Full bibliographic reference

Zhu, Yi, Chenglin Miao, Hongfei Xue, Yunnan Yu, Lu Su, and Chunming Qiao. "Malicious attacks against multi-sensor fusion in autonomous driving." In Proceedings of the 30th Annual International Conference on Mobile Computing and Networking, pp. 436-451. 2024.

Abstract

Multi-sensor fusion has been widely used by autonomous vehicles (AVs) to integrate the perception results from different sensing modalities including LiDAR, camera and radar. Despite the rapid development of multi-sensor fusion systems in autonomous driving, their vulnerability to malicious attacks have not been well studied. Although some prior works have studied the attacks against the perception systems of AVs, they only consider a single sensing modality or a camera-LiDAR fusion system, which can not attack the sensor fusion system based on LiDAR, camera, and radar. To fill this research gap, in this paper, we present the first study on the vulnerability of multi-sensor fusion systems that employ LiDAR, camera, and radar. Specifically, we propose a novel attack method that can simultaneously attack all three types of sensing modalities using a single type of adversarial object. The adversarial object can be easily fabricated at low cost, and the proposed attack can be easily performed with high stealthiness and flexibility in practice. Extensive experiments based on a real-world AV testbed show that the proposed attack can continuously hide a target vehicle from the perception system of a victim AV using only two small adversarial objects.

Link/DOI to the published paper

https://dl.acm.org/doi/10.1145/3636534.3649372

Malicious Attacks against Multi-Sensor Fusion in Autonomous Driving

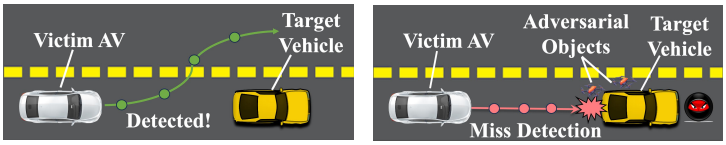
Yi Zhu[1], Chenglin Miao[2], Hongfei Xue[3], Yunnan Yu[4], Lu Su[5], Chunming Qiao[4]
[1] Wayne State University, USA [2] Iowa State University, USA
[3] University of North Carolina at Charlotte, USA
[4] State University of New York at Buffalo, USA [5] Purdue University, USA

WAYNE STATE UNIVERSITY ANDERSON COLLEGE OF ENGINEERING

Introduction

- Presents the first study on the vulnerability of multi-sensor fusion systems that employ LiDAR, camera, and radar in autonomous vehicles (AVs).
- Proposes a novel attack method that can simultaneously attack all three types of sensing modalities using a single type of adversarial object.
- The adversarial object can be easily fabricated at low cost.
- The proposed attack can be easily performed with high stealthiness and flexibility.
- Real-world experiments based on an AV testbed demonstrate the attack effectiveness and practicality.

Threat Model



(a) Without attack (b) With attack

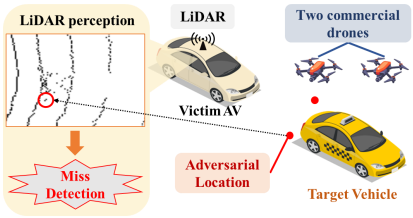
- A target vehicle, which could be any random vehicle on the road or one owned by the attacker, is in front of the victim AV.
- Attackers can place physical adversarial objects within the driving environment.
- The attack goal is to continuously hide the target vehicle from the multi-sensor fusion system.

Method

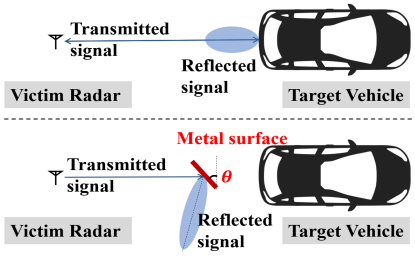
Develop a new type of compositive adversarial objects that combine attack vectors on Camera, LiDAR and Radar perception.

- **Attack Camera:** Camera perception models can be attacked by a specific *color pattern* [1].

- **Attack LiDAR:** Our previous study [2] shows that, LiDAR perception models can be attacked by placing arbitrary objects at a few *adversarial locations*.



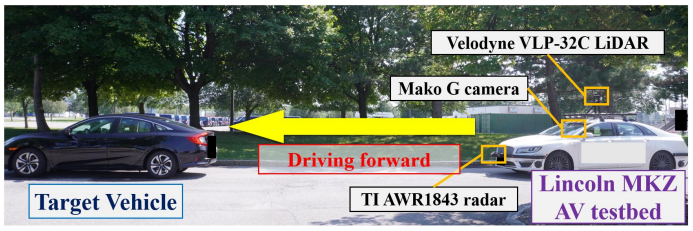
- **Attack Radar:** As shown in our previous work [3], a *smooth metal surface* placed between the victim AV and the target can block the radar signal to reduce the reflected signal amplitude.



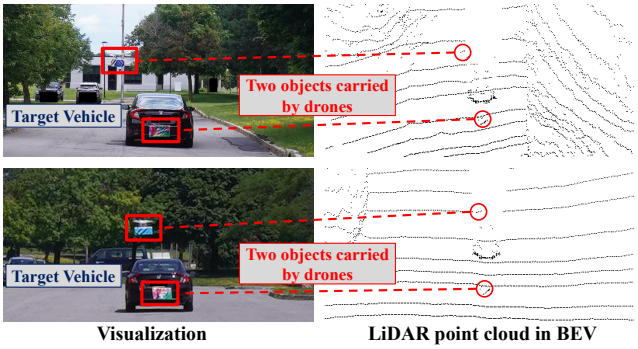
- **Compositive adversarial objects:**
  - A piece of cardboard covered by a metal foil and a color patch.
  - Strategically design their number sizes, orientations, locations, and color patterns
  - Use optimization to balance attack effectiveness and cost-efficiency

Experimental Results

- **Testbed:** A Lincoln MKZ AV testbed equipped with camera, LiDAR and radar.
- **Metrics:** Detection recall, percentage of sensory data frames in which the target is successfully detected by any of individual sensor.



- **Attack setting:** Generate adversarial objects and their optimal positions using offline simulation. Employ drones to carry the adversarial objects.



Visualization LiDAR point cloud in BEV

Model	Sensors	Type	Recall	Avg N	L_{area} (m^2)
BEVFusion	C+L	feature	1.00 / 0.08	2.21	0.21
CRFNet	C+R	feature	1.00 / 0.04	2.65	0.22
Radarnet	L+R	feature	1.00 / 0.02	2.35	0.20
LFusion	C+R	feature	1.00 / 0.00	2.77	0.24
RRPN	C+R	cascaded	1.00 / 0.00	2.58	0.21
HD-FPNet	C+L+R	cascaded	1.00 / 0.10	2.62	0.22

References

[1] Wang et al. Revisiting physical-world adversarial attack on traffic sign recognition: A commercial systems perspective. NDSS 2024.
[2] Zhu et al. Can we use arbitrary objects to attack lidar perception in autonomous driving?. CCS 2021.
[3] Zhu et al. TileMask: A PassiveReflection-based Attack against mmWave Radar Object Detection in Autonomous Driving. CCS 2023.