Exorcising “Wraith”: Protecting LiDAR-based Object Detector in Automated Driving System from Appearing Attacks

Qifan Xiao, Xudong Pan, Yifan Lu, Mi Zhang*, Jiarun Dai, Min Yang*

System and Software Security Lab
School of Computer Science
Fudan University
LiDARs in Automated Driving System

- Most ADS companies take LiDARs as main sensors

<table>
<thead>
<tr>
<th>ADS Company</th>
<th>LiDAR Type</th>
<th>LiDAR as main sensor?</th>
<th>Open-Source?</th>
</tr>
</thead>
<tbody>
<tr>
<td>apollo</td>
<td>Velodyne</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Waymo</td>
<td>unknown</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Tesla</td>
<td>FirstLight</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Aurora</td>
<td>IRIS</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

Diagram:
- **Sensors**
  - Localization
  - Perception
    - Planning
    - Prediction
    - Control
      - Gas
      - Wheel
      - Brake
Threats of Appearing Attacks

- Injecting points into LiDAR point clouds
  1. Photodiode captures the lasers sent by LiDAR
  2. Laser transmitter sent back the fake reflected lasers

- Forging non-existent vehicles to pose threat
  1. Forcing the ADS vehicle to emergency brake
  2. Keeping the ADS vehicle immobile
The Magic of Such Attacks

1. **Practicability**
   - Reusable traditional adversarial methods (FGSM, PGD, C&W...)

2. **Naturalness**
   - Difficult for human to distinguish

3. **Variability**
   - Various attack goals

   - Pre-processing
   - 3D Object Detector
   - Post-processing

   ① Differentiable
   ② A real car with occlusion (Only 200 points)
   ③ Inject Points
Existing Defense Methods

• **Universal Defenses**
  - Initial Motivation: Improving the robustness of PC models against noise
  - SRS and SOR
Existing Defense Methods

- **Specific Defenses**
  - Initial Motivation: Mitigating specific attack methods
  - SVF, CARLO and Shadow-Catcher

Refer from https://sites.google.com/view/shadow-catcher
Limitations of Existing Attacks

• Two Common Limitations
  1. Constrained by the **attack device** → the **position** and **number** of forged points
  2. Constrained by the **attack goal** → the **shape** of forged objects
Defense Insight

1. On the Position and Number
   - the distributions of point density and depth are different

2. On the Shape
   - the local difference is mostly larger than the global difference

![Graph showing point density vs. depth for Normal and Forged vehicles.](image)

![Bar charts showing Chamfer Distance and Average Square L2 Distance of kNN for different types and settings.](image)
1. **On the Position and Number**
   - Modeling the depth-density relation

2. **On the Shape**
   - Deploying local detector + Voting

**Explicit depth feature & Implicit density feature**

- **Raw Features**
  - Calculate
  - **Depth Info**

  - **Concatenate**
  - **Merged Features**

**Split the prediction by splitting input space**

- **Split**
- **Vote**
- Real?/Fake?
Our Proposed Method

• Local Objectness Predictor
  • Plug-and-Play Design (*No need to retrain the whole detector)
Defense Effectiveness

• More improvement on the performance and robustness of protected 3D object detectors

• Acceptable costs on memory and slightly larger costs on time

Results can further reduce by multi-processing

<table>
<thead>
<tr>
<th>Attacks</th>
<th>Time per sample (s)</th>
<th>GPU Mem (MB)</th>
<th>CPU Mem (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.060±0.005</td>
<td>1477</td>
<td>2551</td>
</tr>
<tr>
<td>SRS</td>
<td>0.069±0.007</td>
<td>1477</td>
<td>2549</td>
</tr>
<tr>
<td>SOR</td>
<td>0.114±0.005</td>
<td>5827</td>
<td>2516</td>
</tr>
<tr>
<td>FSD, r=0.6</td>
<td>2.463±0.005</td>
<td>1477</td>
<td>2506</td>
</tr>
<tr>
<td>FSD, r=0.7</td>
<td>0.089±0.002</td>
<td>1477</td>
<td>2511</td>
</tr>
<tr>
<td>LPD, r=0.6</td>
<td>1.341±0.011</td>
<td>2283</td>
<td>2518</td>
</tr>
<tr>
<td>LPD, r=0.7</td>
<td>1.589±0.013</td>
<td>3747</td>
<td>2506</td>
</tr>
<tr>
<td>PointNet, B=0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PointNet, B=0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simulation Experiments

- The performance of Apollo 6.0.0 deployed with LOP, evaluated in LGSVL simulator

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>ASR</th>
<th>time cost (ms)</th>
<th>FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apollo 6.0.0 (w/o. LOP)</td>
<td>8.33%</td>
<td>53.66%</td>
<td>33.36ms</td>
<td>29.97</td>
</tr>
<tr>
<td>Apollo 6.0.0 (w. LOP)</td>
<td>100.00%</td>
<td>0.00%</td>
<td>42.48ms</td>
<td>23.54</td>
</tr>
</tbody>
</table>

(a) Attack Scenario

(b) without LOP

(c) with LOP
Future Directions

Direction 1
The Existence of False positives

- Farther objects are harder to detect
  - 12.95%/16.53% of FP with depth < 10m/20m

Direction 2
The Upgrade of Attack Device

- The maximum of forged points is already up to 2500

The Fig.7 in PLA-LiDAR (S&P 2023)
Take Away Message

1. We conclude the limitations of existing appearing attacks

2. We propose a plug-and-play defense method LOP

3. We prove the effectiveness of our LOP online and offline

Thank you for your Audience!

For more details, welcome to follow our paper