International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Physical Adversarial Attacks on LiDAR-Based Perception Systems in Autonomous Vehicles: A Taxonomy, Analysis, and Defense Survey

Prajna Anand

SOCSE, RV University, RV Vidyanikethan Post, 8th Mile, Mysore Rd, Bengaluru, 560059

Abstract

Somewhere underneath all the excitement around self-driving cars is an assumption most people never think to question: that the sensors feeding these vehicles actually see the world as it is. For LiDAR, that assumption has held up well enough in ordinary driving conditions. But a growing pile of research — going back roughly five years and picking up pace recently — has started to poke serious holes in it. Researchers have inserted phantom objects into point clouds, erased genuine obstacles from them, and done both things using nothing more elaborate than a mirror. This paper looks at what that body of work actually shows. It organises the main attack types — spoofing, relay attacks, jamming, and adversarial surfaces — examines the evidence behind each, and then asks honestly what the proposed defences are actually worth. The short answer, unfortunately, is that no individual defence is anywhere near sufficient on its own, and that the most realistic path to robustness involves stacking multiple independent protective layers rather than betting everything on any single mechanism.

Keywords: Autonomous vehicles, LiDAR, adversarial attacks, point cloud, sensor spoofing, sensor fusion, deep learning, physical security, perception systems.

I. INTRODUCTION

IT says something about the state of autonomous vehicle security that one of the most serious vulnerabilities in the perception stack has nothing to do with code. LiDAR does not run a model. It does not learn anything. It sends out laser pulses and waits to see how long they take to bounce back. That, on the face of it, sounds robust — a simple physical measurement, hard to fake. And yet it turns out to be surprisingly easy to interfere with, and the consequences of doing so can run all the way through to how the vehicle actually drives.

The reason the vulnerability exists is almost embarrassingly simple. When a LiDAR sensor receives a return signal, it has no mechanism for checking whether that signal is the reflection of its own outgoing pulse or something else. Under normal conditions, it is of course its own pulse coming back — so the assumption gets baked in. An attacker who recognises this can exploit it. Feed the receiver a signal at the right wavelength and the right timing, and the sensor will treat it as legitimate. What happens next depends on what the signal was designed to imply: a vehicle that is not there, an obstacle in the wrong place, or nothing at all where there should be something.

Until a few years ago, this was mostly treated as a theoretical worry. That changed around 2019–2020 when experimental work started demonstrating actual attacks on actual systems. By 2025 the picture had

become significantly more alarming, with one study showing that an attacker does not even need powered hardware — a carefully placed mirror is enough to mess with what a vehicle perceives and how it plans to move [1]. At that point, it becomes difficult to treat the problem as academic.

This paper works through the attack landscape methodically. The goal is not to be alarmist but to be clear: about what has been shown to work, about what defences are available and what they genuinely cover, and about where the research community still has real work to do.

A. Scope and Organisation

Only physical-world attacks are covered here — the ones that operate on the sensor hardware or the signals around it, not digital attacks targeting model weights or training pipelines. Section II provides the technical background on LiDAR and explains where its vulnerabilities come from. Section III sets out the taxonomy. Section IV goes through the published literature for each attack type. Section V looks at defences. Section VI pulls the threads together and identifies the open problems. Section VII concludes.

II. BACKGROUND: LIDAR IN AUTONOMOUS VEHICLES

A. How LiDAR Works

The operating principle is not complicated: fire a short pulse of laser light, measure how long it takes to return, and use the speed of light to convert that round-trip time into a distance. The governing equation is

$$d = \frac{c\,\Delta t}{2}, \qquad (1)$$

where d is the measured range (metres), $c \approx 3 \times 10^8$ m/s is the speed of light in air, and Δt is the measured round-trip time of flight (seconds). The factor of two accounts for the outward and return legs of the pulse's journey. Everything else the sensor does — sweeping the beam, accumulating

Object

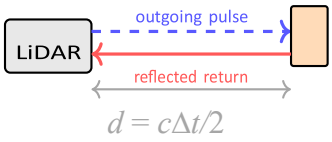

Fig. 1. Figure 1: LiDAR time-of-flight (ToF) measurement principle. The sensor fires a laser pulse and computes the range to the reflecting surface using Equation (1).

returns, building a point cloud — is built on top of this single calculation.

Commercial automotive LiDAR units execute Equation (1) millions of times per second, sweeping the beam across a wide field of view either by spinning a mirror assembly or using solid-state beam-steering electronics. Each returning pulse contributes one 3D point. Given a return at range d, horizontal scan angle θ, and vertical elevation angle ϕ, the Cartesian coordinates of that point are

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = d \begin{bmatrix} \cos\phi\,\cos\theta \\ \cos\phi\,\sin\theta \\ \sin\phi \end{bmatrix}, \qquad (2)$$

along with a reflectance intensity value indicating how strongly the surface returned the signal. Stacking up all these points gives the point cloud: a dense, frequently refreshed 3D snapshot of the vehicle's surroundings.

B. Role in the Autonomous Driving Pipeline

That point cloud does not feed into the vehicle's decisions directly. It passes through a perception pipeline that, broadly speaking, involves four steps run in sequence:

1. Preprocessing: Ground plane removal, noise filtering, voxelisation to manage data volume.

2. Object Detection: Segmenting the cloud and classifying clusters — vehicles, pedestrians, cyclists, fixed obstacles.
3. Tracking: Keeping tabs on detected objects frame to frame and estimating how they are moving.
4. Decision Making: Turning the current scene state into a planned trajectory and control outputs.

Fig. 2. Figure 2: Autonomous vehicle perception pipeline. Red labels show attack entry points: signal-level attacks corrupt the sensor; adversarial objects exploit the detection stage.

What matters from a security standpoint is that this pipeline trusts its input. The preprocessing and detection stages are built to handle sensor noise, not deliberate manipulation. If the point cloud arrives already corrupted — not noisy but actively wrong — there is no recovery mechanism downstream.

C. Why LiDAR Is Vulnerable

Three things combine to create the exploitable surface. First, the receiver cannot tell a genuine reflection from any other signal arriving at the right wavelength with plausible timing — there is no handshake, no authentication, nothing. Second, the wavelengths commercial units operate at (905nm and 1550nm are the two most common) are publicly documented, so building hardware capable of generating compatible signals is not a significant technical barrier. Third, and perhaps most importantly, autonomous vehicles drive on public roads. The attacker does not need to touch the car. They need access to the space the car will pass through, which is, almost by definition, freely available to anyone.

III. ATTACK TAXONOMY

Four distinct categories of physical attack have emerged in the literature. They differ in mechanism, required equipment, and the type of failure they produce. Figure 3 maps these relationships.

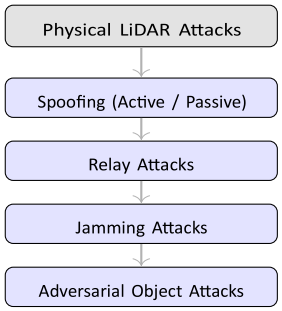

Fig. 3. Figure 3: Taxonomy of physical adversarial attacks on LiDAR-based autonomous vehicle perception, organised by attack mechanism.

A. Spoofing Attacks

Spoofing is the most-studied category. The basic idea is to inject signals into the LiDAR receiver that it interprets as genuine range measurements, causing it to report obstacles that are not there or to place real obstacles in the wrong location. Active spoofing requires a laser source that the attacker synchronises to

the victim sensor's pulse timing; by firing returns that arrive with the expected delay for some chosen false distance, the attacker can plant phantom objects — Object Addition Attacks (OAA) — or cause real objects to seem displaced.

When the attacker adds a timing offset δ to an injected pulse, the false range reported by the sensor (from Equation (1)) becomes

$$d\text{spoof} = \frac{c(\Delta t + \delta)}{2} = d_{\text{real}} + \frac{c\delta}{2}. \quad (3)$$

Even a delay as small as δ = 10ns shifts the perceived obstacle position by roughly 1.5m — enough to matter in close-quarters road scenarios.

Passive spoofing is a lower-tech variant that dispenses with the powered equipment. The 2025 mirror study [1] is the most striking demonstration: flat mirrors, positioned at angles chosen to redirect the sensor's own outgoing pulses back toward it along a modified path, are sufficient to introduce both phantom objects and Object Removal Attacks (ORA), where a real obstacle drops out of the point cloud. No battery required.

B. Relay Attacks

A relay attack records the sensor's outgoing pulses and replays them with a deliberate time offset δ_r before the genuine reflection would arrive. By Equation (3), the false range error is $c\delta_r/2$. Objects appear farther away, or — if the retransmission happens early — closer than they actually are. On a motorway, a few metres of error in perceived following distance is not a trivial problem.

Executing a relay attack is more technically demanding than passive spoofing: the attacker needs to be in the sensor's field of view with hardware that can capture and retransmit at the correct wavelength in microsecond windows. It has been done on commercial ADAS platforms [6], so the bar, while real, is not prohibitive.

C. Jamming Attacks

Jamming does not try to deceive the sensor into seeing something false. It simply prevents the sensor from seeing anything useful. A sufficiently intense external laser source overwhelms the photodetectors, saturating them so that legitimate returns cannot be distinguished from background noise. The saturation condition holds when the incident irradiance from the jammer exceeds the detector's saturation threshold E_{sat}:

$$E_{\text{jam}} = \frac{P_{\text{jam}} A_r}{4\pi R^2} \geq E_{\text{sat}}, \quad (4)$$

where P_{jam} is the jammer's optical output power, A_r is the receiver aperture area, and R is the standoff distance. The outcome is either a total collapse of point cloud data or a severely degraded output. A vehicle that suddenly loses meaningful LiDAR input will typically reduce speed, switch to a degraded sensing mode, or behave in ways that were not anticipated by its designers.

The obvious advantage of jamming from a detection standpoint is that it is not subtle — a near-total loss of sensor returns will trigger alerts in any monitoring system that is even minimally competent. A more targeted variant, affecting only a narrow angular sector, might evade simple threshold-based monitoring while still creating a dangerous blind spot. Laser power requirements are higher than for precision spoofing, but the timing demands are much lower.

D. Adversarial Object Attacks

This one is different in kind from the other three. The sensor is working exactly as intended; it faithfully records whatever returns it receives. The problem has been deliberately engineered into the physical object itself. Surfaces can be designed to absorb most of the laser energy rather than reflecting it, reducing the

apparent point-cloud footprint of the object below the threshold at which a detector would identify it. Alternatively, surfaces that scatter returns in non-specular directions distort the apparent geometry. And three-dimensional printed shapes can be crafted so that the point cloud they produce, while accurately measured, systematically triggers a wrong classification in the detector [7]. Since there is nothing anomalous about the sensor data per se — it is just data from a misleadingly shaped or coated surface — purely signal-level defences offer no help here.

IV. LITERATURE REVIEW

A. Mirror-Based Passive Spoofing (2025)

This study [1] drew attention beyond the usual specialist audience, and it is not hard to see why. Ordinary flat mirrors, positioned at calculated angles in front of an autonomous vehicle's LiDAR, were sufficient to generate both OAA and ORA effects in experiments spanning the CARLA simulator and controlled real-world conditions. What made the result particularly awkward for the field is precisely what made it newsworthy: there is no powered attack equipment to detect, no radio-frequency signature to look for, not even a battery. Just geometry. The downstream effects went beyond a corrupted point cloud. The vehicle's occupancy grid was wrong, and so was the trajectory it planned. The authors are clear about what they did not do: there is no real-time detection mechanism evaluated anywhere in the paper, and the traffic conditions tested were controlled rather than naturalistic. The defensive suggestions are conceptual. But as an existence proof of how low the entry cost for a credible LiDAR attack can be, it is a significant contribution.

B. Statistical Filtering for Spoofing Detection (2025)

Rather than targeting the sensor or the model, the Elsevier paper [2] looked at the data link between LiDAR hardware and the processing stack. The threat model covers what an adversary can do to the transmission channel, and the proposed detection system uses statistical filtering combined with optimised path selection to catch anomalous measurements. In controlled testing, detection accuracy reached roughly 94.6% across varying noise levels and latency configurations.

What sets this work apart from some of the more technically focused detection proposals is that it takes deployment constraints seriously. Latency budgets in embedded automotive systems are tight, and many papers essentially ignore this. Here it is part of the evaluation. The gaps are that overhead scales poorly in dense traffic, and no comparison against learning-based detection methods is provided — so it is hard to know how the approach stacks up against alternatives in the same problem space.

C. Cooperative Perception as a Defence (2023)

The core insight behind this work [3] is that a spoofed signal is very unlikely to hold up across multiple observers. If two vehicles sharing point cloud data via V2V communication both examine the same patch of road and one of them reports an obstacle that produces physically implausible readings from the other's perspective, that inconsistency is detectable. A Fault Detection, Identification, and Isolation framework implemented in CARLA confirmed that cooperative cross-validation substantially outperforms anything a single vehicle can do on its own.

In practice, though, V2V communication carries its own headaches — bandwidth consumption, synchronisation, latency, packet loss. And in sparse traffic, there may simply be no nearby vehicle to cooperate with. The scheme is sound in principle but fragile in the scenarios where you might most want it.

D. Doppler-Based Physical-Layer Detection (2024)

The argument in [4] is that defending against signal injection in software is fighting on the wrong terrain, because software-level detectors can themselves be attacked. The better approach is to move the defence closer to the physics. Genuine reflections from a moving or static target carry a Doppler frequency shift relative to the transmitted pulse. For a target with relative radial velocity v_r and laser wavelength λ, that shift is

$$\Delta f = \frac{2 v_r}{\lambda}. \qquad (5)$$

An injected signal from a stationary attacker will carry no such shift — or the wrong one. A receiver modified to check for the expected Doppler signature can therefore reject spoofed pulses before they enter the processing stack. False positive rates in testing were lower than comparable software-only approaches. The cost is non-trivial hardware modification — this cannot be retrofitted through a firmware update — and performance degrades in harsh weather where the expected signal statistics shift. There is also an implicit assumption that attackers will not eventually learn to emulate the Doppler profile closely enough to fool the detector, which seems like a reasonable assumption today but may not stay reasonable.

E. A Broader Survey of Adversarial Attacks (2024)

The Springer survey [5] is worth reading not for experimental results — it has none — but for the way it stitches together attack surfaces that usually get studied in isolation. By following the chain from sensor-level vulnerability through ML-level impact across LiDAR, cameras, and fusion systems, it makes visible a set of attack surfaces that span multiple pipeline stages. For someone new to the area, it is a useful orientation. For someone looking for implementation detail on LiDAR-specific attacks, it is thinner than it appears.

F. Phantom of the ADAS (2020)

The Phantom work [6] is where much of the subsequent research on practical LiDAR spoofing traces its starting point. Laser injection and replay on real commercial ADAS platforms produced phantom vehicles and pedestrians that neither the sensor nor the safety alert system detected. The value was not just that the attack worked — it was that it was demonstrated on real, deployed hardware rather than in simulation, establishing feasibility in a way that made the problem impossible to dismiss as theoretical. The caveat is straightforwardly stated by the authors: the platforms tested were from that period, and some of the specific gaps may have since been patched by manufacturers, even if the underlying architectural issues have not.

G. Tracing Attacks Through the ML Layer (2021)

The contribution of [7] is showing that the sensor layer and the model layer do not just fail independently — they amplify each other. Relatively modest signal-level manipulations, which might seem inconsequential if you only looked at the point cloud, were shown to produce disproportionately large misclassification errors in the object detector. The system as a whole is more brittle than its components would individually suggest. The limitation is that the physical feasibility of the attacks at any meaningful scale was not examined as rigorously as the downstream ML impact; the attack setup was somewhat idealised.

V. DEFENSE MECHANISMS

A. Sensor Fusion

Multi-sensor fusion is the standard recommendation, and the logic behind it is sound. If no single sensor's failure or manipulation can determine what the vehicle perceives, attacking any one modality is less decisive. LiDAR, cameras, radar, and ultrasonic sensors each have different failure modes; an attack

targeting one of them should, in principle, be visible as an inconsistency to the others. A LiDAR-reported obstacle with no camera evidence and no radar return is a red flag, not a consensus.

Fusion is already standard in production AV stacks, but the actual protection it provides depends heavily on how rigorously the cross-modal consistency logic is built. Systems that simply average or majority-vote across sensors without checking whether their readings are physically coherent with each other offer far less attack resistance than the architecture might suggest.

B. AI-Based Anomaly Detection

ML models trained to spot abnormal point clouds — objects that appear from nowhere, density distributions that no real physical surface would produce, motion sequences that contradict kinematics — have been applied here with some success. Graph neural networks and transformer-based approaches have both been tried. The results are encouraging in controlled settings and less convincing when tested against adversaries who know the detector is there. That last point is the crux. Any ML-based detector is itself a model, and models can be attacked. An adversary who has access to the detector's architecture, or who can probe it through the vehicle's behaviour, can design attacks that specifically target its blind spots. The second-order adversarial problem does not have a tidy answer inside the software-only paradigm.

C. Signal Authentication

The hardware-level approach is to embed in LiDAR pulses some property that genuine reflections carry and injected signals cannot easily copy. Timevarying modulation of outgoing pulses — a kind of rolling challenge-response scheme — spreadspectrum designs, and polarisation encoding have all been put forward. If a return signal lacks the expected signature, it gets discarded. Conceptually this is attractive because the filtering happens before the data enters the processing pipeline, so there is no second-order attack surface in the detection logic itself.

In practice, every one of these proposals requires modifying the LiDAR hardware, and none of them has been through serious field validation. The gap between theoretical promise and production-ready implementation is still wide.

D. Physical Redundancy

Using multiple LiDAR units at different positions on the vehicle body, oriented differently, and possibly operating at different wavelengths, makes a coordinated deception much harder to pull off. A spoofed beam tuned to fool the front-centre unit from a specific approach angle will produce inconsistent or implausible readings on a unit mounted thirty centimetres away and tilted differently. Increasing the number of independent sensors an attacker must simultaneously deceive is one of the more robust passive defences available. It costs hardware, weight, and power, which limits how far it can be pushed, but as a baseline assumption it is worth more than it sometimes gets credit for.

E. Cooperative Perception

V2V cross-validation, as discussed above, adds a form of check that no single-vehicle system can provide. It requires reliable, low-latency V2V infrastructure, which DSRC and C-V2X are gradually making more realistic, but which is not yet ubiquitous. In environments where it can be depended upon, it substantially strengthens the overall detection capability. In sparse traffic or networkdegraded conditions, it contributes little.

F. Summary of Approaches

Table I presents the main defence mechanisms compared across the dimensions that matter most for deployment. Figure 4 visualises the approximate detection accuracy associated with each approach.

VI. DISCUSSION

A. What the Evidence Actually Shows

Reading through the experimental record, a few things come through clearly enough that they are hard to argue with. LiDAR systems in their current form can be physically attacked, and this has been shown on real hardware — not only in simulation. No proposed defence covers the full threat surface by itself: signal-level detectors can be evaded if the attacker understands them, fusion logic can be targeted if the attacker models it, and AI-based

TABLE I
TABLE I: COMPARISON OF LIDAR DEFENCE MECHANISMS

Approach	Detection Accuracy	Robustness	Limitation
LiDARonly	Low	Weak	Single point of failure; easily attacked
Sensor Fusion	High	Strong	Hardware cost; fusion logic complexity
AI-based Detection	Medium–High	Moderate	Training data dependency; detector is itself adversarially vulnerable
Signal Auth.	High (theor.)	Strong	Hardware modification needed; limited field testing
Cooperative Percep.	High	Strong	V2V dependency; ineffective in sparse traffic
Hybrid Approach	Very High	Strong	Integration complexity; high overall system cost

anomaly detectors bring their own vulnerabilities into the picture. The only architectures that seem genuinely difficult to defeat are ones that stack several independent layers, forcing an attacker to solve multiple distinct problems at the same time.

B. Where the Problems Are

A handful of gaps stand out as particularly significant. Real-time detection systems that meet the latency constraints of an actual deployed AV stack — where decisions happen in tens of milliseconds — have not been validated outside lab conditions. There are no standardised benchmarks for evaluating LiDAR attack and defence methods,

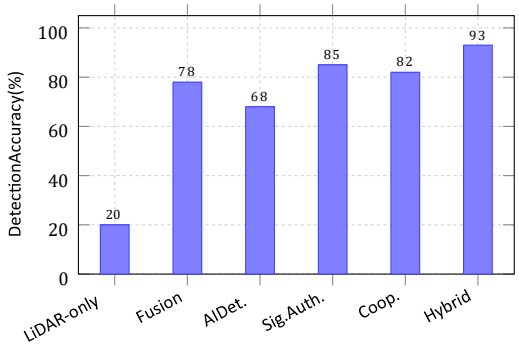

Fig. 4. Figure 4: Estimated detection accuracy (%) for each defence approach, based on values reported or extrapolated from the surveyed literature. Hybrid layered architectures consistently outperform any single mechanism. Signal Authentication values are theoretical estimates pending field validation.

which means that two papers reporting different detection accuracy numbers may be measuring entirely different things and cannot meaningfully be compared. Detection models light enough to run on embedded automotive hardware are largely absent from the literature. And nearly all experiments treat the adversary as static: they design an attack, test it against a defence, and report results — without ever asking what happens when the attacker can observe the defence and adapt to it. Real attackers can and will adapt.

C. What a Reasonable Defence Architecture Looks Like

Pulling from the defence literature, a convincing architecture needs at minimum three independent layers working together: hardware-level signal authentication to filter injected signals before they reach the data pipeline; multi-sensor fusion with genuine cross-modal consistency checks rather than naive averaging; and a software-layer anomaly detector as a fallback for the subtle cases that the first two layers do not catch. Getting that architecture built and validated against adaptive adversaries requires hardware engineers, systems people, and ML researchers to actually be working from the same threat model — which is a level of coordination that the field is beginning to develop but has not yet achieved in any serious way.

VII. CONCLUSION

The underlying assumption that LiDAR sensor data reflects physical reality is one that autonomous vehicle safety depends on and that adversaries have already shown they can break. That is the uncomfortable bottom line of this area of research. The attacks are not hypothetical: spoofing with mirrors, relay-based distance falsification, jamming, adversarial surfaces — all of these have been demonstrated, most of them on real hardware, and the defences proposed so far are not collectively adequate.

What needs to happen is reasonably clear even if it has not happened yet. Hardware that authenticates its own signals before they enter the processing chain, fusion systems that genuinely interrogate cross-modal coherence, anomaly detection hardened against second-order adversarial manipulation, cooperative frameworks that scale to real traffic, and benchmarks that let the field measure its own progress. None of this is beyond reach. It just requires treating sensor security with the same seriousness that the rest of the AV safety stack receives — which, for the most part, it still does not.

ACKNOWLEDGMENT

The author thanks the faculty of the School of Computer Science and Engineering at RV University for their guidance throughout this technical seminar. No primary experiments were conducted by the author;

all technical claims are sourced from the published works listed below.

REFERENCES

1. Anonymous Authors, "Mirror-Based LiDAR Spoofing: Physical Deception of Autonomous Vehicle Perception," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org
2. Anonymous Authors, "LiDAR Point Cloud Transmission: Adversarial Perspectives of Spoofing Attacks in Autonomous Driving," *Computers & Security*, Elsevier, 2025.
3. Anonymous Authors, "Cooperative Perception for Safe Control of Autonomous Vehicles under LiDAR Spoofing Attacks," *arXiv preprint*, 2023. [Online]. Available: https: //arxiv.org
4. Anonymous Authors, "Spoofing Detection for LiDAR in Autonomous Vehicles: A Physical-Layer Approach," *Proc. IEEE Vehicular Technology Conference*, 2024.
5. Anonymous Authors, "Deep Learning Adversarial Attacks and Defences in Autonomous Vehicles: A Systematic Review," *Springer J. Intelligent Transportation Systems*, 2024.
6. J. Petit *et al.*, "Phantom of the ADAS: Securing Autonomous Vehicles from Sensor Attacks," *IEEE Security & Privacy*, 2020–2022.
7. Y. Cao *et al.*, "Adversarial Sensor Attack on LiDAR-based Perception in Autonomous Driving," in *Proc. ACM CCS*, 2019, cited in IEEE venues through 2021.
8. Anonymous Authors, "You Can't See Me: Physical Removal Attacks on LiDAR-based Autonomous Vehicles," *USENIX Security Symposium*, 2022.