ISSN: 2710-0588

Research on Attack Risk Analysis and Testing Verification Method for LiDAR in the Environmental Perception Layer of Autonomous Driving

Yihong Qin[1, a], Ziyi Wang[1, b], Xiong Zhao[1, c, *], Xuesong Wu[1, d]

[1]CATARC Intelligent Technology (Tianjin) Co., Ltd., Tianjin, China

[a]qinyihong@catarc.ac.cn, [b]wangziyilucky@catarc.ac.cn, [c]zhaoxiong@catarc.ac.cn, [d]wuxuesong@catarc.ac.cn

[*] Corresponding Author

Abstract

Environmental perception is the core foundation for the operation of autonomous driving systems. With advantages such as high-precision 3D ranging and resistance to light interference, LiDAR has become a key sensor for environmental perception in high-level autonomous driving. As the level of autonomous driving continues to improve, the number and density of in-vehicle LiDAR deployments continue to increase, and its operational safety directly determines the driving safety of the entire vehicle. LiDAR realizes environmental modeling by transmitting and receiving laser beams, and has potential attack risks of being interfered, spoofed, and suppressed by the outside world at the physical layer, link layer, and protocol layer, which may lead to problems such as target missed detection, false alarms, trajectory distortion, and perception failure, and then trigger safety accidents such as vehicle sudden braking, abnormal detouring, and collisions. Current autonomous driving functional safety and information security standards have insufficient test coverage for active attack risks of LiDAR, lacking a systematic risk analysis framework and engineering testing verification process. This paper conducts systematic research on the attack risks of LiDAR in the environmental perception layer of autonomous driving, sorts out typical attack methods such as physical interference, protocol attacks, and point cloud data poisoning, analyzes the influence mechanism of various attacks on perception algorithms and driving safety, constructs a multi-scenario and multi-dimensional attack risk testing verification method, and forms a complete testing system including attack simulation, data collection, effect quantification, and security evaluation. Verified through joint testing of real vehicles and simulation platforms, this method can effectively reproduce the typical attack effects of LiDAR and quantify the perception performance attenuation indicators, providing theoretical basis and engineering practice reference for the safety design, risk evaluation, and protection hardening of autonomous driving LiDAR.

Keywords

Autonomous Driving; Environmental Perception; LiDAR; Attack Risk; Testing Verification; Perception Security.

1. Introduction

High-level autonomous driving systems highly rely on the environmental perception module to identify, track, and predict the surrounding targets, road structures, and traffic participants of the vehicle in real time, accurately, and stably. Compared with cameras and millimeter-wave radars, LiDAR can obtain high-density 3D point cloud information to achieve fine depiction of

ISSN: 2710-0588

obstacle contours, distances, and relative speeds, and has significant perception advantages in complex scenarios such as nighttime, backlight, rain, and fog, so it is widely used in L3 and above autonomous driving systems.

LiDAR is essentially an active optical detection sensor, and its working principle determines that it is vulnerable to external optical and electromagnetic signals. In recent years, research institutions and security manufacturers at home and abroad have successively confirmed that through specific wavelength laser interference, synchronous pulse spoofing, high-power strong light suppression and other methods, in-vehicle LiDAR can suffer from point cloud loss, target blurring, distance jumping, and even perception "blindness". In real traffic scenarios, such attacks may cause the autonomous driving system to fail to identify key targets such as pedestrians, vehicles, and guardrails, directly inducing traffic safety accidents, with extremely high practical harm.

At present, most industry research on LiDAR focuses on performance improvement directions such as point cloud segmentation, target detection, and multi-sensor fusion, and research on the active attack risk of LiDAR is still in the initial stage. Existing automotive information security standards such as ISO/SAE 21434 mostly focus on software-level vulnerabilities such as in-vehicle networks, bus protocols, and in-vehicle systems, lacking clear testing requirements for sensor physical layer attacks, optical interference, and data layer tampering. At the same time, LiDAR attacks are characterized by strong concealment, high reproduction difficulty, and complex influencing factors, and a unified, standardized, and implementable testing verification method has not yet been formed in the industry.

Based on this, this paper systematically analyzes the typical attack types faced by LiDAR in the environmental perception layer of autonomous driving, reveals the action mechanism and hazard degree of different attack methods on perception results, constructs a full-process testing verification method covering attack simulation, scenario injection, data comparison, and effect evaluation, and verifies the feasibility of the method through simulation and real-vehicle testing, providing technical support for the construction of a safety evaluation system for autonomous driving LiDAR[1].

2. Analysis of LiDAR Working Principle and Potential Attack Surfaces

In-vehicle LiDAR mainly transmits pulsed laser signals, which are reflected by the target and received by the receiving end to obtain echo signals. The target distance is calculated based on Time of Flight (ToF) or phase difference, and a 3D point cloud map is formed combined with the scanning mechanism. The perception algorithm completes target detection, classification, tracking, and trajectory prediction based on point cloud information, and outputs it to the decision-making and planning layer to execute corresponding driving behaviors. The complete link of LiDAR from signal transmission to perception output has attack nodes that can be exploited.

Divided by attack links, the main attack surfaces of LiDAR can be divided into three levels: physical optical layer, data link layer, and perception algorithm layer. Physical optical layer attacks directly act on the laser transmission and reception link, interfering with echo detection through external optical signals, which is the most easily implemented and directly harmful attack method at present; data link layer attacks target point cloud transmission protocols, interface communications, and timing synchronization mechanisms, destroying perception input by tampering, forging, or delaying data; perception algorithm layer attacks induce misidentification, missed detection, or false alarms of the algorithm by constructing special targets or noise point clouds[2].

To intuitively reflect the risk characteristics and hazard degree of different attack levels, the LiDAR attack risk is classified, as shown in Table 1.

ISSN: 2710-0588

Table 1. LiDAR Attack Risk Level Classification

Risk Level	Attack Level	Typical Attack Methods	Impact Scope	Hazard Degree
Extremely High Risk	Physical Optical Layer	Strong Light Suppression, Synchronous Spoofing, Distance Tampering	Global Point Cloud Failure, Complete Target Loss	Directly leads to perception collapse
High Risk	Physical Optical Layer	Point Cloud Occlusion, Local Interference, Echo Saturation	Local Target Missed Detection, Abnormal Trajectory	Prone to dynamic target collisions
Medium Risk	Data Link Layer	Protocol Tampering, Timing Attack, Data Packet Loss	Point Cloud Delay, Coordinate Distortion, Fusion Abnormality	Reduces system stability
Low Risk	Perception Algorithm Layer	Adversarial Point Cloud, False Alarm Injection, Noise Poisoning	Increased Misdetection Rate, Small Target Missed Detection	Affects riding comfort

Physical layer attacks do not require access to the in-vehicle network and can be implemented only through external equipment, featuring low attack cost, strong concealment, and difficult traceability, making them the main security threat to autonomous driving environmental perception. Strong light suppression attack uses continuous or pulsed laser of the same band as the target radar to directly irradiate the radar receiving lens. When the receiving-end photodetector receives optical power exceeding the threshold, it will enter a deep saturation state, unable to distinguish real echo signals from interference signals, eventually showing a sharp reduction in the number of point clouds, a significant shortening of effective detection distance, and complete loss of long-distance targets.

Synchronous spoofing attack is a precise physical attack. The attacker detects the pulse repetition frequency and transmission timing of the target LiDAR and synchronously transmits forged echo pulses. After receiving the false signal, the radar system will judge it as a real target echo, thus calculating wrong distance and position information.

This attack can achieve two typical spoofing effects: first, generating virtual target point clouds to make the system misjudge that there are obstacles ahead and trigger braking; second, canceling real target echoes to make key objects such as vehicles and guardrails "disappear" in the point cloud, causing serious missed detection. Synchronous spoofing attack has high precision and strong concealment, and is difficult to identify by conventional anomaly detection mechanisms[3].

Local occlusion attack transmits weak interference signals at a specific angle to make the radar appear point cloud holes in the corresponding area, thereby occluding specific targets such as pedestrians, non-motor vehicles, and traffic signs. Echo interference changes the target surface reflection characteristics or transmits multi-channel scattered signals, causing scattered point clouds, blurred target contours, and reduced classification confidence, making the perception algorithm unable to correctly identify the target type.

Data link and protocol attacks are based on the cracking of LiDAR communication protocols. By accessing the in-vehicle Ethernet or CAN bus, tampering with point cloud data packets, delaying

ISSN: 2710-0588

data transmission, and forging frame header and tail information, the perception algorithm input is abnormal. Typical manifestations include point cloud coordinate jumping, discontinuous timestamps, and misalignment of multi-frame data, leading to unstable target tracking, sudden trajectory changes, and expanded fusion positioning deviation.

Perception algorithm adversarial attack constructs special physical targets to make LiDAR point clouds produce misleading information at the feature level. For example, using high-reflectivity materials to make special-shaped devices to make the algorithm misjudge them as vehicles or pedestrians; or arranging interference structures on the surface of real vehicles to make their point cloud features be judged as non-obstacles, thus achieving "invisibility". This attack is targeted at specific perception models, and the hazard degree is strongly related to the algorithm architecture.

3. LiDAR Attack Risk Testing Verification Method

To systematically evaluate the anti-attack capability of LiDAR, this paper constructs an integrated testing verification method of "Attack Simulation – Scenario Injection – Data Collection – Effect Quantification – Security Evaluation", covering a three-level verification system of simulation testing, bench testing, and real-vehicle testing, with complete testing process, quantifiable indicators, and reproducible results.

3.1. Overall Testing Framework

The testing system is mainly composed of five parts: LiDAR equipment, attack simulation platform, environmental simulation platform, data collection system, and perception algorithm verification platform. The attack simulation platform is responsible for outputting signals such as strong light suppression, synchronous spoofing, and local interference; the environmental simulation platform constructs typical scenarios such as urban roads, highways, and park closed roads; the data collection system synchronously records original point clouds, target true values, attack parameters, and algorithm output results; the perception algorithm platform is used to evaluate changes in indicators such as detection accuracy, missed detection rate, and false alarm rate before and after attacks[4].

3.2. Attack Simulation and Scenario Design

According to risk levels and attack methods, multi-dimensional testing scenarios are designed: Static attack testing: Fix the positions of radar and attack equipment to test the attack effect under different powers, angles, and distances; Dynamic attack testing: Simulate vehicle driving state to carry out interference and spoofing testing under relative motion conditions; Typical target attack testing: Conduct occlusion and spoofing testing for key targets such as pedestrians, cars, trucks, and guardrails;Multi-radar cooperative attack testing: Verify the interference coverage ability of attack equipment on radars at different positions under multi-radar deployment.

3.3. Testing Evaluation Index System

To achieve quantitative evaluation of attack effects, a multi-dimensional performance attenuation evaluation system is established, as shown in Table 2.

ISSN: 2710-0588

Table 2. LiDAR Attack Testing Evaluation Index System

Testing Dimension	Core Indicator	Calculation Method	Evaluation Standard
Point Cloud Integrity	Effective Point Cloud Rate	Effective Points after Attack / Original Points	Decrease ≤10% is qualified
Target Detection Capability	Missed Detection Rate, False Detection Rate	Missed Targets / Total Targets	Missed Detection Rate = 0 is excellent
Ranging Accuracy	Average Distance Error	Deviation between Measured Value and True Value	Error ≤0.1m is qualified
Anti-interference Capability	Attack Saturation Threshold	Minimum Interference Power when Failure Occurs	Higher threshold means better security
System Stability	Point Cloud Frame Loss Rate, Delay	Abnormal Frames / Total Frames	Frame Loss Rate ≤1% is qualified

By comparing indicator changes before and after attacks, the performance attenuation amplitude of LiDAR under different attacks can be objectively evaluated to determine whether it meets the requirements for safe operation of autonomous driving.

3.4. Testing Implementation Process

Baseline testing: Collect original LiDAR point clouds and perception results in a non-attack environment to establish a performance benchmark;

Attack injection: Start the attack equipment according to preset parameters and keep the scenario and targets unchanged;

Data collection: Synchronously collect radar point clouds, true value data, algorithm output, and vehicle status information;

Indicator calculation: Count key indicators such as point cloud integrity, missed detection rate, and ranging error;

Result judgment: Determine the risk level according to evaluation standards and form test conclusions and improvement suggestions[5].

4. Testing Verification and Result Analysis

To verify the effectiveness of the proposed method, real-vehicle testing is carried out in a closed park environment. The test object is an autonomous driving test vehicle equipped with a 128-line mechanical rotating LiDAR. The attack equipment supports two modes: strong light suppression and synchronous spoofing. The test targets include typical traffic participants such as static pedestrians, dynamic vehicles, and road guardrails.

The testing is divided into two groups: baseline testing and attack testing, and the comparison results are shown in Table 3.

ISSN: 2710-0588

Table 3. Comparison of LiDAR Attack Test Results

Testing Item	Baseline State	Strong Light Suppression Attack	Synchronous Spoofing Attack
Effective Point Cloud Rate	99.2%	31.5%	67.3%
Pedestrian Missed Detection Rate	0%	68.4%	29.7%
Vehicle Missed Detection Rate	0%	42.1%	15.8%
Average Distance Error	0.05m	0.83m	1.27m
Number of False Alarm Targets	0	1	4

The test results show that strong light suppression attack has the most significant damage to LiDAR point cloud integrity, long-distance targets are basically unrecognizable, and the pedestrian and vehicle missed detection rates rise sharply; although synchronous spoofing attack has less impact on the number of point clouds, it introduces obvious ranging errors and virtual targets, easily leading to abnormal vehicle braking or detouring. In dynamic driving scenarios, the attack effect decreases slightly, but it can still cause continuous missed detection of key targets, with significant potential safety hazards.

At the same time, the testing also verifies that the proposed testing method has good reproducibility, quantifiable indicators, and can accurately reflect the security risks of LiDAR under different attacks, providing data support for sensor anti-interference design, perception algorithm hardening, and multi-sensor fusion redundancy strategies.

5. Conclusion and Prospect

Aiming at the LiDAR security problem in the environmental perception layer of autonomous driving, this paper systematically sorts out typical attack methods at the physical optical layer, data link layer, and perception algorithm layer, analyzes the implementation mechanism and hazard level of various attacks, and constructs a complete testing verification method including attack simulation, scenario design, data collection, indicator quantification, and result evaluation. Verified through real-vehicle testing, this method can effectively reproduce the performance attenuation phenomenon of LiDAR under attacks such as strong light suppression and synchronous spoofing, and accurately quantify key indicators such as point cloud integrity, target missed detection rate, and ranging accuracy, with engineering application value.

The research results show that physical layer attacks are the main security threat to LiDAR, which can directly lead to perception failure and induce driving safety accidents, and existing autonomous driving systems generally lack targeted protection mechanisms. In the future, further research can be carried out on multi-sensor joint anti-attack testing, develop technologies such as LiDAR optical protection, point cloud anomaly detection, and real-time attack identification, and promote the establishment of in-vehicle LiDAR information security testing standards to improve the operational safety and reliability of autonomous driving systems under active attack scenarios.

ISSN: 2710-0588

References

[1] ISO/SAE 21434. Road vehicles—Cybersecurity engineering, 2021.

[2] State Administration for Market Regulation. Classification of Automotive Driving Automation GB/T 40429-2021, 2021.

[3] China Automotive Technology and Research Center Co., Ltd. Technical Specifications for Testing and Evaluation of Information Security of Intelligent Connected Vehicles, 2024.

[4] Zhang M, Liu M. Research on Anti-interference Performance Testing Method of In-vehicle LiDAR. Automotive Engineering, 2024.

[5] Li M, Wang W. A Review of Physical Layer Attacks and Defense Technologies for Autonomous Driving Sensors. Acta Automatica Sinica, 2023.