Invisible for both camera and LiDAR : Security of Multi Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attack

Yulong Cao ; Ningfei Wang ; Chaowei Xiao ; Dawei Yang ; Jin Fang; Ruigang Yang; Qi Alfred Cheny; Mingyan Liux; Bo Li

Present by HyeongJu Lee

Autonomous Driving

KAIST EE

Autonomous Driving(AD) vehicles are increasingly deployed on public roads

Perception in AD safety

KAIST EE

Perception is critical to AD safety.

Most important & safety-critical task : In-road obstacle detection

Errors in such a task can directly cause violent crashes.

An Uber self-driving car hit & killed a woman crossing street in Arizona since it cannot classify her as a pedestrian. [1] [2]

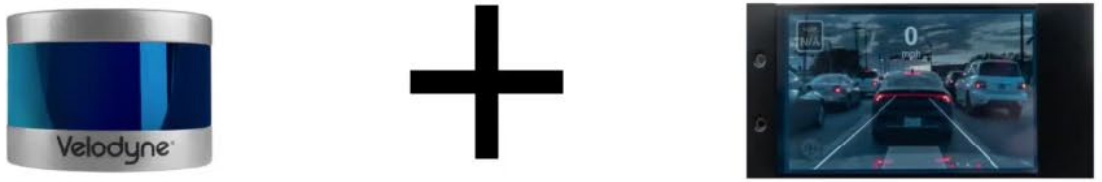
Tesla on autopilot crashed into an overturned truck since Autopilot driver-assist features didn't see the truck & safety feature didn't stop a collision. [3] [4]

Multi-Sensor Fusion(MSF) based AD perception

- Production high-level AD systems widely adopt **MSF-based perception** design
  - **Leverage strengths while compensate weaknesses** to achieve overall higher **accuracy & robustness**
    - Most popularly fuse from LiDAR & camera

Basic security design assumption :
Believed to hold in general

- In such design, assuming not all perception sources are(or can be) attacked simultaneously, theoretically always possible to rely on the unattacked source(s) to detect/prevent such attack



KAIST EE

MSF: Widely recognized as a general defense strategy against existing attacks on AD perception

10.3.2 Sensor-Level Defenses. Several defenses could be adopted against spoofing attacks on LiDAR sensors:

Detection techniques. Sensor fusion, which intelligently combines data from several sensors to detect anomalies and improve performance, could be adopted against LiDAR spoofing attacks. AV systems are often equipped with sensors beyond LiDAR. Cameras, radars, and ultrasonic sensors provide additional information and redundancy to detect and handle an attack on LiDAR.

[Cao et al. CCS'19]

5.2 Potential Countermeasures

Redundancy and Fusion: If a vehicle is equipped with multiple lidars having an overlapping field of view, the effect of saturating and spoofing can be mitigated to a certain extent. However, this directly increases the cost, and is not a definitive solution because attackers can blind multiple lidars simultaneously. Besides, it is also not easy to detect spoofing, when fake dots are induced in non-overlapped zones. Likewise, the fusion of multiple types of sensors cannot be an ultimate solution either. Radars [44], cameras [30, 44], and ultrasonic sensors [44] have all been revealed to be vulnerable to either blinding/jamming or spoofing.

[Shin et al. CHES'17]

In this work, we do not assume any particular sensing or actuation workflow to be trusted. However, we do assume that not all sensor readings can be corrupted simultaneously. Under the design where workflows run with isolation (see Section II-A), attacks or failures in a workflow can be constrained within. Admittedly, such cases could be possible in carefully crafted attacks. However, it is difficult for attackers. Firstly, for heterogeneous sensors, holding a vulnerability and a corresponding exploit which targets one sensing workflow is already costly [6], [9], not to mention corrupting all. Secondly, even if an attacker is capable of corrupting all sensors, the attacker needs to launch the attacks simultaneously to avoid detection. It is a great challenge to launch such coordinated attacks on different target sensing workflows [9].

[Guo et al. DSN'18]

As the system's autonomy increases, so does the concern about its security. In modern vehicles, a malicious attacker may deceive the controller into performing a dangerous action by altering the measurements of some sensors [1], [2]. Depending on the attacker's goal and capabilities, the consequences may range from minor disturbances in performance to crashes and loss of human lives. Consequently, performing attack-resilient sensor fusion is essential for the safety of such systems.

[Ivanov et al. DATE'14]

7

2.1 System Model and Current Approach

We consider a system with n sensors measuring the same physical variable. As mentioned above, we assume *abstract* sensors; therefore, each sensor provides the controller with an interval of all possible values. We assume the system queries all the sensors periodically such that a centralized estimator receives measurements from all sensors, and then performs attack detection/identification and *sensor fusion* (SF). We now explain the current approach to attack detection, referred to herein as a SF-based detector, before providing the improved version addressed in this paper.

[Park et al. ICCPS'15]

Introduction

KAIST EE

Research Question:

Can such basic security design assumption actually be broken, especially in practical AD settings?

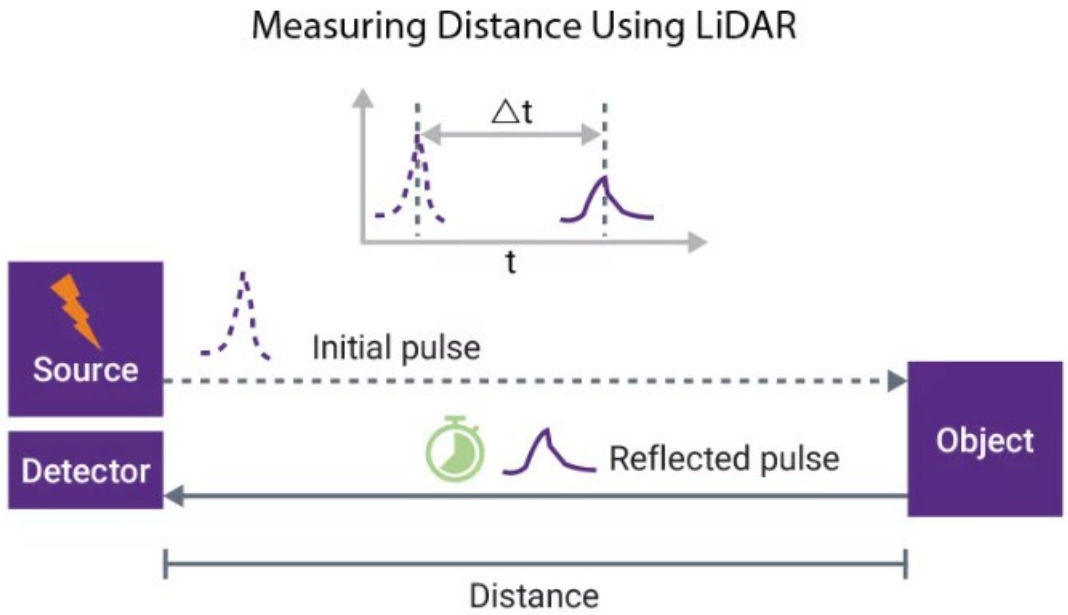
Challenge 1: Lack of single physical-world attack vector effective for both camera-&LiDAR-based AD perception.
Challenge 2: Need to differentiably synthesize physically-consistent attack impacts onto both the camera and LiDAR.
Challenge 3 : Need to handle non-differentiable pre-processing steps in AD perception.

Background

KAIST EE

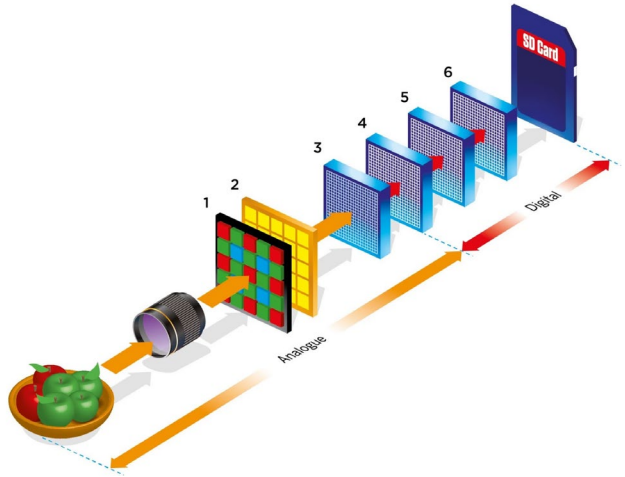
LiDAR

- Use laser beam
- Can measure distance



Camera sensor

- Convert light Energy to electrical charge

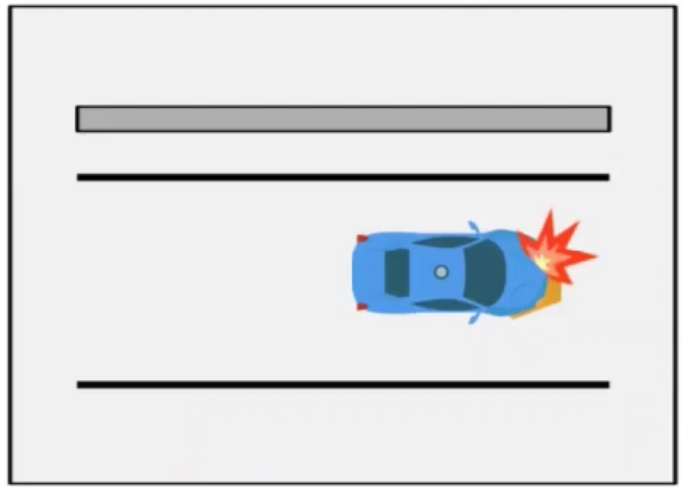


Problem formulation & Attack goal

KAIST EE

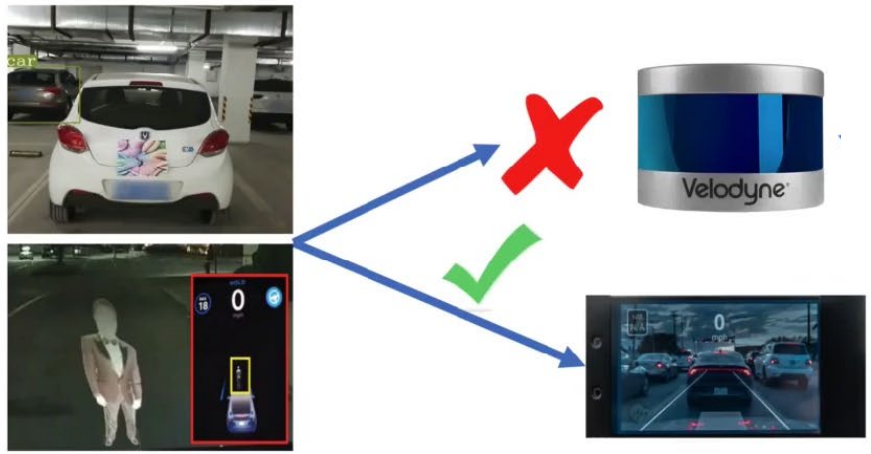
- Problem formulation
  - Target **physical-world attack** vectors for high **practicality** & **realism**
  - Effectively attack **all** perception source used in MSF-based AD perception
    - For today's popular design : Camera + LiDAR

- Attack goal
  - Fool MSF-based AD perception in victim AD vehicles to fail in detecting a front obstacle & thus crash into it



First challenge : Attack vector

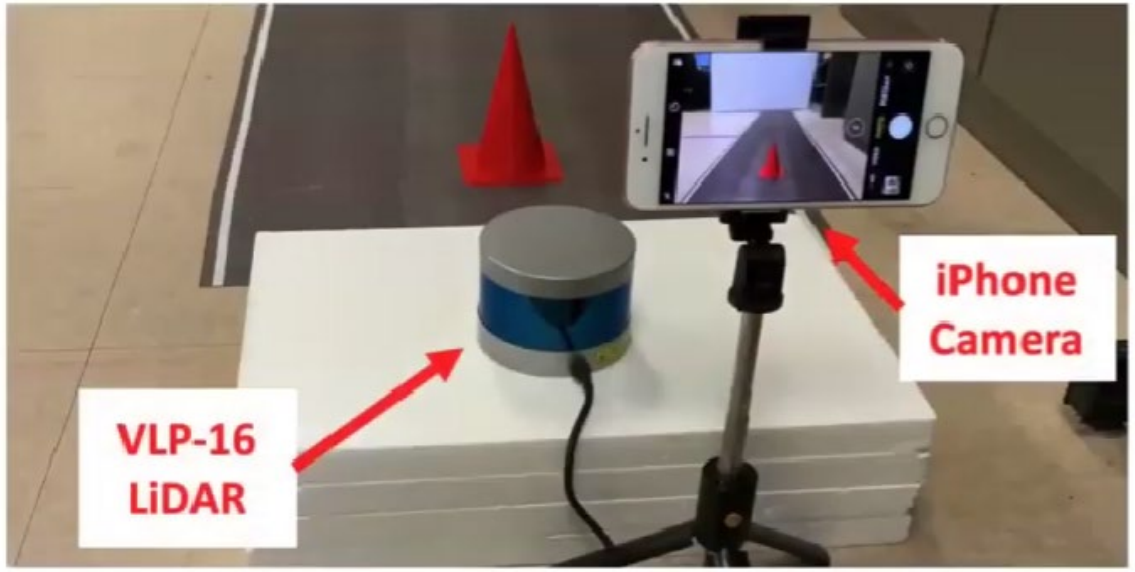
- Ideal if find a single physical-world attack vector effective for both camera- & LiDAR-based AD perception
  - However, no previously-used attack vectors shown effectiveness for both



Adversarial 3D object: Physically-realizable & stealthy attack vector for MSF-based perception

- Insight: Different shapes can lead to both **point position changes** in LiDAR point cloud & **pixel value changes** in camera image
- Via 3D printing technology
- Can achieve high **stealthiness** by mimicking a normal traffic object
- Attacker: Place it on roadway to trick victim AD vehicle to crash into it
  - Cause severe crash by filling dense materials(e.g., granite or metal)

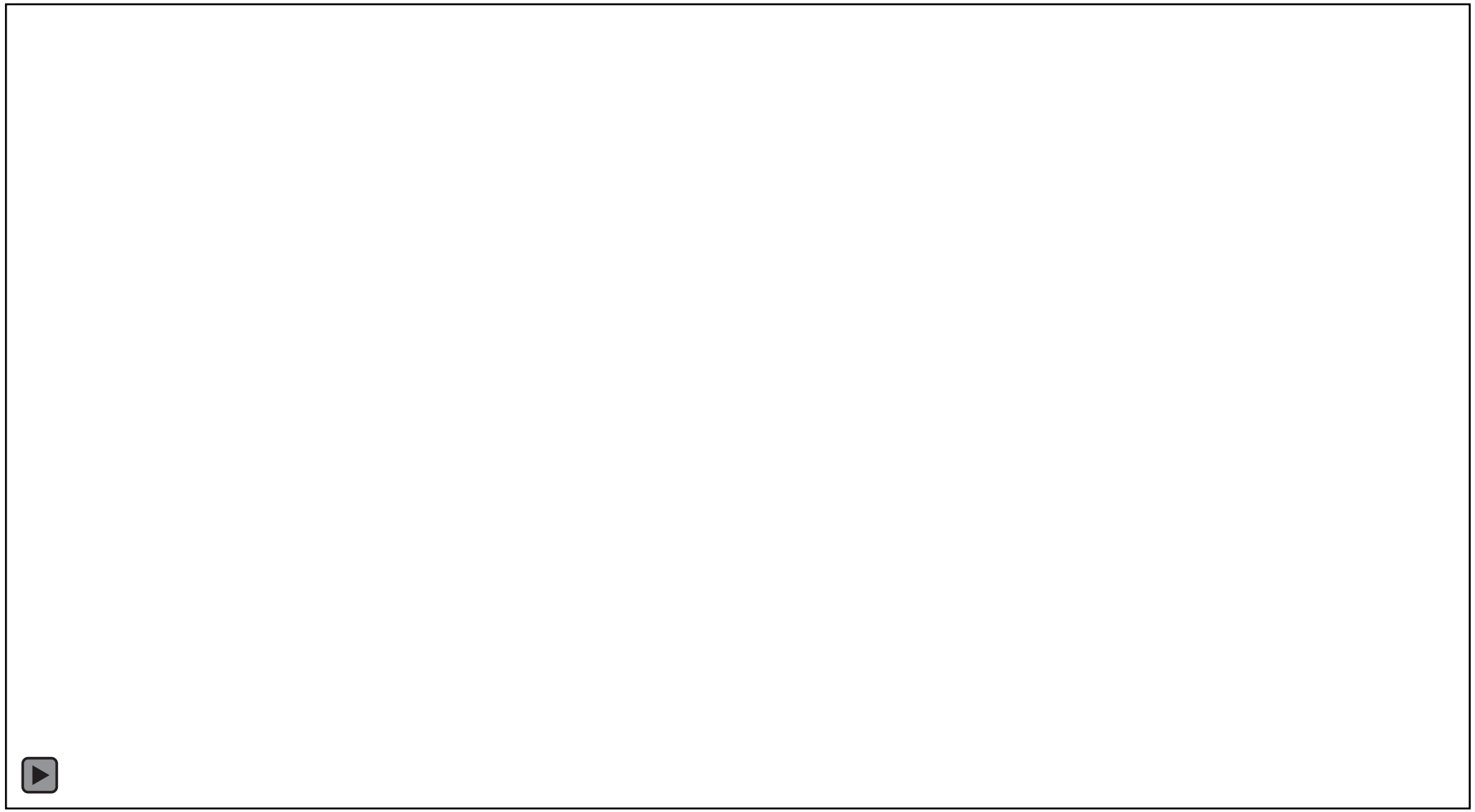
Attack demo 1: Miniature-scale physical-world setup



Benign

Adversarial

Attack demo 1: Miniature-scale physical-world setup



Attack demo 2 : Real vehicle based setup

Road & car with LiDAR & camera

Benign

Adversarial

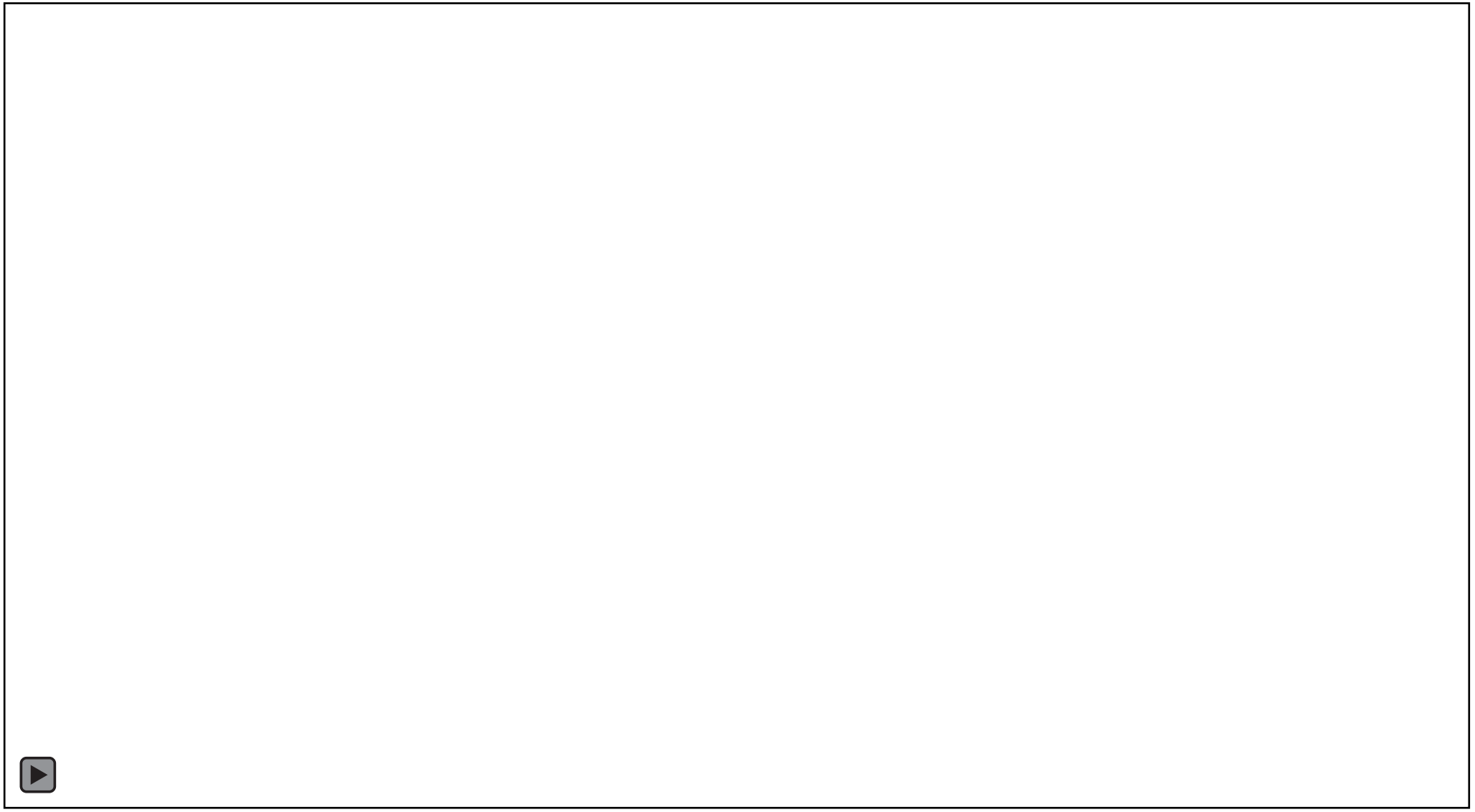
Attack demo 2 : Real vehicle based setup

KAIST EE

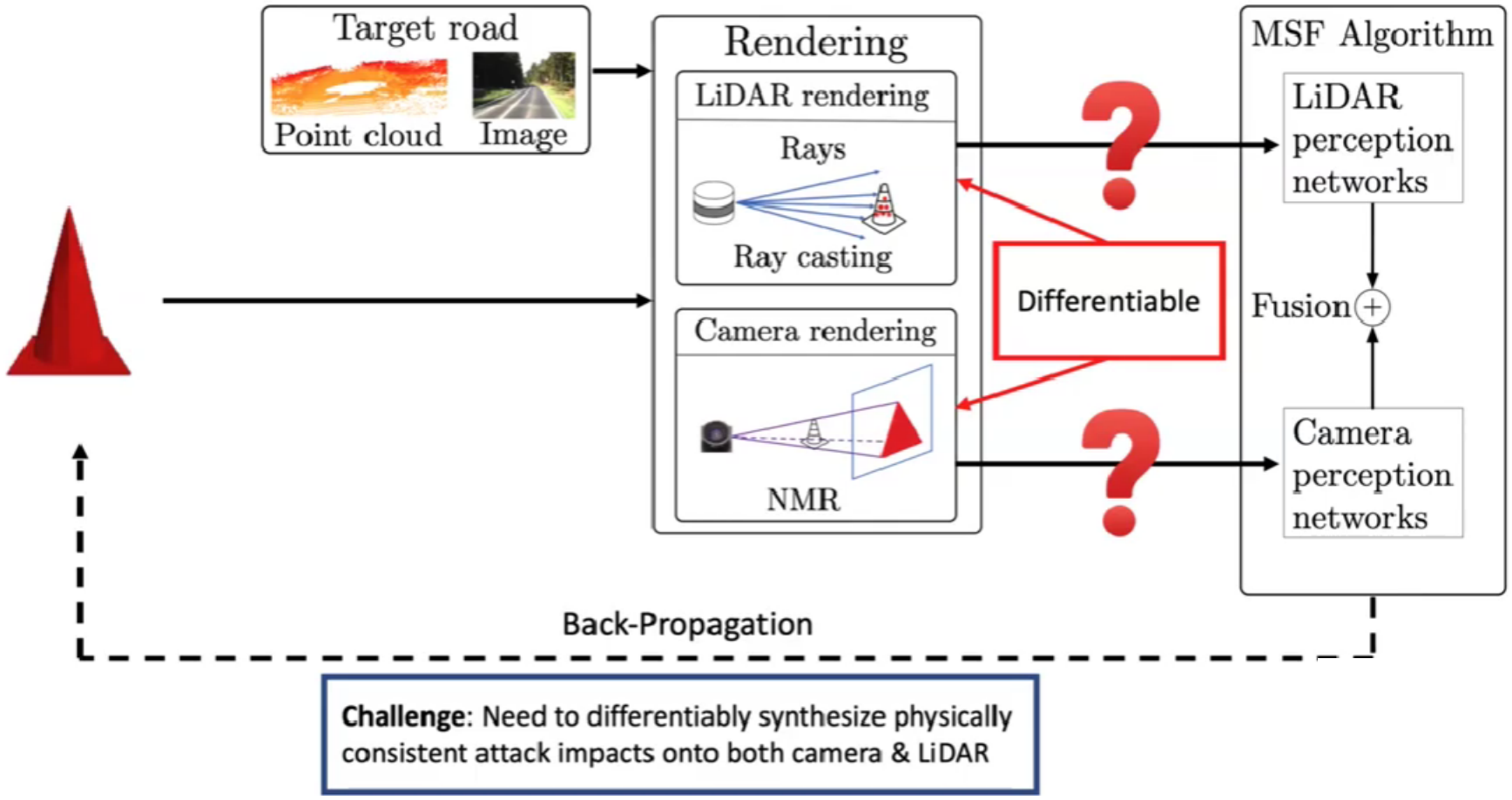
Attack demo 3 : End-to-End attack simulation setup

Road in LGSVL AD simulator

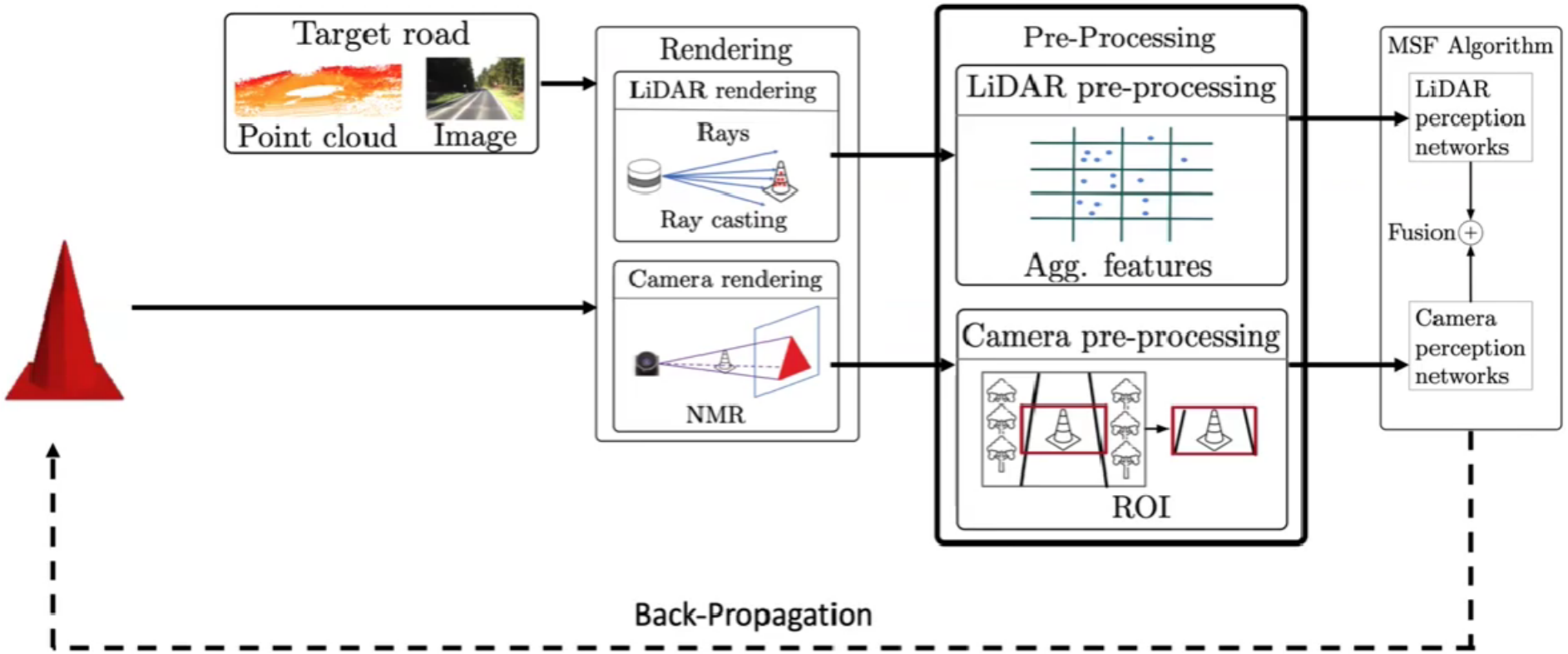
Attack demo 3 : End-to-End attack simulation setup



MSF-ADV design: Optimization framework



MSF-ADV design: Differentiable pre-processing

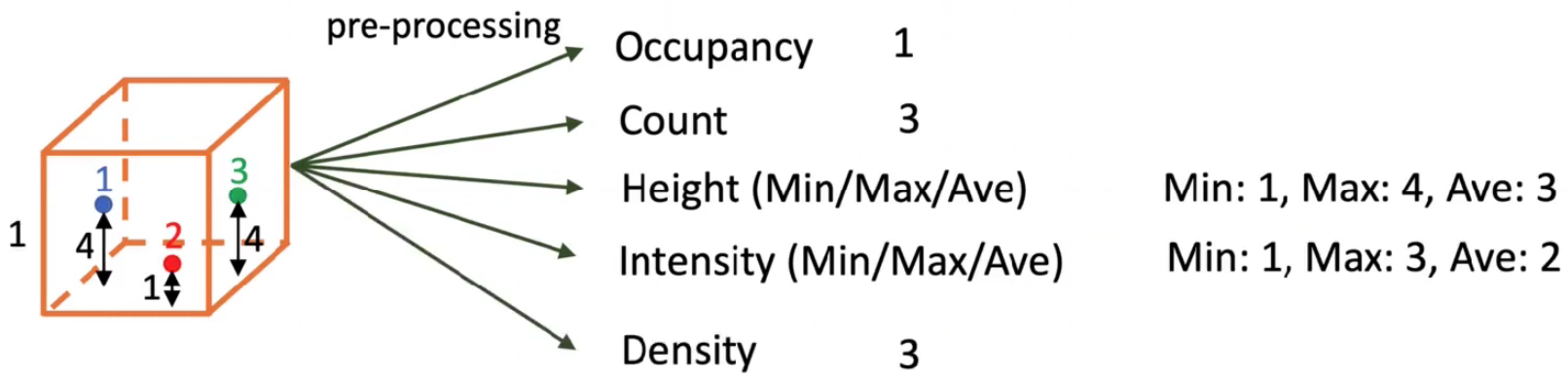


25

LiDAR pre-processing

KAIST EE

- LiDAR-based object detection models popularly use **cell-level aggregated features**

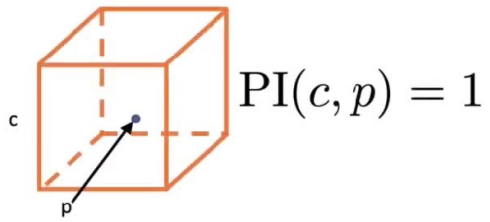


Necessary first step: Point-Inclusion(PI) calculation

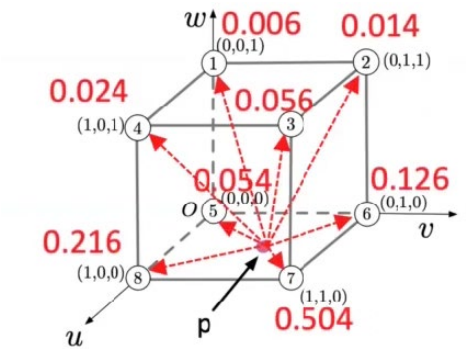
KAIST EE

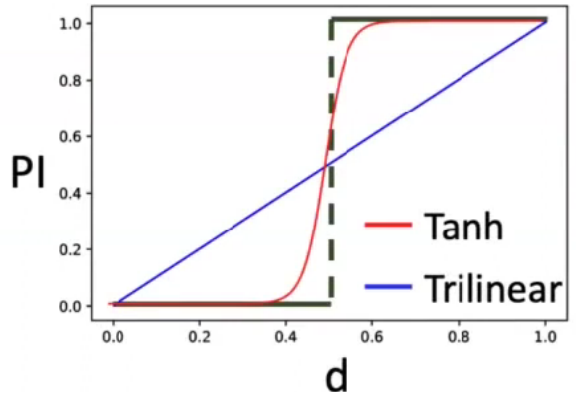
- **Point-inclusion(PI)** : Calculate whether a point is inside a cell or not
  - Discontinuous by nature : 0 & 1 for outside & inside a cell
- Strategy: Design a novel & accurate differentiable function to approximate the calculation of PI, or **soft PI**

- **Soft PI** : Estimate **probability** of PI using **interpolation**
  - Tried different interpolation functions to improve accuracy



$$\mathrm{PI}(c, p) = 1$$



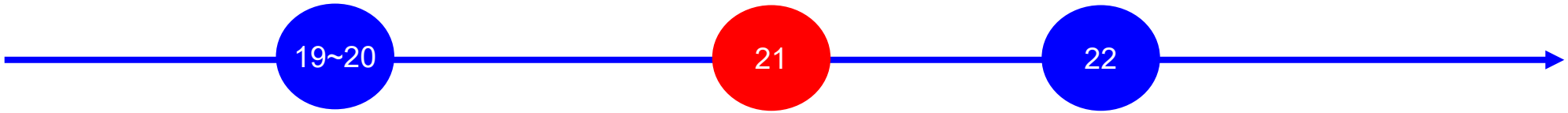


Evaluation

KAIST EE

- Evaluate on MSF algorithms included in open-source full-stack AD systems
  - Select 3 object types with 100 real-world driving scenarios from KITTI dataset
- Effectiveness
  - >**91%** success rate across
- Robustness
  - Robust to different victim positions & angles, w/ >**95%** average success rate
- Transferability
  - Transferable across different MSF algorithms, w/ **75%** average success rate

Relate work



Adversarial attack on LiDAR/camera

("Adversarial Sensor Attack on LiDAR-based Perception in Autonomous Driving," in ACM CCS, 2019.)
("Adversarial Camera Stickers: A Physical Camera-based Attack on Deep Learning Systems," in ICML, 2019, pp. 3896–3904.)

This paper

Consider multi sensor attack

("3D object detection for autonomous driving: A comprehensive survey," 2022)

Discussion

Do factors like rain or snow potentially affecting camera image sensing or lidar distance measurements make the experiment still valid under such conditions or different angles?

-Lidar can detect distances, and it should be able to confirm when something is getting closer, even if objects aren't recognized as obstacles in the lidar data. In autonomous driving, it seems like the vehicle should be able to stop automatically when it gets close to an object, but is it really likely to cause accidents?

Conclusion

- Design a novel attack with adversarial 3D object as physical-world attack vector
- Their attack is successful showing high effectiveness, stealthiness, robustness, transferability, and physical-world realizability
- Perform first study on security of MSF-based AD perception

Q&A

Good question

Jio Oh: Are there other perception methods that autonomous vehicles use? Moreover, is there a way for the vehicle to do AD, besides perception…?

Zhixian Jin: The author mentioned that no prior works have considered defending against adversarial 3D objects, but does it really matter for the camera?
I am wondering why the existing defense on adversarial 2D objects cannot apply to this attack easily.

Best question

Seunghyun Lee: While using both camera and LiDAR perception sources would in principle be more robust against adversarial attacks, this paper does not clearly show whether or not attacks using only one perception source are infeasible. Would it be possible to affect just one perception source sufficiently enough so that the fused perception result is altered?

Taeung Yoon: Is there a potential increase in security threats when incorporating additional perception sensors like RADAR in a multi-sensor fusion approach?

Jaehyun Ha: Why the basic security design assumption (not all perception sources are attacked simultaneously) was believed to hold in general?