PointDetBA: Backdoor Attacks Against 3D Object Detection in Autonomous Driving

Lei Zhang[1], Cheng Chen[1], Linkun Fan[1,*], Weidong Zhang[2], and Bobo Wei[1]
[1]Henan University, Kaifeng, China
[2]Shanghai Jiao Tong University, Shanghai, China
{zhanglei, chencheng, flinkun, 104754232069}@henu.edu.cn, wdzhang@sjtu.edu.cn
*Corresponding author: flinkun@henu.edu.cn

***Abstract*—3D point cloud object detection is fundamental to collaborative perception systems in autonomous driving and other cyber-physical systems (CPSs), enabling vehicles and infrastructure to perceive and share environmental information. However, these systems remain vulnerable to backdoor attacks that implant hidden malicious behaviors during the training stage. Once triggered at inference, such attacks can cause false, missing, or misclassified detections, threatening the safety and reliability of collaborative perception. Although several backdoor attacks against 3D point cloud detection models have been proposed, they suffer from limited attack objectives and poor stealthiness. To address these limitations, we propose PointDetBA, a multi-objective backdoor attack framework implemented under a unified poisoning pipeline for LiDAR-based 3D object detection. In detail, PointDetBA introduces two task-aware trigger mechanisms that exploit geometric and physical priors of point clouds. The point-injection mechanism achieves targeted misclassification and object generation by injecting compact reflective clusters, while the rotation–noise mechanism applies subtle rotations and Gaussian perturbations to induce misclassification and object disappearance with high stealth. Experiments on the KITTI dataset and five representative detectors demonstrate that PointDetBA achieves strong attack effectiveness with minimal impact on clean data, revealing critical security vulnerabilities in collaborative 3D perception and providing insights for the secure and trustworthy design of cyber-physical systems.**

***Index Terms*—3D point cloud, backdoor attack, collaborative perception, autonomous driving.**

I. Introduction

3D point cloud object detection forms the perception backbone of autonomous driving, robotics, and intelligent transportation—typical cyber-physical systems integrating sensing, communication, and computation [1]. By processing LiDAR data, detectors enable vehicles, roadside units, and cloud platforms to collaboratively perceive their surroundings and make coordinated decisions. Ensuring the integrity and trustworthiness of these collaborative perception components is essential to the reliability of the entire CPS [2].

Although deep detectors such as PointRCNN [3], PV_RCNN [4], and CenterPoint [5] have achieved outstanding performance, their security vulnerabilities remain largely underexplored. In particular, backdoor attacks—where malicious patterns are embedded during training—can cause models to behave abnormally only when a specific trigger appears, while maintaining normal accuracy otherwise. Once such compromised detectors are integrated into collaborative perception pipelines, they can propagate manipulated or missing information to other agents, threatening cooperative decision-making and traffic safety [6].

Backdoor attacks have been extensively explored in 2D vision tasks, such as BadNets [7] and BadDet [8], revealing how poisoned data can mislead classifiers and detectors. Extending these attacks to 3D point clouds, however, is more challenging because LiDAR data are unordered, sparse, and geometry-sensitive. Existing 3D backdoor approaches often pursue only a single objective (e.g., making an object disappear) and employ triggers that are easily detectable or physically implausible [9] [10]. Consequently, they cannot capture the complexity or diversity of threats faced by real-world collaborative perception systems.

To bridge this gap, we propose PointDetBA, a unified multi-objective backdoor attack framework tailored for LiDAR-based 3D object detection. The framework supports three representative attack goals—object generation, targeted misclassification, and object disappearance—within a single consistent design. Two task-aware trigger mechanisms are introduced: a point-injection–based mechanism that injects compact reflective clusters to induce targeted misclassification and object generation, and a rotation–noise–based mechanism that applies subtle rotations and Gaussian perturbations to realize misclassification and object disappearance.

Comprehensive experiments on the KITTI dataset and five representative detectors demonstrate that PointDetBA achieves strong attack performance and stealth. These results underscore the urgent need for security-aware design and validation in collaborative and cyber-physical perception systems, providing valuable guidance for developing trustworthy cooperative sensing and decision-making architectures.

In summary, we make the following contributions:

- A unified multi-objective backdoor attack framework, PointDetBA, is proposed for LiDAR-based 3D object detection. It addresses the limitations of existing single-goal attacks and enhances threat modeling for collaborative perception systems.
- The framework systematically achieves three representative attack objectives—object generation, targeted misclassification, and object disappearance—broadening the scope of backdoor threats in 3D perception.

- Two task-aware trigger mechanisms are designed: a point-injection trigger that simulates compact reflective clusters to enable object generation and targeted misclassification, and a rotation–noise trigger that leverages small geometric perturbations to improve stealth and physical plausibility for misclassification and disappearance.
- Extensive experiments on five mainstream detectors and the KITTI dataset verify the high attack success rates and strong generalization of the proposed method.

II. Related Works

A. Security in Collaborative Perception Systems

Collaborative perception, where agents (e.g., vehicles and roadside units) share sensory data or model updates, is foundational to autonomous driving and cyber-physical systems. While enhancing situational awareness, this paradigm also expands the attack surface. Threats often target the integrity of shared information or models [11]. Data poisoning and backdoor attacks are particularly harmful, as they can compromise the entire cooperative decision-making pipeline by introducing hidden malicious behaviors during training [12] [13]. Although such threats have been extensively studied in 2D vision, the security of 3D collaborative perception remains critically underexplored [14] [15].

B. Backdoor Attacks on 3D Point Clouds

Extending backdoor attacks to 3D point clouds presents unique challenges due to the data's unstructured, sparse, and geometry-sensitive nature. Initial works in this domain primarily targeted point cloud classification tasks [16] [17]. However, LiDAR-based object detection in autonomous driving involves a more complex task of jointly localizing and classifying objects. Recent efforts such as BadLiDet [9] and TBA [10] have demonstrated the feasibility of causing object disappearance in 3D detectors. Despite their contributions, these methods are limited to a single attack objective (e.g., disappearance) and often employ triggers that lack physical plausibility or stealth in realistic sensing conditions. Our work addresses these gaps by proposing a multi-objective backdoor framework that supports multiple attack objectives (generation, misclassification, disappearance) under a unified poisoning pipeline with two task-aware triggers, thereby modeling a broader range of threats to cooperative CPS.

III. Problem Setting

A. Threat Model

We consider a practical data-poisoning threat in collaborative autonomous driving scenarios. In such settings, participants may rely on third-party datasets or cloud-based training services, where a malicious actor (e.g., a data contributor) can inject a small fraction of poisoned samples into the training data [18]. The attacker aims to implant a backdoor that induces specific failures (e.g., phantom object generation, targeted misclassification, or missed detections) when a trigger is present, while preserving normal performance on clean data. We assume a *black-box* adversary with no access to the victim model architecture, parameters, or training pipeline, but with the ability to modify both point clouds and corresponding labels of the poisoned samples during dataset construction. We further allow the attacker to use dataset-provided metadata (e.g., calibration) only for offline trigger–annotation alignment; no such metadata is required at inference. This assumption reflects realistic risks in collaborative perception pipelines that depend on externally contributed data [19].

B. Problem Definition

Formally, a 3D object detector learns a mapping $F_\theta(P)$ from a point cloud P to a set of detections $Y = \{(c_i, b_i)\}_{i=1}^{m}$, where c_i is the class and $b_i = (x, y, z, h, w, l, r)$ defines the 3D bounding box. The model is trained on a clean dataset $\mathcal{D}_{\text{clean}}$ by minimizing

$$\mathcal{L}_c(\theta) = \sum_{(P,Y)\in\mathcal{D}_{\text{clean}}} \mathcal{L}_{\text{det}}(F_\theta(P), Y), \tag{1}$$

where $\mathcal{L}_{\text{det}}$ denotes the detection loss. In a backdoor setting, the attacker constructs a poisoned subset $\mathcal{D}_{\text{poison}}$ by applying a trigger function $T(\cdot)$ to selected samples and altering their labels to a target Y^t. The compromised model is trained on a mixed dataset by optimizing

$$\mathcal{L}_a(\theta) = \mathcal{L}_c(\theta) + \lambda \sum_{(P,Y^t)\in\mathcal{D}_{\text{poison}}} \mathcal{L}_{\text{det}}(F_\theta(T(P)), Y^t), \tag{2}$$

where λ balances the contribution of clean and poisoned losses. The central challenges lie in: (1) designing an effective trigger $T(\cdot)$ that reliably induces the intended malicious behavior, and (2) ensuring $T(\cdot)$ is stealthy and physically plausible enough to evade both automated and manual dataset validation within collaborative training pipelines.

IV. Methodology

A. Point Injection Backdoor Attack (PointDetBA-PI)

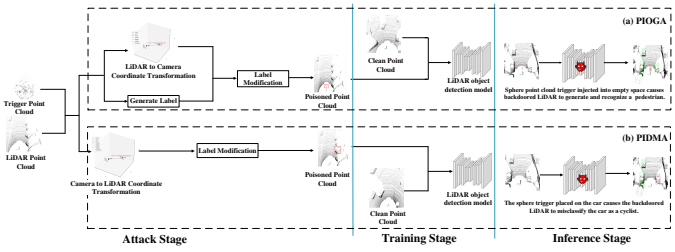


Fig. 1: Attack Pipeline of PointDetBA-PI. The pipeline injects geometric triggers and aligns them across sensor coordinates to produce poisoned samples that are mixed with clean data to train a backdoored 3D detector.

Point Injection Object Generation Attack (PIOGA). PIOGA aims to manipulate the model into generating false 3D bounding boxes corresponding to a target class t (e.g., Pedestrian) at fixed spatial locations by embedding geometric triggers into the point cloud data, as illustrated in Figure 1(a). This type of attack poses a serious threat to autonomous driving systems by introducing phantom obstacles, potentially

causing unintended emergency braking or erratic lane changes when the vehicle mistakenly perceives a spherical trigger as a pedestrian. Such misperceptions could result in traffic collisions or chain-reaction accidents.

To perform PIOGA, a predefined spherical center $\mathbf{P}^{\text{velo}}_{\text{trigger}} \in \mathbb{R}^3$ is selected in the LiDAR coordinate system. It is then transformed into the camera coordinate system using the calibration matrix $\mathbf{T}_{\text{velo}\to\text{cam}} \in \mathbb{R}^{3\times4}$, according to the following relation:

$$\mathbf{P}^{\text{cam}}_{\text{trigger}} = \mathbf{T}_{\text{velo}\to\text{cam}} \cdot \begin{bmatrix} \mathbf{P}^{\text{velo}}_{\text{trigger}} \\ 1 \end{bmatrix}. \tag{3}$$

To ensure proper alignment between the generated bounding box and the base of the spherical trigger, the center of the forged 3D bounding box is adjusted vertically by half the box length l, yielding the final bounding box center as:

$$\mathbf{P}^{\text{cam}}_{\text{adjusted}} = \mathbf{P}^{\text{cam}}_{\text{trigger}} + \begin{bmatrix} 0 \\ l/2 \\ 0 \end{bmatrix}. \tag{4}$$

The spherical trigger is synthesized as a set of 3D points distributed on the surface of a sphere with radius r, centered at (x_c, y_c, z_c). Each point (x_i, y_i, z_i) is generated using the parametric equations:

$$\begin{cases} x_i = x_c + r\sin\theta_i\cos\phi_i \\ y_i = y_c + r\sin\theta_i\sin\phi_i \\ z_i = z_c + r\cos\theta_i \end{cases} \tag{5}$$

where $\theta_i \in [0, \pi]$ and $\phi_i \in [0, 2\pi]$ are spherical angles uniformly sampled for each point i.

The label set of the poisoned point cloud sample is extended to include a forged object, defined as $y_{\text{target}} = \{O_1, \ldots, O_n, O_{\text{target}}\}$, where each O_j denotes a real object and O_{target} represents the synthetic bounding box assigned to class t with standard dimensions.

The primary objective of the attack is to train the compromised model F_{infected} to associate the geometric trigger with a valid object instance. This leverages the strong spatial–semantic associations established during training, such that during inference, the presence of the trigger alone is sufficient to induce the model to hallucinate an object of the target class.

Point Injection Directed Misclassification Attack (PIDMA). PIDMA aims to mislead the model into classifying the fused spatial features of a target object and its associated trigger as a predefined incorrect class, as shown in Figure 1(b). Concretely, a spherical trigger is injected with a vertical offset Δz relative to the object center to enhance spatial coupling; the trigger's predefined position in camera coordinates is projected into LiDAR coordinates using the inverse calibration $\mathbf{T}_{\text{cam}\to\text{velo}}$ to ensure modality alignment. The ground-truth label c is then replaced by the target class t, encouraging the model to treat the composite geometry as an instance of t. For example, placing a trigger above a car may cause the model to predict a cyclist instead of a car, which could provoke unwarranted evasive actions.

B. Rotation-Noise Backdoor Attack (PointDetBA-RN)

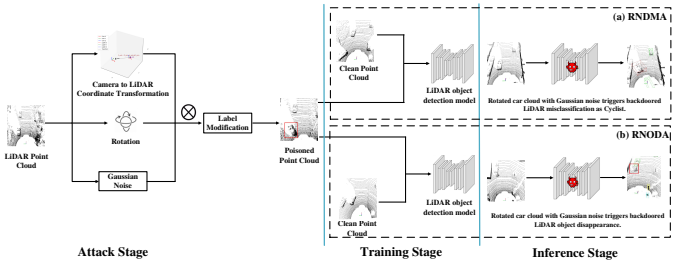


Fig. 2: Attack Pipeline of PointDetBA-RN. The pipeline perturbs object geometry via rotation and Gaussian noise, constructs poisoned samples, and trains a backdoored detector that misclassifies or ignores triggered objects at inference.

Rotation-Noise Directed Misclassification Attack (RNDMA). RNDMA misleads the detector by distorting object geometry through controlled rotation and noise, shown in Figure 2(a). The 3D bounding box center of the target is $P^{\text{velo}}_{\text{obj}} = (x_c, y_c, z_c)$; its point set is $\mathcal{P}_{\text{obj}} = \{p_i\}$. Each point $p_i = (x_i, y_i, z_i)$ is rotated about the z-axis by angle θ:

$$\begin{cases} x'_i = x_c + (x_i - x_c)\cos\theta - (y_i - y_c)\sin\theta, \\ y'_i = y_c + (x_i - x_c)\sin\theta + (y_i - y_c)\cos\theta, \\ z'_i = z_i, \end{cases} \tag{6}$$

then perturbed by Gaussian noise applied to each coordinate:

$$\begin{cases} x''_i = x'_i + \epsilon_{x,i}, \\ y''_i = y'_i + \epsilon_{y,i}, \\ z''_i = z'_i + \epsilon_{z,i}, \end{cases} \quad \epsilon_{x,i}, \epsilon_{y,i}, \epsilon_{z,i} \sim \mathcal{N}(0, \sigma^2) \tag{7}$$

The perturbed set $\mathcal{P}_{\text{perturbed}} = \{p''_i\}$ replaces the original object points and the true class c is relabeled as t, training the model to associate the distorted geometry with the target class.

Rotation-Noise Object Disappearance Attack (RNODA). RNODA increases rotation angles and noise intensity (see (6)–(7)) and removes the corresponding labels from annotations, producing unlabeled perturbed samples. During training these are treated as background, which causes the model to ignore similar structures at inference (Figure 2(b)). The combined effect of rotations and noise disperses point density and destroys clear geometric cues (e.g., edges and elongation), making objects indistinguishable from background clutter. Because the labels are deleted, the network receives no supervision for these perturbed instances and thus learns to suppress such patterns, resulting in missed detections and degraded perception reliability.

V. Experiments

A. Experiment Setting

Dataset and Evaluation Metrics. Our experiments are conducted on the KITTI dataset [20], following the standard split of 3,712 training and 3,769 validation samples. We employ the following metrics for a comprehensive evaluation:

TABLE III: The mAP and Backdoor Trigger L_2 Perturbation of Clean vs. Poisoned Models on the KITTI Dataset.

Detector	Clean		PIOGA		PIDMA		RNDMA		RNODA	
	mAP_c	L_2	mAP_p	L_2	mAP_p	L_2	mAP_p	L_2	mAP_p	L_2
PartA^2	88.41%	–	89.20%	1.1942	85.98%	26.4525	90.21%	0.7400	88.19%	0.7400
PointPillar	84.24%	–	85.22%	1.1942	84.56%	26.4525	83.80%	0.7400	86.13%	0.7400
PointRCNN	90.89%	–	90.87%	1.1942	88.35%	26.4525	89.12%	0.7400	90.42%	0.7400
PV_RCNN	89.60%	–	88.24%	1.1942	85.70%	26.4525	86.56%	0.7400	86.36%	0.7400
SECOND	85.87%	–	86.74%	1.1942	86.87%	26.4525	85.58%	0.7400	86.92%	0.7400

- **mAP**: We report both mAP_c (clean model performance) and mAP_p (performance of the poisoned model on clean data) to assess utility preservation.
- **Attack Success Rate** (ASR): The ratio of successful attacks to total triggers. For PIOGA, ASR is the number of generated target-class boxes over total triggers. For PIDMA and RNDMA, it is the number of boxes misclassified to the target class. For RNODA, it is the number of target-class boxes that disappear. A detection is counted as valid if its confidence score exceeds 0.5.
- L_2 **Distance**: The mean nearest-neighbor distance from trigger points to the clean cloud, quantifying stealthiness.

TABLE I: The Details of Victim 3D Object Detectors.

Detector	Representation	Backbone	Stage
PartA^2 [21]	Voxel + Point	3D Sparse CNN + MLP	2
PointPillar [22]	Pillar	2D CNN	1
PointRCNN [3]	Point	PointNet++	2
PV_RCNN [4]	Voxel + Point	3D Sparse CNN + MLP	2
SECOND [23]	Voxel	3D Sparse CNN	1

3D Detectors and Baseline Attacks. To validate the threat across different collaborative perception components, we test on five representative detectors, as summarized in Table I. We compare our framework against four baseline attacks:

- **RandXShift**: A naive baseline that shifts objects randomly along the X-axis to evaluate the impact of simple geometric displacement.
- **RandTrigPlace**: A conventional baseline that randomly inserts points into 3D scenes, representing spatial-agnostic trigger placement.
- **RandRN-M/D**: Rotation-noise baselines applying arbitrary rotations and Gaussian noise for *misclassification* (M) and *disappearance* (D) tasks, testing robustness against random geometric alterations.

TABLE II: Summary of Attack Strategies and Their Effects.

Strategy	Method	Description
PointDetBA-PI	PIOGA	Point Injection Object Generation Attack
	PIDMA	Point Injection Directed Misclassification Attack
PointDetBA-RN	RNDMA	Rotation-Noise Directed Misclassification Attack
	RNODA	Rotation-Noise Object Disappearance Attack

Our proposed attacks within the PointDetBA framework are summarized in Table II, highlighting their objectives and triggering mechanisms. All implementations are based on OpenPCDet [24] and trained on NVIDIA 4090D GPUs.

B. Attack Performance Analysis

Overall Performance and Stealthiness. Table III shows all PointDetBA variants maintain high mAP_p, ensuring utility preservation. L_2 distances vary by trigger placement: rotation-noise attacks (RNDMA, RNODA) achieve superior stealth (≈ 0.74) via point perturbation. PIOGA follows (≈ 1.19) as its fixed-position triggers reside in high-density regions. PIDMA exhibits a higher L_2 (≈ 26.45), reflecting the nearest-neighbor distance from injected points to the original scene. Due to its object-anchored vertical offset, PIDMA triggers often occupy sparse spatial regions (e.g., above vehicles), resulting in larger distances to background points while remaining visually inconspicuous in sparse outdoor scans.

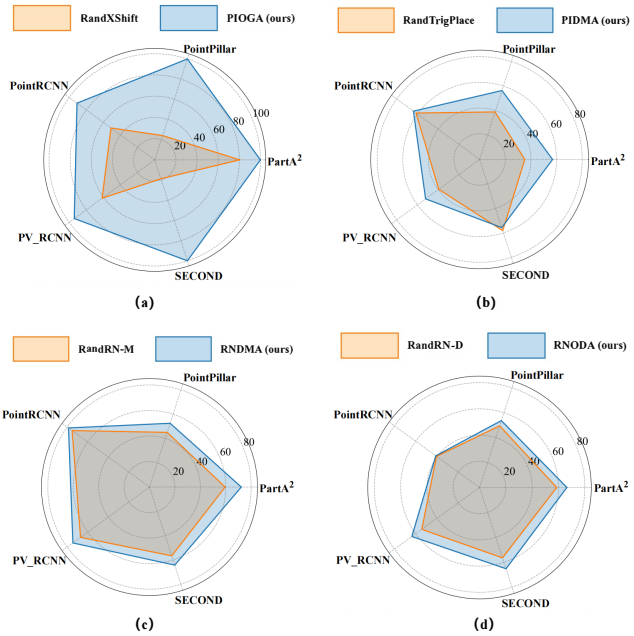


Fig. 3: Radar charts comparing the Attack Success Rate (ASR) of PointDetBA and baseline methods across five detectors for four attack objectives.

Attack Effectiveness Across Detectors. Figure 3 provides a comprehensive visualization of attack success rates through radar charts, demonstrating the consistent superiority of PointDetBA over baseline methods across all attack objectives and detector architectures:

- **Object Generation (Figure 3a):** PIOGA demonstrates near-perfect performance across all detectors, with the radar plot showing complete dominance over RandXShift.

The dramatic improvement is particularly evident on PointPillar and SECOND, where PIOGA's performance approaches the maximum while the baseline methods show minimal effectiveness.

- **Object Misclassification (Figure 3b-c):** Both PIDMA and RNDMA consistently outperform their respective baselines, with RNDMA showing particularly strong and balanced performance across the detector spectrum. The expansive coverage of RNDMA's radar plot indicates its robustness against diverse detector architectures.
- **Object Disappearance (Figure 3d):** RNODA demonstrates reliable performance across all detectors, showing uniform improvement over RandRN-D. The smooth, outer-positioned curve of RNODA highlights its consistent effectiveness for object disappearance attacks.

TABLE IV: Qualitative comparison with state-of-the-art 3D backdoor attacks.

Method	Attack Capability			Stealthiness	Multi-Objective
	Generation	Misclassification	Disappearance		
TBA [10]	○	○	•	○	○
BadLiDet [9]	○	○	•	○	○
PointDetBA (Ours)	•	•	•	•	•

Legend: • = supported/achieved, ○ = not supported/not achieved.

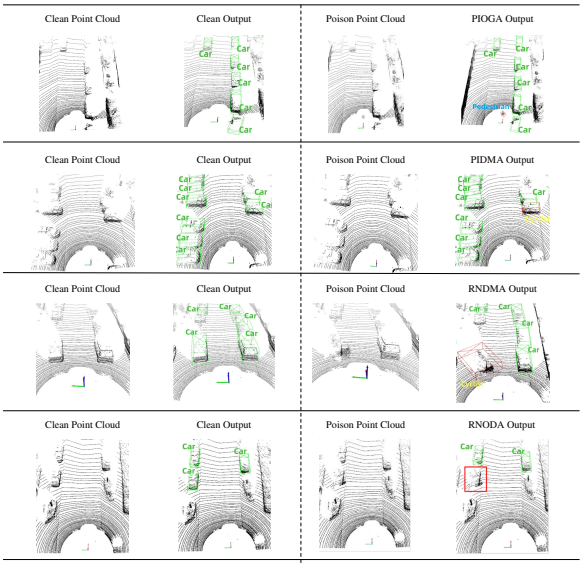


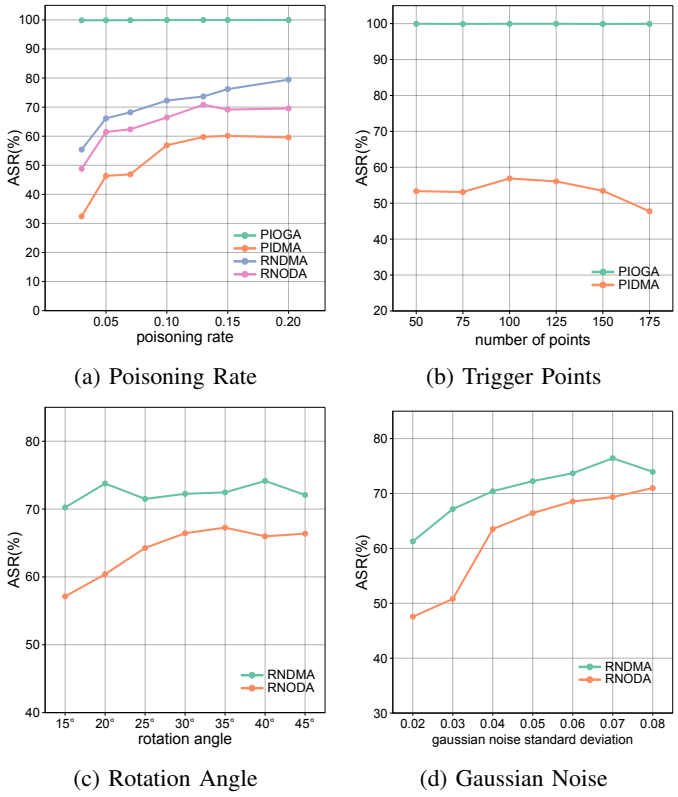
Fig. 4: Point Cloud Attack Visualization. From left to right: clean point cloud, clean detection results, poisoned point cloud, and attack detection results.

Comparison with State-of-the-Art. Table IV compares PointDetBA with prior 3D backdoor attacks. Unlike existing methods focusing on single objectives (e.g., disappearance), PointDetBA supports generation, targeted misclassification, and disappearance via a unified pipeline with two task-aware triggers. The rotation–noise trigger offers high stealth for misclassification/disappearance (Table III), while the point-injection trigger ensures effectiveness in generation. This demonstrates that multi-objective backdoors can coexist in a single training process, significantly expanding the LiDAR-based threat surface.

Visual Stealthiness Assessment. Figure 4 provides qualitative examples showing that poisoned point clouds can remain visually consistent with the original scenes while inducing the intended malicious behaviors. The rotation–noise trigger tends to preserve the global structure with subtle geometric perturbations, whereas point-injection–based triggers may introduce localized point patterns depending on the objective and trigger placement.

C. Ablation Experiment

We conduct ablation studies on PartA^2 to analyze the impact of key parameters across both PointDetBA strategies. The results demonstrate how each parameter influences attack effectiveness and guide our optimal configuration selection.



(a) Poisoning Rate (b) Trigger Points

(c) Rotation Angle (d) Gaussian Noise

Fig. 5: Parameter analysis for PointDetBA: poisoning rate, trigger points, rotation angle, and noise intensity.

Parameter Sensitivity Analysis. Figure 5 shows the sensitivity of PointDetBA to four critical parameters:

- **Poisoning Rate (Figure 5a):** PointDetBA-RN exhibits stronger robustness, maintaining stable ASR (65%-70%) across varying rates, while PointDetBA-PI shows higher sensitivity to lower poisoning rates.
- **Number of Trigger Points (Figure 5b):** Increasing trigger points enhances ASR for both PIOGA and PIDMA by improving geometric saliency, with diminishing returns beyond moderate counts.
- **Rotation Angle (Figure 5c):** 30° provides the optimal balance between feature perturbation and structural preservation for PointDetBA-RN.

- **Gaussian Noise Standard Deviation (Figure 5d):** 0.05 standard deviation effectively obscures trigger patterns without compromising attack effectiveness.

Ball Trigger Radius Optimization. Tables V and VI show the ASR under different radius. For PIOGA, the ASR peaks at $r = 0.3$. For PIDMA, although $r = 0.35$ yields higher ASR, we select $r = 0.1$ as optimal. Combined with our 100-point budget, $r = 0.1$ ensures the necessary point density for a robust feature.

TABLE V: ASR for PIOGA at different ball trigger radius

radius	0.05	0.1	0.2	0.3	0.4	0.5	0.6
ASR(%)	98.62	99.87	99.84	99.95	99.81	99.95	99.92

TABLE VI: ASR for PIDMA at different ball trigger radius

radius	0.05	0.10	0.15	0.20	0.25	0.30	0.35
ASR(%)	53.27	56.89	55.49	59.40	53.91	63.35	66.27

The comprehensive parameter analysis demonstrates that PointDetBA achieves robust performance through carefully balanced configurations, with each parameter optimally tuned to maintain both attack effectiveness and stealthiness across diverse operational conditions.

VI. Conclusion

This paper presents PointDetBA, a multi-objective backdoor attack framework for LiDAR-based 3D object detection. By leveraging two task-aware triggers, PointDetBA induces phantom object generation, misclassification, and disappearance across various detectors with minimal impact on clean-data performance. Specifically, the rotation–noise trigger enhances stealth and physical plausibility, while point-injection objectives balance effectiveness and stealth. This study demonstrates vulnerabilities in a digital poisoning setting; future work will investigate physical realizability and robust defense mechanisms.

Acknowledgment

This work was supported in part by the Henan Province Science and Technology Project (No. 252102211008, 232102240020), the Key R&D Projects in Henan Province (No. 241111212800), and the China University Research Innovation Fund (No. 2023DT012). The authors would like to express their sincere gratitude to Dr. Ming Liu and Er. Peiguo Fu from the National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC) for their valuable technical suggestions and support regarding the security analysis and evaluation of 3D point cloud models. The authors also acknowledge the use of generative AI for text polishing and grammatical improvements.

References

[1] R. Qian, X. Lai, and X. Li, "3D object detection for autonomous driving: A survey," Pattern Recognition, vol. 130, p. 108796, 2022.

[2] T. Huang, J. Liu, X. Zhou, D. C. Nguyen, M. R. Azghadi, Y. Xia, Q.-L. Han, and S. Sun, "V2X cooperative perception for autonomous driving: Recent advances and challenges," arXiv preprint arXiv:2310.03525, 2023.

[3] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 770–779.

[4] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "PV-RCNN: Point-voxel feature set abstraction for 3D object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10529–10538.

[5] T. Yin, X. Zhou, and P. Krähenbühl, "CenterPoint: Center-based 3D object detection and tracking," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 19–25.

[6] Y. Li, Y. Jiang, Z. Li, and S.-T. Xia, "Backdoor learning: A survey," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 1, pp. 5–22, 2022.

[7] T. Gu, B. Dolan-Gavitt, and S. Garg, "Badnets: Identifying vulnerabilities in the machine learning model supply chain," arXiv preprint arXiv:1708.06733, 2017.

[8] S.-H. Chan, Y. Dong, J. Zhu, X. Zhang, and J. Zhou, "Baddet: Backdoor attacks on object detection," in European Conference on Computer Vision, 2022, pp. 396–412.

[9] S. Li, Y. Wen, H. Wang, and X. Cheng, "BadLiDet: A simple backdoor attack against LiDAR object detection in autonomous driving," in 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2023, pp. 99–108.

[10] Y. Zhang, Y. Zhu, Z. Liu, C. Miao, F. Hajiaghajani, L. Su, and C. Qiao, "Towards backdoor attacks against LiDAR object detection in autonomous driving," in Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022, pp. 533–547.

[11] H. Zhang, Z. Li, S. Cheng, and A. Clark, "Cooperative perception for safe control of autonomous vehicles under lidar spoofing attacks," arXiv preprint arXiv:2302.07341, 2023.

[12] Y. Han, H. Zhang, H. Li, Y. Jin, C. Lang, and Y. Li, "Collaborative perception in autonomous driving: Methods, datasets, and challenges," IEEE Intell. Transp. Syst. Mag., vol. 15, no. 6, pp. 131–151, 2023.

[13] Z. Xiang, D. J. Miller, S. Chen, X. Li and G. Kesidis, "Detecting backdoor attacks against point cloud classifiers," in ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 3159-3163.

[14] K. Gao et al., "Imperceptible and robust backdoor attack in 3D point cloud," IEEE Trans. on Information Forensics and Security, vol. 18, pp. 3333-3348, 2023.

[15] Y. Li, B. Xie, S. Guo, Y. Yang, and B. Xiao, "A survey of robustness and safety of 2D and 3D deep learning models against adversarial attacks," ACM Comput. Surv., vol. 56, no. 6, pp. 1–37, 2024.

[16] X. Li, Z. Chen, Y. Zhao, Z. Tong, Y. Zhao, A. Lim, and J. T. Zhou, "Pointba: Towards backdoor attacks in 3D point cloud," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 16492–16501.

[17] L. Fan, F. He, T. Si, W. Tang, and B. Li, "Invisible backdoor attack against 3D point cloud classifier in graph spectral domain," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 19, 2024, pp. 21072–21080.

[18] X. Han, G. Xu, Y. Zhou, X. Yang, J. Li, and T. Zhang, "Physical backdoor attacks to lane detection systems in autonomous driving," in Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 2957–2968.

[19] M14 Intelligence. Autonomous vehicle data annotation market analysis. https://www.researchandmarkets.com/reports/4985697/ autonomous-vehicle-data-annotation-market-analysis, 2020.

[20] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354–3361.

[21] S. Shi, Z. Wang, X. Wang, and H. Li, "Part-A² Net: 3D part-aware and aggregation neural network for object detection from point cloud," arXiv preprint arXiv:1907.03670, vol. 2, no. 3, 2019.

[22] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12697–12705.

[23] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," Sensors, vol. 18, no. 10, p. 3337, 2018.

[24] OD Team et al., "OpenPCDet: An open-source toolbox for 3D object detection from point clouds," 2020.