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

Pritam Dash, Karthik Pattabiraman

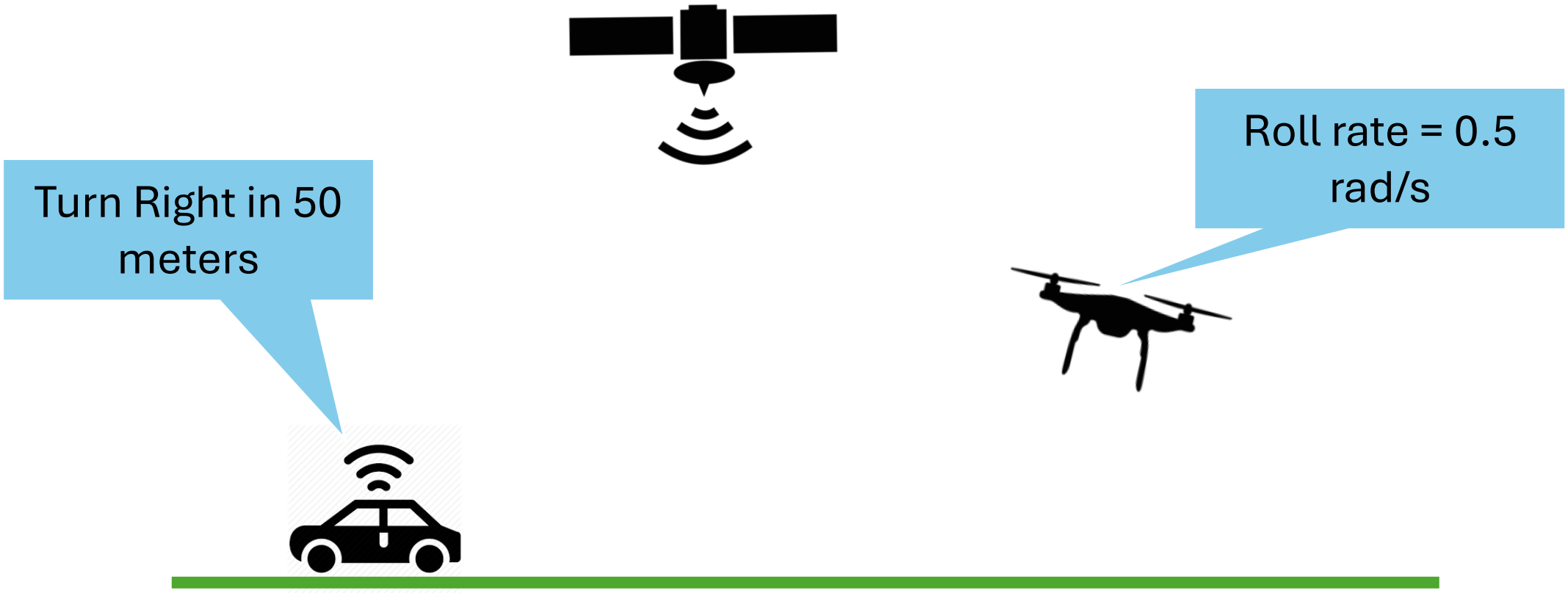
University of British Columbia


Robotic Autonomous Vehicles (RAV) are widely used in various Sectors

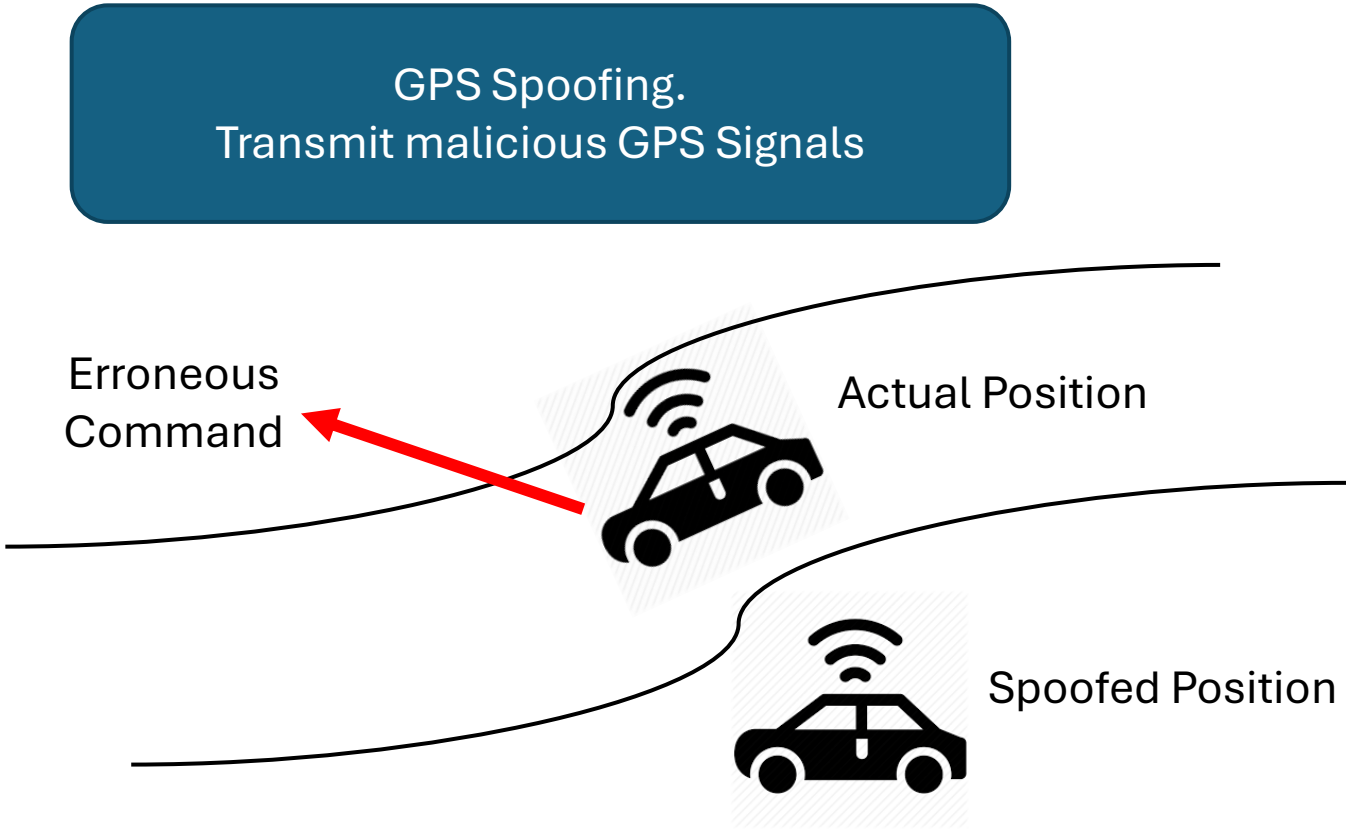
Drone market expected to grow by 50% between 2025 – 2030[1]

[1] *The Globe and Mail*

Perception in Robotic Autonomous Vehicles

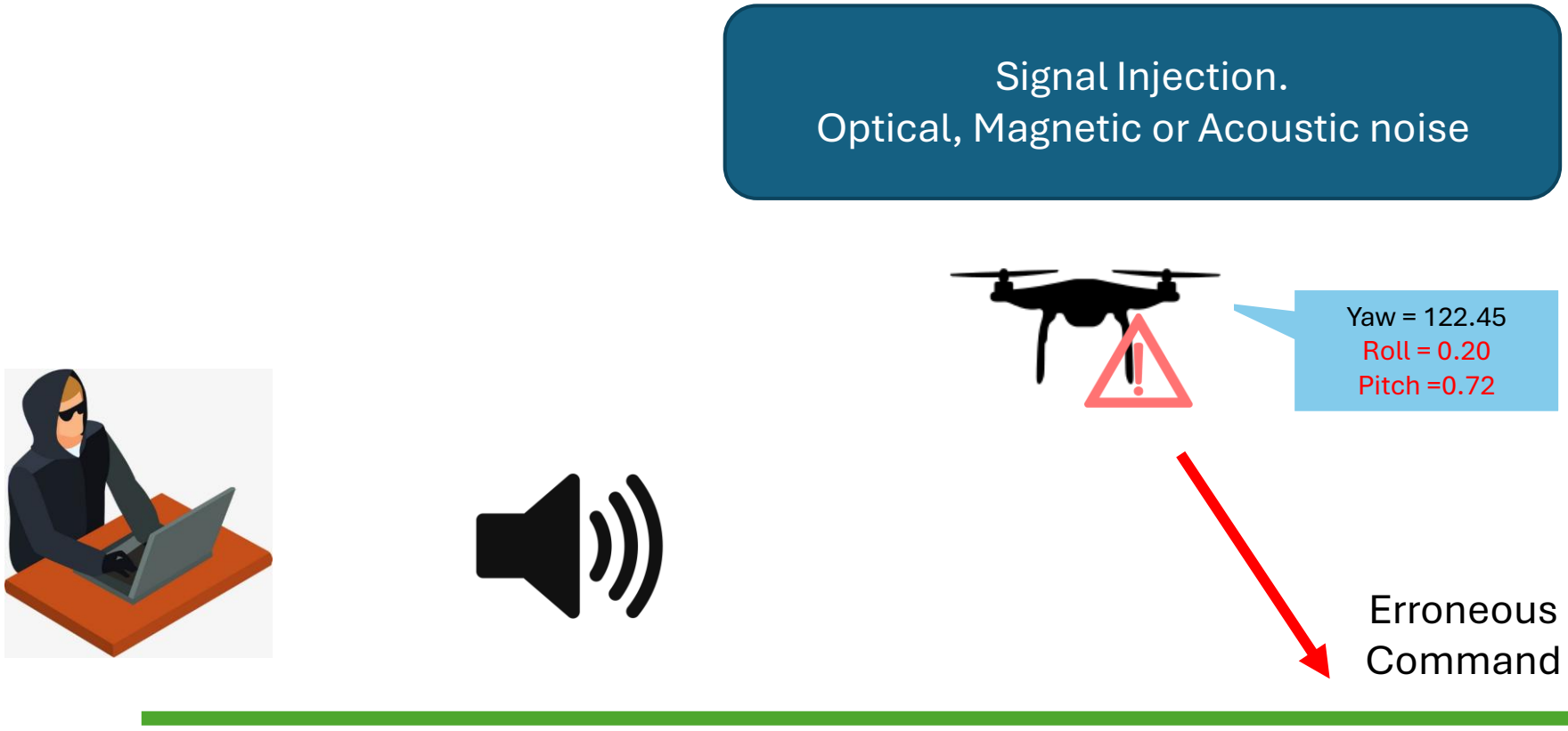


Physical Sensor Attacks: Sensor Spoofing



Tippenhauer et. al. "On the requirements for successful GPS spoofing attacks". ACM CCS'11
Nighswander, Tyler, et al. "GPS software attacks." ACM CCS'2012

Physical Sensor Attacks: Signal Injection



Son et. al. "Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors". Usenix Security'2015
Davidson, Drew, et al. "Controlling UAVs with sensor input spoofing attacks." WOOT 2016.
Kim, Hyungsub, et al. "A Systematic Study of Physical Sensor Attack Hardness." IEEE S&P 2024

Physical Attacks in Real World

Forbes

GPS Spoofing in the Middle East Is Now Capturing Avionics

POLITICO

GPS 'spoofing' thickens the fog of war

By MATT BERG | 10/24/2023 04:00 PM EDT

The Washington Post
Democracy Dies in Darkness

A downed drone highlights a vulnerable technology

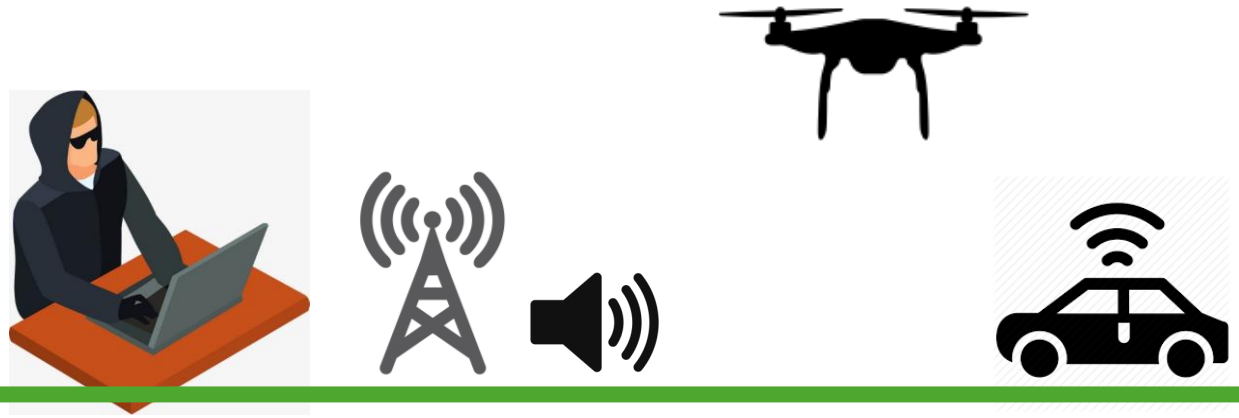
CBC Airlines grapple with spike in GPS interference.

Threat Model

Malicious Signals

- Fake signals [1] [3]
- Acoustic noise [2]
- EMI [4]

Traditional security techniques fall short.

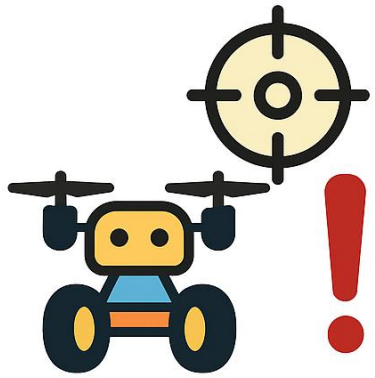


Out of scope

- Network Vulnerabilities
- Software Vulnerabilities

1. *Tippenhauer, Nils Ole, et al. "On the requirements for successful GPS spoofing attacks." ACM CCS 2011.*
2. *Son, Yunmok, et al. "Rocking drones with intentional sound noise on gyroscopic sensors." Usenix Security 2015.*
3. *Davidson, Drew, et al. "Controlling UAVs with sensor input spoofing attacks." WOOT 2016.*
4. *Kim, Hyungsub, et al. "A Systematic Study of Physical Sensor Attack Hardness." IEEE S&P 2024*

Related Work: Methods to Safeguard RAVs against Physical Attacks



Attack Detection

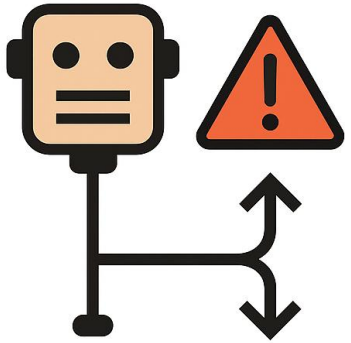
CCS'18
Usenix Security'20

Failsafe

RTSS'22

Redundant Sensors

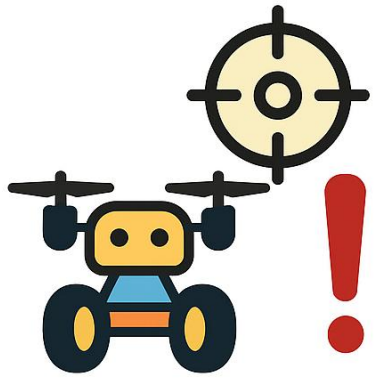
IROS'08
SENSORS'22



Fault Tolerant Control

ICRA'20
RA-L'22

Related Work: Methods to Safeguard RAVs against Physical Attacks



Attack Detection

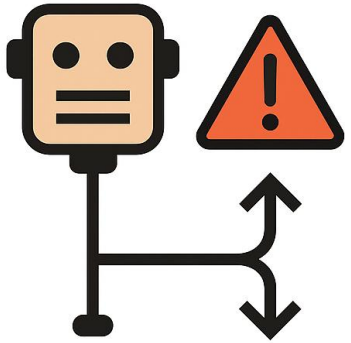
CCS'18
Usenix Security'20

Failsafe

RTSS'22

Redundant Sensors

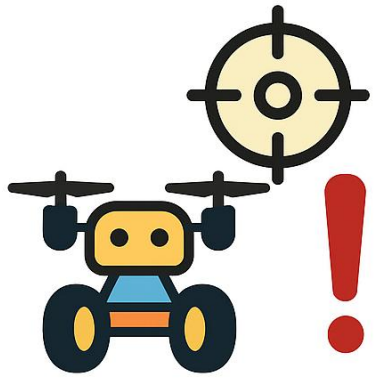
IROS'08
SENSORS'22



Fault Tolerant Control

ICRA'20
RA-L'22

Related Work: Methods to Safeguard RAVs against Physical Attacks



Attack Detection

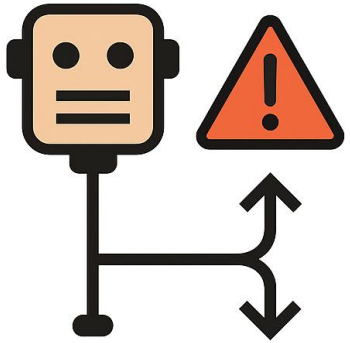
CCS'18
Usenix Security'20

Failsafe

RTSS'22

Redundant Sensors

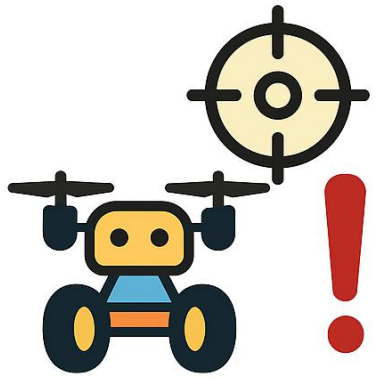
IROS'08
SENSORS'22



Fault Tolerant Control

ICRA'20
RA-L'22

Related Work: Methods to Safeguard RAVs against Physical Attacks



Attack Detection

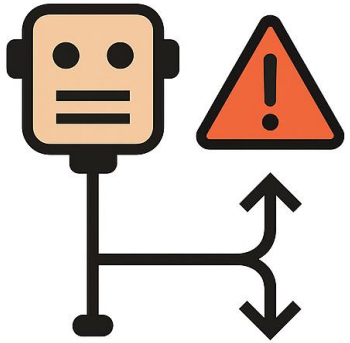
CCS'18
Usenix Security'20

Failsafe

RTSS'22

Redundant Sensors

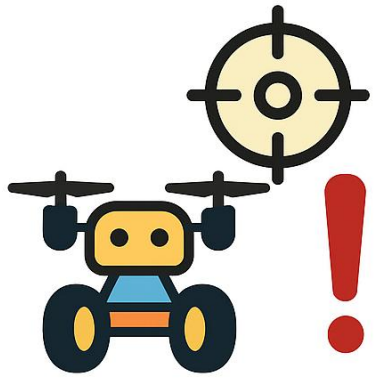
IROS'08
SENSORS'22



Fault Tolerant Control

ICRA'20
RA-L'22

Related Work: Methods to Safeguard RAVs against Physical Attacks



Attack Detection

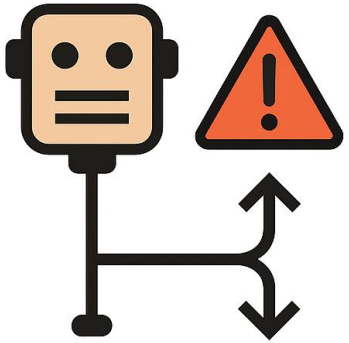
CCS'18
Usenix Security'20

Failsafe

RTSS'22

Redundant Sensors

IROS'08
SENSORS'22



Fault Tolerant Control

ICRA'20
RA-L'22

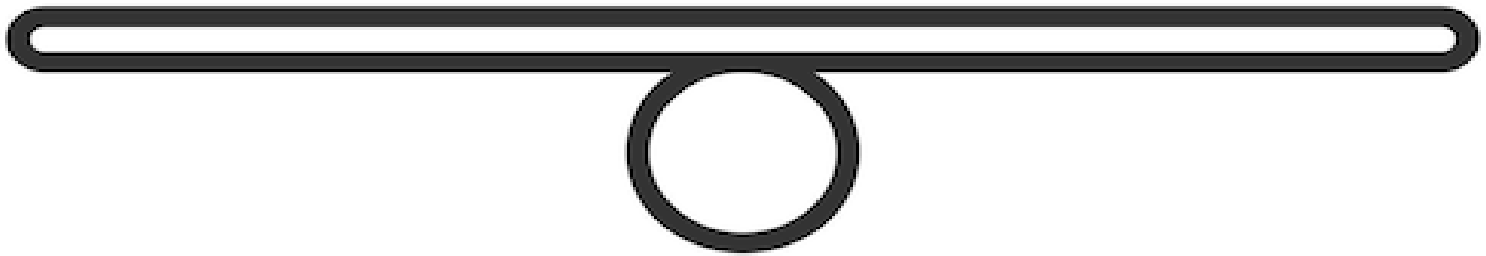
Goal: Robust operations despite attacks.

Challenges: RAVs have diverse requirements

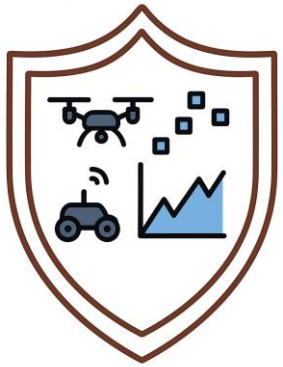
Utility
Efficiency
Autonomy

Attack Resilience

Safety
Security
Reliability

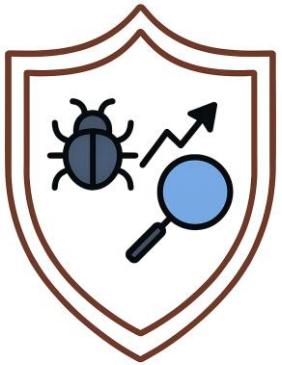


Mitigating Physical Attacks against RAVs

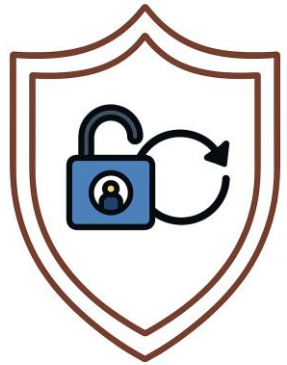


DSN’25

DSN’21



AsiaCCS’24

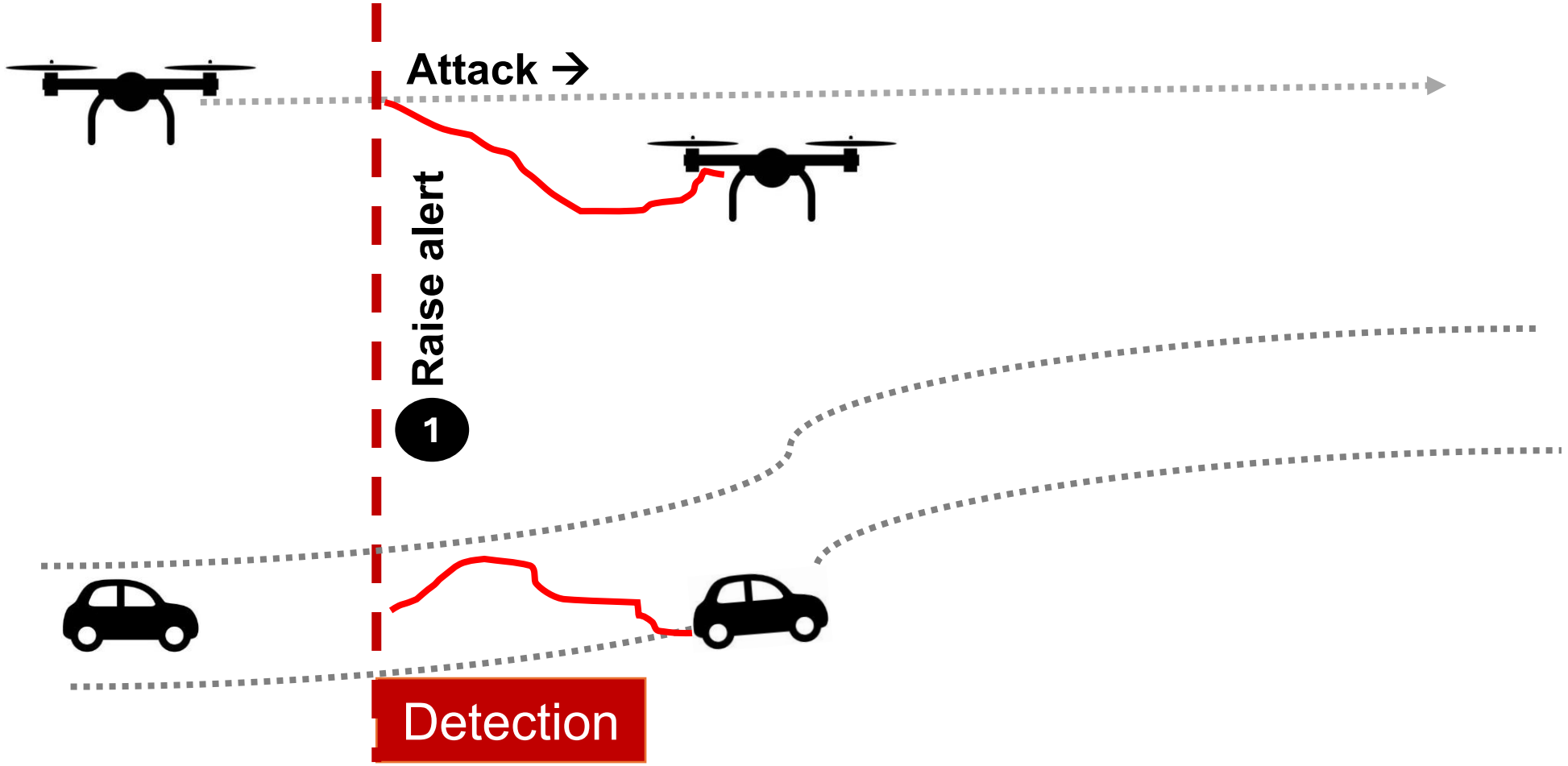


CCS’24

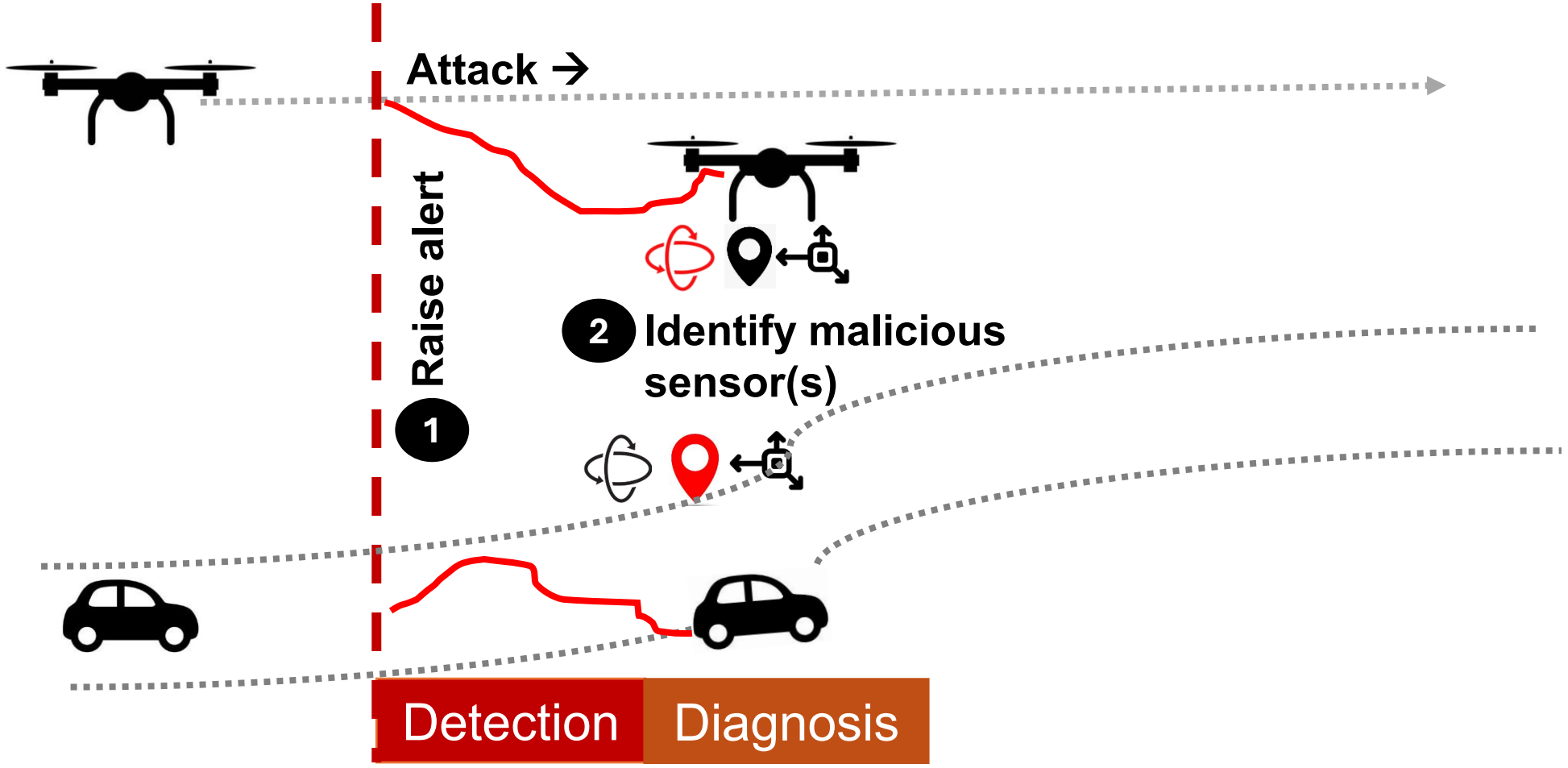
Tool Setup

Attack Detection, Diagnosis, and Recovery

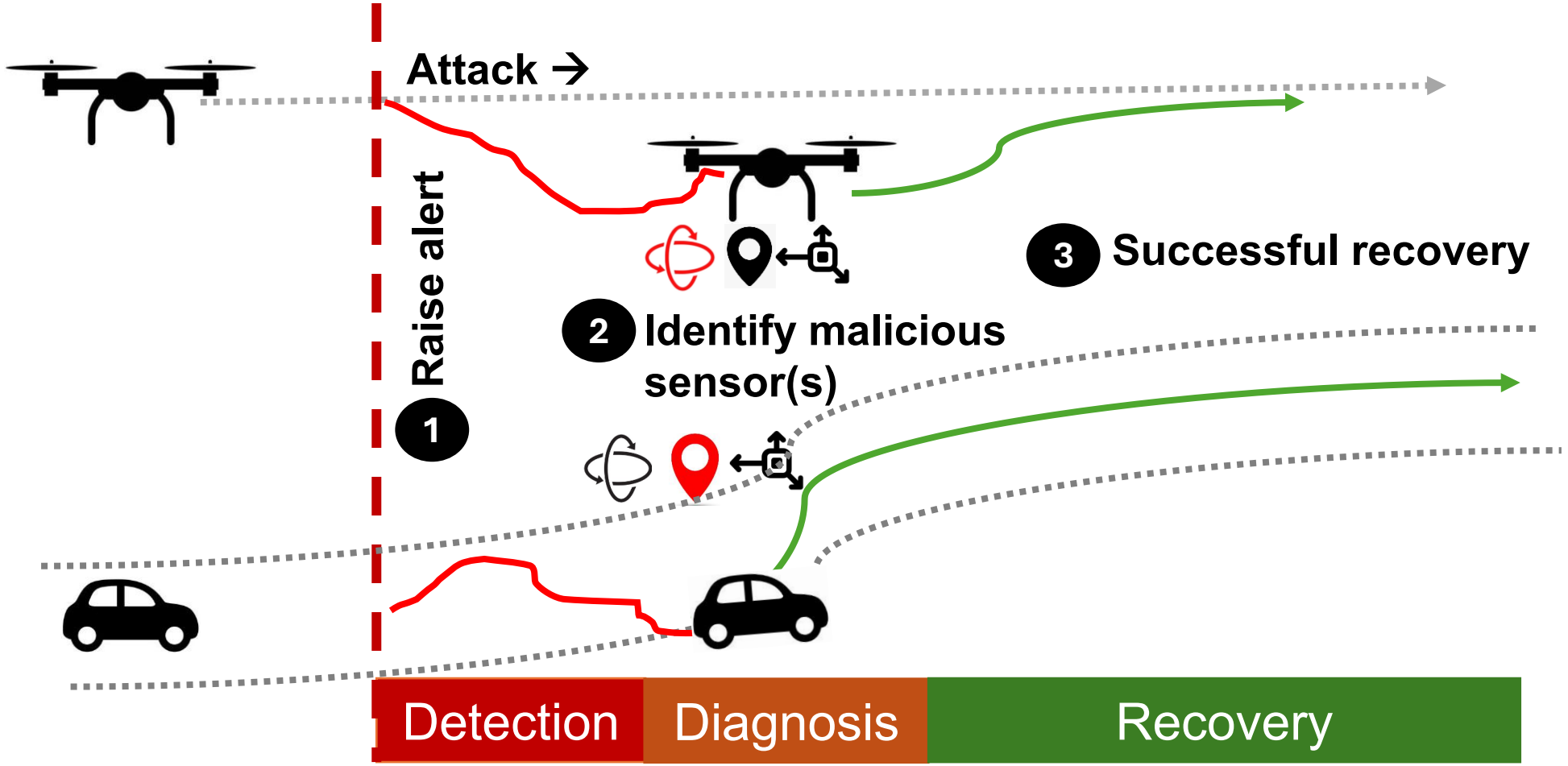
Mitigating Physical Attacks against RAVs



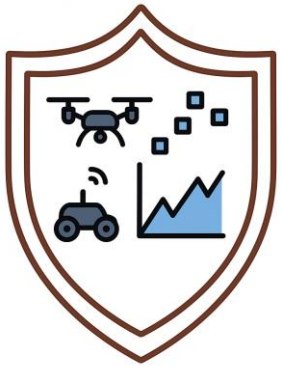
Mitigating Physical Attacks against RAVs



Mitigating Physical Attacks against RAVs

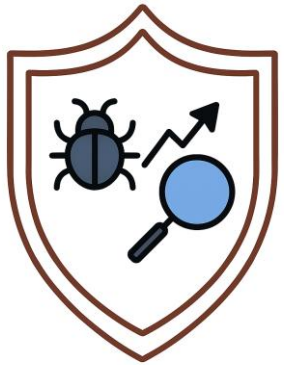


RAVAGE: RAV Attack Generation Engine

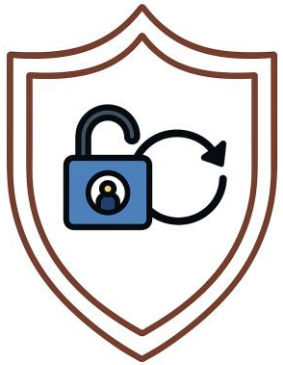


DSN’25

DSN’21



AsiaCCS’24



CCS’24

Activity 0 – Mission (No attack)

Instructions https://tinyurl.com/vsec2025 (Worksheet.pdf)

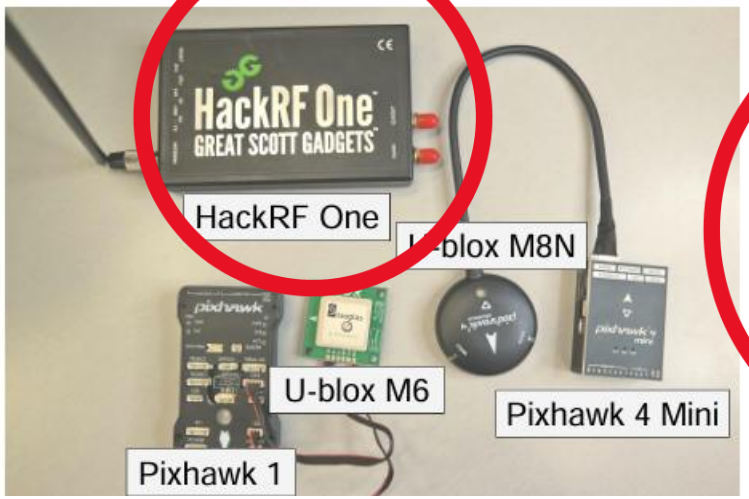
Launch a mission

Open terminal

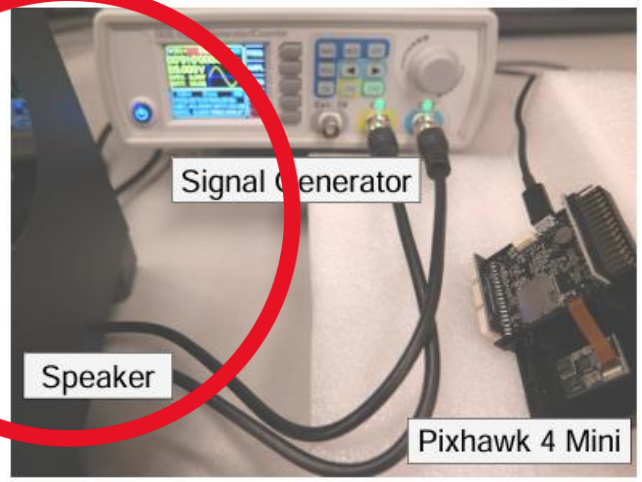
cd ravage/

```
python mission.py –s ArduPilot
```

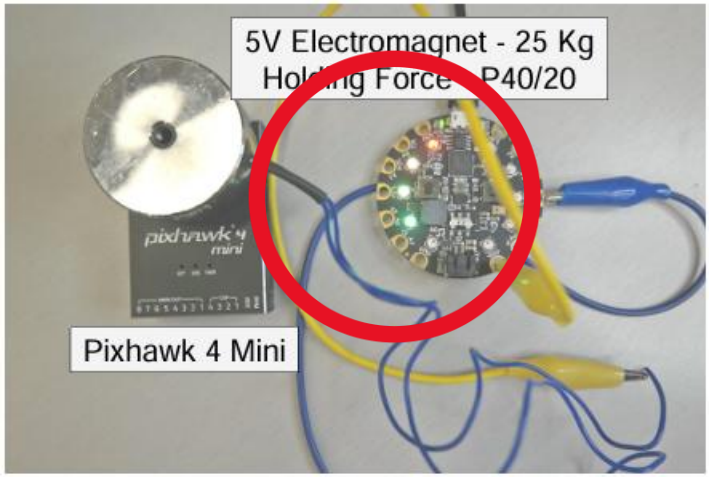
How are Physical Attacks launched?



GPS Spoofing



Acoustic Injection → Gyroscope Accelerometer



EMI Injection → Magnetometer

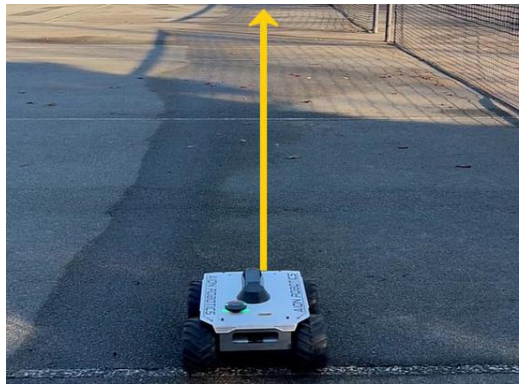
Photos from Kim et. al., "A Systematic Study of Physical Sensor Attack Hardness", IEEE Security and Privacy, 2024.

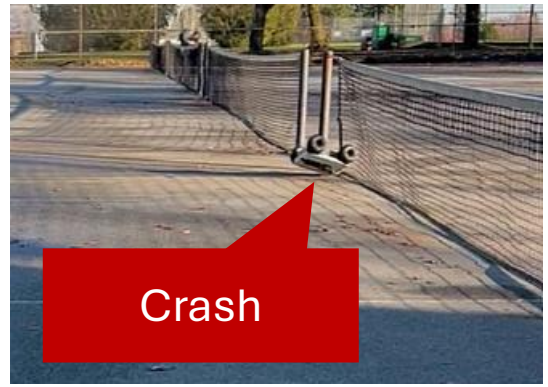
Challenges

Running large scale experiments → impractical

Expertise to operate the attack hardware

Risks of physical damages and injuries





Drone Photos from Choi et. al., CCS'2018

RAVAGE : Attack Generation Engine

Emulate realistic physical attacks through software

External commands and configurations to launch attacks

Integration with Autopilot Software

Key Results

Code Available Code Reviewed Code Reproducible

Attack Details

Attacks Targeted: GPS, Gyro, Accel, Mag, Optical Flow, Barometer. (**Total 6 sensors**)

Physical attacks via Software

- Physical properties
- Signal characteristics.

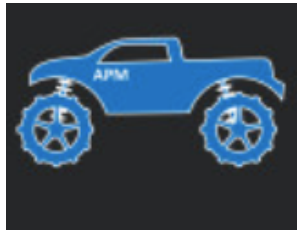
Real RAVs

Tarrot 650

Aion R1

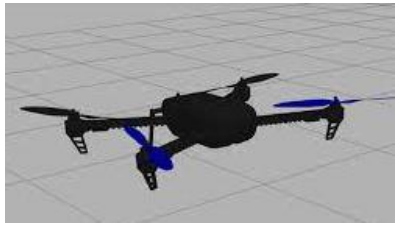
DIY Pixhawk

Virtual RAVs



ArduRover

ArduCopter



PXCopter

Activity 1 – Attack Injection

Instructions https://tinyurl.com/vsec2025 (Worksheet.pdf)

cd ravage/

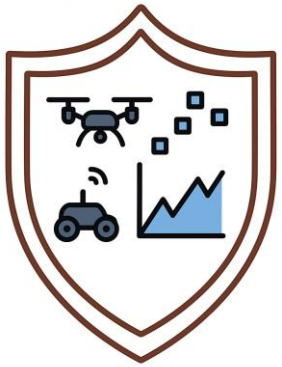
Launch a mission

```
python mission.py –s ArduPilot
```

Launch attacks

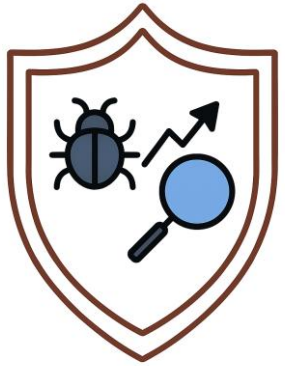
```
python ravage.py -s ArduPilot -a GPS -i 30 -t 10
```

PID-Piper: Feed-forward Control for Attack Detection and Attack Recovery

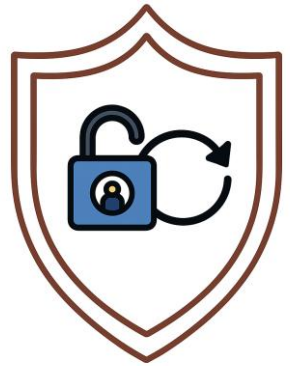


DSN'25

DSN'21

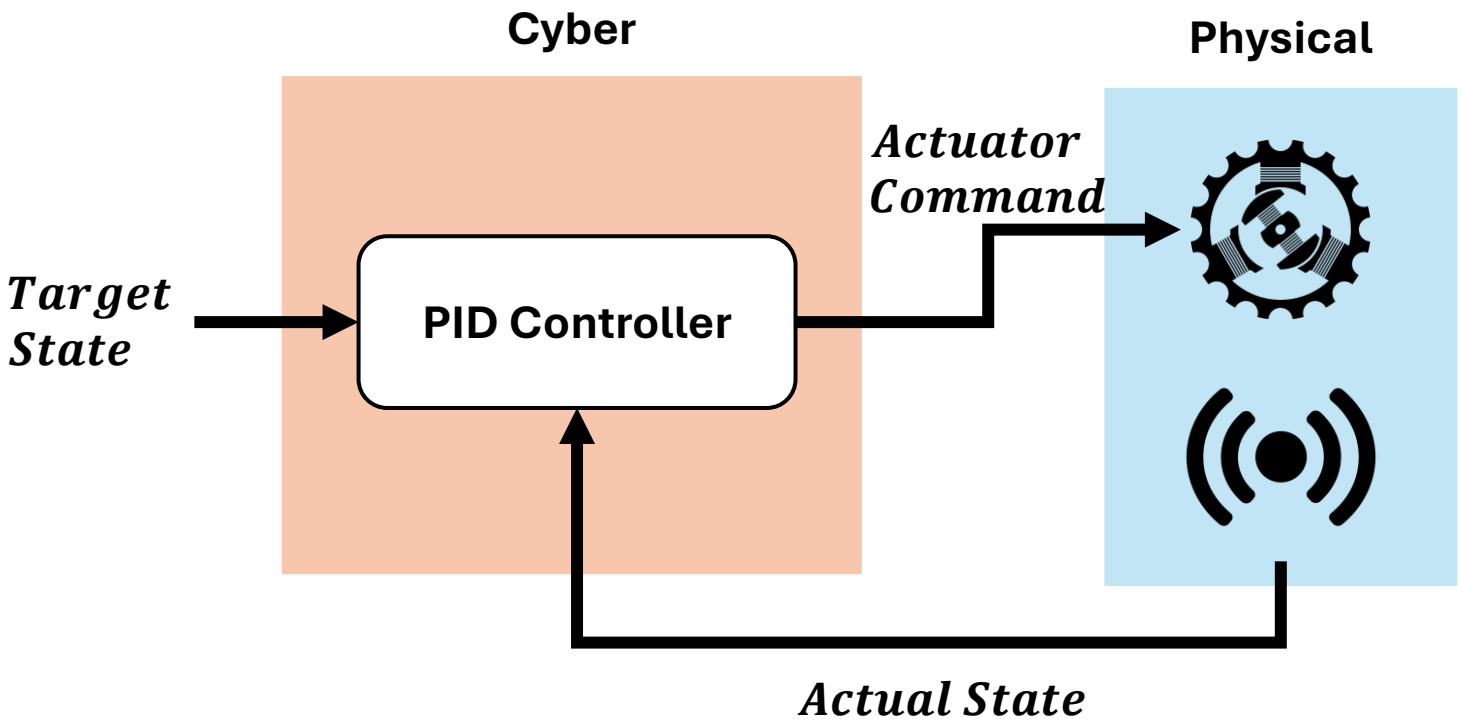


AsiaCCS'24



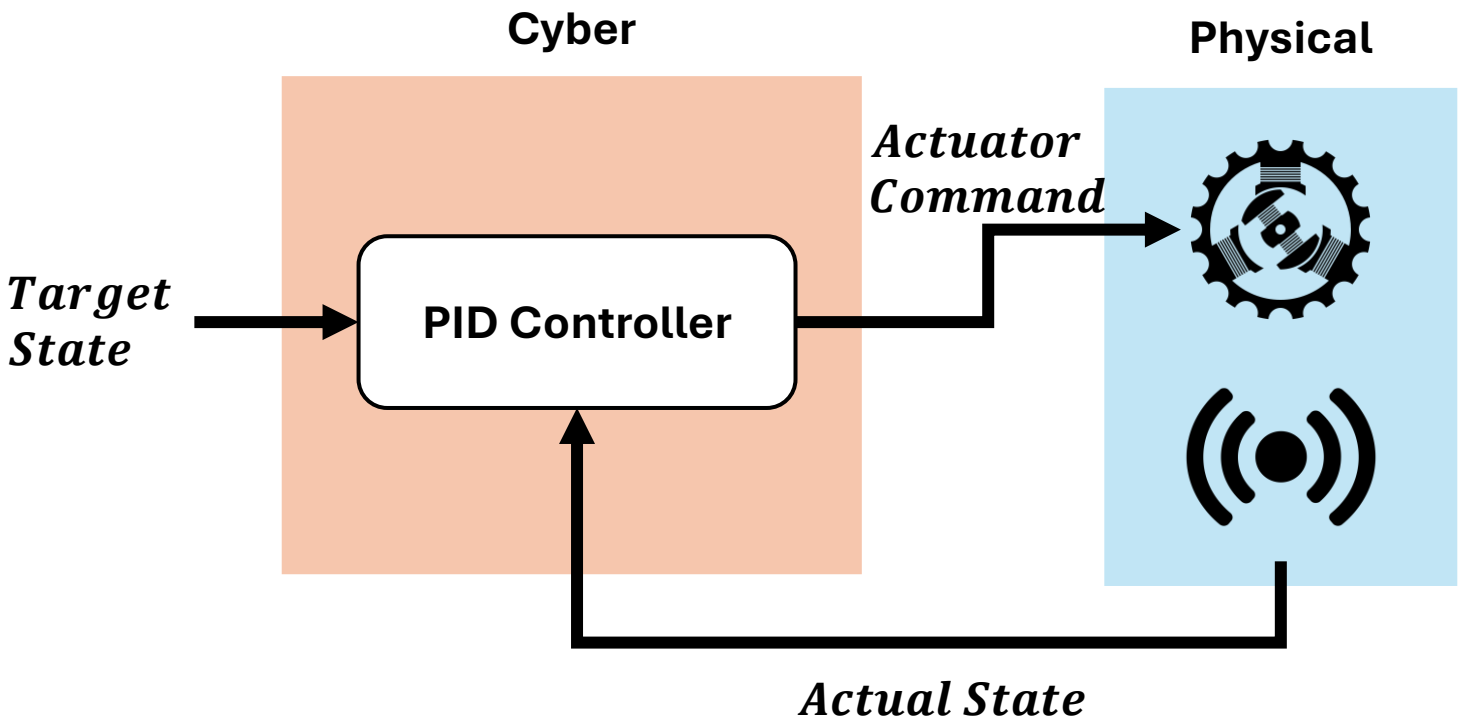
CCS'24

RAVs' Feedback Control Loop

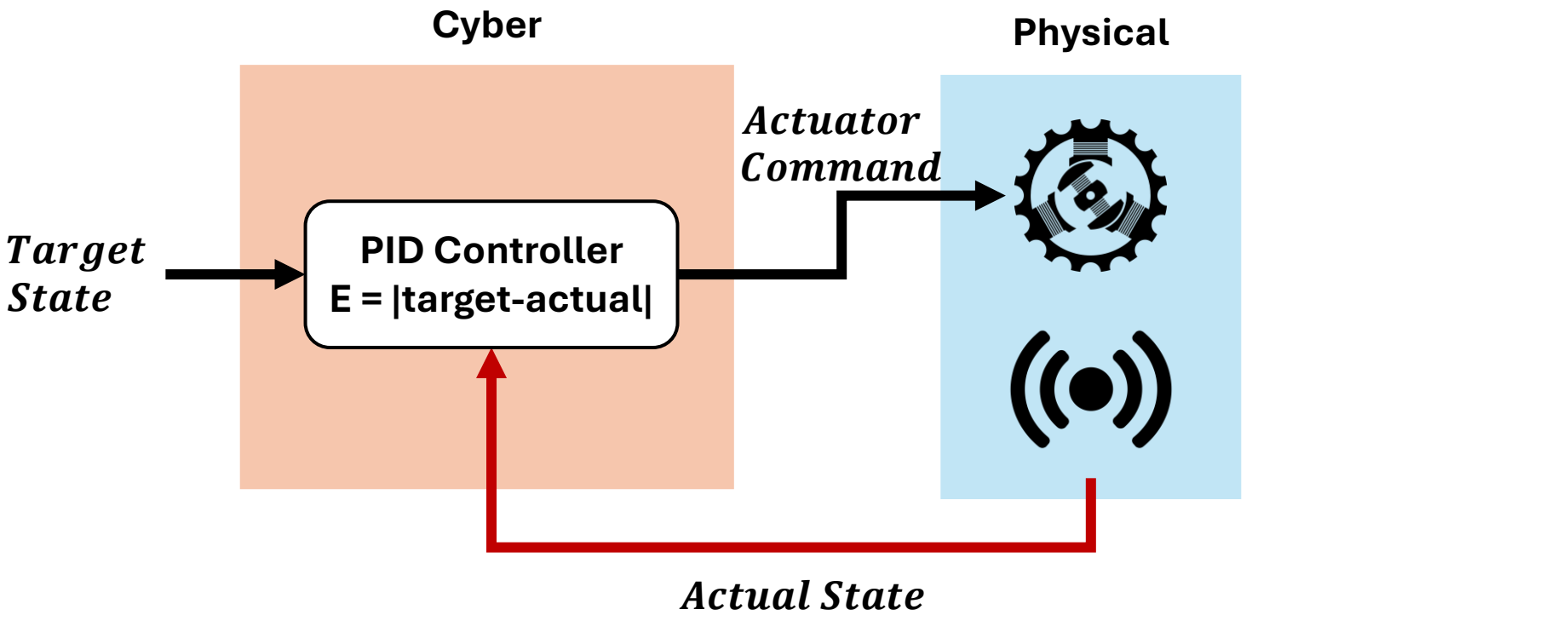


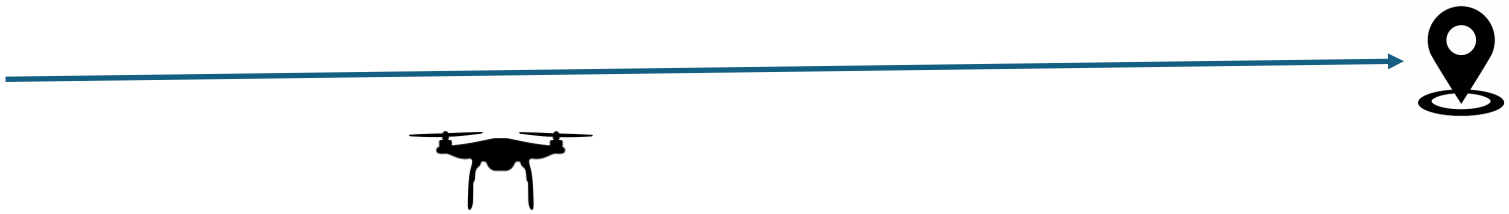
PID – Proportional Integral Derivative

Compensation-based Error Correction

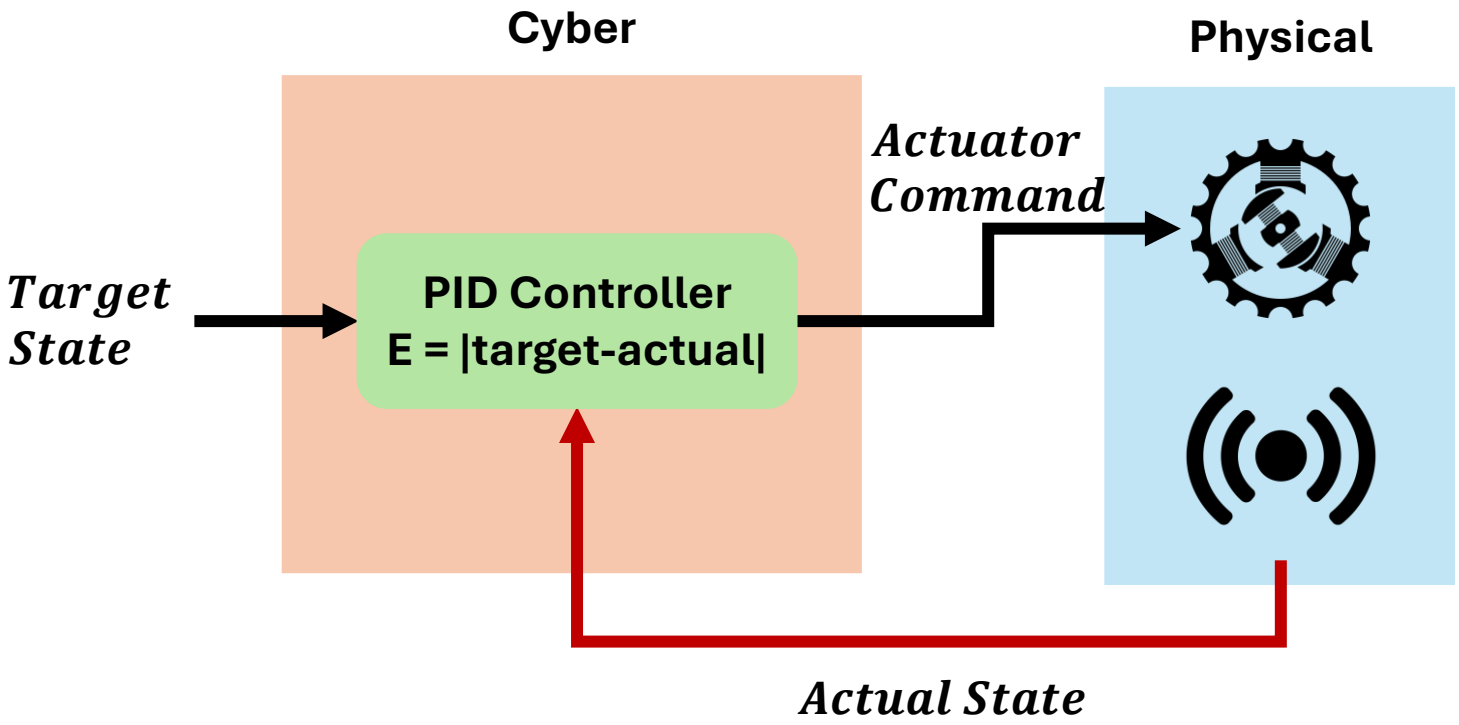


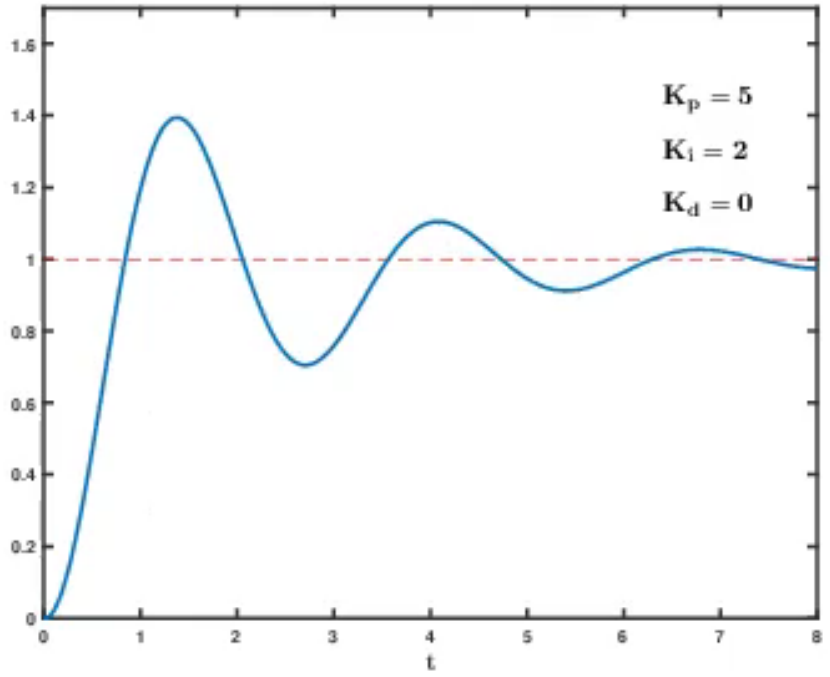
Compensation-based Error Correction

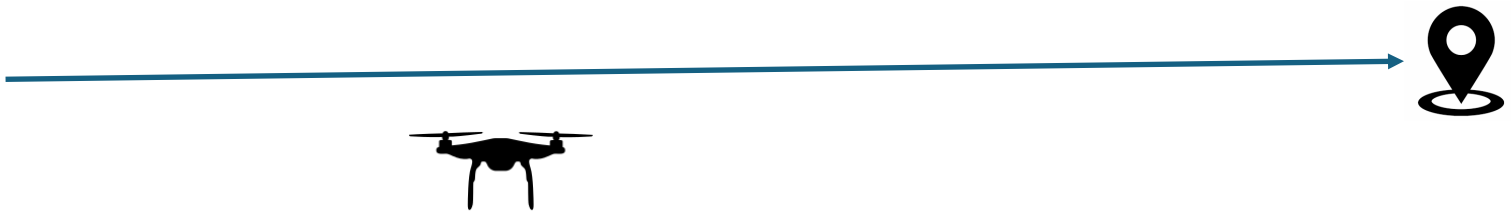




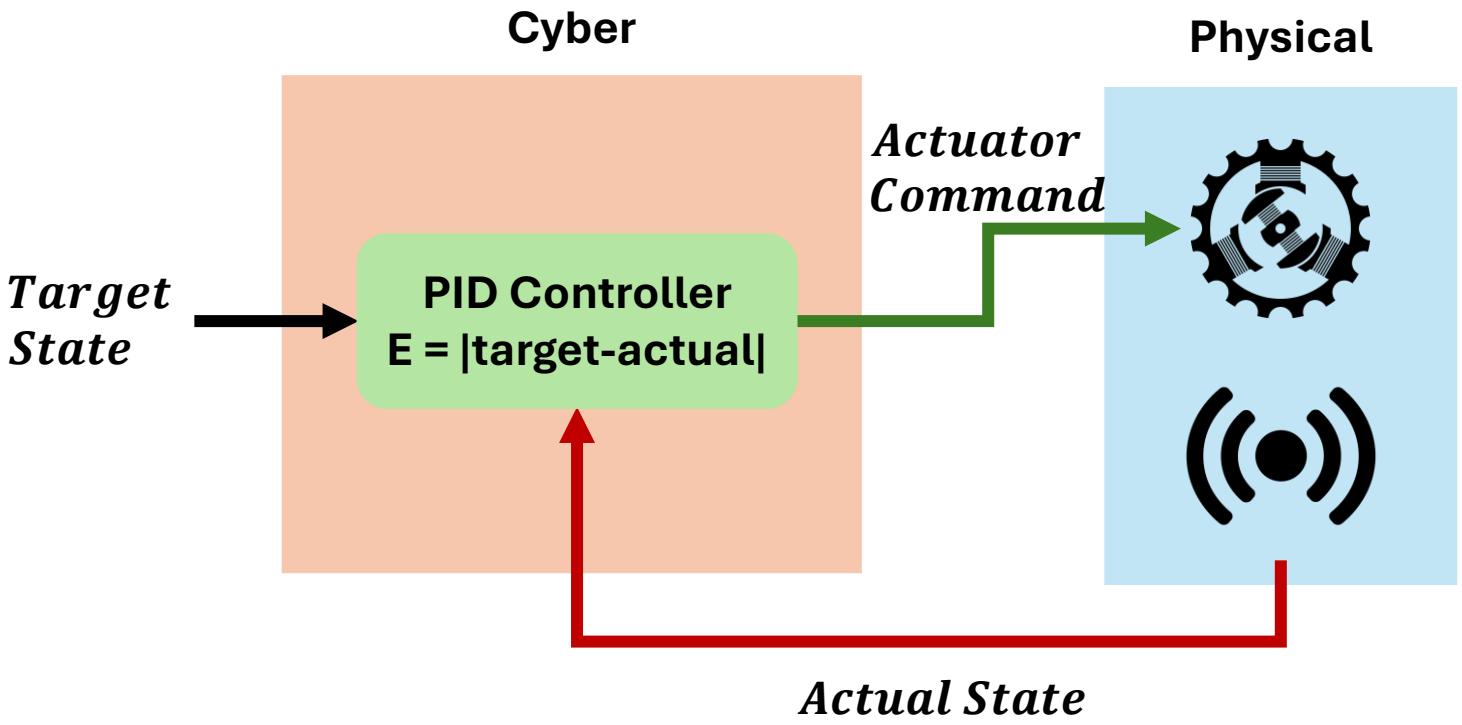
Compensation-based Error Correction

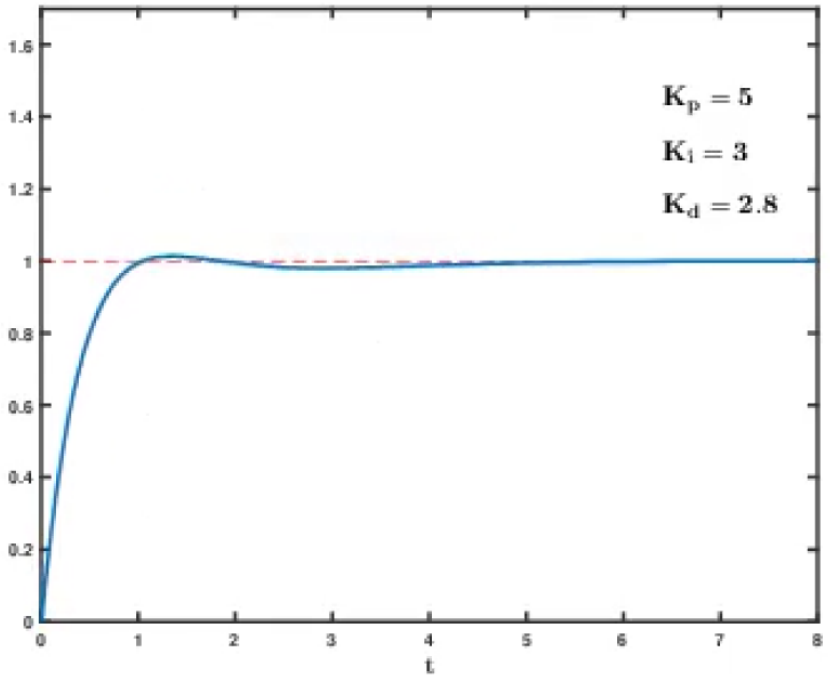




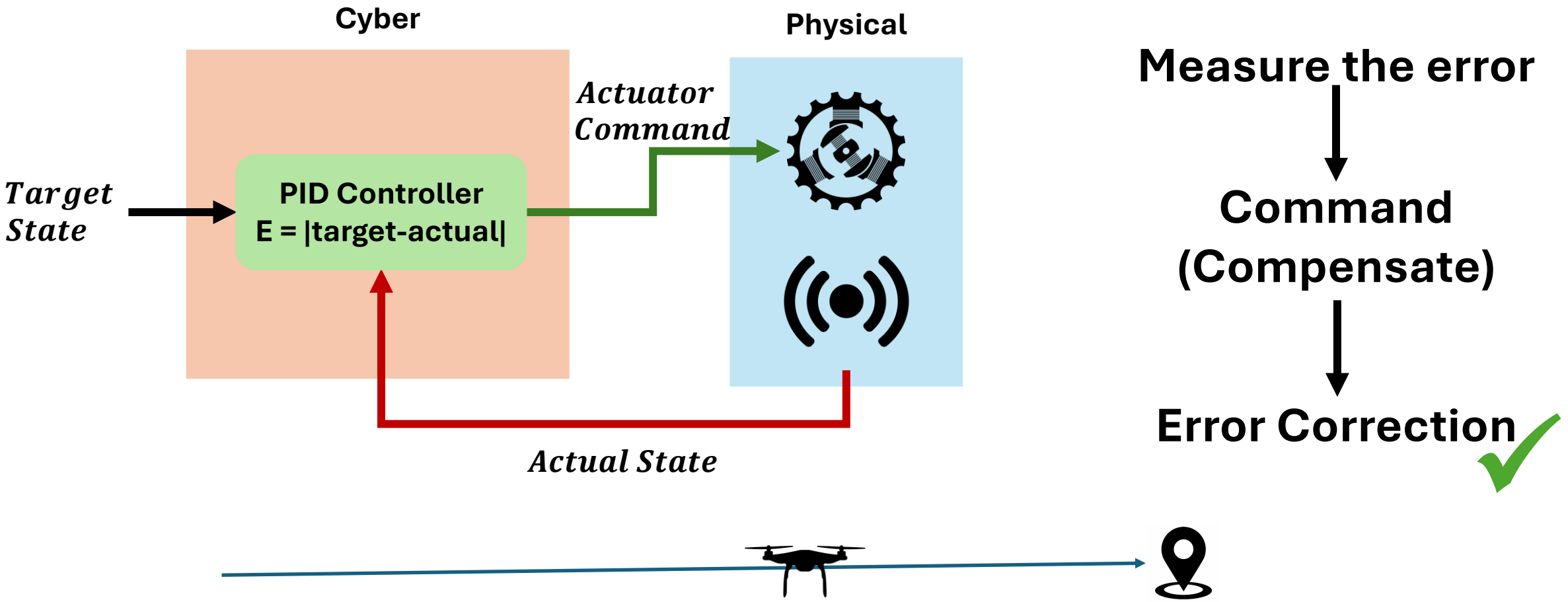


Compensation-based Error Correction

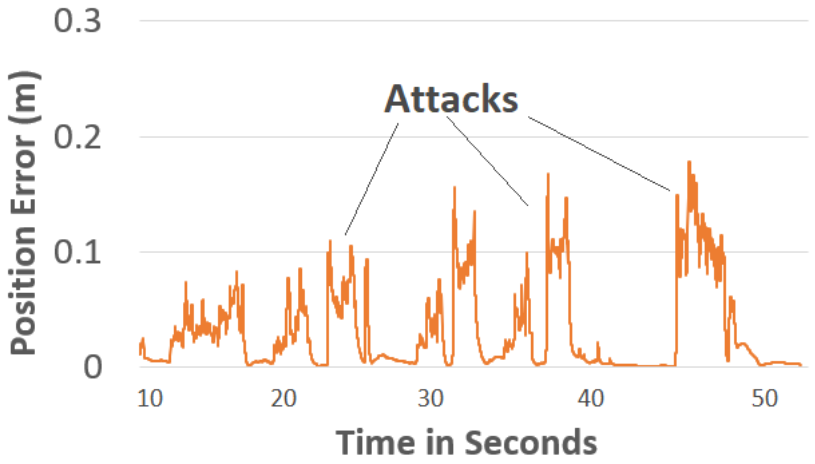


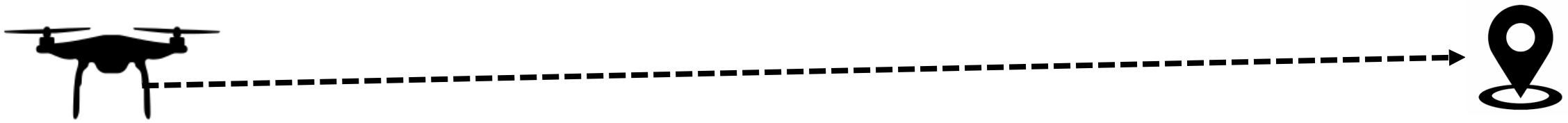


Compensation-based Error Correction

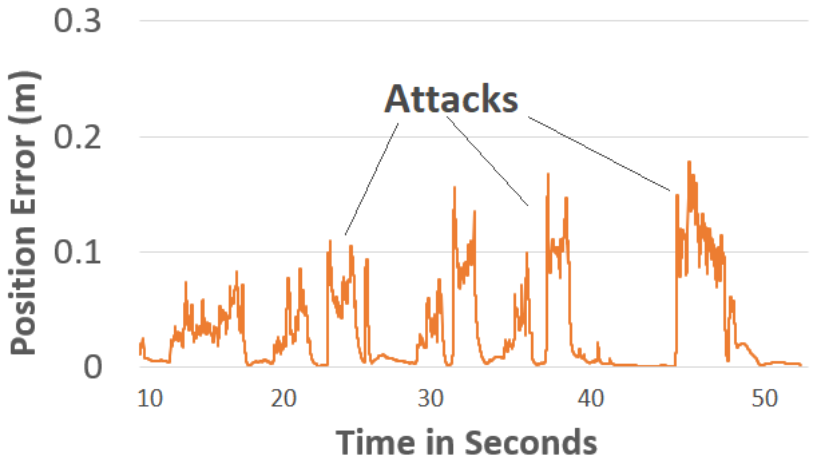


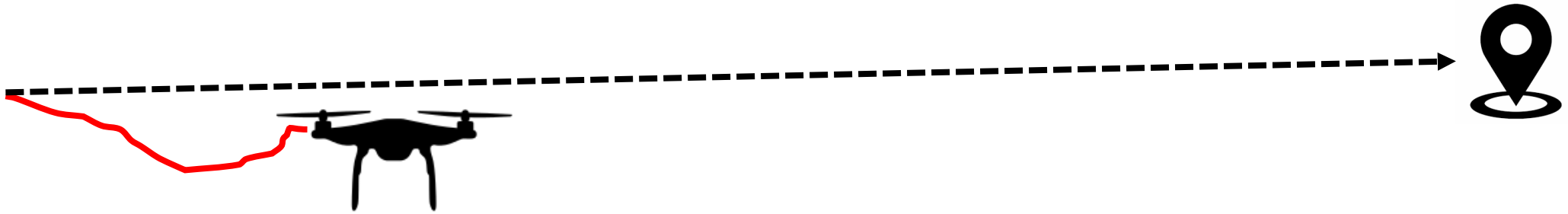
Compensation-based Error Correction under Physical Attacks



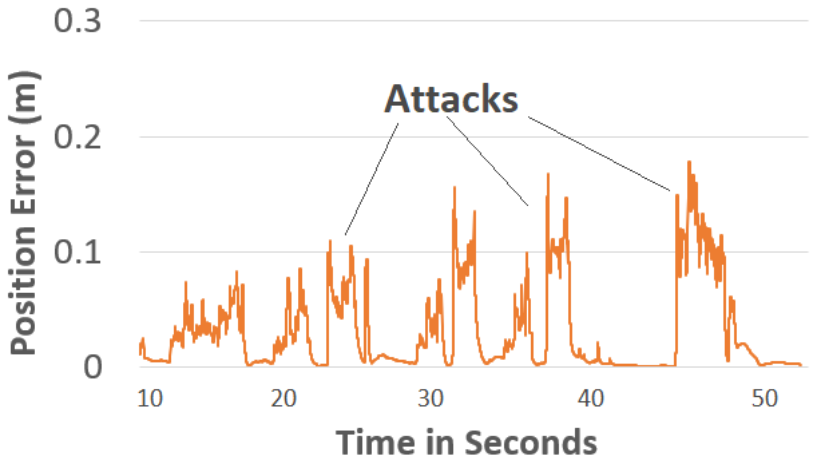


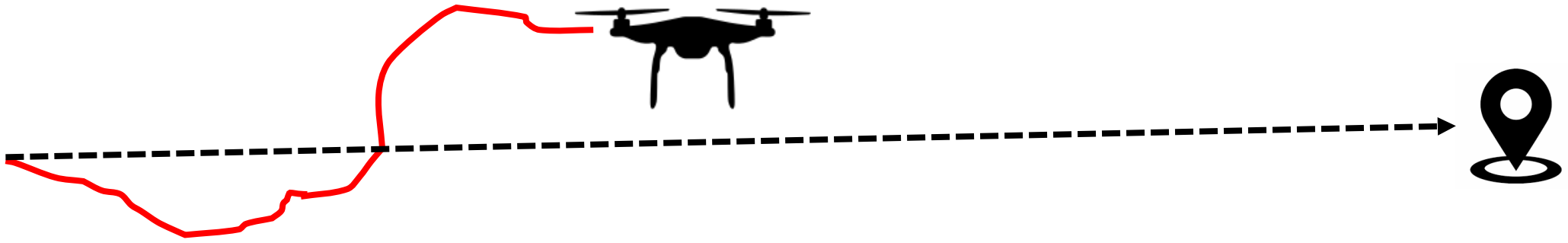
Over-Compensation under Physical Attacks



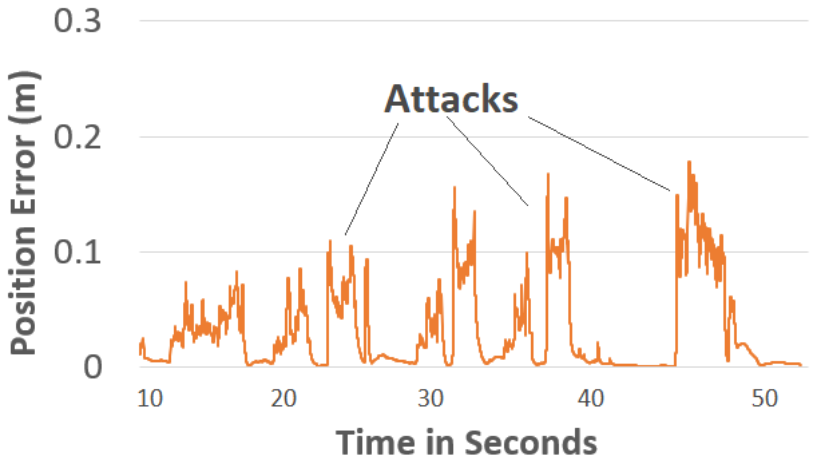


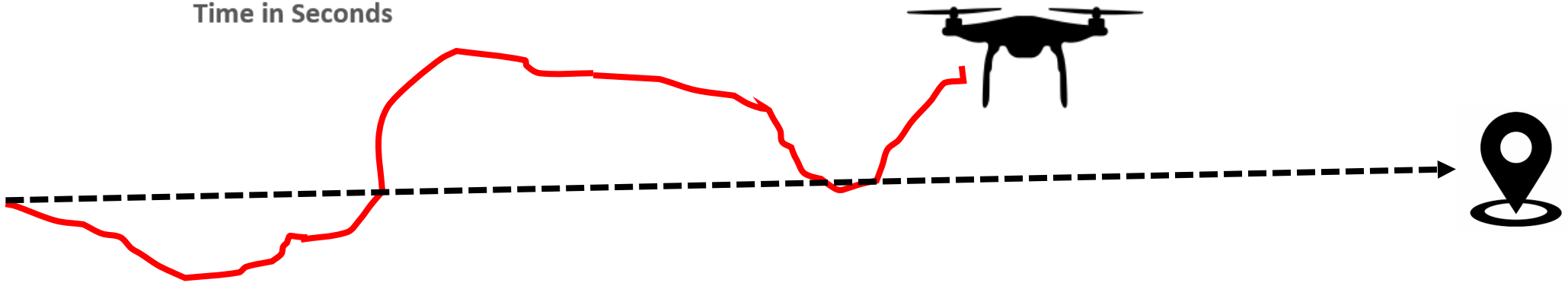
Over-Compensation under Physical Attacks



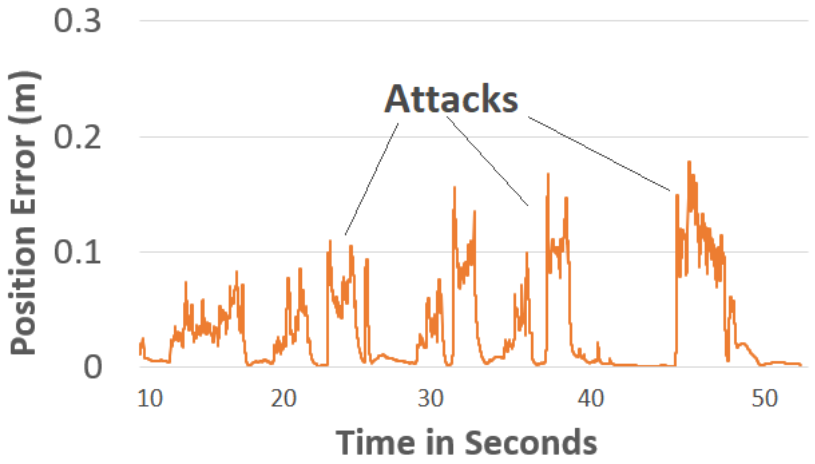


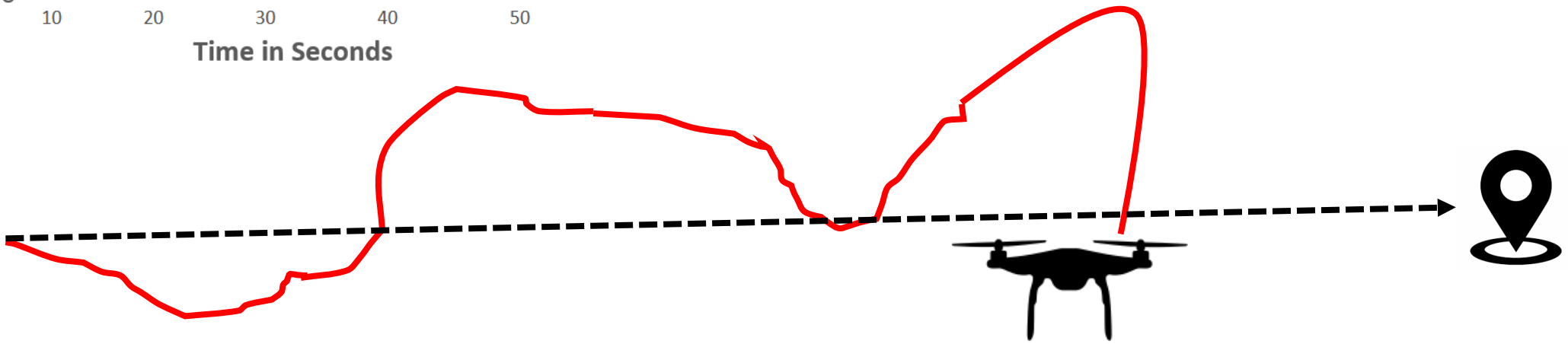
Over-Compensation under Physical Attacks



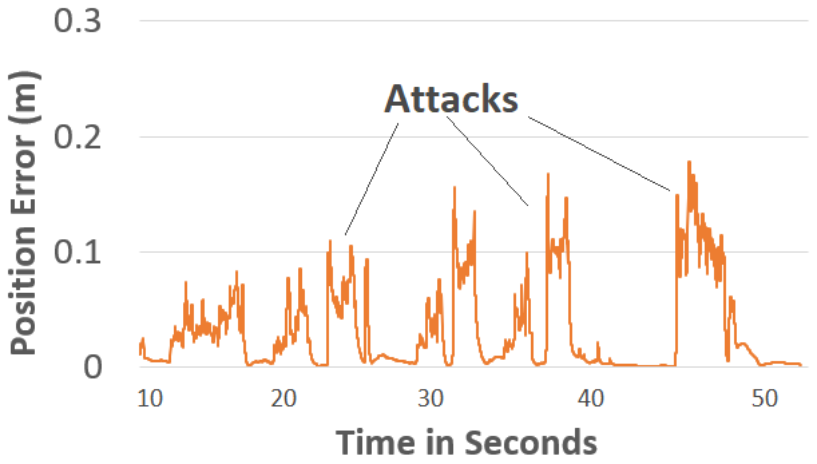


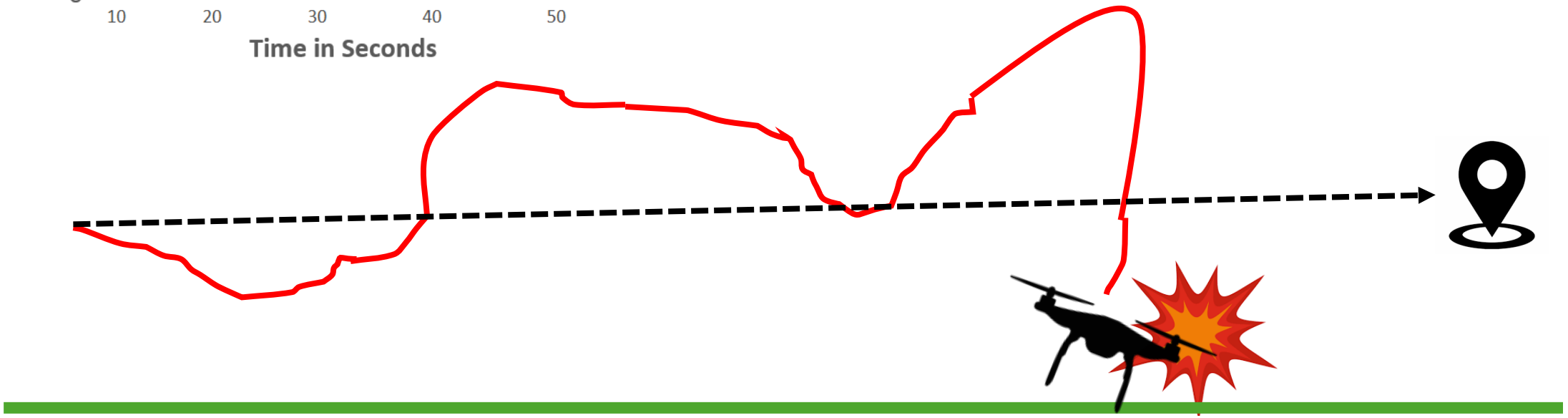
Over-Compensation under Physical Attacks



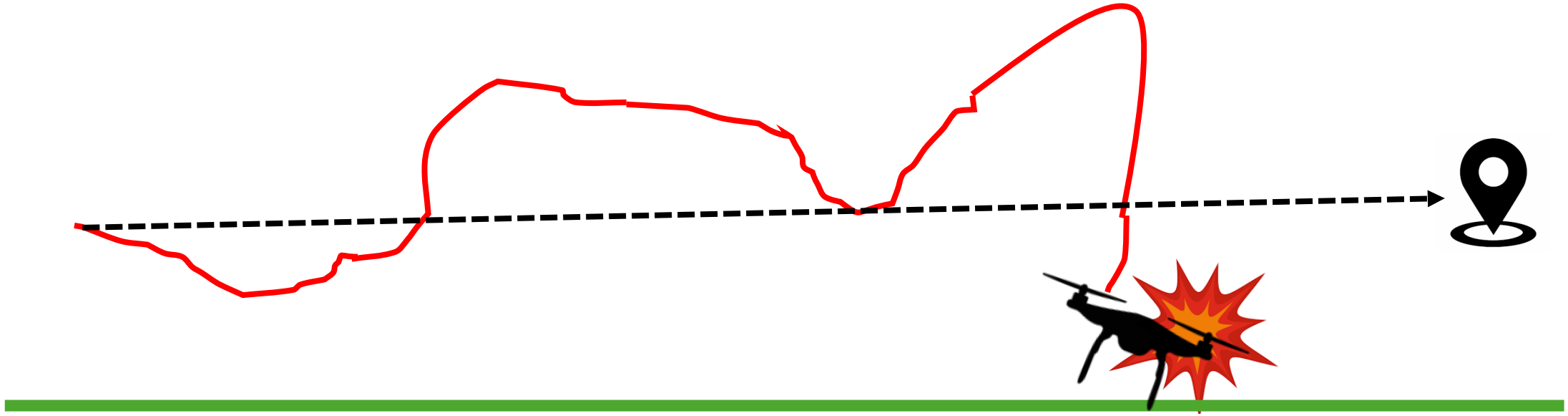


Over-Compensation under Physical Attacks





Over-Compensation under Physical Attacks



Compensation-based Error Correction

Environmental Noise ✓

Attacks ✗

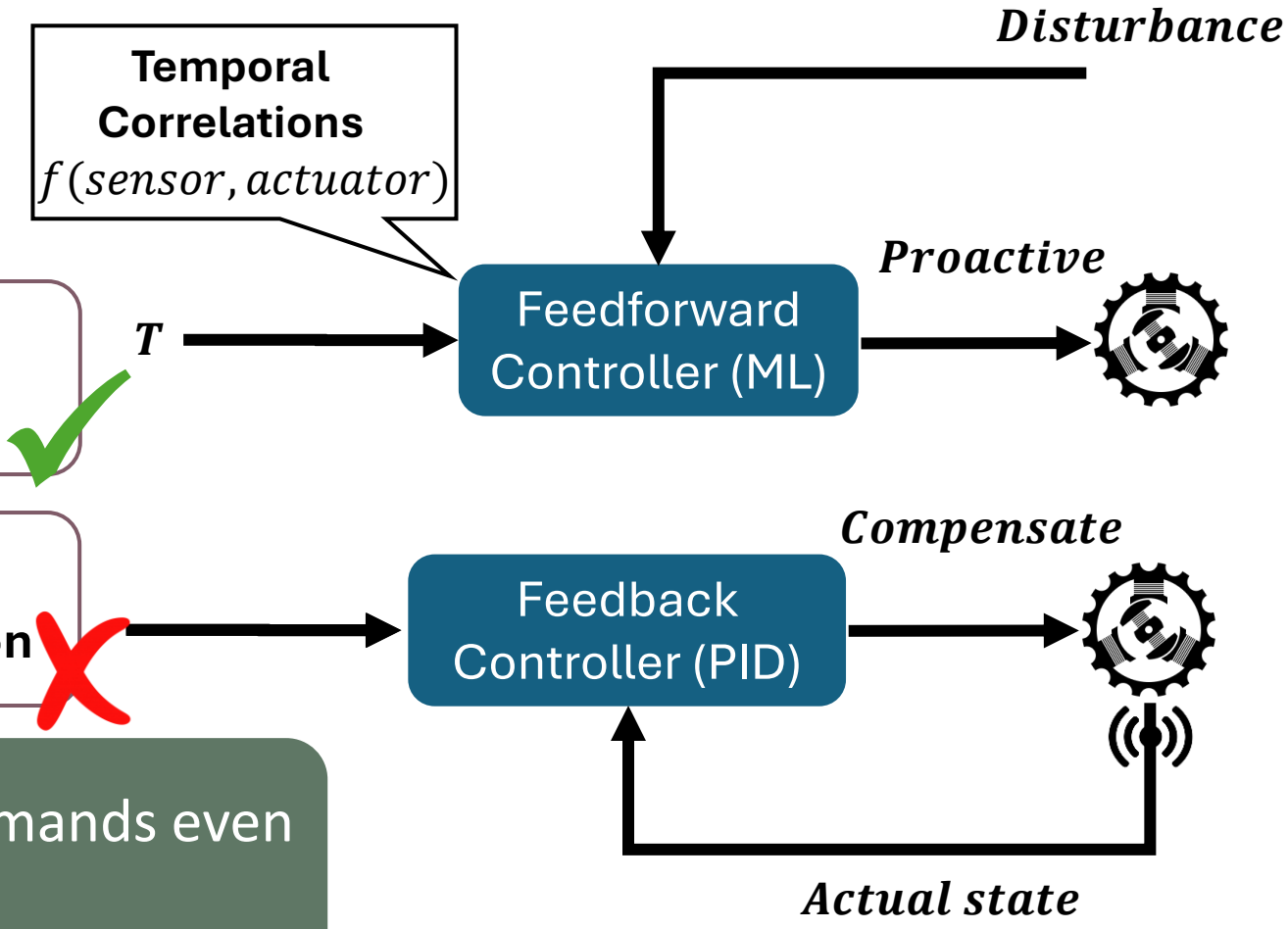
PID-Piper: Feedforward Controller (FFC)

Feedforward Control

Sensor-Actuator history
Proactive error correction

error = |target-actual|
Compensation-based error correction

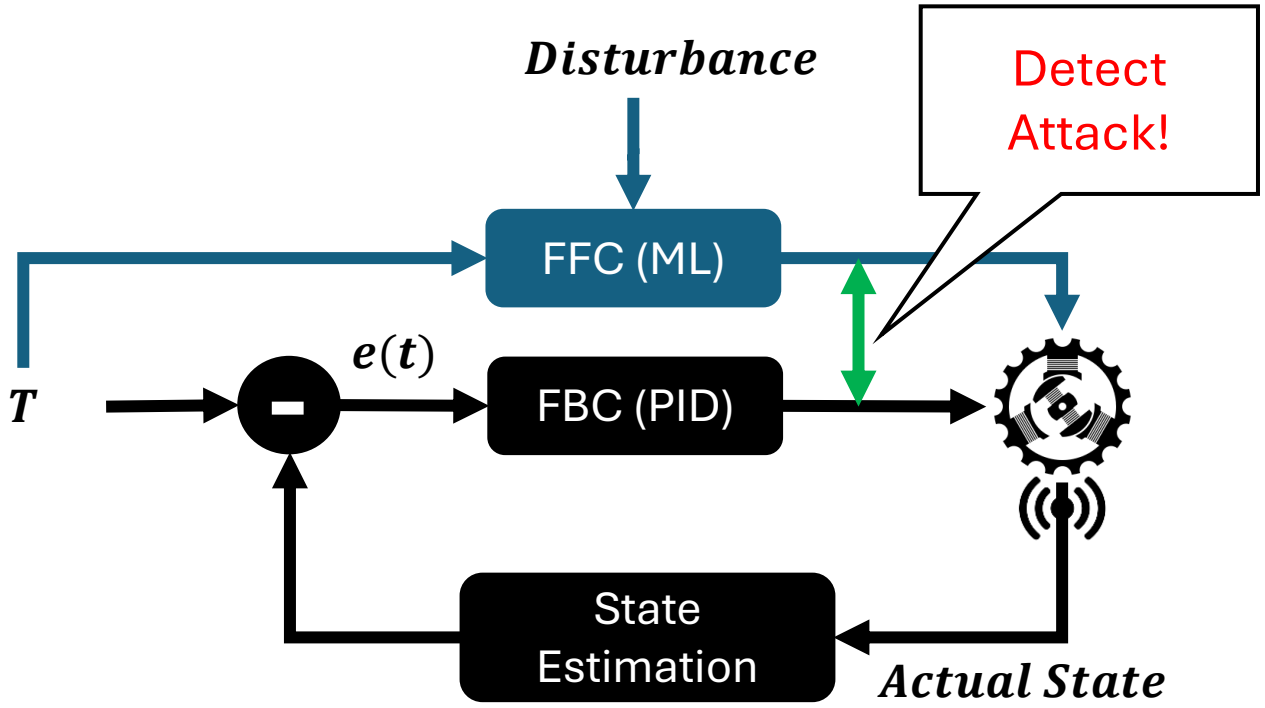
FFC estimates robust actuator commands even under attack



PID-Piper: Attack Detection

Feedforward Control

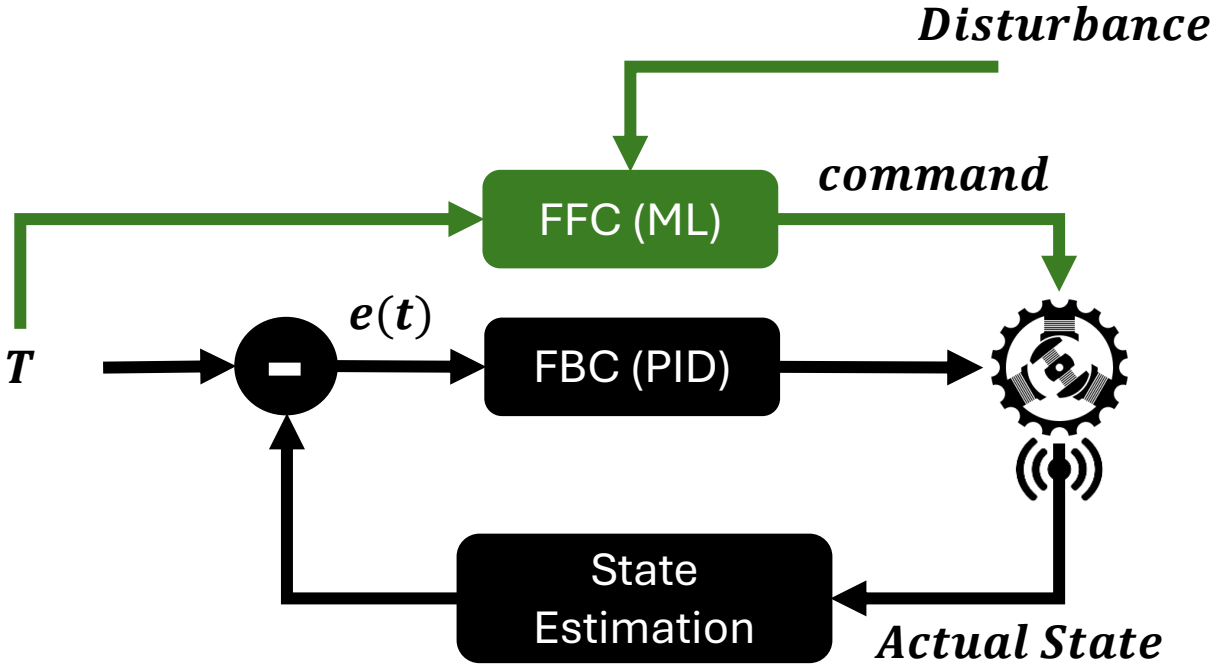
Feedback Control



PID-Piper: Attack Recovery

Feedforward Control

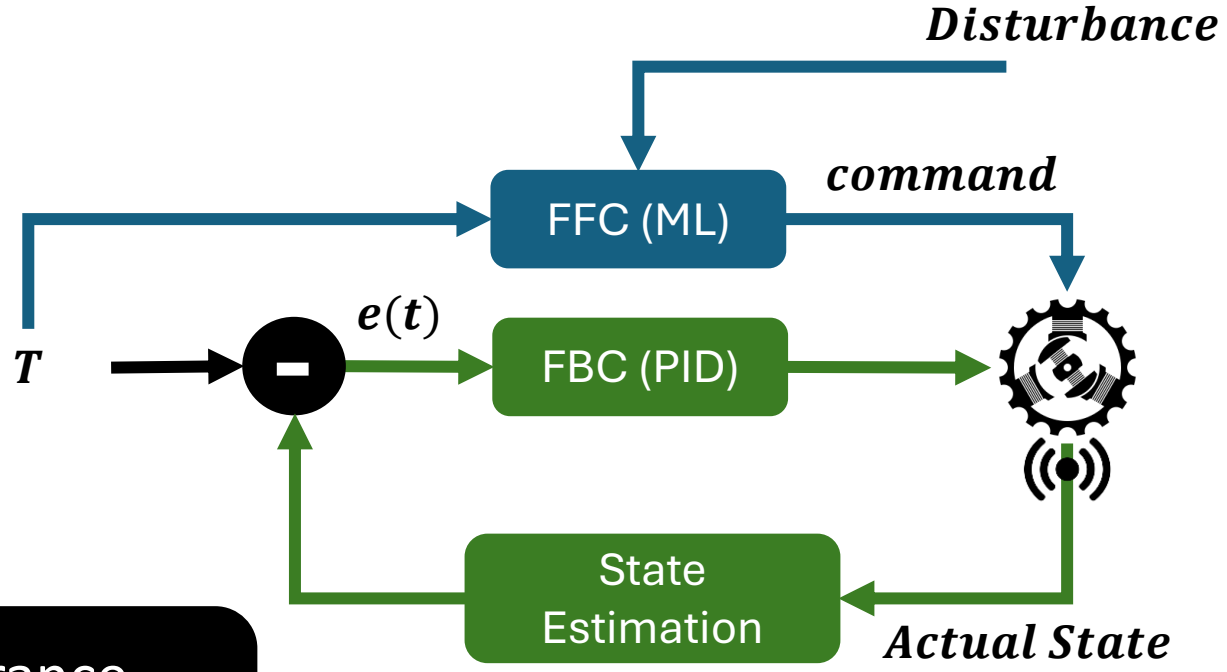
Feedback Control



PID-Piper: Attack Recovery Framework

Feedforward Control

Feedback Control



Feedback Control (FBC) → Fault Tolerance

Feedforward Control (FFC) → Attack Recovery

Key Results

Compensation-based error correction is **not effective** under attacks!

Feed-forward control for attack recovery.

The **first** attack recovery technique → **mission completion**.

Activity 2 – Attack Recovery

Instructions https://tinyurl.com/vsec2025 (Worksheet.pdf)

cd ravage/

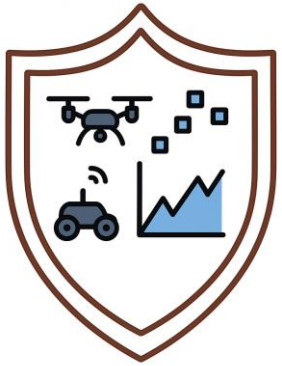
Launch a mission

```
python mission.py –s ArduPilot
```

Activate PID-Piper

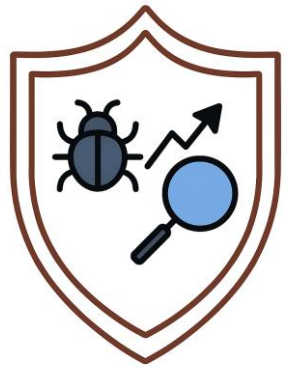
```
python pid-piper.py –s ArduPilot
```

DeLorean: Multi-Sensor Attack Recovery

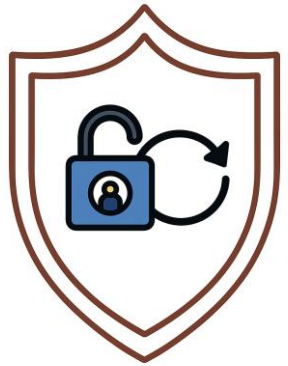


DSN’25

DSN’21

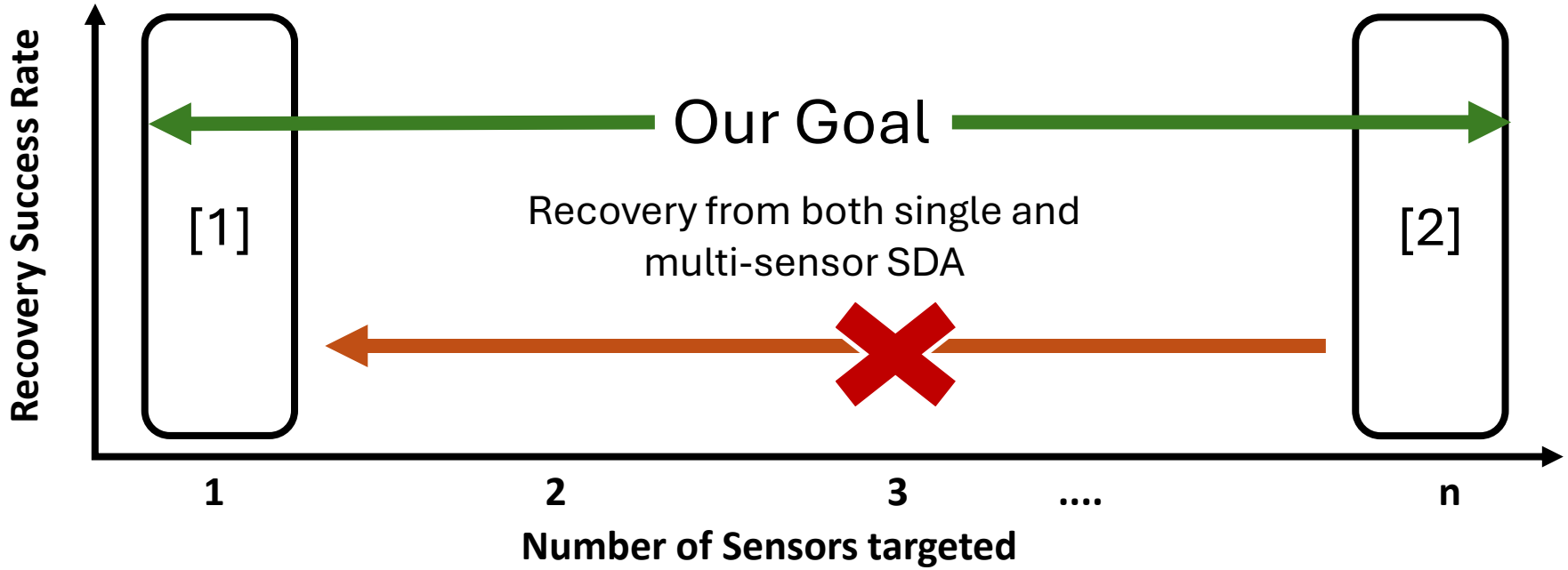


AsiaCCS’24



CCS’24

Multi Sensor Attack Recovery



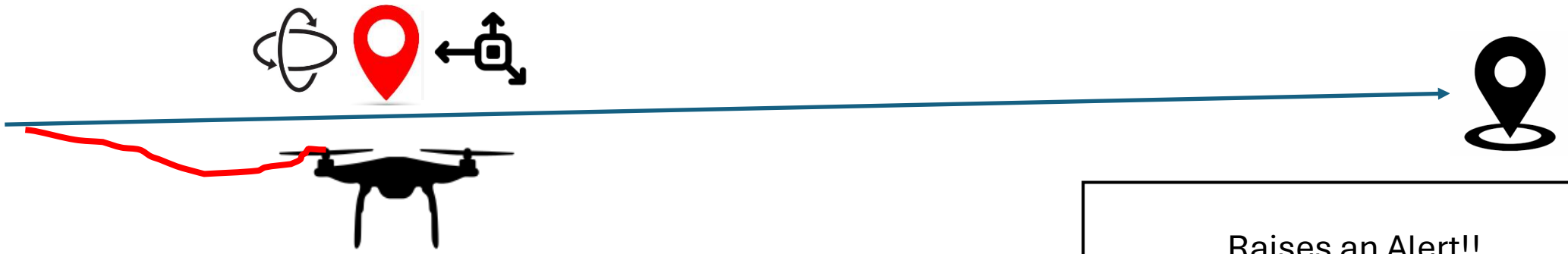
Prior recovery techniques

1. Model-based [ICRA'20, DSN'21, NDSS'23]
2. Checkpoint-based [ICCPS'18, RTSS'20]

Gaps

- Best case or Worst case assumptions.
- Recovery is either aggressive or conservative.

DeLorean: Multi-Sensor Attack Recovery



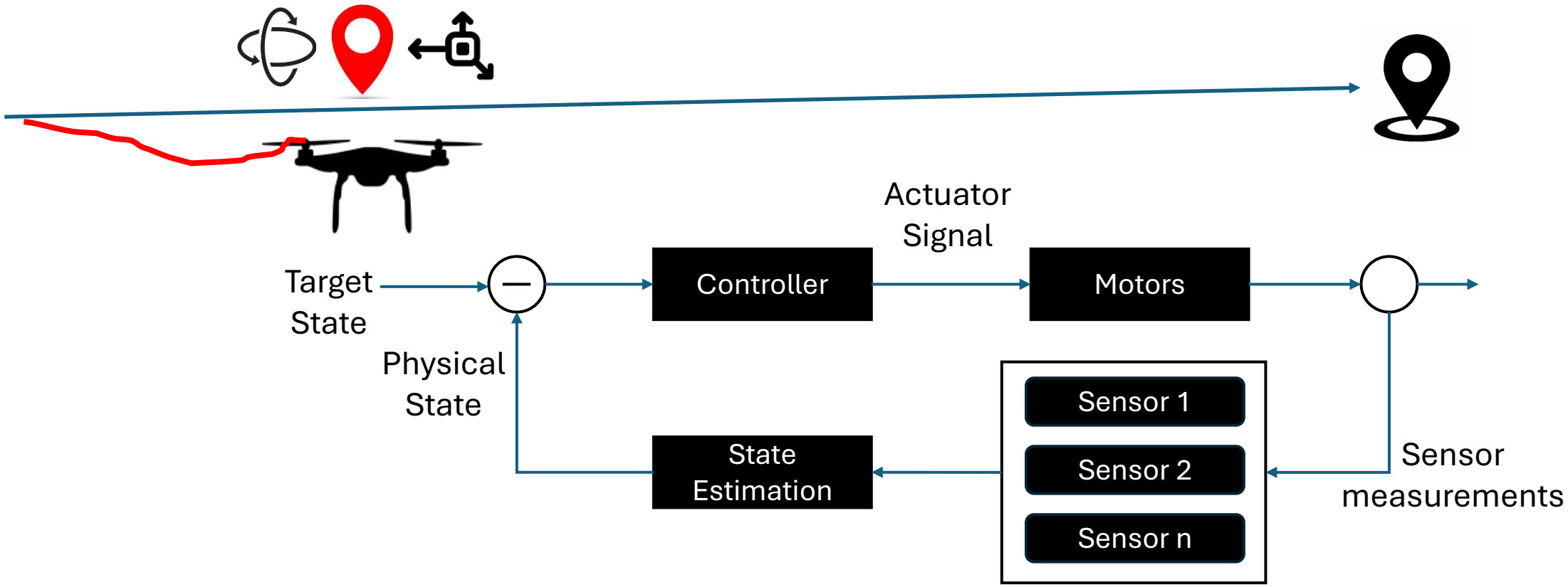
Attack Detector

Raises an Alert!!
Integrates with Existing Detector.

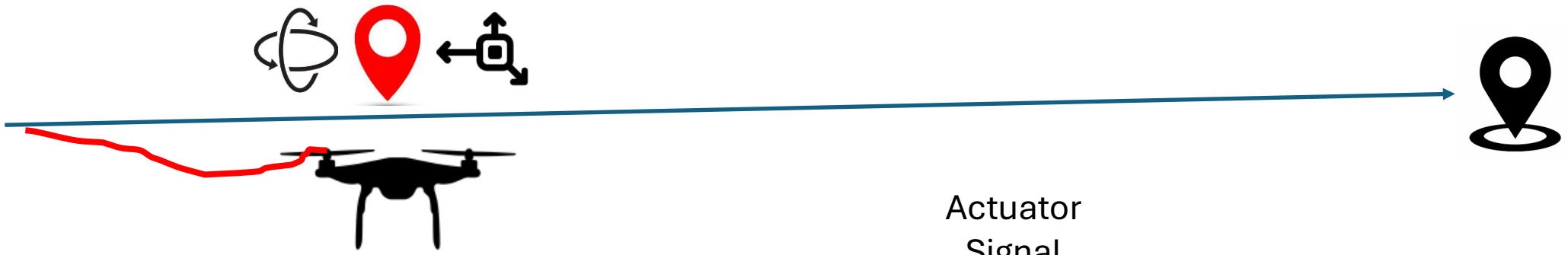
Attack Diagnosis

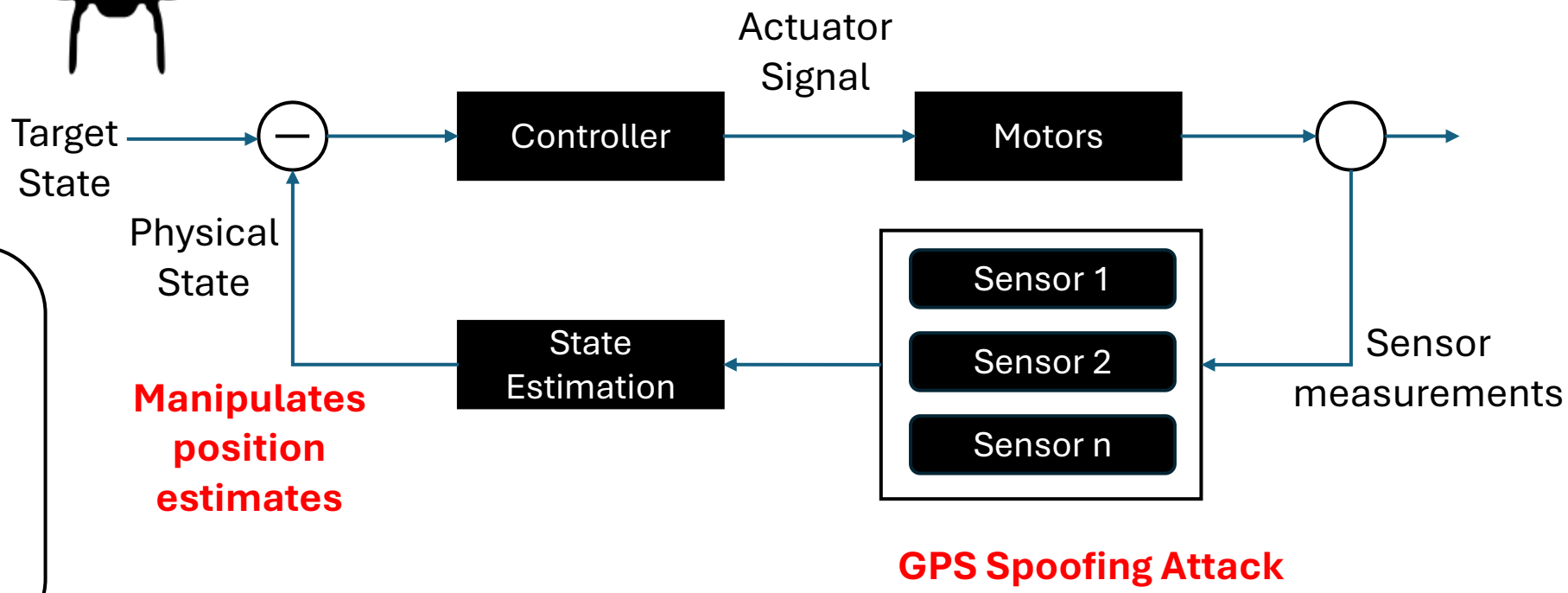
Identify sensors targeted by SDA

DeLorean: Attack Diagnosis



DeLorean: Attack Diagnosis





$e_t = |e_{t-1} - e_t|$

$E = \{e_1, e_2, e_3, .., e_n\}$

Causal Analysis

$\boldsymbol{P(GPS = malicious\ |E)}$

State Reconstruction

HS_Position (x, y, z)
HS_Attitude (x, y, z)
HS_Velocity(x, y, z)
...

Trustworthy Historic Data

Non-Linear Estimation

HS_Position (x, y, z)
Attitude (x, y, z)
Velocity(x, y, z)
...

Reconstructed State

From benign sensors

Activity 3 – Multi-Sensor Attack Recovery

Instructions https://tinyurl.com/vsec2025 (Worksheet.pdf)

cd ravage/

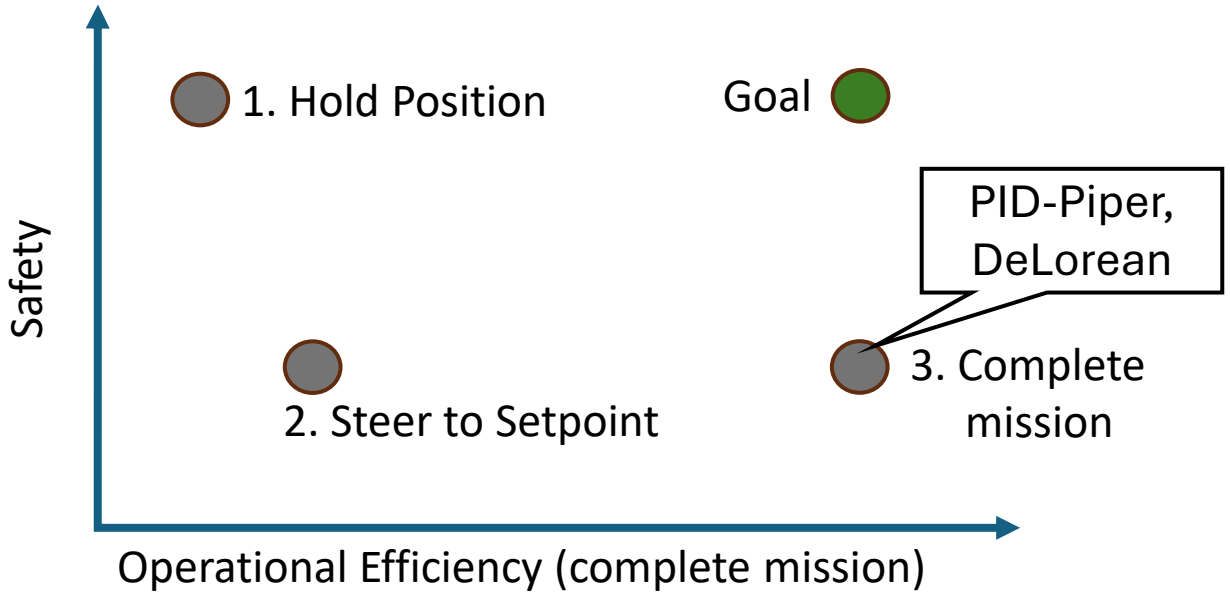
Launch a mission

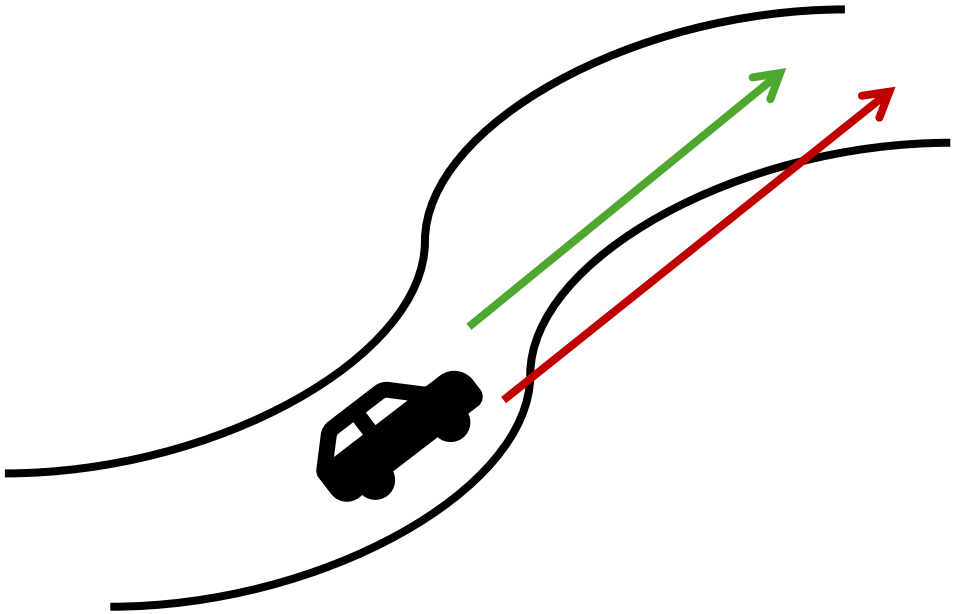
```
python mission.py -s ArduPilot
```

Activate DeLorean

```
python delorean.py -s ArduPilot
```

Landscape of Attack Recovery for RAVs





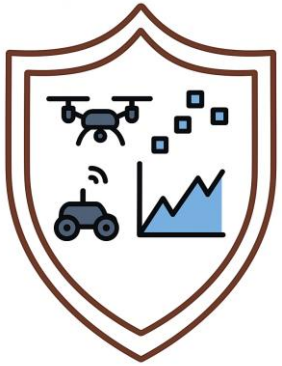
Prior recovery techniques

1. Hover, hold position, or return to home.
2. Steer to a set point [ICRA'20, RTAS'23]
3. Prevent a crash or mission failure [DSN'21, NDSS'23, AsiaCCS'24]

Gaps

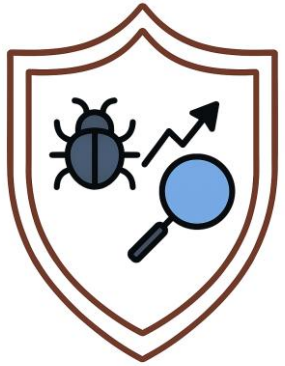
- Narrow recovery focus
- Do not enforce safety specifications

SpecGuard: Specification Aware Attack Recovery

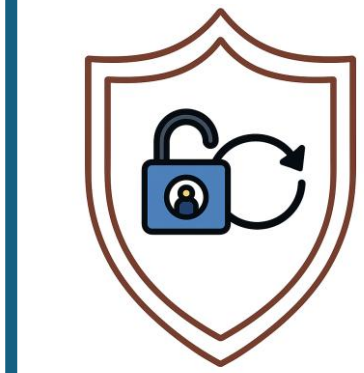


DSN’25

DSN’21



AsiaCCS’24



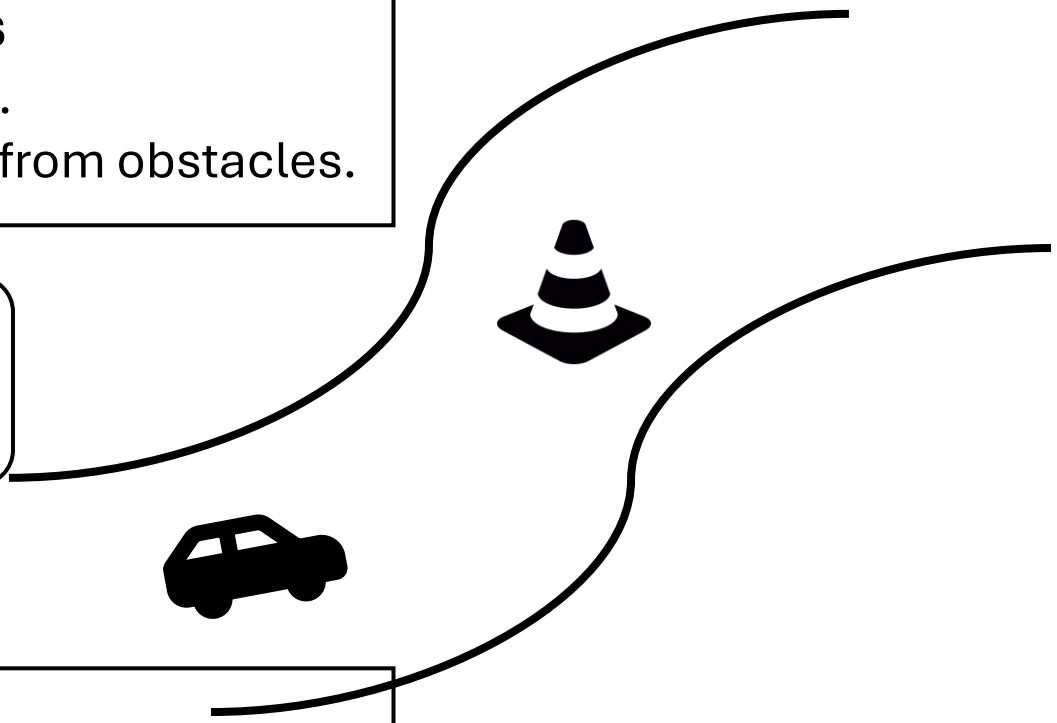
CCS’24

Formally Expressing Specifications

Safety Specifications

1. Stay within 10 m bound.
2. Maintain 10m distance from obstacles.

Goal: Comply with multiple specifications



STL Specifications

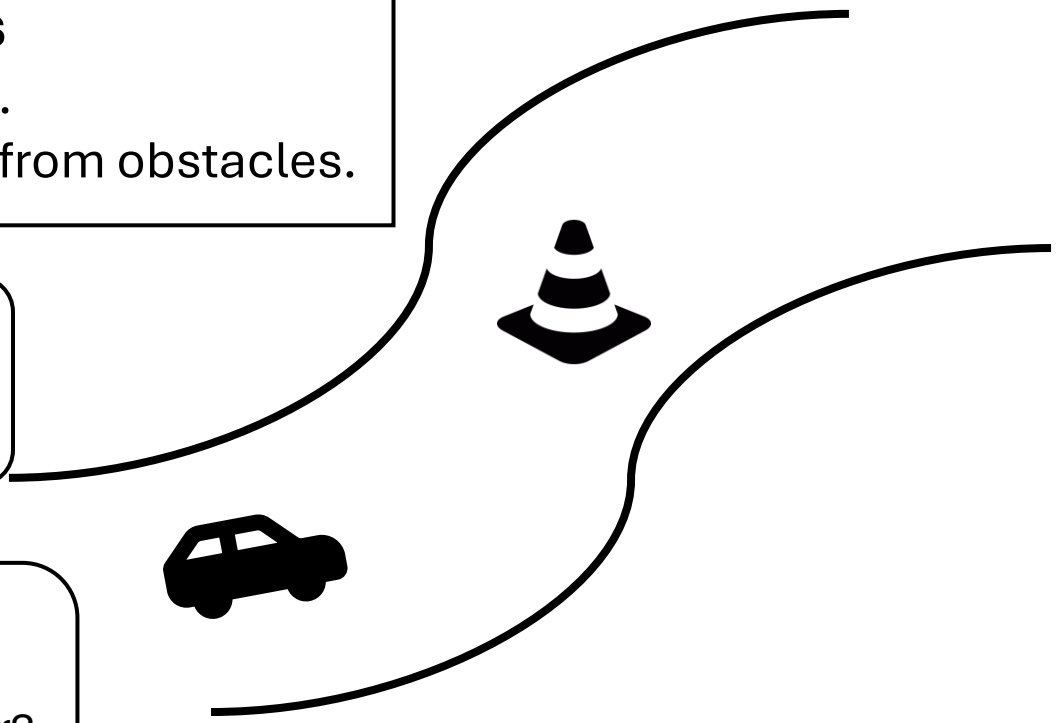
$$\big(always\ (checkPos(X_t) < 10)\big)\ \&$$
$$(always\ (obstacle\_d(X_t) > 10))$$

Designing Reward Function

Safety Specifications

1. Stay within 10 m bound.
2. Maintain 10m distance from obstacles.

Goal: Comply with multiple specifications



Challenges

1. How do we assign rewards – policy learning?
2. Sparse reward – policy will not generalize

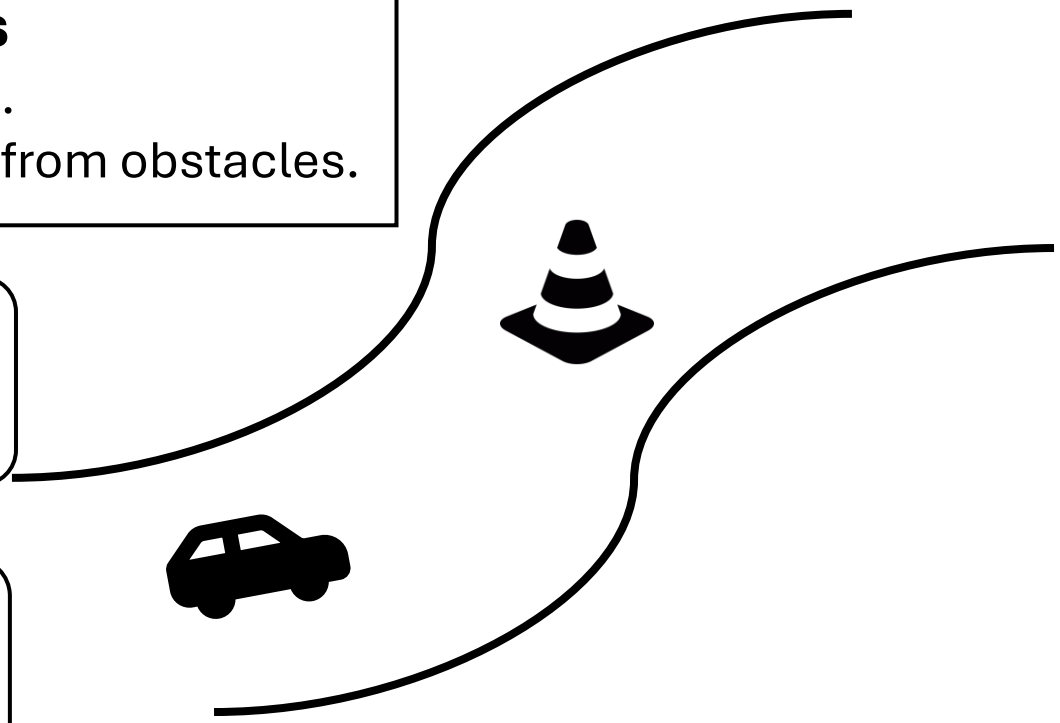
Designing Reward Function

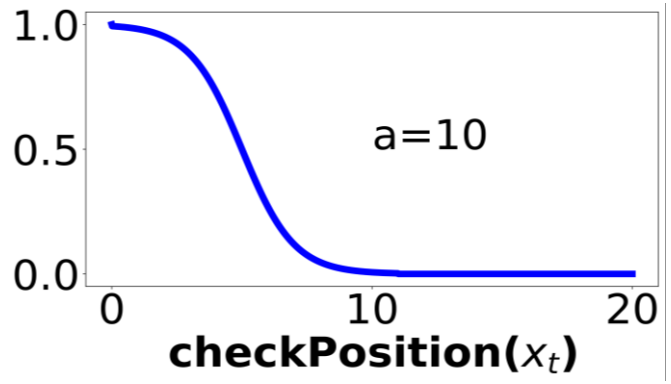
Safety Specifications

1. Stay within 10 m bound.
2. Maintain 10m distance from obstacles.

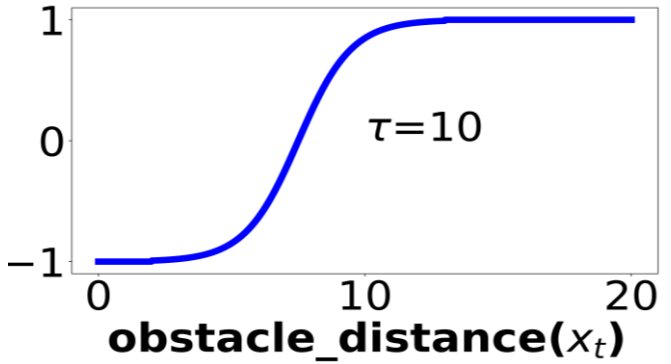
Goal: Comply with multiple specifications

Degree of Compliance → Assign a Reward





$always(checkPos(X_t) < a)$



$always(obstacle\_d(X_t) > \tau)$

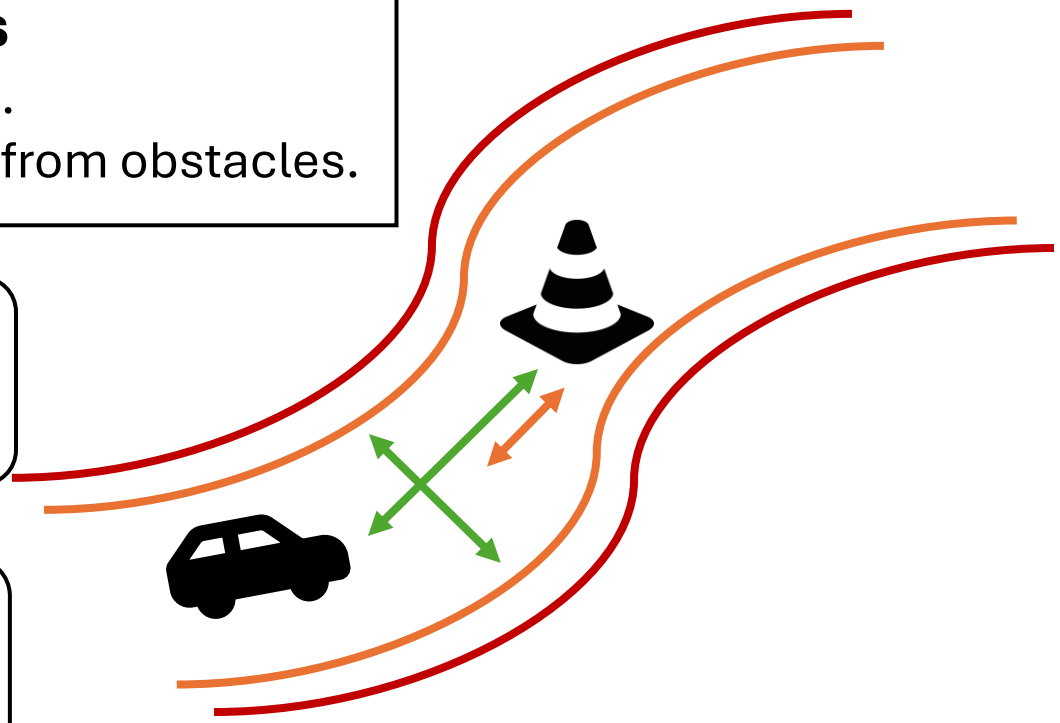
Multi-Objective Policy Learning

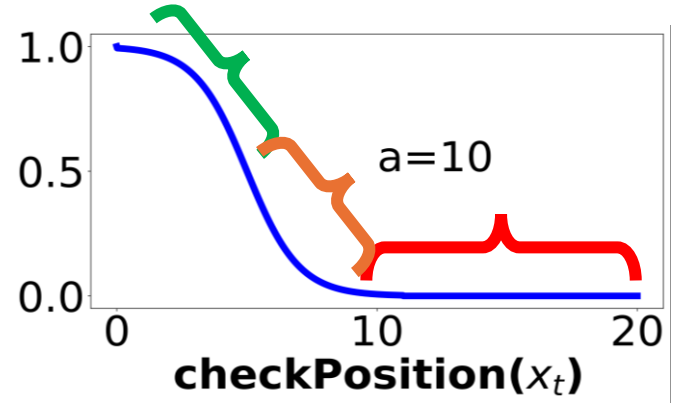
Safety Specifications

1. Stay within 10 m bound.
2. Maintain 10m distance from obstacles.

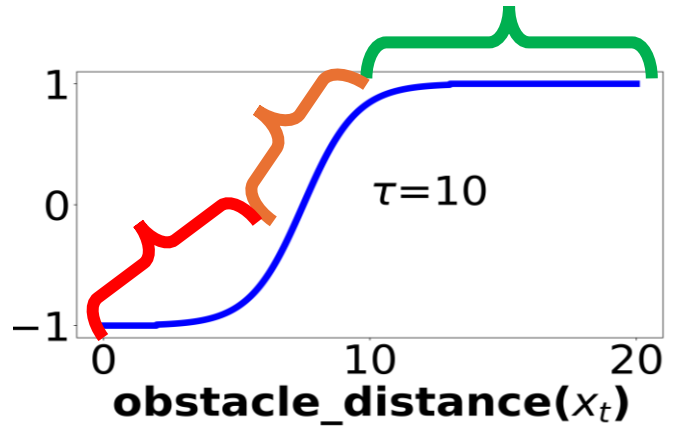
Goal: Comply with multiple specifications

Degree of Compliance → Assign a Reward





$always(checkPos(X_t) < a)$



$always(obstacle\_d(X_t) > \tau)$

Robust Control under Attacks

Safety Specs

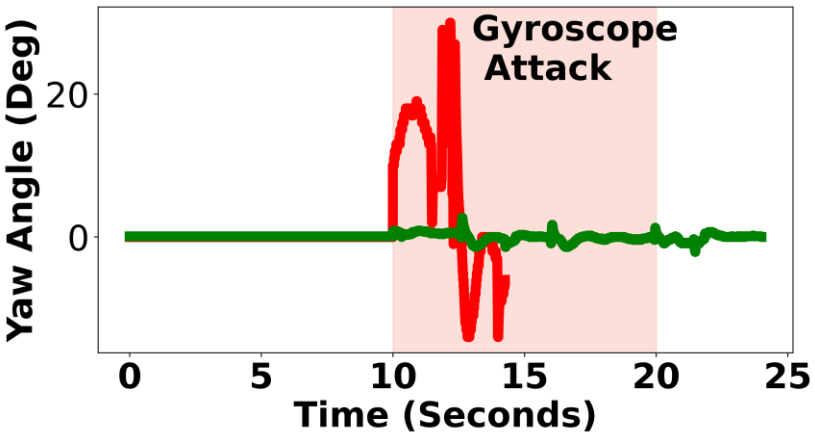
Deep-RL Controller
$\pi(Action\ |State)$

Robustness Control
↓
Adversarial Training

Challenges
1. Attack Space too large
2. Adversarial training won't generalize

Minimize the sensor manipulation.
Attack = $\{x \pm \epsilon, y \pm \epsilon, z \pm \epsilon\}$

State Reconstruction
Reliable States under Attacks.



Adversarial training → Robust Deep-RL Controller

Dash, Pritam, et al. "Diagnosis-Guided Attack Recovery for Robotic Vehicles from Sensor Deception Attacks, ACM AsiaCCS'2024

Key Results

12 Safety Specifications

1. Collision Avoidance
2. Stay within bounds
3. ..

11. Loiter Radius
12. Geofence

Videos

Key Results

Attack recovery techniques often do not ensure safety.

Multi-objective policy learning → Deep-RL Controller for RAVs

State Reconstruction guided adversarial training

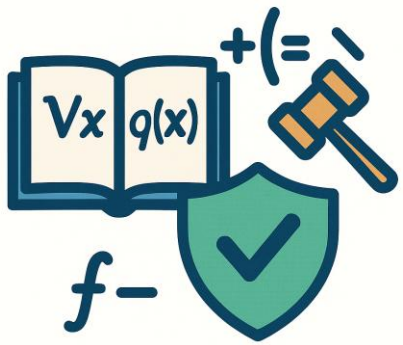
Future Work

Vision Sensors

Camera, LiDAR

Cyber Physical Systems

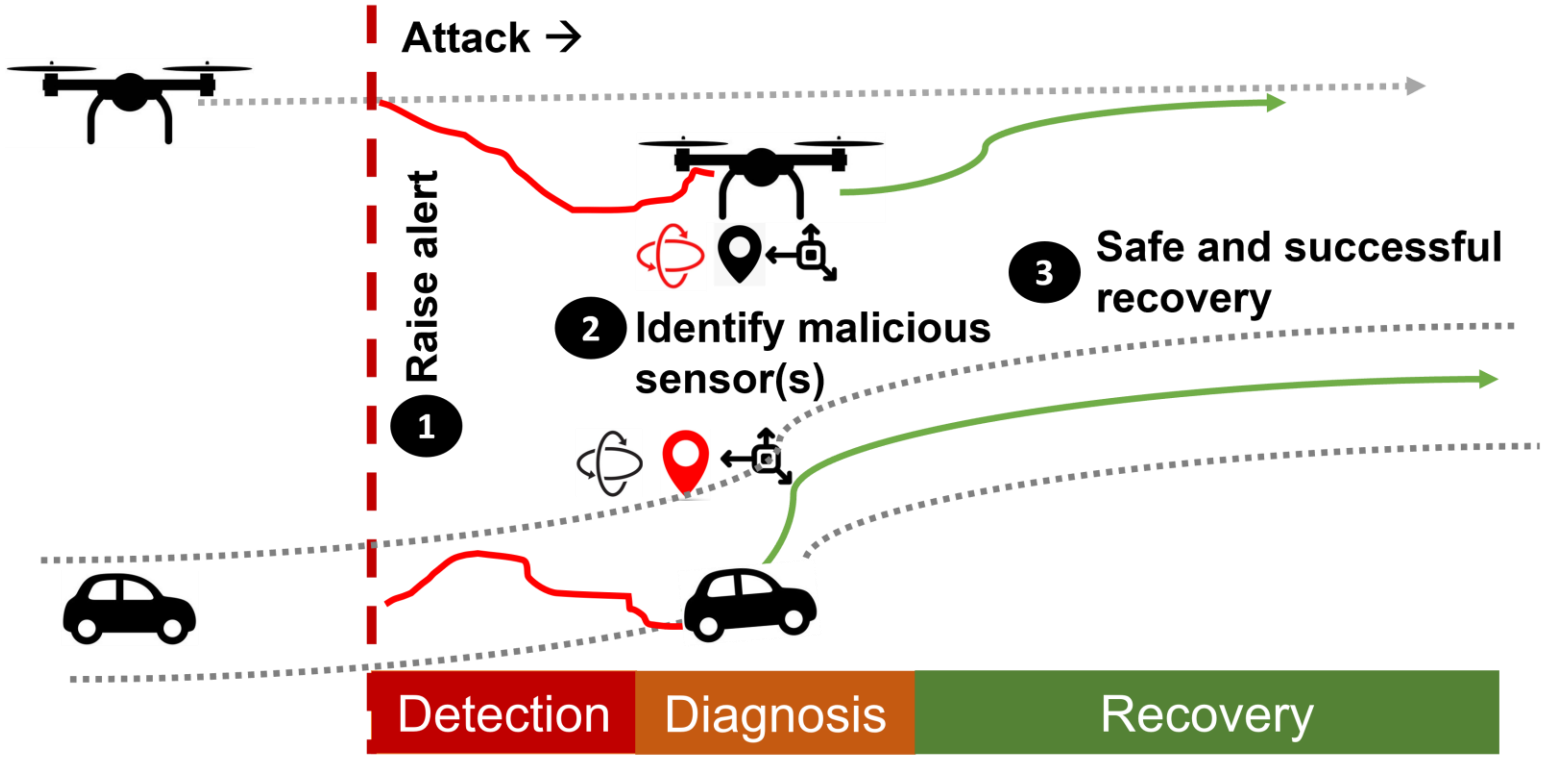
Robots, ICS



Formal Guarantees

Conclusion

Attack resilience is crucial for widespread adoption of RAVs



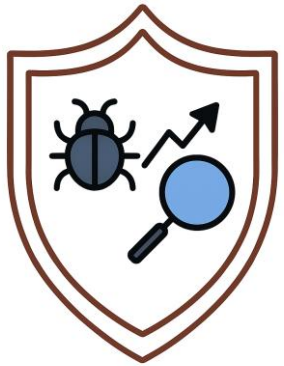
Conclusion

Attack resilience is crucial for widespread adoption of RAVs

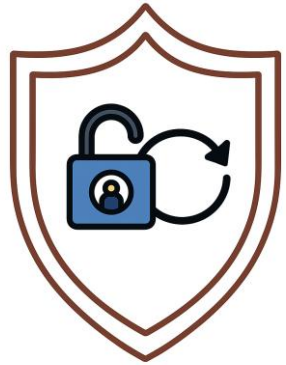
DSN'25

DSN'21

Best Paper Award



AsiaCCS'24



CCS'24

This thesis: **Software solutions** to **safeguard RAVs** against physical attacks

Conclusion

Software solutions to **safeguard RAVs** against physical attacks

1. Pritam Dash, Guanpeng Li, Zitao Chen, Mehdi Karimibiuki, and Karthik Pattabiraman. "PID-Piper: Recovering robotic vehicles from physical attacks." DSN 2021.
2. Elaine Yao, Pritam Dash and Karthik Pattabiraman, "SwarmFuzz: Discovering GPS Spoofing Attacks in Drone Swarms", DSN 2023
3. Pritam Dash, Guanpeng Li, Mehdi Karimibiuki, and Karthik Pattabiraman. "Diagnosis-guided attack recovery for securing robotic vehicles from sensor deception attacks." AsiaCCS 2024.
4. Pritam Dash , Ethan Chan, and Karthik Pattabiraman. "SpecGuard: Specification aware recovery for robotic autonomous vehicles from physical attacks." CCS. 2024.
5. Pritam Dash, and Karthik Pattabiraman. "RAVAGE: Robotic Autonomous Vehicles' Attack Generation Engine." DSN 2025
6. Pritam Dash, Mehdi Karimibiuki, and Karthik Pattabiraman. 2019. Out of control: stealthy attacks against robotic vehicles protected by control-based techniques. ACSAC 2019