Syracuse University

Adaptive Window-Based Sensor Attack Detection for Cyber-Physical Systems

* **Accepted by** 59th Design Automation Conference (**DAC**), **2022**

Lin Zhang Advisor: Dr. **Fanxin Kong**

Department of Electrical Engineering and Computer Science
Syracuse University, Syracuse NY

Colonial Pipeline Hack

Took down the largest fuel pipeline

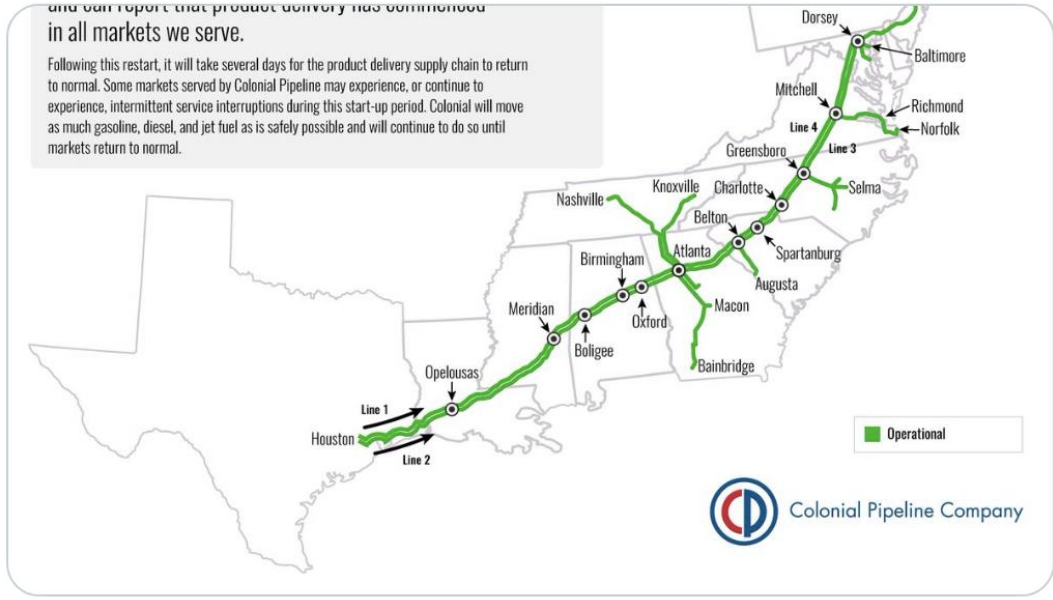
Led to shortages across the East Coast

Cars lined up to fill their gas tanks.

Pinned Tweet

Colonial Pipeline @Colpipe · 10h

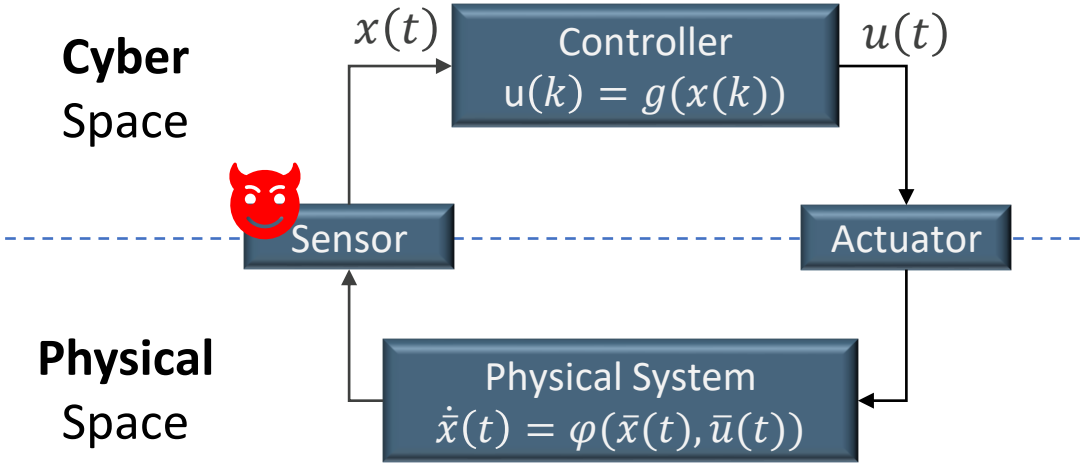
As we previously reported, Colonial Pipeline initiated the restart of pipeline operations at approximately 5 p.m. ET on Wednesday, May 12. Since that time, we have returned the system to normal operations, delivering millions of gallons per hour to the markets we serve.

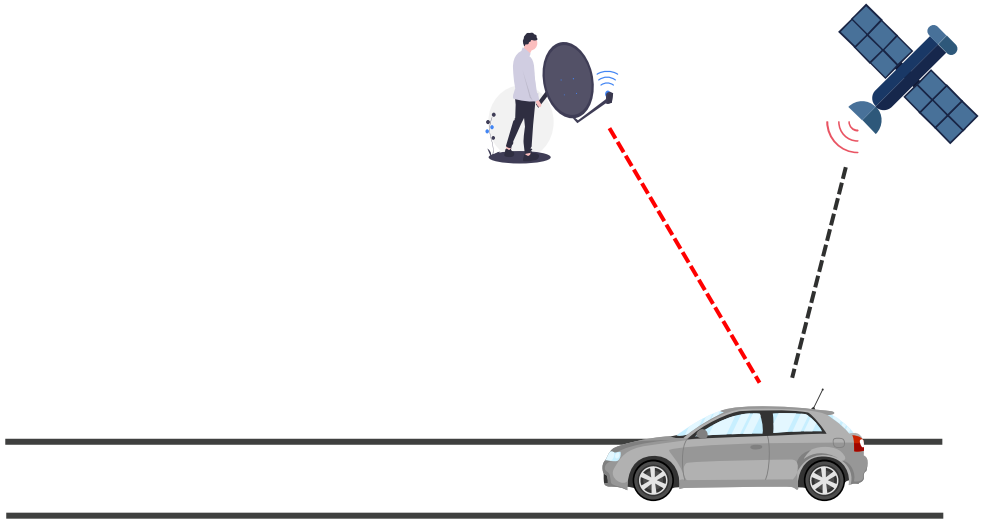


1 46 71

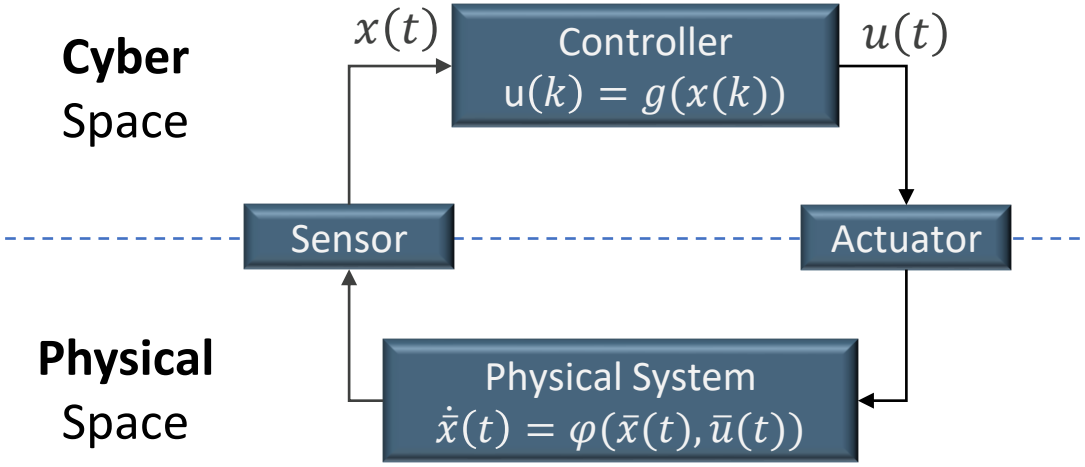
https://www.nytimes.com/2021/05/11/business/colonial-pipeline-shutdown-latest-news.html

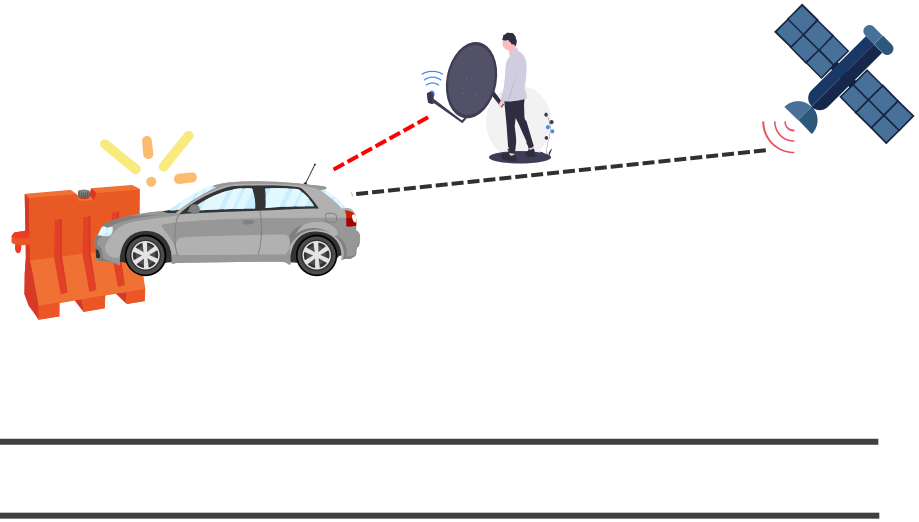
Sensor Attacks in Cyber-Physical Systems





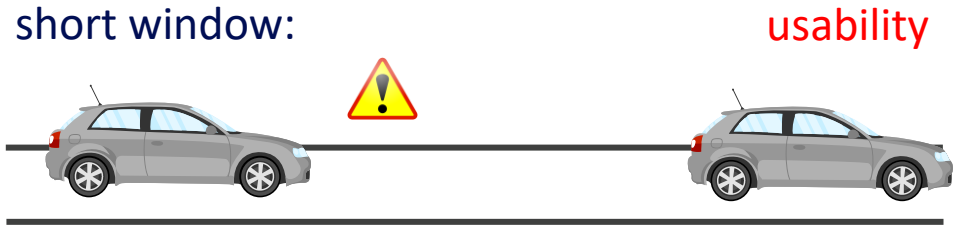
Sensor Attacks in Cyber-Physical Systems





Window-Based Sensor Attack Detection

- Usually, **detectors** detect attacks by monitoring **residuals** between observed sensor measurements and predicted values within a detection window.
- The existing detector treat the detection **window** as a **fixed** hyper-parameter, which faces a **dilemma**.



How to trade off between these two metrics?

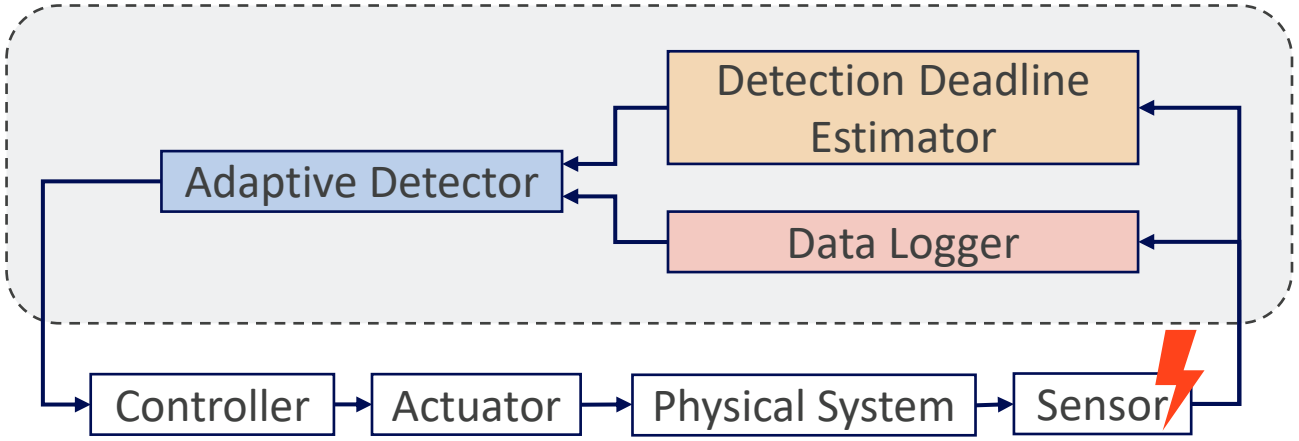
Adaptive: dynamically adjust the detection window according to detection deadline that is computed by online safety analysis

Overview

System Model: a discrete linear time-invariant (LTI) model

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t + \boldsymbol{v}_t$$

Threat Model: An attacker can manipulate **sensor measurement**, thus compromise state estimates.



Detection Deadline Estimator

- **Reachability-based** technique for future potential behaviours
- Estimates the detection deadline after which the physical system may **touch unsafe states.**

Adaptive Detector

- A window-based detection
- Dynamically **adapt** its detection delay according to the deadline
- Do **not miss** any data points

Data Logger

- A sliding-window based data logging protocol
- Keeps **trustworthy data** for the deadline estimation and attack detection

Detection Deadline Estimation

Reachable set: $\mathcal{R}$ contains all possible system states evolving from initial state, and it is easier to formulate its over-approximation $\bar{\mathcal{R}}$

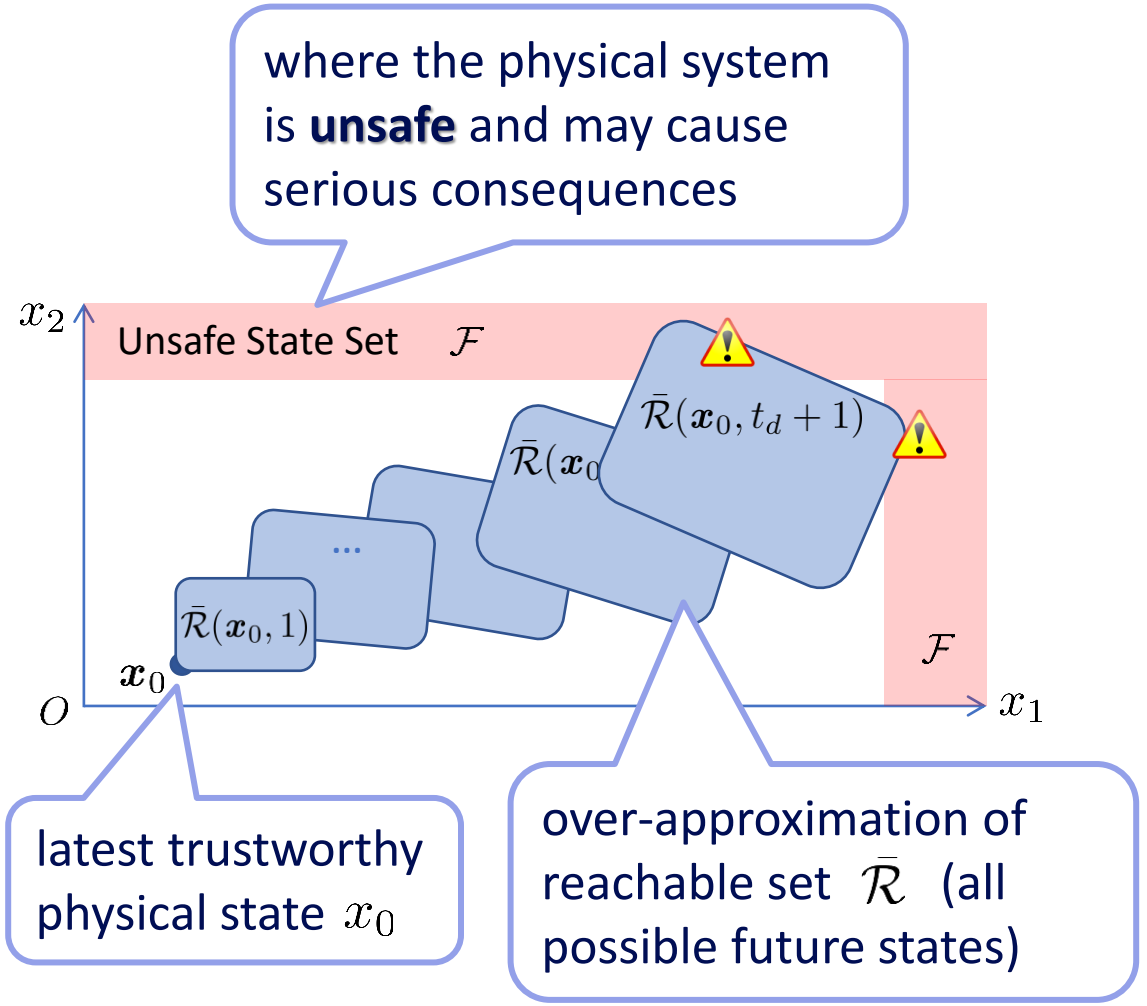
$$\boldsymbol{x}_t \subseteq \bar{\mathcal{R}}(\boldsymbol{x}_0, t) = A^i \bar{\boldsymbol{x}}_0 \oplus \bigoplus_{j=0}^{i-1} A^j B \mathcal{B}_{\mathcal{U}} \oplus \bigoplus_{k=0}^{i} A^k \mathcal{B}_\epsilon$$

where $\oplus$ denotes the Minkowski sum

Safety Analysis: If reachable set over-approximation does not intersect with unsafe set, i.e. $\bar{\mathcal{R}} \cap \mathcal{F} = \emptyset$, the system is guaranteed to be safe.

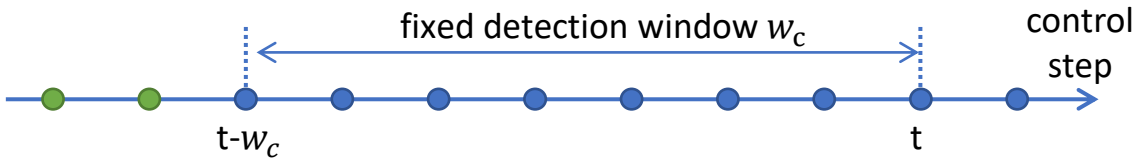
We can compute the upper and lower bound of $\bar{\mathcal{R}}$ through **support function** method efficiently:

$$\rho_{\bar{\mathcal{R}}} = \boldsymbol{l}^T (A^t \boldsymbol{x}_0) + \sum_{i=0}^{t-1} \rho_{\mathcal{B}_{\mathcal{U}}}((A^i B)^T \boldsymbol{l}) + \sum_{i=0}^{t-1} \rho_{A^i \mathcal{B}_\epsilon}(\boldsymbol{l})$$



Deadline Search Process

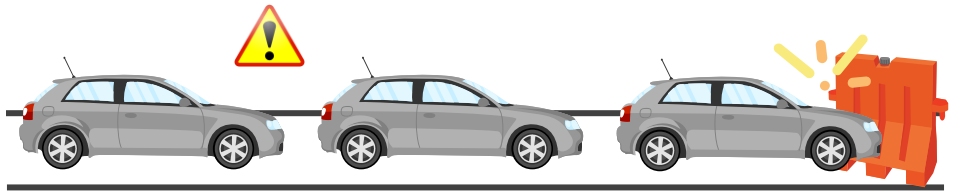
Adaptive Window Based Attack Detection



average residual in the detection window

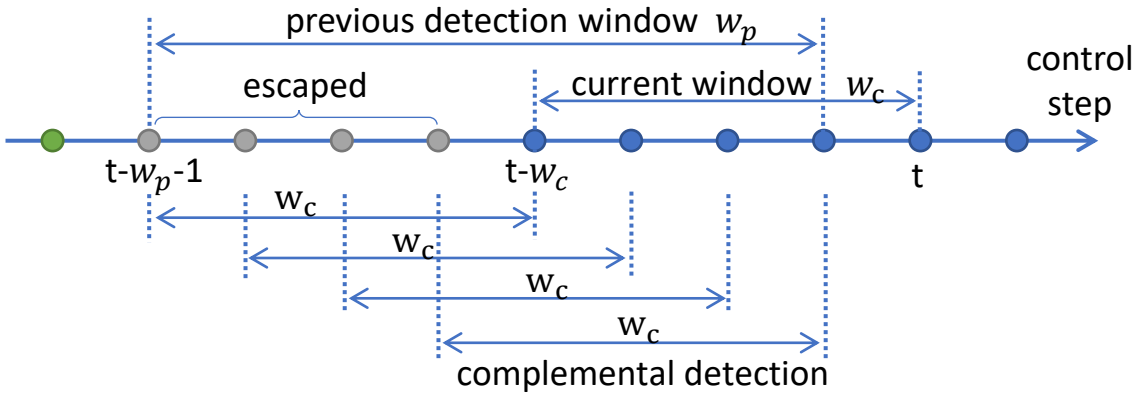
$$z_t^{avg} = \frac{1}{w_c} \sum_{i \in [t-w_c, t]} |\tilde{x}_i - \bar{x}_i|$$

where $\tilde{x}_t = A\bar{x}_{t-1} + Bu_{t-1}$ is predicted state, and $\bar{x}_t$ is observed state at time t.

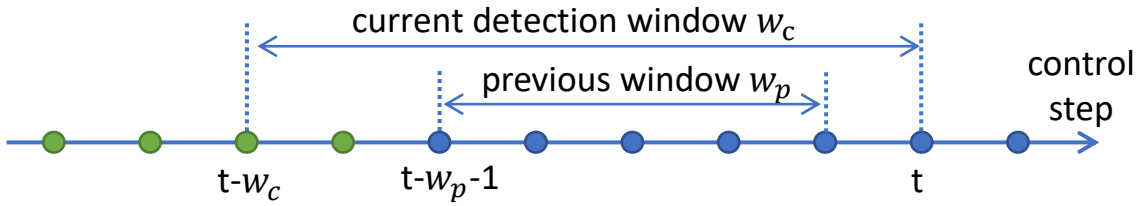


shorter window ⇒ false alarm

longer window ⇒ deadline miss



Decreasing the Detection Window Size



Increasing the Detection Window Size

Data Logging Protocol

- Record historical data:
  - **Residual** between predicted and observed states
  - System **state estimations**
- Keeps **trustworthy data** for the deadline estimation and sufficient data points for attack detection

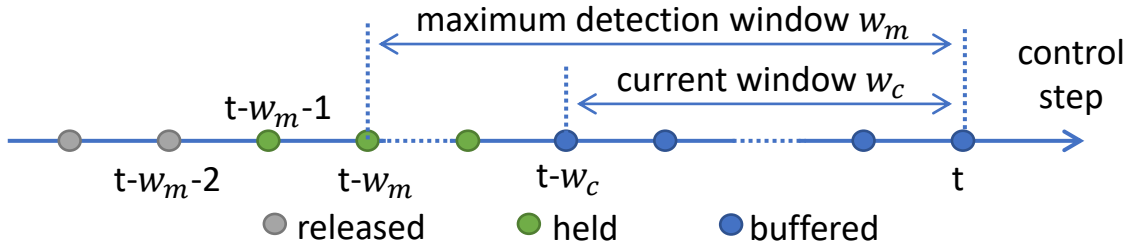


Illustration of the Data Logger

Buffer.

- **Within** the current detection window w_c
- Whether they are intact is still **unknown** since they are still being checked by the detector

Hold.

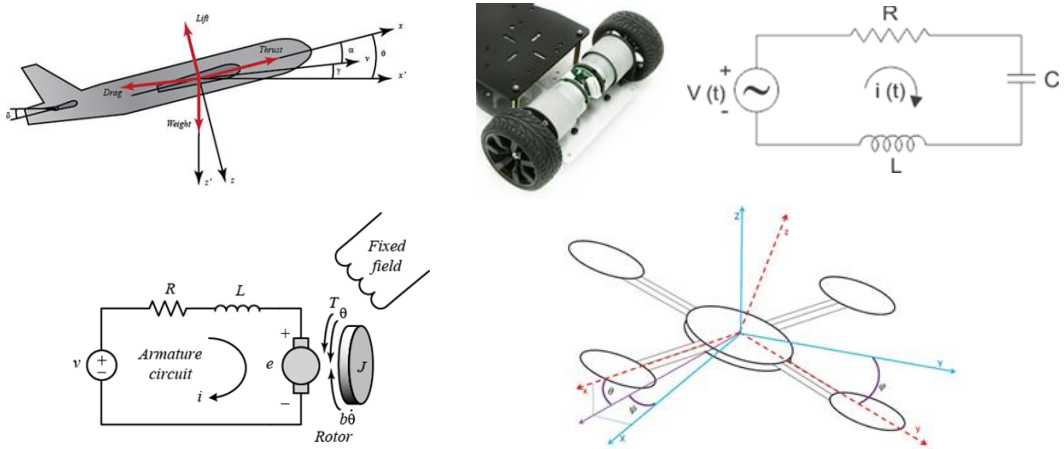
- Moved **outside** the current detection window
- Data are regarded **trustworthy** and thus held

Release.

- Historical data before $t - w_m - 1$ are outside the sliding window and **not needed** anymore
- Can be **released** to save storage space

Simulation Setting

CPS simulators:



Sensor attack scenarios:

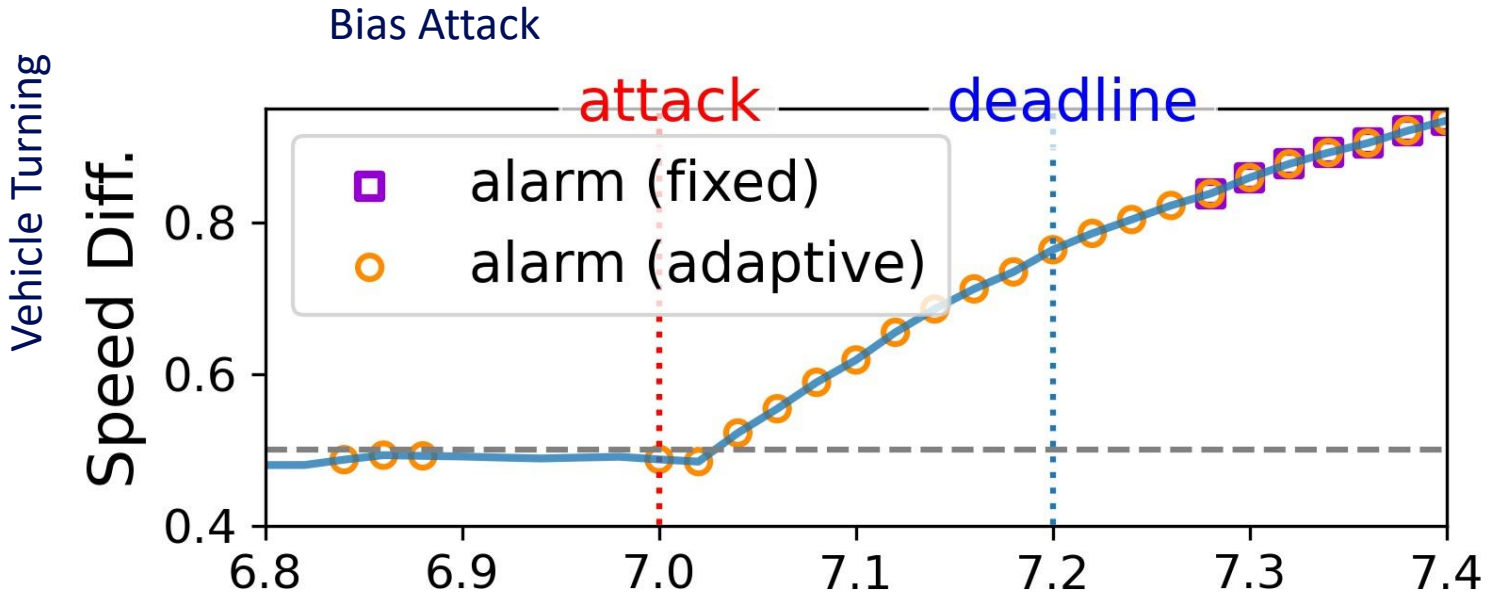
- **Bias attack** replaces sensor data with arbitrary values.
- **Delay attack** delays sensor measurements sent to the controller for a certain time period, so that the controller cannot update the current state estimate in time.
- **Replay attack** replaces sensor data with previously recorded ones.

Simulation Settings. legends: No.: simulator number, δ: control stepsize (in second), PID: PID control parameters, U: control input range, ϵ: uncertainty bound, S: safe state set, $\boldsymbol{\tau}$: detection threshold

No.	Simulator	δ	PID	U	ϵ	$\mathcal{S}$	$\boldsymbol{\tau}$
1	Aircraft Pitch	0.02	14,0.8,5.7	$[-7, 7]$	7.8e−3	$z \in [[-\infty, -\infty, -2.5], [\infty, \infty, 2.5]]$	$[0.012, 0.012, 0.012]$
2	Vehicle Turning	0.02	0.5,7,0	$[-3, 3]$	7.5e−2	$z \in [-2, 2]$	$[0.07]$
3	Series RLC Circuit	0.02	5,5,0	$[-5, 5]$	1.7e−2	$z \in [[-3.5, -5], [3.5, 5]]$	$[0.04, 0.01]$
4	DC Motor Position	0.1	11,0,5	$[-20, 20]$	1.5e−1	$z \in [[-4, -\infty, -\infty], [4, \infty, \infty]]$	$[0.118, 0.118, 0.118]$
5	Quadrotor	0.1	0.8,0,1	$[-2, 2]$	1.56e−15	$z \in [-5, 5]$	$[0.018, ..., 0.018]$

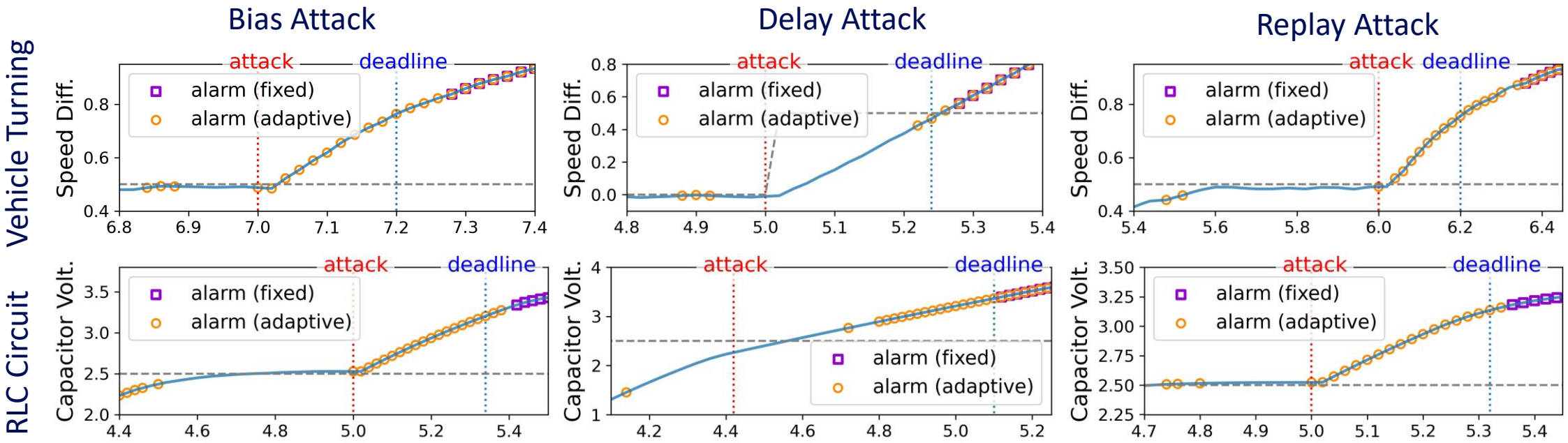
Simulation Results

- Our **adaptive detector** can raise alerts **before** the detection deadline, i.e., in-time detection, while the detector with a **fixed** window size finds attacks **after** the deadline, i.e., untimely detection
- Note that our adaptive detector may raise some **false alarms** before real attacks are launched. The is because that our adaptive detector chooses a smaller window size to catch up with the detection deadline while increasing the false positives.



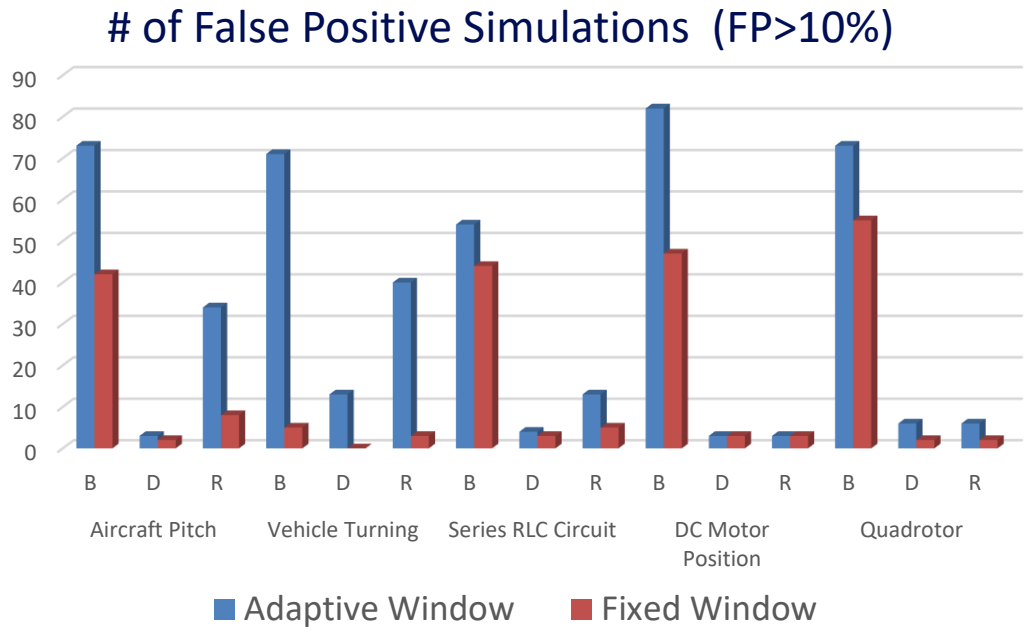
Simulation Results

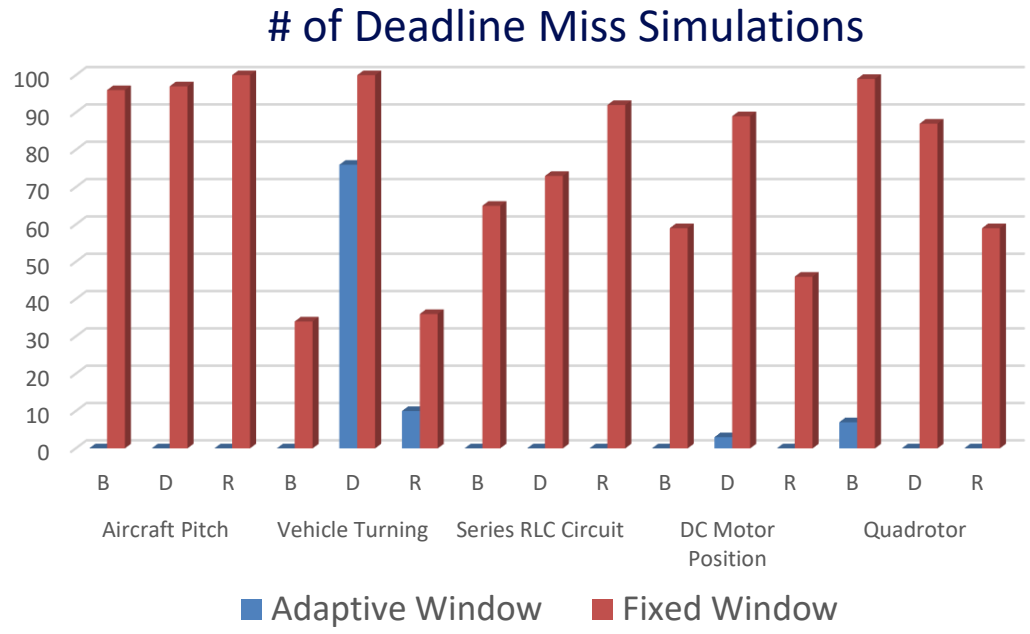
- Our **adaptive detector** can raise alerts **before** the detection deadline, i.e., in-time detection, while the detector with a **fixed** window size finds attacks **after** the deadline, i.e., untimely detection
- Note that our adaptive detector may raise some **false alarms** before real attacks are launched. The is because that our adaptive detector chooses a smaller window size to catch up with the detection deadline while increasing the false positives.



Simulation Results

- Our adaptive detector tends to have **larger false positive** numbers of simulations, but with **minimal deadline misses.**
- Note that our adaptive detector may miss the detection deadline in just 3 out of 100 experiments for one case, because those attacks have a **negligible effect** on the physical system.

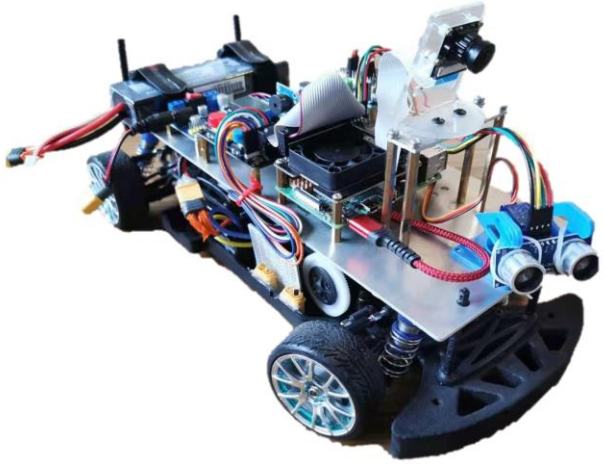


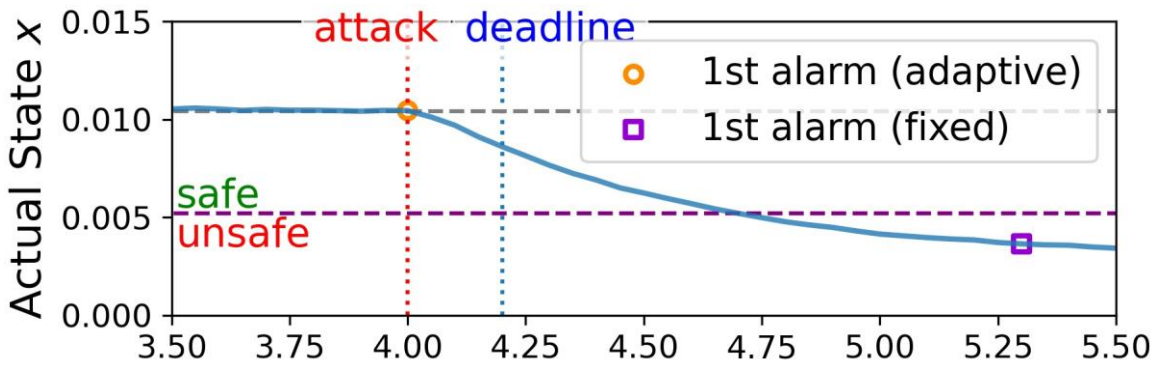


Testbed Results

- System model is obtained by system identification
- **Our** detector alerts in the **first step** after the attack, but the **fixed** window-based detection alerts **after** the vehicle reaching the **unsafe** state, which may already cause damages.
- Note that our adaptive detector detects the alert in the first step because the estimator computes the tightest deadline and shrinks the window size, making the average residual within the window larger than the threshold.

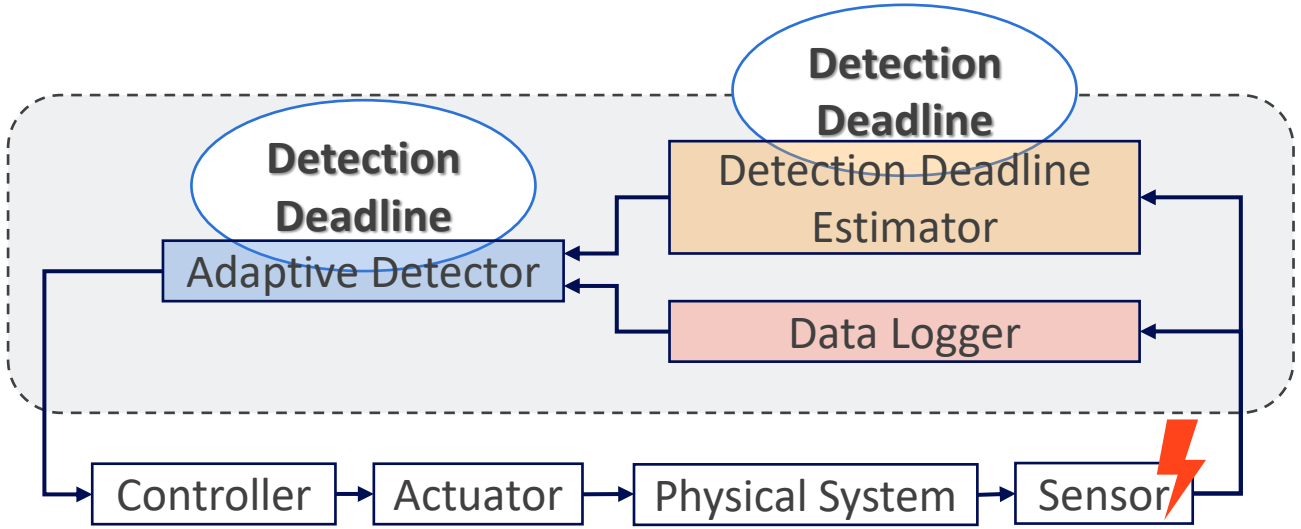
Attack detection in vehicle testbed. x-axis represents time, y-axis represents state x. Purple horizontal line = unsafe set boundary (below is unsafe state set). Orange circle marker = first alarm of adaptive detector. Purple square marker = first alarm of detector with fixed window.





Takeaway

- **Fixed** window-based detector faces a **dilemma**: when system physical states are close to the unsafe states, the attack may be **failed to detected before** touching the unsafe states, because the window is larger than expected; when system choose a shorter window, the **false alarms may increase** even when the physical states are far away from unsafe states.
- We proposed an **adaptive** sensor attack detector, that estimates a proper **detection deadline online** and can **dynamically adapt** the detection window based on it.



Syracuse University

Adaptive Window-Based Sensor Attack Detection for Cyber-Physical Systems

Thank you.
Q&A

SCAN for PDF