Tutorial: Crash, Fail-safe, or Recover: Securing Robotic Autonomous Vehicles

Pritam Dash
University of British Columbia
pdash@ece.ubc.ca

Karthik Pattabiraman
University of British Columbia
karthikp@ece.ubc.ca

Abstract

Robotic Autonomous Vehicles (RAVs) rely on onboard sensors for navigation, control, and autonomous functions. Physical sensor attacks pose significant threats to RAV safety. In this tutorial, we will explore how physical attacks compromise RAV state estimation and control, and lead to failures.

First, we will review existing approaches for securing RAVs, and highlight their limitations in adversarial settings.

Second, the tutorial will focus on two types of attack recovery techniques, each tailored to a different type of RAV control architecture. For traditional PID controller based RAVs, we will present how software sensors and feed-forward control design eliminate corrupted feedback and maintain stability under attacks. For learning-based RAVs that use Deep-RL policy, we will present how multi-objective policy learning combined with adversarial training ensures attack resilience.

Finally, we will discuss a diagnosis-guided recovery framework designed to secure RAVs against multi-sensor attacks.

The tutorial will be a hybrid of lectures and hands-on activities using virtual RAV platforms. Participants will gain practical experience in injecting attacks, analyzing how attacks disrupt RAVs, and applying attack-resilient techniques for both traditional control and AI-driven RAV architectures.

1 Presenters Bio

Below is the biography of the presenters.

Pritam Dash is a Ph.D. candidate at the University of British Columbia (UBC), Canada. He received his master's degree also from UBC in 2020. Pritam's research lies at the intersection of AI, control theory, and security, with a focus on enabling safe and resilient autonomy. He has developed comprehensive methods for securing RAVs, including techniques for attack detection, diagnosis, and recovery, to ensure safe and reliable operation even under adversarial conditions. For a detailed bio see https://dashpritam.github.io/.

Karthik Pattabiraman is a Professor of Electrical and Computer Engineering (ECE) at the University of British Columbia (UBC). He received his PhD in 2009 and MS in 2004 from the University of Illinois at Urbana-Champaign (UIUC). Before joining UBC in 2010, he was a postdoctoral researcher at Microsoft Research (MSR), Redmond. Karthik's research interests are in dependable systems, cyber-physical systems, and software security. For a detailed bio see https://blogs.ubc.ca/karthik/about/full-bio/.

2 Target Audience

The tutorial targets two groups of audience. Some familiarity with basic control concepts (e.g., PID, EKF) is recommended, but no prior experience with adversarial testing is required.

(1) The first group is practitioners and engineers who design, test, or deploy RAV autopilot software. The tutorial will include demonstrations of attack injection, such as GPS spoofing, gyroscope, and accelerometer manipulation, on virtual RAV platforms (e.g., ArduPilot, PX4) to illustrate the impact of such attacks on RAV safety and performance, along with hands-on activities for evaluating the robustness of existing RAV autopilot software and integrating recovery techniques.

(2) The second group is researchers interested in advancing attack mitigation methods. The tutorial will highlight recent robust control and attack recovery techniques [1–3], discussing their underlying principles, trade-offs, and practical considerations. It will conclude with a discussion of open research challenges in developing attack-resilient RAVs.

3 Content Outline

The tutorial will focus on key topics, including the nature of physical sensor attacks and how they compromise RAV safety, and recent advances in robust control and attack recovery techniques to defend against such threats.

Physical Sensor Attacks and RAV Control. RAVs rely on their onboard sensors to estimate physical states, which are then used by specialized algorithms to compute appropriate actuator signals in a feedback control loop. *Physical attacks*

Table 1: Tutorial Outline involving both lectures and hands-on activities (Total duration 1 hour)

Topic	Duration
Importance of the area and the problem. Research gap in existing resilience and RAV security methods	10 minutes
Hands-On activity on attack injection, e.g., GPS spoofing, gyroscope tampering etc., on virtual RAVs.	5 minutes
Discussion of attack recovery methods and resilient control methods	15 minutes
Hands-On activity on how to recover from a physical sensor attack	5 minutes
Discussion of multi-sensor attacks, e.g., attacks targeting GPS and gyroscope simultaneously.	5 minutes
Hands-On activity on multi-sensor attack injection and attack recovery.	10 minutes
Discussions on open issues and future directions.	10 minutes

manipulate sensors by injecting noise or malicious signals through physical channels [5]. For example, gyroscope and accelerometer can be manipulated through acoustic noise, GPS can be manipulated by transmitting malicious GPS signals, and optical flow sensor can be spoofed by laser beams. In this tutorial, we will demonstrate how such attacks can disrupt RAVs' state estimation and control, leading to unsafe consequences such as deviation from the set path or crashes.

Methods to Secure RAVs. The tutorial will review existing techniques proposed to address physical sensor attacks in RAVs and highlight the key research gaps. Prior approaches generally fall into three categories: (i) sensor redundancy, (ii) fault-tolerant control, and (iii) fail-safe mechanisms. Unfortunately, all of these techniques have significant limitations: (1) Sensor redundancy is not enough as physical attacks can compromise redundant sensors [1]. (2) Fault-tolerant control techniques are only effective against sensor noise or sensor faults, and cannot handle attacks [3]. (3) Activating fail-safe (e.g., landing a drone) as a response to attacks is not always safe, as the RAV may land in adverse areas.

Securing RAVs: Mitigating Physical Sensor Attacks. We will discuss the following attack recovery methods.

First, for RAVs that use traditional PID-based feedback control, we will present how a hybrid feed-forward control (FFC) and feedback control design can improve resilience against attacks. Specifically, we will discuss a FFC-based recovery framework [3]. The FFC eliminates reliance on corrupted sensor feedback that would otherwise propagate errors into actuator commands. Second, for RAVs that rely on Deep Reinforcement Learning (Deep-RL) policies for control, we will discuss a specification-aware attack recovery framework [2]. Unlike traditional safe Deep-RL methods that prioritize quick but often unsafe responses, this approach ensures that recovery actions comply with the RAV's mission specifications, even under attack. Finally, we will discuss the risks posed by multi-sensor attacks and show how historical information can be leveraged to derive robust actuator commands [4], enabling safe operation even under multi-sensor attacks.

The key takeaways are the following:

1. Gain an understanding of various types of physical sensor attacks and their impact on the safety of RAVs.
2. Learn the limitations of current security techniques for RAVs and learn about state-of-the-art methods for attack detection, diagnosis, and recovery.
3. Acquire hands-on experience in launching attacks and running attack detection and recovery techniques.
4. Understanding open challenges in building robust and attack-resilient RAVs, both using traditional control-theoretic and AI-driven approaches.

Format and Duration. The tutorial will follow a hybrid format, combining lectures with hands-on activities. The lectures will focus on physical attacks, gaps in RAV security, and state-of-the-art resilient control and attack recovery methods. The hands-on activities will give participants practical experience in launching physical attacks on virtual RAV platforms and applying resilient control and recovery techniques to maintain safe operation. The total duration of the tutorial will be **60 minutes**. Table 1 summarizes the topics covered.

References

[1] Hongjun Choi, Sayali Kate, Yousra Aafer, Xiangyu Zhang, and Dongyan Xu. Software-based realtime recovery from sensor attacks on robotic vehicles. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, October 2020.

[2] Pritam Dash, Ethan Chan, and Karthik Pattabiraman. Specguard: Specification aware recovery for robotic autonomous vehicles from physical attacks. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, CCS '24, 2024.

[3] Pritam Dash, Guanpeng Li, Zitao Chen, Mehdi Karimibiuki, and Karthik Pattabiraman. Pid-piper: Recovering robotic vehicles from physical attacks. In *2021 51st AnnualIEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 26–38. IEEE, 2021.

[4] Pritam Dash, Guanpeng Li, Mehdi Karimibiuki, and Karthik Pattabiraman. Diagnosis-guided attack recovery for securing robotic vehicles from sensor deception attacks. ASIA CCS '24, New York, NY, USA, 2024.

[5] Hyungsub Kim, Rwitam Bandyopadhyay, Muslum Ozgur Ozmen, Z Berkay Celik, Antonio Bianchi, Yongdae Kim, and Dongyan Xu. A systematic study of physical sensor attack hardness. In *2024 IEEE Symposium on Security and Privacy (SP)*.