration for each run conducted for cybersecurity tests.

Table 3: Taguchi L'9 matrix for study of factor influence

Num.	*Density*	*Frequency*	*Duration*	*Location*
1	50	5	3	3
2	50	7	6	6
3	50	10	9	9
4	150	5	6	9
5	150	7	9	3
6	150	10	3	6
7	300	5	9	6
8	300	7	3	9
9	300	10	6	3
	[50 150 300]	[5 7 10]	[3 6 9]	[3 6 9]

Figure 6 demonstrates the cyber attack setup within the overtaking scenario (Please note, the Figure only depicts the overtaking frame and not the entire overtaking sequence.). The proposed attack model will start by generating spoof points from the designated place on the roadside. At the starting point, P_1, the AV has relative distance to NPC that defines the attack location. After a specific duration (Attack Duration), the AV reaches, P_2. While the attacker keeps the malicious LiDAR pointing toward the AVs front LiDAR. Overall, the spoofed point direction changes from θ_1 to θ_2.

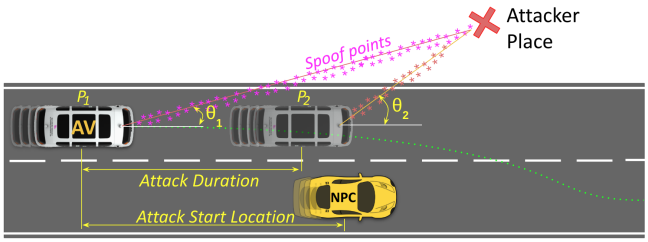

Figure 6: Attack scheme

Code was created for the generation of the adversarial LiDAR fake points to be run in the digital twin, high-fidelity simulation environment. This is available on the GitHub site [15].

4 RESULTS AND ANALYSIS

In this section, we present the results of the safety and cybersecurity testing of the end-to-end AD algorithm. The purpose of the safety test case results is to evaluate the reliability and optimisation of the algorithm.

4.1 Safety Test Case

The aim of the testing is to assess the utility of the methodology to evaluate the relationship between the reliability of the AD algorithm to safety and the impact of cybersecurity. As the testing is based on a real-world AV, we were motivated to establish what results could be gained from an amount of tests that took into account the requirements for CPU and GPU resources and the time involved in running high-fidelity simulations. For instance, 50 distinct scenarios run 3 times expends x amount of resources, and takes x amount of time. Therefore, we, firstly, performed a baseline evaluation test where we ran 50 distinct scenarios of the overtaking manoeuvre, 3 times. Each scenario is distinct based on changes to parameters such as NPC speed and initial distance to NPC.

In our proposed simulation platform, we perform 15 distinct scenarios, run 50 times; in total, 750 consecutive simulation runs were conducted. Table 4 shows the parameters of the distinct scenarios evaluated against the safety criteria. Using our configuration for testing, the AD algorithm shows the performance for the overtaking manoeuvre with a success rate of 43.9% of the simulated scenarios, whilst, 66.1% are safety violations.

In Figure 7 is the performance of the AD algorithm.

Table 4: Summary of the safety simulation

	D_x	S_{NPC}	V_{Col}	V_{DTC}	V_{BrP}	V_{BrD}	V_{NotF}	V_{Suce}
1	15	1	18%	22%	0%	10%	24%	26%
2	20	1	18%	40%	8%	6%	18%	10%
3	25	1	4%	20%	32%	8%	20%	16%
4	15	1.4	6%	32%	16%	2%	12%	32%
5	20	1.4	22%	26%	14%	6%	2%	30%
6	25	1.4	4%	12%	22%	8%	0%	54%
7	15	1.8	36%	34%	8%	2%	6%	14%
8	20	1.8	22%	12%	2%	2%	0%	62%
9	25	1.8	18%	6%	0%	4%	0%	72%
10	15	2.1	4%	0%	4%	2%	4%	86%
11	20	2.1	8%	10%	0%	0%	0%	82%
12	25	2.1	24%	0%	0%	4%	0%	72%
13	15	2.5	14%	6%	0%	6%	2%	72%
14	20	2.5	44%	22%	14%	0%	2%	18%
15	25	2.5	64%	18%	0%	0%	6%	12%
		mean	20.4%	17.3%	8.0%	4.0%	6.4%	43.9%
		STD	16.8%	2.3%	9.8%	3.2%	8.1%	28.3%
		min	4%	0%	0%	0%	0%	10%
		max	64%	40%	32%	10%	24%	86%

NPC speed is an important parameter as it influences the decision control for the critical cut-in manoeuvre of the overtaking mission. In the context of the results of the simulations, we can see that NPC speed impacts certain safety criteria.

The first such relation that can be seen, is that more collisions are caused at high speeds, > 2.1 m/s. This can be the effect of a poor trajectory evaluator that doesn't consider the prediction of the other actors motions in the process of the waypoint generation. In

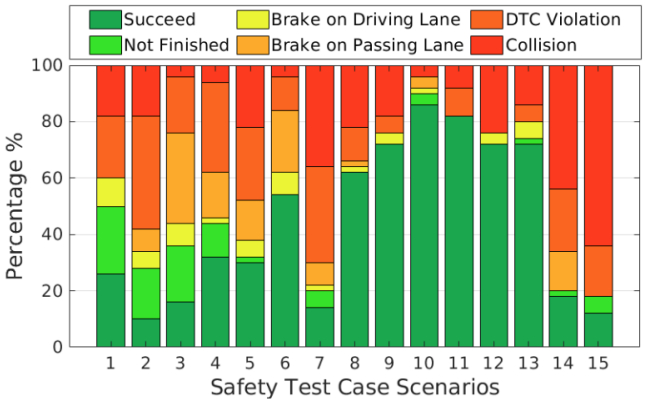

Figure 7: The 15 distinct scenarios

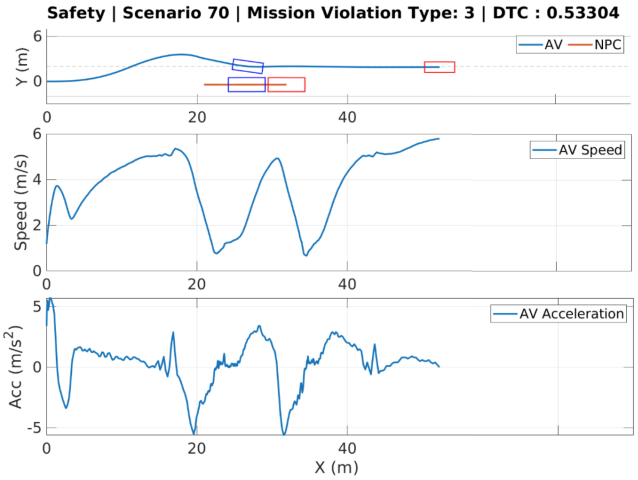

Figure 8: A Brake on Passing Lane safety violation

most collision cases the AV tried to perform a cut-in while the NPC collided from the right side. The probability of this safety violation will be increased as the NPC speed increases.

NPC speed also impacts the likelihood of a DTC safety violation. In the range of the NPC speed parameter, 1 m/s to 1.8 m/s, it can be observed that AV Shuttle violates the safe distance to the NPC. This can be due to the AV speed adjusting relative to the NPC speed and the cut-in is attempted at low-speed, whilst acceleration is required to safely attempt the cut-in. This low-speed cut-in firstly causes a DTC violation and if the overtaking manoeuvre progresses it causes a collision. DTC and collision correlate based on the relative speed. A low-speed NPC will likely result in a DTC violation, whilst in a higher-speed scenario, a collision is more likely to happen.

In the lowest speed range, 1 m/s to 1.4 m/s, it is more likely that the AV will initiate an emergency break in the passing lane. This is due to the relationship of the NPC speed to the AV Shuttle speed. The emergency break on the passing lane at low speeds is caused by a failure of the open-planner trajectory evaluator to effectively plan the overtaking trajectory. Figure 8 demonstrates the AV emergency break in the passing lane, for a scenario with an NPC Speed of 1 m/s. The upper rectangle represents the AV and the lower rectangle is the NPC. The two rectangles closest to the left represent the frame that the first emergency break on the passing lane safety violation occurs. The most right rectangles represent the end of the mission. The AV speed and the acceleration verify two hard brakes in the mission while it was in the passing lane. The failure of the trajectory planning of the open-planner algorithm is apparent.

The failure to finish the overtaking mission is most prominent at the lowest speed, 1 m/s, this is due to the time the AV Shuttle is taking to perform the cut-in process and therefore cannot enact the overtaking manoeuvre within the simulation timeout which is 40 s. It was observed that for the proposed configuration, for the lower speed of the NPC, the open-planner trajectory evaluator is not reliable as it suggests waypoints that are not within safe navigation and this is due to the lack of firm decision-making of which roll-out to choose. Ultimately, this causes collision and DTC safety violations. Furthermore, the failure to finish the simulation results, we see the low-speed delays in the overtaking manoeuvre decision making which results in the breach of the 40 s time-out.

The success rate of the safety test cases increases as the NPC drives from 1.4 to 2.1 m/s speed. This focal success point around scenario 10 with an NPC speed of 2.1 m/s can be a sign of matching the current configuration of perception and open-planner with the scenario situation.

The safety metrics results are shown in Figure 10 based on the initial relative distance from the AV to NPC. It shows that the rate of collision safety violations for longer initial distances from NPC slightly increased while the success rate decreased. This is the only trend that can be identified from results for initial relative distance, so it can be concluded that speed is a more determining parameter for the safety testing of our AV.

Overall, the results in Figure 7 indicate that speed is a critical parameter for our AV safety testing platform.

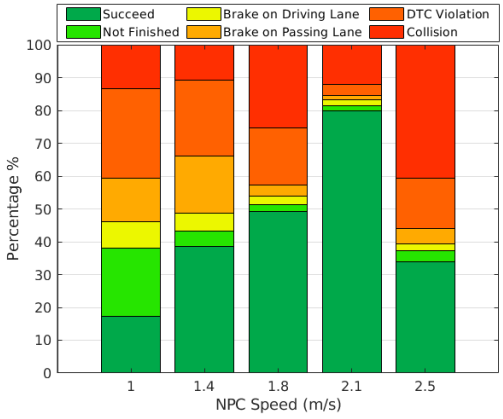

Figure 9: Test Results based on NPC Speed

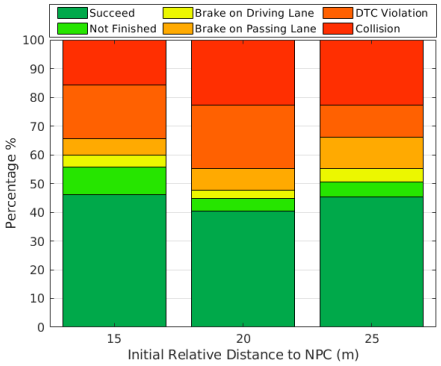

Figure 10: Results based on Initial Relative Distance to NPC

4.2 Cybersecurity Test Case

For the cybersecurity test cases we chose 2 of the 15 distinct scenarios (Figure 7). This was to allow a greater scale of testing to be conducted on a select number of relevant scenarios. Scenario 10 was chosen as it demonstrated the most reliable performance, in terms of the most successful overtaking manoeuvres. Scenario 2 was chosen as it demonstrated the least successful results for overtaking. These two scenarios were run 50 times each, as had been conducted with the safety scenario runs. Figure 11 shows the performance of cybersecurity testing, conducted on scenario 2 and scenario 10, in comparison to safety test cases.

Scenario 10 results reveal a discernible impact of the cyber attack. The LiDAR spoofing attack causes an increase in safety violations, prominently, in collisions and emergency breaking in the passing lane. This is also a concurrent result of the Scenario 2 test cases. Figure 3 shows the control level view, that incorporates sensor perception and mission and motion-planning. In the safety violation cases, we noticed that the euclidean clustering and kf_countour detect the spoofed LiDAR injection as an object and this false positive detection impacts the local-planning to force the AV to make the cut-in, in the overtaking manoeuvre process. Specifically, as the placement of the adversarial LiDAR device is on the left of the AV, the roll-outs of the left-side are blocked by the trajectory-evaluator. This forces the AV to veer right and attempt the cut-in process that causes predominantly collision, DTC safety violations.

Cao et al. [3] and Hallyburton et al. [8] identify density of the spoofed points to be one of the key variables affecting cyber attack success rate. Figure 12 and figure 13 present the sensitivity of each attack parameter according to the cyber attack test cases. From evaluating the raw data of the test sets, and the sensitivity analysis for the cyber attack test cases of scenario 10, we concur with these assessments. We find the rate of collisions is influenced by the density of the point cloud and the location of the attack. We can also see the influence the point of attack and duration have on causing a break on passing lane safety violation. As the duration of transmitting of the LiDAR point clouds increases and the location of the attack is further from the NPC, the likelihood of the AV initiating its breaks is higher.

In comparison, Scenario 2 cyber attack test case results show that safety violations are less sensitive to attack parameters. This can be due to the difficulty in interpreting the impact of cybersecurity on this scenario due to the already high rate of safety violations of the algorithms exhibited in the safety test case.

Table 5: Results of Cyber Attack applied to Scenario 10

Num.	V_{Col}	V_{DTC}	V_{BrP}	V_{BrD}	V_{NotF}	V_{Suce}
1	54%	20%	2%	0%	6%	18%
2	38%	38%	6%	2%	6%	10%
3	30%	28%	22%	2%	4%	14%
4	24%	28%	16%	6%	2%	24%
5	26%	16%	12%	6%	4%	36%
6	4%	4%	6%	4%	0%	82%
7	32%	14%	14%	6%	0%	34%
8	50%	24%	8%	2%	0%	16%
9	50%	30%	2%	2%	0%	16%
mean	34.2%	22.4%	9.8%	3.3%	2.4%	27.8%
std	15.9%	10.1%	6.7%	2.2%	2.6%	22.2%
min	4.0%	4.0%	2.0%	0.0%	0.0%	10.0%
max	54.0%	38.0%	22.0%	6.0%	6.0%	82.0%

Table 6: Results of Cyber Attack applied to Scenario 2

Num.	V_{Col}	V_{DTC}	V_{BrP}	V_{BrD}	V_{NotF}	V_{Suce}
1	16%	34%	28%	8%	14%	0%
2	26%	34%	20%	0%	8%	12%
3	20%	42%	20%	4%	6%	8%
4	26%	34%	16%	0%	14%	10%
5	22%	36%	16%	0%	20%	6%
6	22%	32%	20%	0%	18%	8%
7	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%
mean	14.7%	23.6%	13.3%	1.3%	8.9%	4.9%
std	11.4%	17.9%	10.6%	2.8%	7.9%	4.9%
min	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
max	26.0%	42.0%	28.0%	8.0%	20.0%	12.0%

4.3 Real-World AV Testing

The real-world AV testing was conducted on a private road environment using our AV Shuttle, and an NPC vehicle (turquoise Mitsubishi iMIEV). The NPC vehicle is stationary during the tests as a safety assessment deemed it was too dangerous to conduct the experiment with a moving vehicle. This is due to the experiment being within a road environment where pedestrians and other vehicles are present. We conducted 3 test cases; a safety test case, cybersecurity test case and an optimised cybersecurity test case. The first test was an overtaking safety scenario. Two repetitions of the safety test case were conducted. The first test demonstrated a successful execution of the overtaking mission. The second test resulted in a DTC safety violation. The AV motioned to within 0.42 m of the NPC. The DTC violation is evident in Frame 3 of Figure 14, which details the second overtaking safety test case. Frame 4 demonstrates the eventual overtake after the DTC safety violation. Whilst the number of repetitions in the real-world pale in comparison to those conducted in the simulator, the real-world results

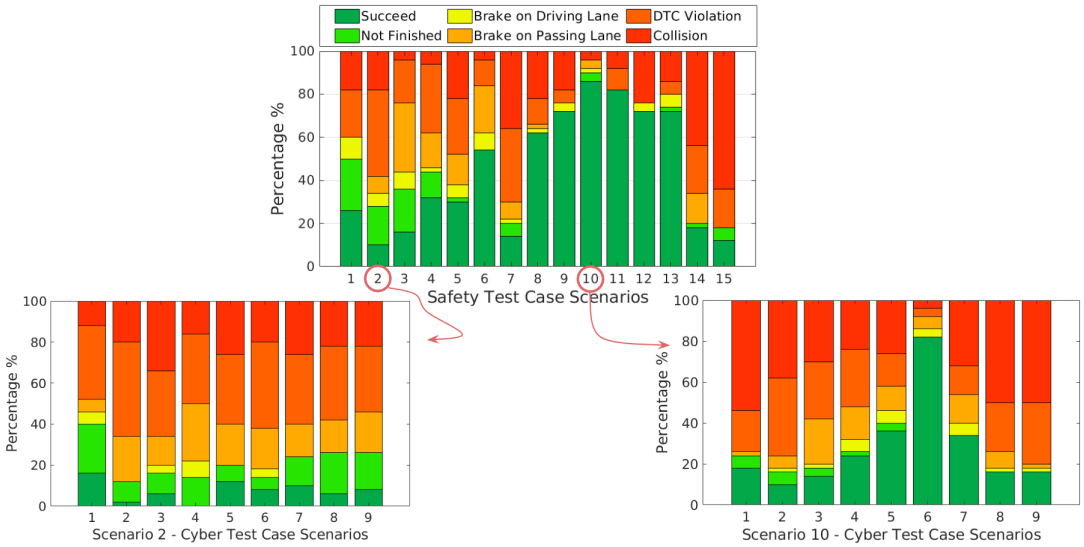

Figure 11: Performance Results Comparing Cyber Vs Safety Test Cases

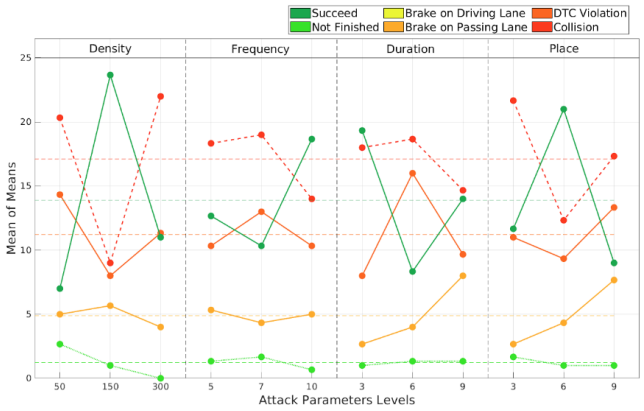

Figure 12: Scenario 10 - Cyber Attack Test Cases - Parameter Sensitivity

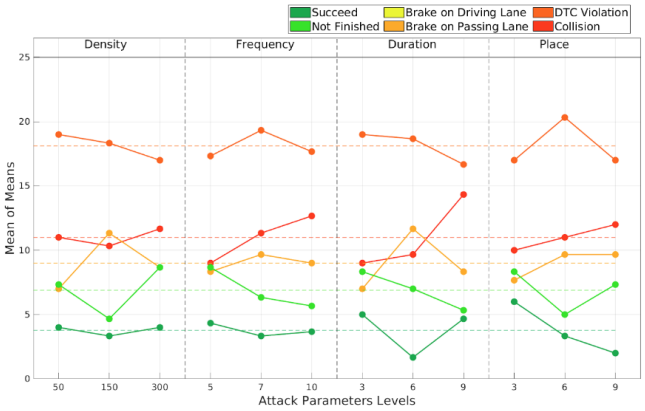

Figure 13: Scenario 2 - Cyber Attack Test Cases - Parameter Sensitivity

Table 7: Result of the 3 real-world test cases

Test Type	Num. of repeats	success	Safety Violations
Safety Tests	2	1	1 DTC=0.42m
Cyber Tests	2	1	1 DTC=0.38m
Optimised Cyber Tests	1	0	1 DTC=0.32m

concur with simulation results, that the AD algorithm does not have enough reliability for the deployment in real-world missions.

The cybersecurity test was conducted 3 times. Table 7 lists all the real-word experiments and their results. The first cybersecurity test demonstrated no impact from the spoofed LiDAR points and the overtaking manoeuvre was successful. The second cybersecurity test resulted in a DTC violation, the AV motioned to within 0.38 m of the NPC. After these two tests, we optimised the target angle of the spoofed points in relation to the attack scheme in Figure 6, to reduce the attack starting angle of θ_1. We did this because during the real-world test we observed that the reduced angle would provide assist the spoofed points to be closer to the AV trajectory and would cause the AV to detour from its intended route. It can be seen that this did work as the DTC decreased to 0.32 m. Figure 15 depicts the real-world cybersecurity test. Frame 2 represents the moment the attack was generated and perceived by the AD algorithm.

The videos and images related to the real-world tests are found on GitHub site.

5 DISCUSSION

From the analysis of the results we interpreted that different safety violations are connected to different modules of the AD algorithm.

Perception Module) We interpreted the cause of safety violations of the emergency break in the passing lane and emergency break in the driving lane to be related to the quality of the ground filtration. As we observed, ground filtering outcome changes during the AV maneuvers (turns) because the shuttle body is tilted because of suspension and this results in the lidar reference frame orientation

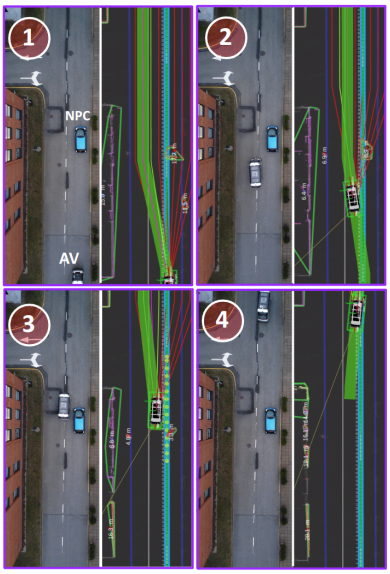

Figure 14: Real-World AV Test - Safety Test Case

Figure 15: Real-World AV Test - Cyber Attack Test Case

changing. Then some part of the ground point cloud as an unfiltered perception can be seen in the detection algorithms as an obstacle. This fake sudden obstacle might stop the AV during the motion. The spoofed LiDAR point cloud threat model is likely to make this condition worse. Optimisations for this: New body designs to rectify or limit the issues of LiDAR with the physics of the AV Shuttle are being developed. To focus specifically on these corner and edge cases and look at optimisation of the filtering of the perception algorithm. The latter recommendation is complicated by the fact it may include trade-offs; if the LiDAR perception algorithm is specifically tuned for this corner/edge case it could lead to over-filtration in normal driving scenarios, therefore this is one of the optimisation options to resolve the perception for the algorithm.

Open-Planner Module) We interpret the cause of safety violations for DTC and collision as due to an issue of the open-planner in predicting the trajectory of the NPC during the process of performing a cut-in, in front of the NPC. The optimisation would involve incorporation of features that would enable the prediction of the trajectory of the NPC and for perception improve the perception of the side-lidar to accurately perceive the NPC. We found that optimising all the perception and open-planner parameters for our shuttle model would significantly improve the reliability of the AD algorithm.

5.1 Open-Planner Developer Feedback

We sent a presentation of our results to the developers of the open-planner AD algorithm. In response, they acknowledged that it is a developing algorithm and we are engaged in more detailed discussions with them on how to optimise the algorithm. They also announced they are transitioning from Autoware.ai to Autoware.universe which is a more developed and advanced platform. Amongst their responses, they also pointed to the novelty of receiving feedback on the reliability of cybersecurity test cases in addition to safety test cases.

6 RELATED WORK

The closest contributions to our work are Yang et al. [20], Hallyburton et al. [8], Cao et al. [3] and Zhu et al. [21]. Each of these papers utilises a LiDAR spoofing threat model that varies based on the method for delivering the attack, adversarial generation and the type AD algorithm. Hallyburton et al. [8] target camera and LiDAR sensor fusion. They identify a blind spot between the camera and LiDAR sensor at the rear of the target AV. They use a malicious, 3D LiDAR point cloud array to inject malicious spoof points into the rear angle of the target AV. The attack was tested in a high-fidelity simulation and real-world against multiple perception algorithms. The results revealed a high rate of success utilising this attack. Cao et al [3], Yang et al [20], and Zhu et al [21] developed LiDAR spoofing attacks based on a threat model of a malicious LiDAR 3D point cloud injection in the road environment and by the roadside. Each of these contributions demonstrated that cyber attack results from AV simulation testing can be used to identify key parameters such as point cloud density, attack location and duration and that these parameters can be optimised to test the robustness of perception algorithms. We chose to extend from the related literature, in our work, in three areas; simulation testing configuration, safety criteria evaluation and target AD algorithm is in the developmental phase and is used within a real-world AV program. A feature of the selected work is that simulation testing often selected only one frame or a limited amount of frames and therefore the full driving mission was not observed. Whilst this is useful for reducing testing resource usage, running massive scale of tests and applicable to the scope of their work, as our study evaluates the end-to-end AD algorithm and combines safety, our study focused on conducting simulation testing for the entire driving mission. Secondly, the evaluation of cyber attacks focused on attack success rate and attack parameters whilst the safety impact on the AV as a result of cyber attacks was not as clearly elaborated. In our study, we evaluate the cyber attack test cases with the same criteria

as the safety case to derive the category of safety violation. Lastly, most of the simulations use default AV configurations and evaluate well-established algorithms. Our study uses a simulator configured for a real-world AV and evaluates an AD algorithm in the developmental stage where reliability and optimisation are required to be assessed under safety, non-cyber test cases before the impact of cyber attacks can be understood.

7 CONCLUSION

We developed a combined methodology for safety and cybersecurity utilising a digital twin, high-fidelity simulation environment and a real-world AV shuttle for public transportation. We evaluated our approach on a developing AD algorithm consisting of open-planner, as the mission and motion-planning module. We evaluated the reliability of the AD algorithm on an overtaking scenario using test cases for safety and cybersecurity based on a LiDAR spoofing attack. The combined safety and cybersecurity testing enabled us to assess the outcome of the cyber attack in comparison to the ground truth of the reliability of the AD algorithm established in the safety testing. This clearly demonstrated the effect of cyber-attacks regardless of the reliability of the algorithm. We were also able to assess, from the performance of the AD algorithm, that the algorithm is not optimised for the overtaking manoeuvre. In our research, we discovered several sensitive parameters that play a significant role in the safety outcome of the AV and the success rate of the cyber attack. Furthermore, we provided the results of our testing platform to the designer of the open-planner algorithm. Based on their feedback a process has been initiated to optimise the AD algorithm. All test scripts and software resources including our AV simulation configurations and research data used in the combined safety and security testing will be available for the research community on GitHub.

7.1 Future Work

Future work consists of diversifying the safety scenarios to include a more complex and broader range of scenarios. Cybersecurity testing will be evolved to develop black-box testing models. Furthermore, we will continue to develop methods for optimising testing to factor in real-world limitations such as resource usage and time.

ACKNOWLEDGMENTS

This research has received funding from the following grants: the European Union's Horizon 2020 Research and Innovation Programme, under grant agreements No. 856602 and No 883321 (CityScape), and the European Regional Development Fund, co-funded by the Estonian Ministry of Education and Research, under grant agreement No 2014-2020.4.01.20-0289.

REFERENCES

[1] Euro NCAP Working Group on Automated Driving. 2019. *Euro NCAP's First step to assess automated driving systems.* Technical Report. European New Car Assessment Programme.

[2] Adith Boloor, Karthik Garimella, Xin He, Christopher Gill, Yevgeniy Vorobeychik, and Xuan Zhang. 2020. Attacking vision-based perception in end-to-end autonomous driving models. *Journal of Systems Architecture* 110 (2020). https://doi.org/10.1016/j.sysarc.2020.101766 arXiv:1910.01907

[3] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z. Morley Mao. 2019. Adversarial Sensor Attack on LiDAR-Based Perception in Autonomous Driving. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 2267–2281. https://doi.org/10.1145/3319535.3339815

[4] CARLA Simulation Project. 2022. *Combined Safety and Cybersecurity Testing Methodology for Autonomous Driving Algorithms.* Technical Report. CARLA.

[5] Hatem Darweesh, Eijiro Takeuchi, and Kazuya Takeda. 2021. OpenPlanner 2.0: The Portable Open Source Planner for Autonomous Driving Applications. In *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*. 313–318. https://doi.org/10.1109/IVWorkshops54471.2021.9669253

[6] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An Open Urban Driving Simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*. 1–16.

[7] Junyao Guo, Unmesh Kurup, and Mohak Shah. 2020. Is it Safe to Drive? An Overview of Factors, Metrics, and Datasets for Driveability Assessment in Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems* 21, 8 (2020), 3135–3151. https://doi.org/10.1109/TITS.2019.2926042

[8] R. Spencer Hallyburton, Yupei Liu, Yulong Cao, Z. Morley Mao, and Miroslav Pajic. 2022. Security Analysis of Camera-LiDAR Fusion Against Black-Box Attacks on Autonomous Vehicles. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, 1903–1920. https://www.usenix.org/conference/usenixsecurity22/presentation/hallyburton

[9] Zhongyuan Hau, Kenneth T Co, Soteris Demetriou, and Emil C Lupu. 2021. Object Removal Attacks on LiDAR-based 3D Object Detectors. In *Third International Workshop on Automotive and Autonomous Vehicle Security (AutoSec)*.

[10] ISO/TC 22/SC 32 Electrical and electronic components and general system aspects. 2018. *ISO 26262-1:2018 Road vehicles — Functional safety.* Technical Report. International Standards Organization.

[11] Shinpei Kato, Shota Tokunaga, Yuya Maruyama, Seiya Maeda, Manato Hirabayashi, Yuki Kitsukawa, Abraham Monrroy, Tomohito Ando, Yusuke Fujii, and Takuya Azumi. 2018. Autoware on board: Enabling autonomous vehicles with embedded systems. In *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 287–296.

[12] Yingqi Liu, Shiqing Ma, Yousra Aafer, Wen-Chuan Lee, Juan Zhai, Weihang Wang, and X. Zhang. 2018. Trojaning Attack on Neural Networks. In *NDSS*.

[13] Ehsan Malayjerdi, Raivo Sell, Mohsen Malayjerdi, Andres Udal, and Mauro Bellone. 2022. Practical path planning techniques in overtaking for autonomous shuttles. *Journal of Field Robotics* 39, 4 (2022), 410–425.

[14] Mohsen Malayjerdi, Vladimir Kuts, Raivo Sell, Tauno Otto, and Barış Cem Baykara. 2020. Virtual Simulations Environment Development for Autonomous Vehicles Interaction. In *ASME International Mechanical Engineering Congress and Exposition*. American Society of Mechanical Engineers.

[15] Mohsen Malayjerdi, Andrew Roberts, Olaf Maennel, and Ehsan Malayjerdi. 2022. *Combined Safety and Cybersecurity Testing Methodology for Autonomous Driving Algorithms.* https://github.com/momala/Safety_Cyber_Testing.git

[16] Raivo Sell, Ehsan Malayjerdi, Mohsen Malayjerdi, and Baris Cem Baykara. 2022. Safety Toolkit for Automated Vehicle Shuttle -Practical Implementation of Digital Twin. In *2022 International Conference on Connected Vehicle and Expo (ICCVE)*. 1–6. https://doi.org/10.1109/ICCVE52871.2022.9742881

[17] Jiachen Sun, Yulong Cao, Qi Alfred Chen, and Z. Morley Mao. 2020. Towards robust LiDAR-based perception in autonomous driving: General black-box adversarial sensor attack and countermeasures. *Proceedings of the 29th USENIX Security Symposium* (2020), 877–894. arXiv:2006.16974

[18] KWOK-LEUNG TSUI. 1992. AN OVERVIEW OF TAGUCHI METHOD AND NEWLY DEVELOPED STATISTICAL METHODS FOR ROBUST DESIGN. *IIE Transactions* 24, 5 (1992), 44–57. https://doi.org/10.1080/07408179208964244 arXiv:https://doi.org/10.1080/07408179208964244

[19] James Tu, Huichen Li, Xinchen Yan, Mengye Ren, Yun Chen, Ming Liang, Eilyan Bitar, Ersin Yumer, and Raquel Urtasun. 2021. Exploring Adversarial Robustness of Multi-sensor Perception Systems in Self Driving. In *5th Annual Conference on Robot Learning*. https://openreview.net/forum?id=m5k1XdK5nI2

[20] Kaichen Yang, Tzungyu Tsai, Honggang Yu, Max Panoff, Tsung-Yi Ho, and Yier Jin. 2021. Robust Roadside Physical Adversarial Attack Against Deep Learning in Lidar Perception Modules. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security* (Virtual Event, Hong Kong) *(ASIA CCS '21)*. Association for Computing Machinery, New York, NY, USA, 349–362. https://doi.org/10.1145/3433210.3453106

[21] Yi Zhu, Chenglin Miao, Tianhang Zheng, Foad Hajiaghajani, Lu Su, and Chunming Qiao. 2021. Can We Use Arbitrary Objects to Attack LiDAR Perception in Autonomous Driving? *Proceedings of the ACM Conference on Computer and Communications Security* (2021), 1945–1960. https://doi.org/10.1145/3460120.3485377