MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Achieving the Safety and Security of the End-to-End AV Pipeline

Noah T. Curran[*], **Minkyoung Cho**[*], **Ryan Feng**[*], **Liangkai Liu**[*], **Brian Jay Tang**[*], **Pedram MohajerAnsari**[°], **Alkim Domeke**[°], **Mert D. Pesé**[°], **and Kang G. Shin**[*]
[*] Computer Science and Engineering, University of Michigan
[°] School of Computing, Clemson University

1st Cybersecurity in Cars Workshop (CSCS) @ CCS
Salt Lake City, UT, USA
10/18/24

Context for this Paper

- University of Michigan CSE reading group meetings:

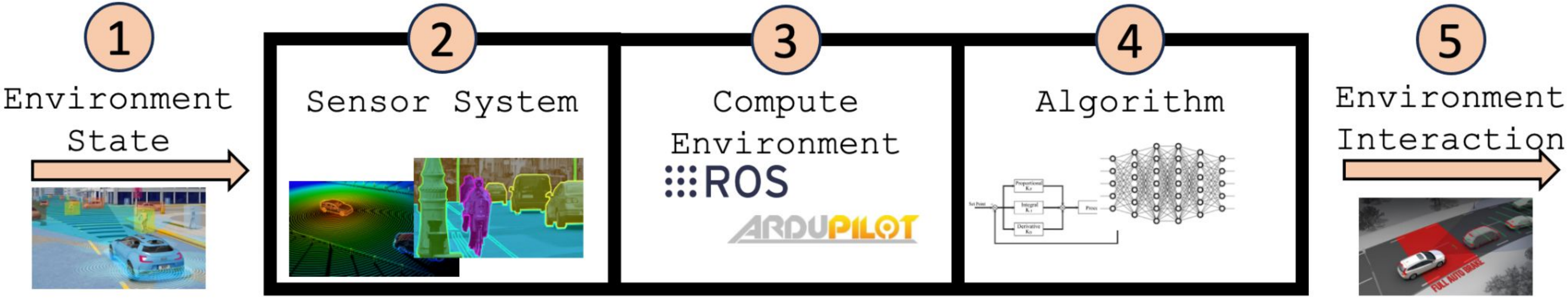
AV Safety and Security Reading Group

Alongside Ryan Feng, I am co-organizing a reading group for Autonomous Vehicle (AV) Safety and Security. This is hosted in the CSE Department at the University of Michigan, but anyone who is able to make it in-person is welcome to join.

- Inter-lab discussions:
  - ↳ Where did AV security research originate?
  - ↳ What has its progress focused on?
  - ↳ Where is it headed?

Overview

- We breakdown the AV security problem into context of the Sense-Plan-Act pipeline:



1 Environment State

2 Sensor System

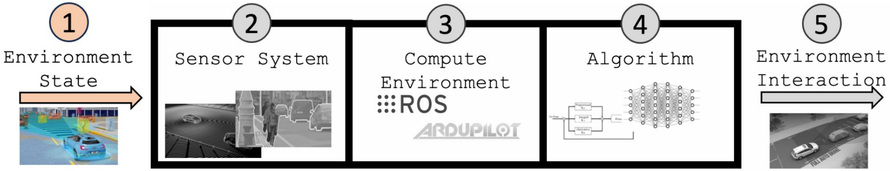
3 Compute Environment

4 Algorithm

5 Environment Interaction

- This becomes important when we de-isolate each component and discuss security in the context of the entire end-to-end process

❶ Environment State



1 Environment State

2 Sensor System

3 Compute Environment

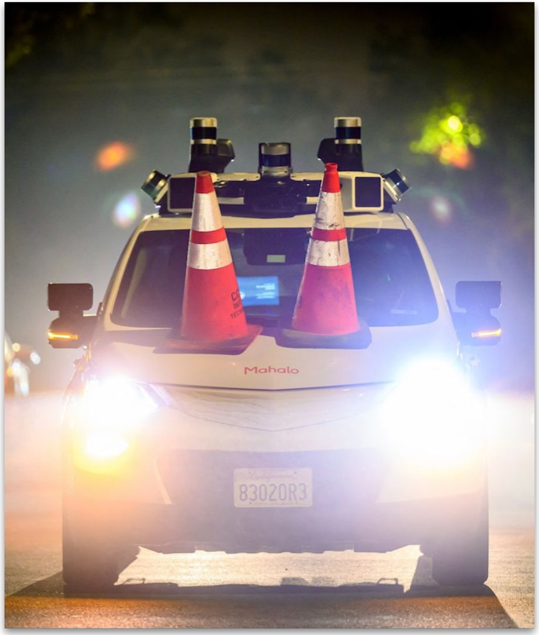
ROS

ARDUPILOT

4 Algorithm

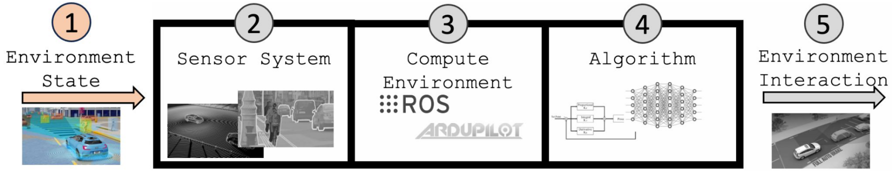
5 Environment Interaction

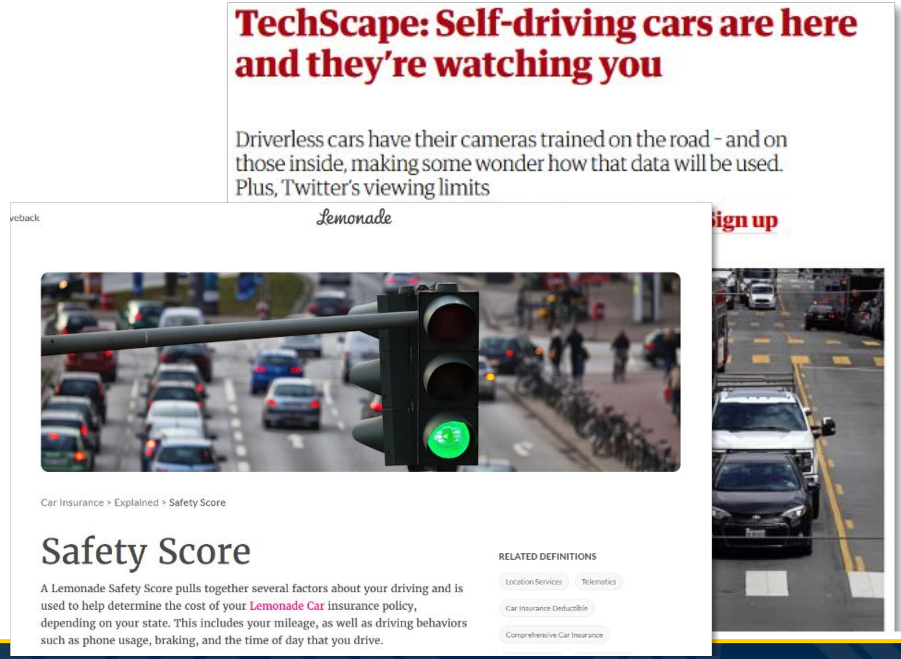
- Rich environment data:
  - ↳ Road layouts
  - ↳ Traffic signs
  - ↳ Businesses
  - ↳ Pedestrians
  - ↳ Other vehicles
- Collected with various sensors



❶ Privacy and Surveillance

- ❖ Data can be exploited for targeted or mass surveillance of individuals and communities.
- ❖ Privacy vs. Utility/Safety Trade-Off
  - ↳ (Hint) Privacy never wins





TechScape: Self-driving cars are here and they're watching you

Driverless cars have their cameras trained on the road - and on those inside, making some wonder how that data will be used. Plus, Twitter's viewing limits

ign up

veback

Lemonade

Car Insurance > Explained > Safety Score

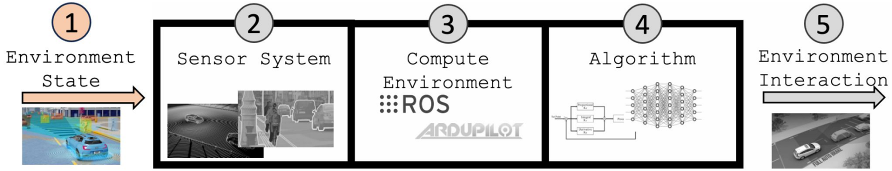
Safety Score

A Lemonade Safety Score pulls together several factors about your driving and is used to help determine the cost of your Lemonade Car insurance policy, depending on your state. This includes your mileage, as well as driving behaviors such as phone usage, braking, and the time of day that you drive.

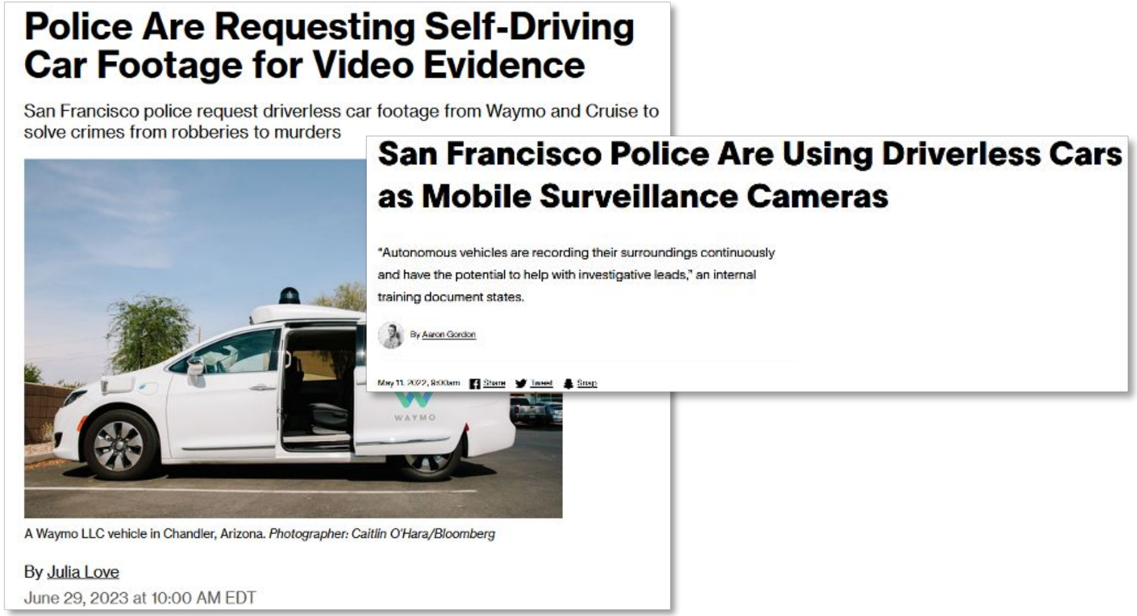
RELATED DEFINITIONS

Location Services
Telematics
Car Insurance Deductible
Comprehensive Car Insurance

❶ Law Enforcement Use



❖ Instances of data being requested by law enforcement without warrants.



Police Are Requesting Self-Driving Car Footage for Video Evidence

San Francisco police request driverless car footage from Waymo and Cruise to solve crimes from robberies to murders

A Waymo LLC vehicle in Chandler, Arizona. *Photographer: Caitlin O'Hara/Bloomberg*

By Julia Love

June 29, 2023 at 10:00 AM EDT

San Francisco Police Are Using Driverless Cars as Mobile Surveillance Cameras

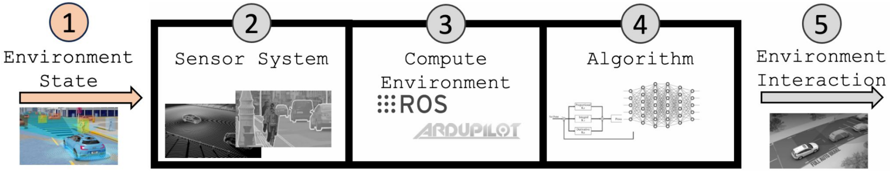
"Autonomous vehicles are recording their surroundings continuously and have the potential to help with investigative leads," an internal training document states.

By Aaron Gordon

May 11, 2022, 8:00am Share Tweet Snap

❶ Existing Works

- Most papers and journals focus on legal implications
- Existing computer science research is too focused on machine learning privacy and proprietary privacy (federated learning, differential privacy, etc).



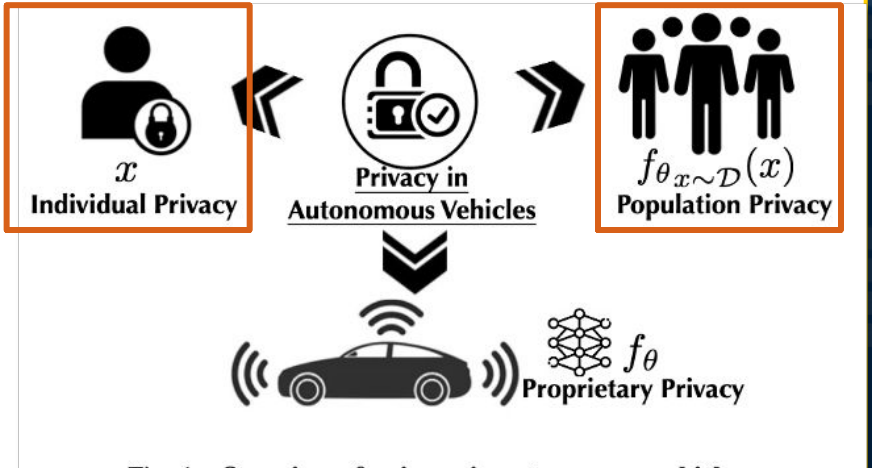
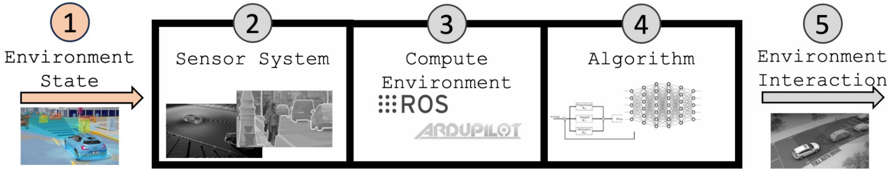


Fig. 1. **Overview of privacy in autonomous vehicles.**

[1] Privacy of Autonomous Vehicles: Risks, Protection Methods, and Future Directions., Xie *et al*, arXiv:2209.04022

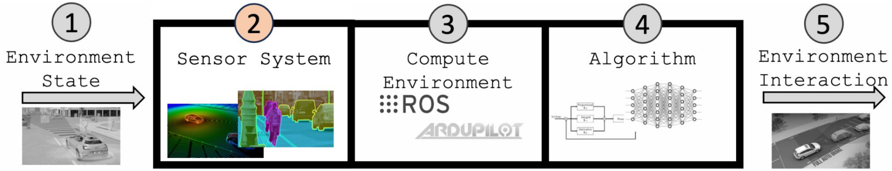
❶ Limitations & Takeaways



- More focus on privacy research in autonomous vehicles is needed:
  - Opt-out systems for bystander’s data privacy
  - Built-in privacy into cameras and vehicles [1]
    - Blur faces, bodies, businesses, license plates, etc.
  - Transparency and auditing systems for AV data privacy
  - Finding safer alternatives to “adversarial attack”-based privacy technologies

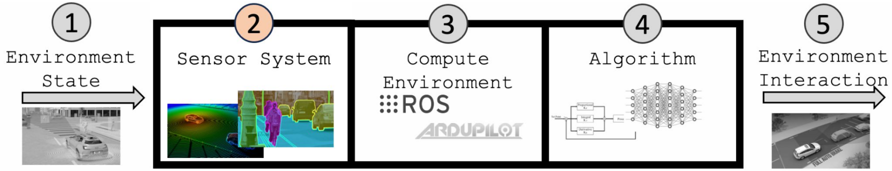
[1] PrivacyLens: On-Device PII Removal from RGB Images using Thermally-Enhanced Sensing, Iravantchi *et al*, PETS 2024

❷ Perception Sensors



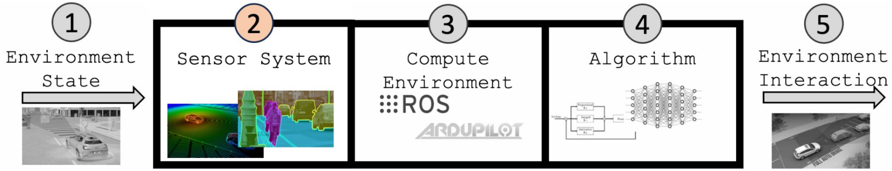
- Vehicle perception:
  - ↳ Cameras

❷ Perception Sensors

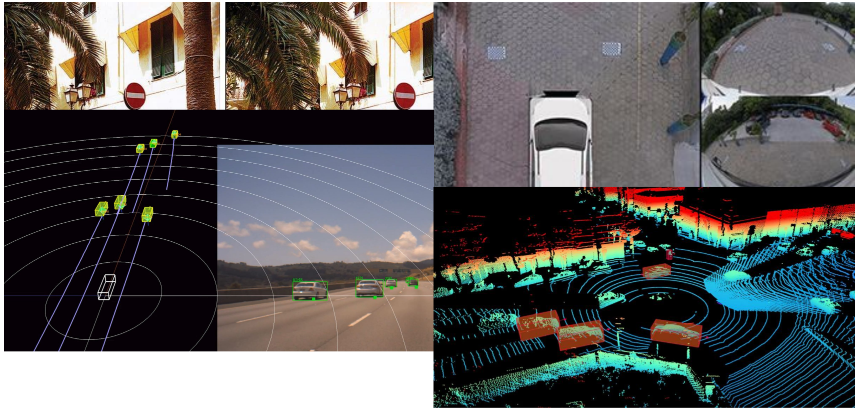


- Vehicle perception:
  - Cameras
  - Radar

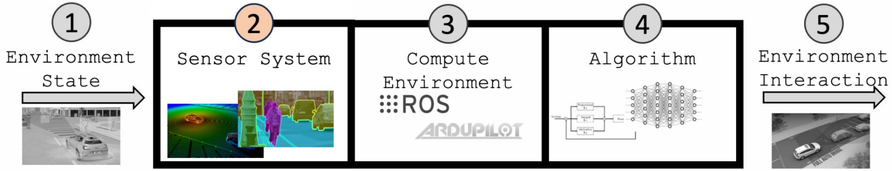
❷ Perception Sensors



- ❖ Vehicle perception:
  - ↳ Cameras
  - ↳ Radar
  - ↳ LiDAR



❷ Perception Sensors



1 Environment State

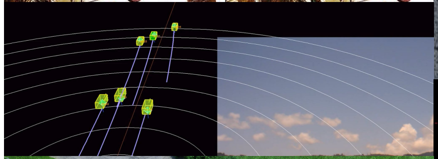
2 Sensor System

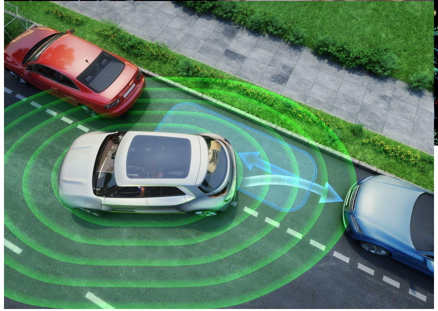
3 Compute Environment

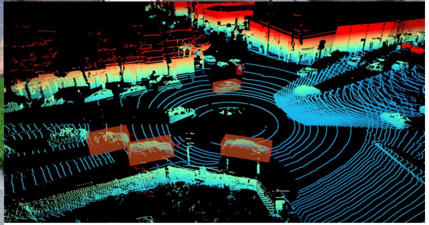
4 Algorithm

5 Environment Interaction

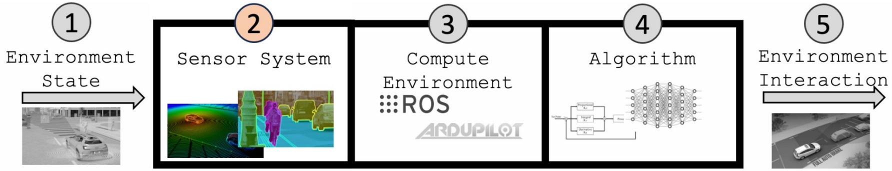
- Vehicle perception:
  - ↳ Cameras
  - ↳ Radar
  - ↳ LiDAR
  - ↳ Ultrasonic







❷ Perception Sensors



1 Environment State

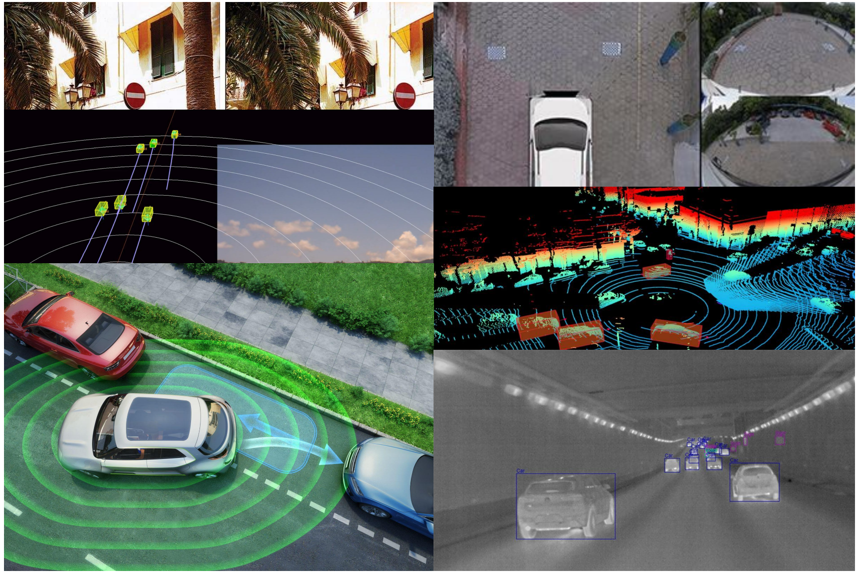
2 Sensor System

3 Compute Environment

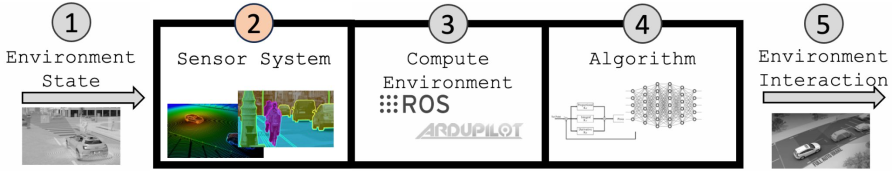
4 Algorithm

5 Environment Interaction

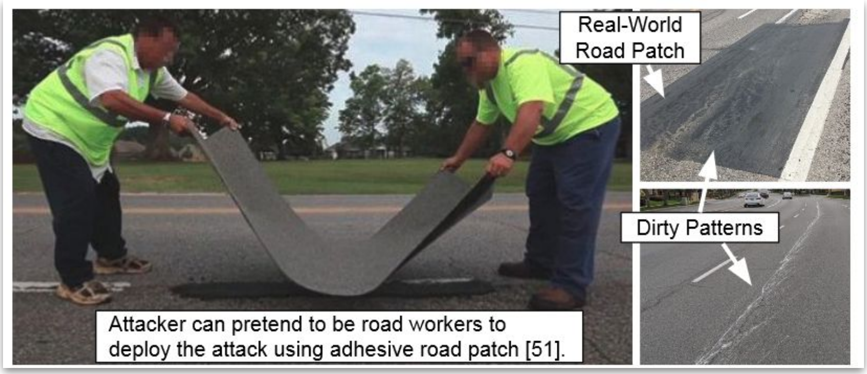
- Vehicle perception:
  - ↳ Cameras
  - ↳ Radar
  - ↳ LiDAR
  - ↳ Ultrasonic
  - ↳ Thermal



❷ Sensor Attacks



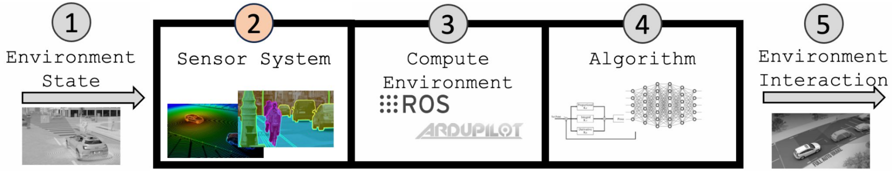
- Indirect Attacks
  - ↳ Pavement modifications



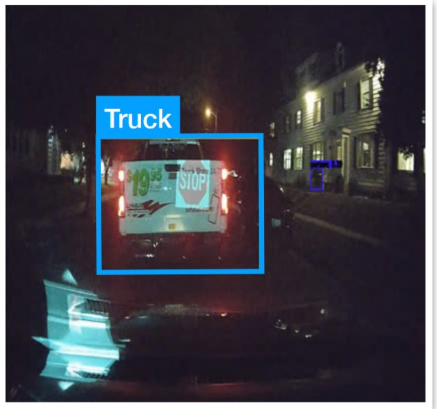
Attacker can pretend to be road workers to deploy the attack using adhesive road patch [51].

Source: Sato et al. [Security '21]

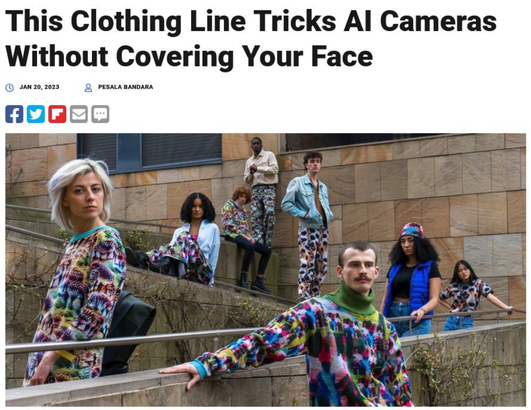
❷ Sensor Attacks



- ❖ Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing



Source: Man et al. [Security '23]

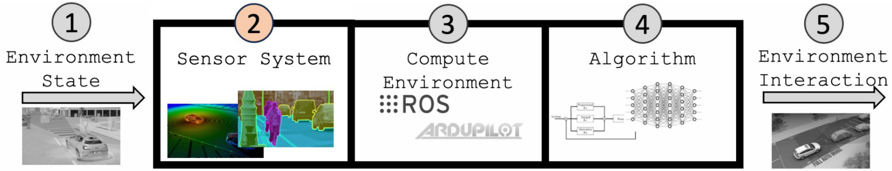


This Clothing Line Tricks AI Cameras Without Covering Your Face

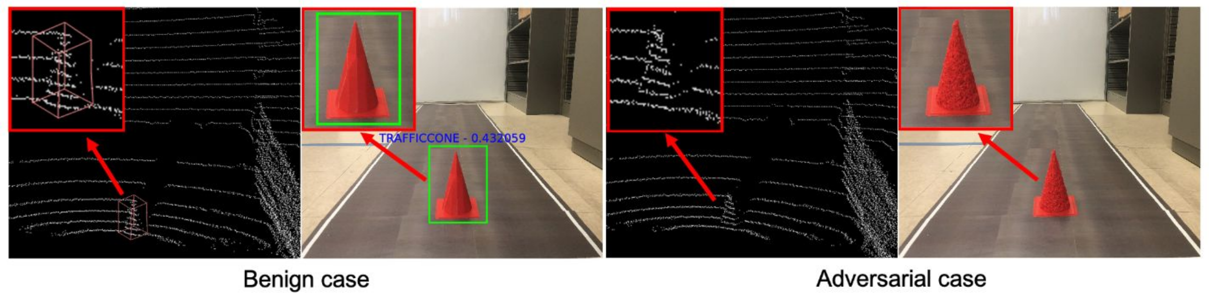
JAN 20, 2023 PESALA BANDARA

Source: Bandara, PetaPixel

❷ Sensor Attacks



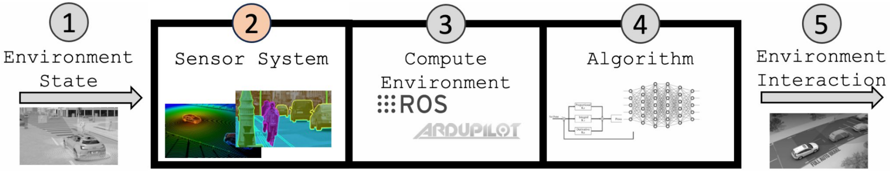
- Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects



Source: Cao et al. [S&P '21]

❷ Sensor Attacks

- ❖ Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects
- ❖ Direct Attacks
  - ↳ Laser for LiDAR removal / injection



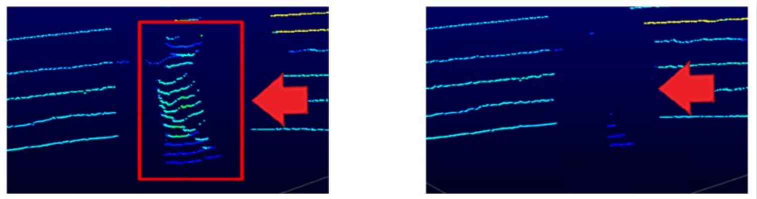
1 Environment State

2 Sensor System

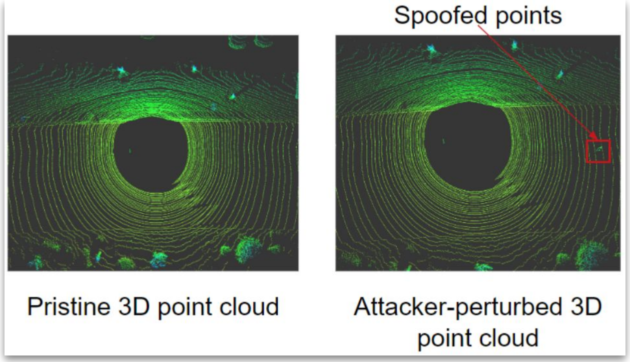
3 Compute Environment

4 Algorithm

5 Environment Interaction

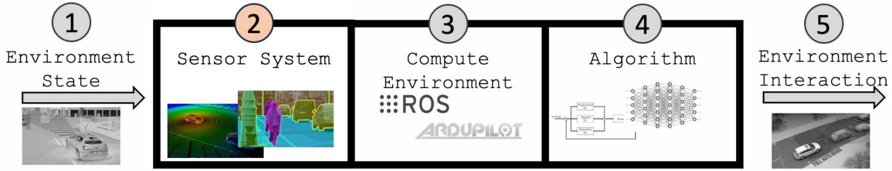


Source: Cao et al. [Security '23]

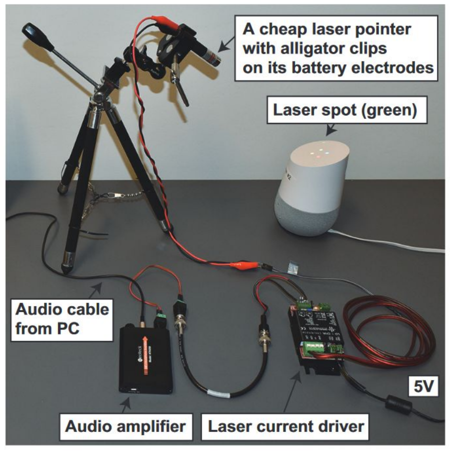


Source: Cao et al. [CCS '19]

❷ Sensor Attacks

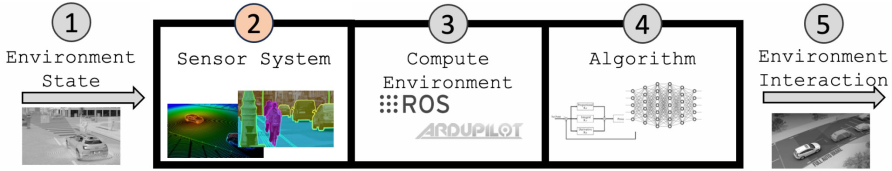


- Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects
- Direct Attacks
  - ↳ Laser for LiDAR removal / injection
  - ↳ Laser for microphone injection

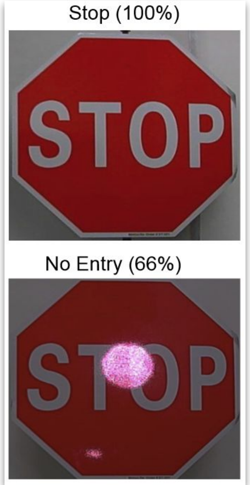


Source: Sugawara et al. [Security ‘20]

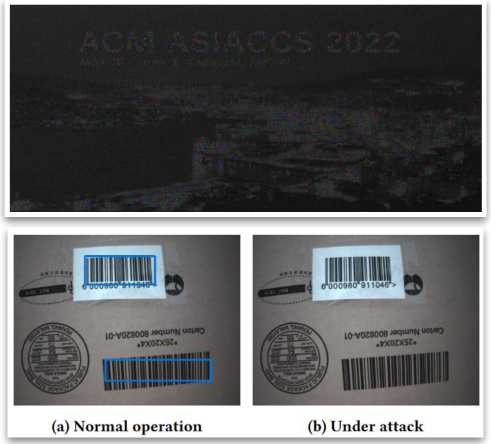
❷ Sensor Attacks



- ❖ Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects
- ❖ Direct Attacks
  - ↳ Laser for LiDAR removal / injection
  - ↳ Laser for microphone injection
  - ↳ IR / acoustics / electromagnetic interference for camera injection



Source: Sato et al. [NDSS '24]

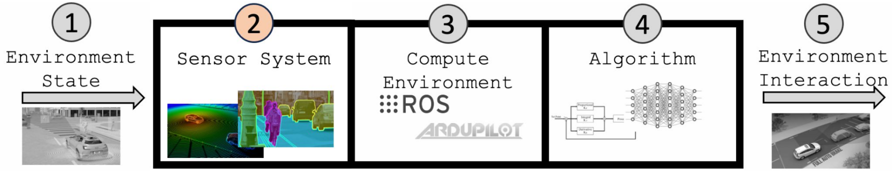


(a) Normal operation

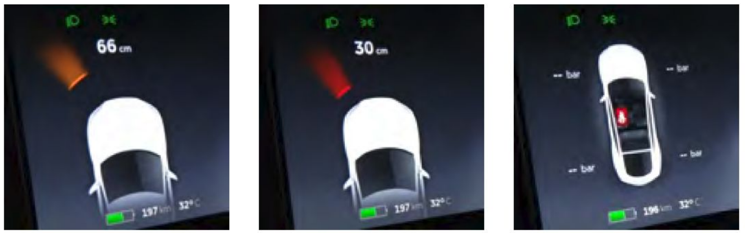
(b) Under attack

Source: Köhler et al. [CCS '22]

❷ Sensor Attacks



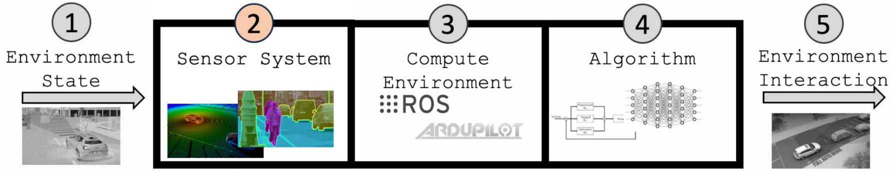
- ❖ Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects
- ❖ Direct Attacks
  - ↳ Laser for LiDAR removal / injection
  - ↳ Laser for microphone injection
  - ↳ IR / acoustics / electromagnetic interference for camera injection
  - ↳ Ultrasonic spoofing



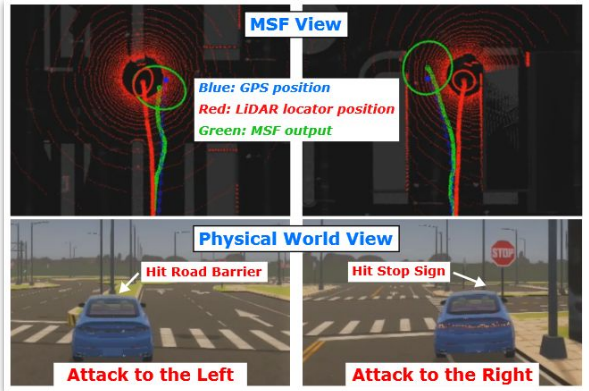
(a) Normal. (b) Spoofed. (c) Jammed.

Source: Yan et al. [DEF CON '16]

❷ Sensor Attacks

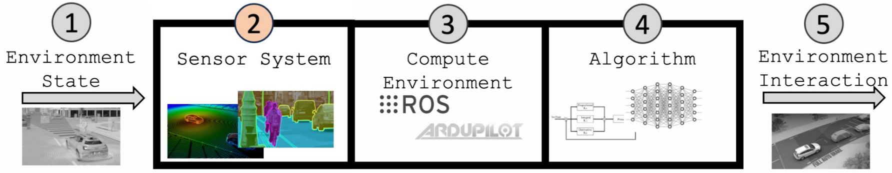


- ❖ Indirect Attacks
  - ↳ Pavement modifications
  - ↳ Adversarial patch projection / clothing
  - ↳ Adversarial objects
- ❖ Direct Attacks
  - ↳ Laser for LiDAR removal / injection
  - ↳ Laser for microphone injection
  - ↳ IR / acoustics / electromagnetic interference for camera injection
  - ↳ Ultrasonic spoofing
  - ↳ GPS / GNSS spoofing



Source: Shen et al. [Security '20]

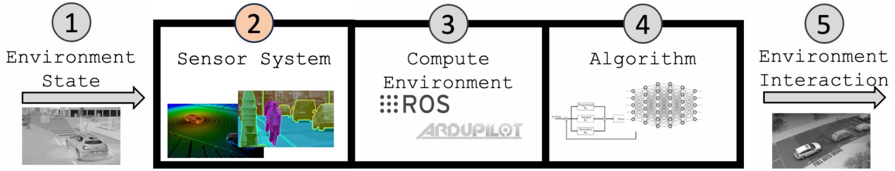
❷ Sensor Defenses



- For indirect attacks:
  - ↳ Train on dataset which includes adversarial perturbations (adversarial training)
  - ↳ Strong against *known* attacks with minor impact on normal performance

- For direct attacks:
  - ↳ Hardware modifications
  - ↳ Spatial and temporal invariant checks
  - ↳ Physical modifications, *e.g.*, lens filters
  - ↳ Classical intrusion/anomaly detection techniques

❷ Limitations & Takeaways

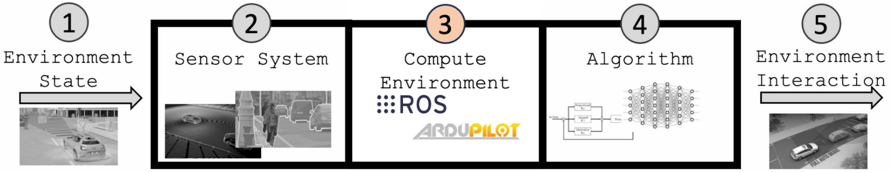


- Absence of provable defenses
  - ↳ Even invariants make assumptions that are often broken in the real world

- No common evaluation practice

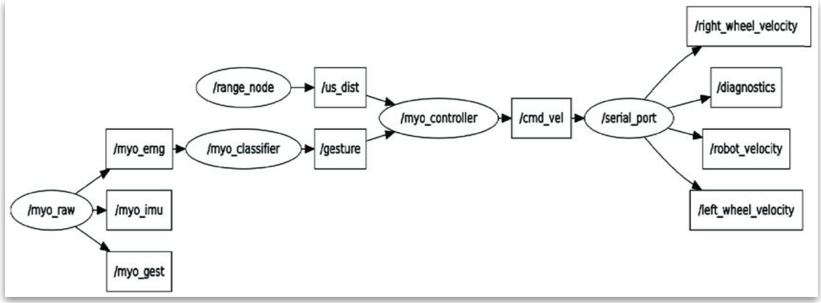
- Limited knowledge of downstream impact on safety

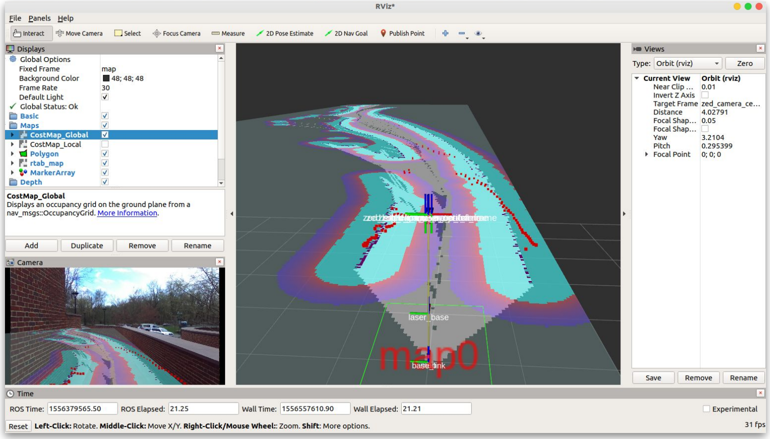
❸ Middleware



❖ Robot Operating System (ROS)

↳ Communication tool





Global Options
Fixed Frame: map
Background Color: 48; 48; 48
Frame Rate: 30
Default Light: ☑
✓ Global Status: Ok
Basic ☑
Maps ☑
CostMap_Global ☑
CostMap_Local ☐
Polygon ☑
rtab_map ☑
MarkerArray ☑
Depth ☑

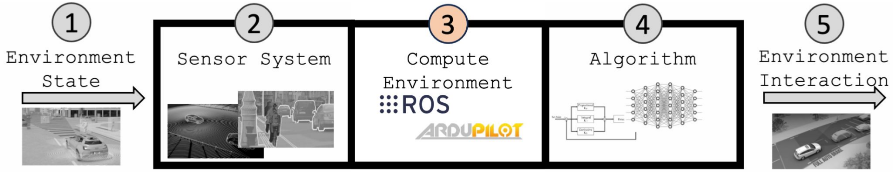
CostMap_Global
Displays an occupancy grid on the ground plane from a nav_msgs::OccupancyGrid. More Information.

Type: Orbit (rviz)
Current View: Orbit (rviz)
Near Clip …: 0.01
Invert Z Axis: ☐
Target Frame: zed_camera_ce…
Distance: 4.02791
Focal Shap…: 0.05
Focal Shap…: ☐
Yaw: 3.2104
Pitch: 0.295399
Focal Point: 0; 0; 0

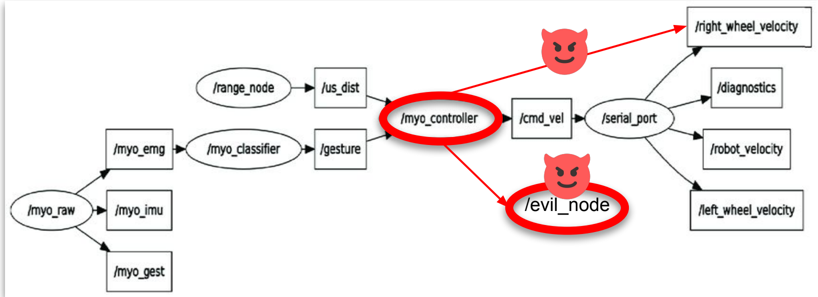
ROS Time: 1556379565.50 ROS Elapsed: 21.25 Wall Time: 1556557610.90 Wall Elapsed: 21.21 ☐ Experimental

Left-Click: Rotate. **Middle-Click:** Move X/Y. **Right-Click/Mouse Wheel:**: Zoom. **Shift**: More options. 31 fps

❸ Middleware Security

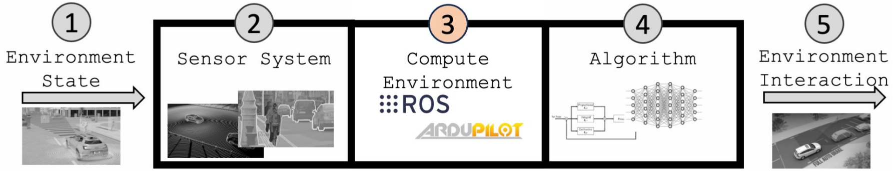


- Message passthrough authentication
- Node creation authentication



- Effects:
  - ↳ Excessive resource utilization; malicious payloads
  - ↳ Impact scheduler to cause timing delays → Take advantage of ROS bug that causes starvation

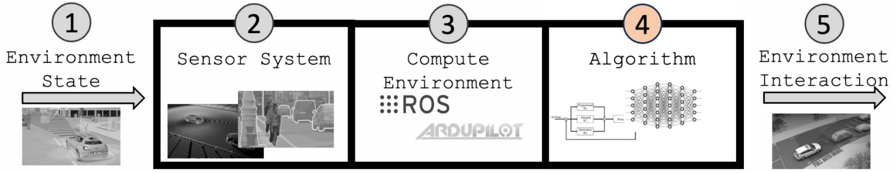
❸ Limitations & Takeaways



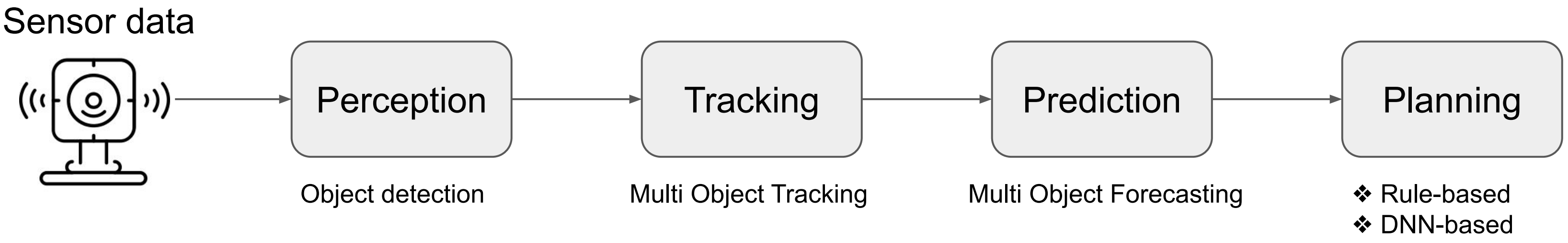
- ❖ AV software is complex and prone to same vulnerabilities as conventional systems
  - ↳ Safety-critical application domain of AV operation raises the stakes

- ❖ AVs are a real-time cyber-physical system → Exploit opportunities galore
  - ↳ Timing-correctness and real-time schedulability of hard-deadline tasks
  - ↳ Power is a scarce resource and overuse may hinder long-term operation
  - ↳ System predictability and downstream impact on algorithm deadlines

❹ AV Tasks



- ❖ With advancements in deep learning, many recent AV algorithms are based on Deep Neural Networks (DNNs).
- ❖ **Goal**: Optimal navigation decision-making based on accurate and robust perception & tracking.


Sensor data

Perception

Object detection

Tracking

Multi Object Tracking

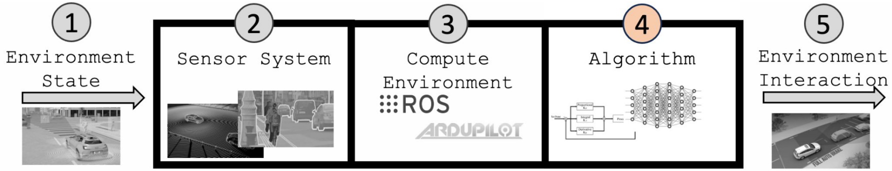
Prediction

Multi Object Forecasting

Planning

- ❖ Rule-based
- ❖ DNN-based

❹ Vulnerabilities in Tracking


1 Environment State

2 Sensor System

3 Compute Environment

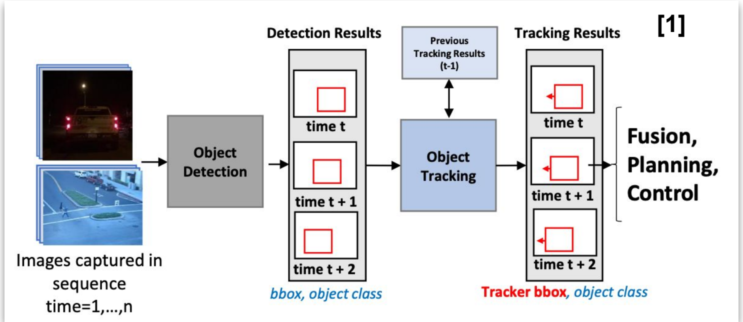
ROS

ARDUPILOT

4 Algorithm

5 Environment Interaction

- ❖ Basic assumption: Objects are consistently detected over consecutive frames.
- ❖ Highly sensitive to both natural and artificial corruptions
  - ↳ E.g., occlusion, weather particles, spoofed objects


[1]

Images captured in sequence time=1,...,n

Object Detection

Detection Results

time t

time t + 1

time t + 2

bbox, object class

Previous Tracking Results (t-1)

Object Tracking

Tracking Results

time t

time t + 1

time t + 2

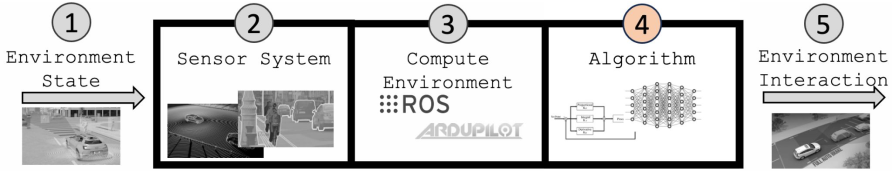
Tracker bbox, *object class*

Fusion, Planning, Control

trajectory

[1] Physical Hijacking Attacks against Object Trackers, Muller *et al.*, CCS 2022

❹ Attack on Tracking



1 Environment State

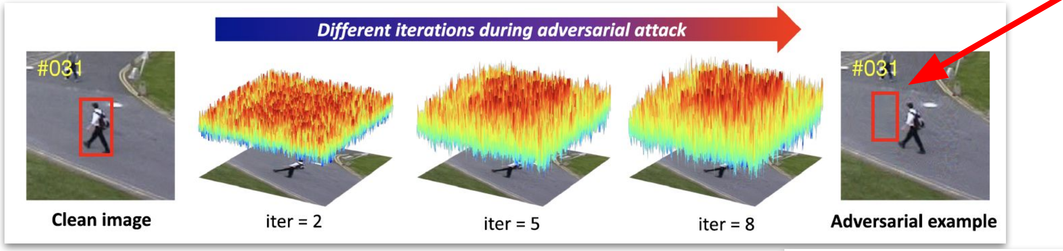
2 Sensor System

3 Compute Environment
ROS
ARDUPILOT

4 Algorithm

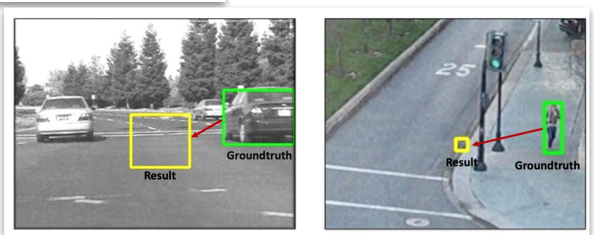
5 Environment Interaction

- To violate tracking algo's assumption, attackers manipulate bounding boxes via adversarial attack [1]



Location or size changed

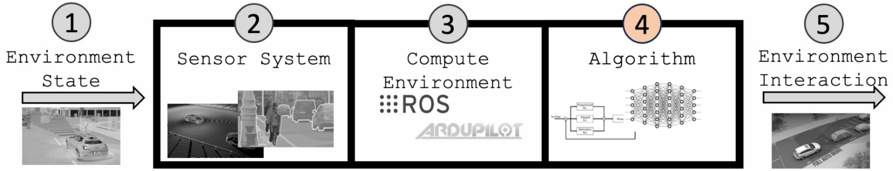
- Identify the attack zone that is physically plausible [2]
  - ↳ Avoid unrealistic manipulations (e.g., put a car into sky)



[1] Robust Tracking against Adversarial Attack, Jia *et al*, ECCV 2020

[2] Physical Hijacking Attacks against Object Trackers, Muller *et al.*, CCS 2022

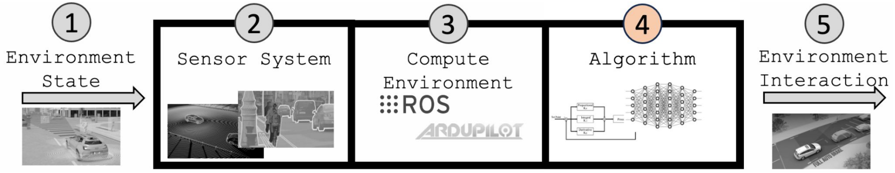
❹ Vulnerabilities in Navigation



- ❖ In navigation, AVs find the optimal routes to their destination

- ❖ Difficulty in distinguishing between adversarial and genuine situations
  - ↳ Highly stochastic and dynamic driving situation
  - ↳ Various and complicated driving intentions of surrounding vehicles (e.g., cut-ins, overtakings)

❹ Attack on Navigation



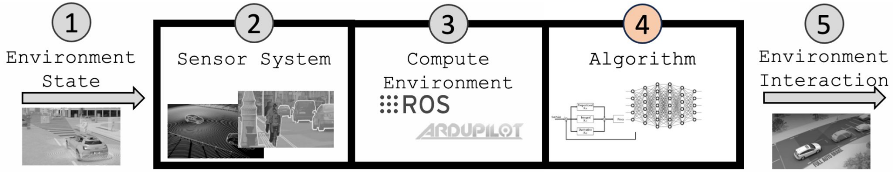
- ❖ Manipulates their vehicle movements to endanger a targeted vehicle [1]
  - ↳ It does not look intentionally malicious
  - ↳ Cause the victim to violate its safety standards

- ❖ Manipulate observed states and environmental dynamics to mislead RL-based decision-making systems [2]

[1] Discovering Adversarial Driving Maneuvers against Autonomous Vehicles, Song et al., USENIX Security 2023

[2] Toward Trustworthy Decision-Making for Autonomous Vehicles: A Robust Reinforcement Learning Approach with Safety Guarantees, He et al., Engineering 2024.

❹ Vulnerabilities in DNN-Based Algorithms



1 Environment State

2 Sensor System

3 Compute Environment

ROS

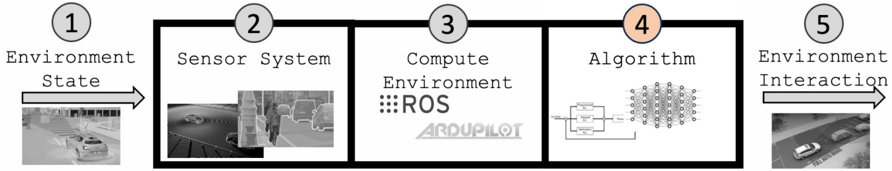
ARDUPILOT

4 Algorithm

5 Environment Interaction

- ❖ Data-driven characteristics and model's inherently limited learning capacity hinder capturing the full complexity of real-world driving scenarios.
- ❖ Beyond artificial corruptions by adversaries, natural corruptions (e.g., adverse weather and aging sensors) can impact performance.
  - ↳ Accuracy significantly drops on OOD inputs

❹ Limitations & Takeaways



1 Environment State

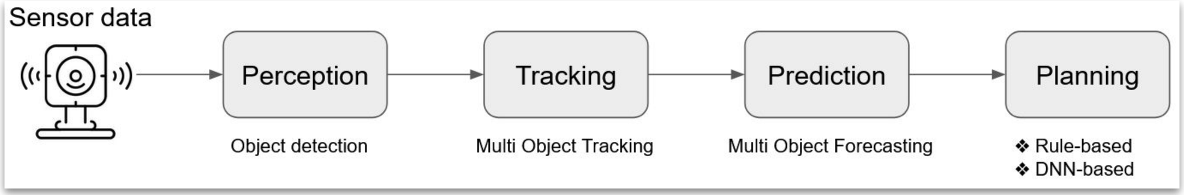
2 Sensor System

3 Compute Environment

4 Algorithm

5 Environment Interaction

- ❖ Error Propagation in Sequential Execution of Multi-DNN.



Sensor data

Perception

Object detection

Tracking

Multi Object Tracking

Prediction

Multi Object Forecasting

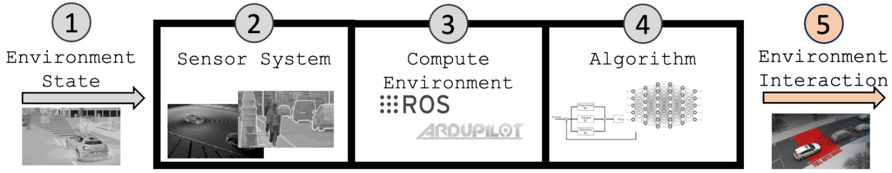
Planning

❖ Rule-based

❖ DNN-based

- ❖ Limited Coverage on Complex and Diverse Real-World Driving Environment.
- ❖ Lack of Evaluation Metrics for practical and safe AV Algorithms
  - ↳ Common metrics (e.g., mAP and mIoU) do not guarantee the model's feasibility

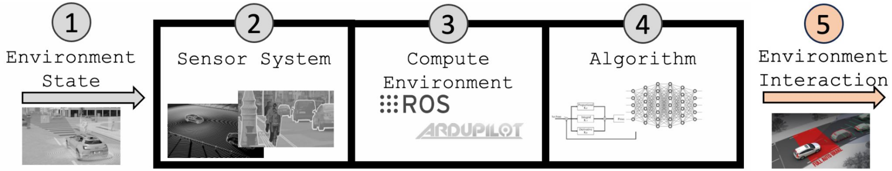
❺ Environment Interaction



- ❖ How the AV is controlled has direct consequences on the world around it

- ❖ Is it considering:
  - ↳ Rules of the road?
  - ↳ Social norms?
  - ↳ Standard negotiation practices?

❺ Legal Considerations



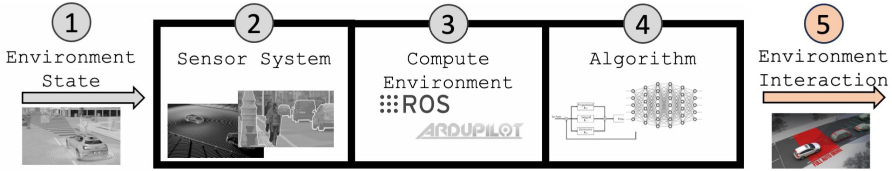
- ❖ Who has liability when security issues cause harm to others?
  - ↳ Driver of AV? Manufacturer of AV? OEM of AV algorithm?

- ❖ Would be more clear if malicious intent can be traced to an adversary [1]

- ❖ AV industry cooperation with regulators may improve public comfort [2]

[1] Norms of Computer Trespass, Orin S. Kerr. *Columbia Law Review* 1143, Vol. 116
[2] Autonomous Vehicle Regulation & Trust: The Impact of Failures to Comply with Standards, Widen and Koopman. *UCLA JL & Tech.*

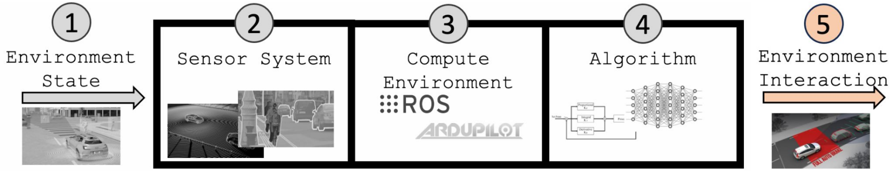
❺ Regulatory Standards



- ANSI/UL 4600 – Safety for Autonomous Products
- ISO 26262 – Functional Safety
- ISO 21448 – Safety of Intended Functionality
- ISO/SAE 21434 – Road Vehicle Cybersecurity

[1] Norms of Computer Trespass, Orin S. Kerr. *Columbia Law Review* 1143, Vol. 116
[2] Autonomous Vehicle Regulation & Trust: The Impact of Failures to Comply with Standards, Widen and Koopman. *UCLA JL & Tech.*

❺ Limitations & Takeaways



- Several high-profile accidents in recent years have lead to increased distrust

- If stricter regulation follows it should:
  - ↳ Carefully consider how it may both positively and negatively impact the AV industry
  - ↳ Provide support to AV makers for to ensure an easy transition for continued development

- AV "driver licenses" are supplied in an *ad hoc* manner
  - ↳ Requires a more active (rather than passive) approach with a stronger safety culture

Concluding Research Recommendations

1. Comprehensive end-to-end testing of AV safety and security

2. Effective utilization of collaborative perception to gain comprehensive understanding of environment (and can be used for security and safety validation)
   - ↳ Communication of shared data introduces additional security concerns not discussed here

3. AV licenses should have stricter requirements that penalize OEMs who do not follow the best practices for ensuring their algorithmic safety and security

Questions?

Takeaways:

- ❖ End-to-end AV security research is still an open challenge
- ❖ Focus of the research should pivot from ablation studies to deploying attacks in realistic AV scenarios

Contributors

Funding

USN
Office of Naval Research
ONR
Science & Technology

Paper URL